STATISTICAL REGRESSION MODELS AND CONTROL CHARTS FOR THE REAL TIME WATER QUALITY NETWORK IN NEWFOUNDLAND

ROBERT RICHARD HARVEY

Statistical Regression Models and Control Charts for the Real Time Water Quality Network in Newfoundland

by

Robert Richard Harvey

A thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirements for the

degree of Master of Engineering

Faculty of Engineering and Applied Science

Memorial University of Newfoundland

2009

St. John's Newfoundland Canada

Abstract

The Newfoundland and Labrador Real Time Water Quality (RTWQ) network was established by the Water Resources Management Division (WRMD) of the Newfoundland and Labrador Department of Environment and Conservation in 2001. Digital sensors continuously record water temperature, pH, specific conductance, dissolved oxygen and stage level for the rivers and streams in the network. This technology is still a relatively new and unfamiliar approach to collecting water quality data in the province. This thesis presents the complete findings of research and development carried out to further enhance the WRMD's capability to work with the RTWQ data in new and innovative ways.

Statistical regression models for predicting water temperature and dissolved oxygen levels are developed for RTWQ stations owned and operated by the WRMD: Humber River, Peter's River, Leary's Brook and Waterford River. A logistic S-shaped model using air temperature can accurately predict water temperature. An exponential model using water temperature can accurately predict dissolved oxygen. Investigations are also carried out into developing statistical regression models using RTWQ data as a surrogate for grab sample chemical concentrations (alkalinity, chloride, etc.). There is more potential to develop these grab sample regression models for urban rivers than there is for rural rivers.

Investigations are made into designing statistical process control charts for monitoring the RTWQ data. Modifications to the traditional process control chart need to be made so that it can monitor the highly autocorrelated RTWQ data.

Acknowledgements

I would like to extend my sincerest thanks to Dr. Leonard Lye who has been a wonderful professor, supervisor and mentor over the past two years of study and research. His guidance through the sometimes overwhelming world of statistics will always be appreciated. I would also like to thank Dr. Ali Khan for giving me the opportunity to get involved with the cutting edge monitoring of water quality being carried out by the Water Resources Management Division here in Newfoundland and Labrador. A special thanks is also warranted to his talented staff of engineers and technicians - Michael Colbert, Leona Hyde, and Renee Paterson in particular took time out of what were already very busy days to pass along their intimate knowledge of the real time network.

For financial support I would like to thank the Natural Sciences and Engineering Research Council of Canada (NSERC), the Institute for Biodiversity, Ecosystem Science and Sustainability (IBES), the Water Resources Management Division of the Department of Environment and Conservation, the Associate Dean's Office, and the School of Graduate Studies at Memorial University.

I would like to thank my wonderful parents, Richard and Verna, who have supported me through these many years of schooling. And finally, to Rebecca - thank you for providing me with laughter after those long days in front of the computer.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

WRMD	Water Resources Management Division
RTWQ	Real Time Water Quality
USGS	United States Geological Survey
GOES	Geostationary Orbiting Satellite
WT	Water Temperature
DO	Dissolved Oxygen
SC	Specific Conductance
TDS	Total Dissolved Solids
ST	Stage
DS	Hydrolab Datasonde Real Time Sensor
MS	Hydrolab Minisonde Real Time Sensor
ADRS	Automatic Data Retrieval System

Chapter One

An Introduction to this Research

1.1 Aim

This thesis focuses on three different areas of research: (1) development of regression models for predicting water temperature and dissolved oxygen for stations in the Newfoundland real time water quality monitoring network, (2) determination of the potential for using real time measurements of water quality as a surrogate for water quality measurements obtained through manual grab sample collection, and (3) determination of the potential for implementing statistical process control charts for monitoring data collected by the real time network.

1.2 Scope

This first chapter provides background information on the history of water quality monitoring for resource management. The transition from manual collection monitoring programs to real time monitoring programs is provided in great detail. A discussion of the need to carry out new and innovative research in the areas of regression modeling and statistical process control conclude the chapter.

1.3 What is Water Quality Monitoring?

Freshwater is essential for the survival of humanity and the natural environment around us. Unfortunately, aquatic ecosystems world-wide are threatened by rapid human population growth and increased industrialization. Even here in Canada, a country largely considered to be unspoiled by the negative effects of today's industrialized society, some of our once pristine water bodies are now unfit even for recreational purposes. Fortunately, the urgent need to implement effective water quality monitoring programs across the country has been recognized in recent years.

The phrase "water quality monitoring" refers to the process of collecting samples, taking measurements and recording various physical, chemical and biological characteristics of a water body so that the suitability of the water for sustaining various uses can be assessed. This suitability for a particular use is determined by comparing the collected data against a set of requirements for the physical, chemical and biological characteristics of the water - i.e. restrictions on the range of pH levels for a river, lake or stream supporting invertebrate communities (Bartram and Ballance, 1996).

The quality of a water body can be affected by a wide range of influences - both natural (geological, hydrological and climatic) and anthropic (population growth and industry). Water quality monitoring programs are designed to use a variety of data collection methods to bring together enough data so that a complete understanding of the health of the water body being studied is achieved.

The data collected in water quality monitoring programs is normally grouped under four different categories: (1) physical and chemical parameters, (2) major ions, (3) nutrients, and (4) trace elements and metals. A list of some of the more commonly sampled indicators of water quality in each of these categories is shown in Table 1.1.

Table 1.1 - Commonly Sampled Indicators of Water Quality

Physical and Chemical Parameters	Major Ions	Nutrients	Trace Elements and Metals	
Turbidity	Calcium	Nitrogen	Aluminum	Lead
Color	Sodium	Nitrate and Nitrite	Arsenic	Manganese
Dissolved Oxygen	Magnesium	Phosphorus	Barium	Mercury
pH	Potassium	Silica	Beryllium	Molybdenum
Conductivity	Sulphate	Dissolved Organic Carbon	Cadmium	Selenium
	Chloride		Cobalt	Strontium
	Flouride		Chromium	Zinc
			Copper	Vanadium
			Iron	

Source: Department of Environment and Conservation (2003)

1.4 The Technology Used for Water Quality Monitoring

The technology used for collecting water quality data has gone through a great deal of change over the past 60 years. Prior to the 1950s, researchers in search of water quality data were forced to carry a vast array of analogue measurement tools and sample collection equipment out to a field site. Once at the site they were free to use the equipment to make a few measurements and collect samples but at the end of the day everything needed to be gathered up and carried back out. Time and labor costs associated with this type of data collection were high and monitoring efforts were usually limited to infrequent measurements of easy to reach water bodies. The biggest downfall of this type of monitoring was the inability to collect enough useful data to provide a complete understanding of the site being studied on both spatial and temporal scales (Glasgow et al., 2004). Although it was possible to use the collected data to gain insight into water quality, it was not possible to use the data to identify trends in water quality over time or determine the environmental conditions that trigger events that caused poor water quality.

In the late 1950s the digital era gained momentum and the opportunity to incorporate digital technology into monitoring programs was widely embraced. The older analogue instruments were shelved and replaced by more accurate in-stream sensors with the capability of digitally measuring multiple water quality parameters at the same time. Advances in data storage technologies made it possible to connect the new digital equipment to data loggers that could be programmed to collect and store repeat measurements made at time intervals small enough that the resulting dataset would represent a continuous record of water quality over a deployment period (Teillet et al., 2002). Once this continuous record was downloaded from the data loggers it could be used to investigate trends and examine how water quality at a site changes over time. For the first time, it was possible to look at an important water quality indicator and see how it might change from one day to the next.

Even with those new digital advances researchers were still forced to personally travel out to the rivers and streams to retrieve the data from the data loggers. In the mid 1970s a push was made by researchers at the United States Geological Survey (USGS) to reduce the amount of man power needed to collect the water quality data when a series of studies were carried out exploring the viability of 3 different data-relay systems for 9 monitoring stations in Florida. Streamflow and rainfall data were remotely collected at the monitoring stations using digital in-stream sensors and the collected data was relayed back to USGS offices using three different methods: (1) telephone landlines, (2) the Landsat-1 satellite operated by the National Aeronautics and Space Administration, and (3) the GOES-1 satellite operated by the United States National Oceanic and Atmospheric Administration. This relayed data was then compared to data downloaded on site in terms of quality and accuracy. Although the data relayed using the telephone landlines was disappointingly inaccurate, the use of satellites did show some considerable promise for monitoring (Turner and Woodham, 1980).

When the Landsat-1 satellite was launched in 1972 it was equipped with a data collection system that could retrieve information from remote individually equipped ground stations and then relay the data to central acquisition stations. Rainfall from the field stations in Florida was monitored over a period of seventeen months and Landsat-1 was used to transmit the data at three minute intervals. The data that was transmitted successfully was available within twelve hours after transmission from the field stations. Similar to the telephone landline system, there were problems relaying the data using Landsat-1 - satellite orbit effects and hardware malfunction onboard the satellite severely limited the usefulness of the data (Turner and Woodham, 1980).

When GOES-1 was launched in 1975 it was outfitted with a data collection system capable of relaying the rainfall data collected at the stations in Florida. Unlike the Landsat-1, the GOES-1 satellite's data collection system was based on a convertible data collection platform that stored the data in an on-board memory and then transmitted the data at three hour intervals. USGS researchers could then obtain this data from the satellite about eight hours after the initial time of transmission. GOES-1 proved to be much more successful in relaying the data then Landsat-1 as the only data lost was due to malfunctions in hardware at the monitoring stations that could easily be fixed.

In the late 1980s, the USGS built on the success of their previous research and established the first real time water quality monitoring network in the United States - the USGS Water Quality Watch. At that point in time monitoring and communications technology had advanced far enough that in-stream sensors could be used to obtain a continuous record of water quality and orbiting satellites could be used to retrieve the data from the sensors without having to visit the site in person. The network grew to prominence in the 1990s to the point where over one thousand stations are in operation today (USGS, 2008). The use of real time technology gained in popularity during the 1990s - to the point where the technology is now being used in countries like Canada, India, Haiti, and Egypt (Khan, et al., 2008) to only name a few.

1.5 Water Quality Monitoring in Newfoundland

The province of Newfoundland and Labrador is located on the Atlantic coast of northeastern North America. The total population of the province is approximately five hundred thousand - with the majority of people living in the province settled on the island of Newfoundland. The people of Newfoundland are surrounded by a considerably large supply of freshwater but some sources of freshwater have been threatened in recent years as urban centers have grown in size. The pressing need for the implementation of a water quality monitoring program for the protection of Newfoundland and Labrador's freshwater resources was officially recognized on April 29, 1986 when the *Canada-Newfoundland Water Quality Monitoring Agreement* was signed between the Canadian federal government and the provincial government. Under this agreement, federal and provincial water quality monitoring activities in the province are coordinated. This coordination of activities has ensured the assessment of the suitability of water for various beneficial water uses and has helped resource managers develop effective pollution control regulations, water quality guidelines and objectives (Department of Environment and Conservation, 2009a).

The Water Resources Management Division (WRMD) of the Newfoundland and Labrador Department of Environment and Conservation established a real time water quality monitoring pilot station on Leary's Brook in the capital city, St. John's, in late 2001. The WRMD use this station to test the feasibility of acquiring water quality information in the province using equipment similar to that used in the United States. The tests carried out at the pilot station site showed there was a great deal of potential for using real time technology for monitoring water bodies in the province. In 2003 real time monitoring equipment was deployed in the waters of four other rivers across the province and the Newfoundland and Labrador Real Time Water Quality (RTWQ) monitoring network has since grown to include more than 15 stations. Table 1.2 presents a list of the stations currently in operation and the location of these stations are shown in Figure 1.1.

Table 1.2 - Newfoundland and Labrador RTWQ Monitoring Stations

Network	Station Number	Station Name	Installed
Provincial base stations for research managed solely by the WRMD			
	NF02YL0012	Humber River at Humber Village Bridge	December-03
	NF02YO0121	Peters River near Botwood	June-05
	NF02ZM0178	Leary's Brook at Prince Philip Drive	November-01
	NF02ZM0009	Waterford River at Kilbride	July-05
Federal-Provincial partnership between the WRMD and Environment Canada			
	NF02ZE0033	Southwest Brook Below Southwest Pond	November-06
	NF02YG0009	Main River at Paradise Pool	June-07
Industry partnership between the WRMD, Environment Canada and Local Industry			
	NF03NE0008	Voisey's Bay Well After Tailings Dam	July-06
	NF03NE0009	Reid Brook at Outlet of Reid Pond	July-03
	NF03NE0010	Camp Pond Brook Below Camp Pond	July-03
	NF03NE0011	Lower Reid Brook Below Tributary	July-03
	NF03NE0012	Tributary to Reid Brook	July-06
	NF02ZK0023	Rattling Brook Below Bridge	December-06
	NF02YO0190	Tributary to Gills Pond Brook	May-06
	NF02YO0192	East Pond Brook	August-06
	NF02YO0193	Well After Tailings Dam - Duck Pond	June-07
	NF03OD0013	Churchill River Below Metchin River	September-08
	NF03OE0030	Minipi River Below Minipi Lake	September-06
	NF03OE0050	Churchill River Below Lower Muskrat Falls	September-08
	NF03OE0051	Churchill River Below Grizzle Rapids	September-08
	NF02ZH0009	Come by Chance River near Goobies	June-07

Figure 1.1 Google Earth Image of Newfoundland and Labrador RTWQ Network Stations

Author's note - the location of the RTWQ stations can be found using the Google Earth software. The number of stations in operation does change over the years - so please refer to the WRMD website for a list of stations that are in operation:

http://www.env.gov.nl.ca/wrmd/RTWQ/RTWQ_Stations.asp

Stations in the network are classified under three headings: (1) Provincial, (2) Federal-Provincial and (3) Industry. The provincial network consisting of base stations for research managed solely by the WRMD. The federal-provincial network is managed under the joint partnership between the WRMD and Environment Canada. The industry network is managed under joint partnerships between the WRMD, Environment Canada and local industry in the province.

The RTWQ network gives WRMD resource managers the ability to monitor in real time the health of select aquatic ecosystems in the province, identify trends in water quality over time and determine the timing of specific events that threaten water quality. Implementation of the network is considered to have been a great success and other provinces have looked to the WRMD for assistance in bringing real time water quality monitoring to the rest of Canada.

1.6 The Need for New Research

Real time water quality monitoring is still relatively new and unfamiliar to the resource managers in the WRMD and there is a great deal to be learned about drawing out important information from the data collected by the network. In late 2007 the WRMD recognized the need for new research and development in the areas of regression modeling for predicting water quality parameters and the implementation of statistical process control techniques for real time water quality data.

Can regression models be developed for water temperature & dissolved oxygen?

Researchers outside of Newfoundland have found success in using historical records of air temperature and water quality data for developing statistical regression models that can predict important indicators of water quality like water temperature and dissolved oxygen (Crisp and Howson, 1982; Webb, 1987; Stefan and Preud'homme, 1993; Pilgrim and Stefan, 1995; Mohseni et al., 1998; Pilgrim et al., 1998; Webb et al, 2003). However, the same type of models have never been developed for stations in the provincial RTWQ network. The regression models presented in this thesis for predicting water temperature and dissolved oxygen represent the first successful models developed using data collected by stations in the Newfoundland RTWQ network. These models should prove to be a useful resource for the WRMD in that they can be used to understand the influence the surrounding environment has on water quality at the stations.

Can regression models be developed for predicting grab sample water quality?

Researchers within the USGS have had success developing regression models that use real time water quality to predict data water quality data that is normally obtained through the manual collection of water quality grab samples (Christensen et al., 2000; Christensen, 2001). Although the WRMD has experimented with developing similar models for stations in the Newfoundland network, they have had limited success and no work has been made public. It was hoped that the potential for developing grab sample regression models could be determined as these kinds of models would not only save the WRMD time and money but would also give resource managers a more accurate idea of chemical loading levels in a river or stream at any point of time. The regression models presented in this thesis for predicting data collected through grab sampling represent the first time any successful models have been developed for the RTWQ network.

Can control charts be used for monitoring real time water quality data?

The statistical process control chart has been used for many years in the processing and manufacturing industries for quality control purposes. The chart is basically a time series plot of the observations with lines drawn that help managers and engineers identify when any unwanted changes occur in a unit or product. There has been interest in recent years in using these charts for monitoring environmental data (Manly, 1994; MacNally and Hart, 1997; Smeti et al., 2007), but this environmental data tends to be highly autocorrelated . This autocorrelation violates the statistical foundation of the traditional control chart (which were designed only to study independent observations collected from a process) and complicates the design process.

There has been no published work looking at using these control charts for monitoring water quality data of the type collected by the RTWQ network. The sensors collect measurements every 15 minutes or every hour (depending on the location) and this record of data tends to be autocorrelated and rather overwhelming. A historical record for a RTWQ station over the course of a year would contain thousands of lines of water quality data and it is not easy to quickly look at the data line by line to see where water quality might have gone outside of a safe range defined by the water quality guidelines. Various methods of implementing control charts for studying the water quality data are presented in this research.

1.7 Thesis Structure

This thesis has been divided into eight distinct chapters. With background information on water quality monitoring now covered, the following chapter will provide an in-depth look at the four RTWQ monitoring stations studied in this thesis - Humber River, Peter's River, Leary's Brook and Waterford River. An overview of the historical records of data available for these stations is presented.

Chapter three presents the results of developing regression models for predicting mean, maximum and minimum water temperatures for the four real time stations at the daily, weekly and monthly time-scales.

The success of the developed regression models for water temperature is built upon in the fourth chapter - where regression models that use water temperature as an explanatory variable for predicting dissolved oxygen are the main focus.

The fifth chapter focuses on developing regression models that link real time water quality data to measurements of water quality collected through manually collected grab samples. The potential for developing regression models for predicting various physical properties, elements, major ions, metals and nutrients at each station is discussed.

Chapter six shifts the focus away from regression modeling and over to examining the potential for implementing statistical process control techniques for monitoring the real time data collected by the network. Background information on how statistical process control techniques are used for monitoring data in the manufacturing industry is first presented. This is followed by a literature review of recent research into using these techniques on environmental data. Results from investigations into the implementation of seven different control chart techniques are presented.

The seventh and eighth chapters conclude this thesis with a review of the major findings in this research and a note on ways to make future work carried out by the Water Resources Management Division in the areas of regression analysis and control charts as straightforward as possible.

Chapter Two

An Overview of the RTWQ Stations

2.1 Scope

This chapter presents an in-depth look at the real time monitoring network in Newfoundland. A description of the setup of the network, the kinds of equipment used for collecting data and the historical records of data collected at each station is provided.

2.2 Setup of the Network

The RTWQ network has been designed around the same framework used for real time monitoring in the United States - where sensors in the water collect the data, the data is relayed to an orbiting satellite, and the data is then retrieved by resource managers from a central repository. Figure 2.1 illustrates the sequence of events that transpire from the time the sensor collects the data to when the data is retrieved by the WRMD for analysis.

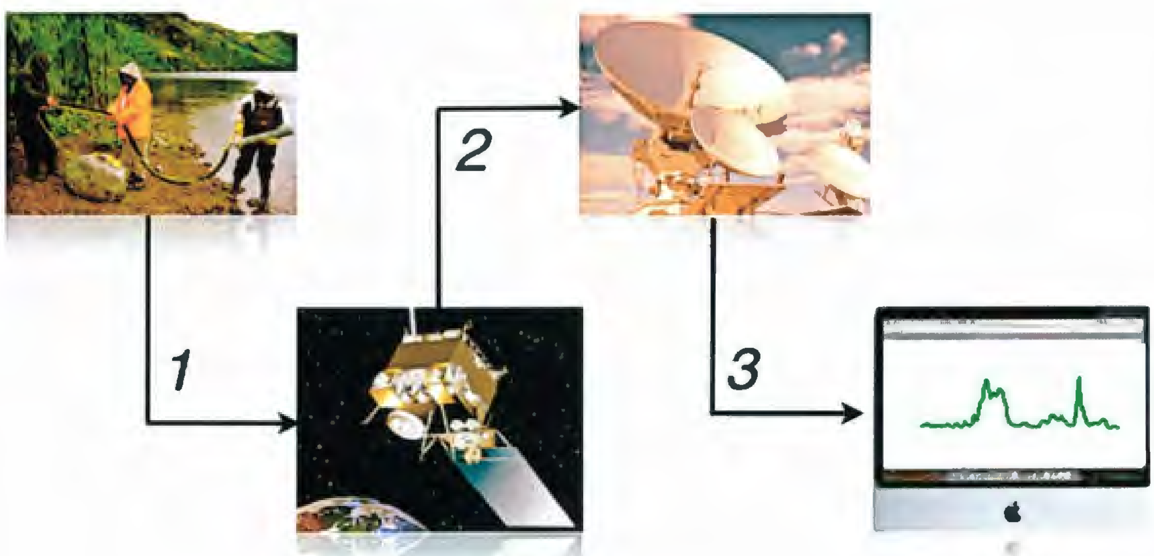


Figure 2.1 How the RTWQ Network Collects Water Quality Data

In stage one a real time sensor installed in the water to obtain measurements of important indicators of water quality like water temperature, pH, specific conductance, dissolved oxygen and turbidity. Sensors have been installed by the WRMD in rivers and streams across the province so that a representative sample of water quality can be obtained. The number of measurements collected by the sensor varies depending on the location of the water body. For rural stations the sensor will collect one measurement every hour, while sensors installed in rivers near urban centers will collect one measurement every fifteen minutes.

In stage two the communication at the monitoring station transmits the water quality every three hours to the GOES satellite.

In stage three the GOES satellite relays the water quality data to a central depository located in Maryland, USA - known as the National Environmental Satellite Data Information System. This central depository is owned and operated by the United States National Oceanic and Atmospheric Administration.

In stage four the water quality data stored at the central depository is automatically downloaded, processed and distributed using an Automatic Data Retrieval System (ADRS) designed by the Department of Environment and Conservation. This ADRS software is setup to automatically upload the data to the WRMD internet site so the general public, resource managers and industry representatives can go on-line to view real time plots of water quality parameters.

2.3 Data Collected by the Network

The water quality data collected at each monitoring station in the RTWQ network is classified under one of four headings:

1. Real time measurements of water temperature, pH, specific conductance, dissolved oxygen and turbidity obtained using the Hydrolab Datasonde real time sensor
2. Real time measurements of stage and streamflow obtained from nearby hydrometric monitoring stations operated by Environment Canada
3. Real time measurements of air temperature obtained from nearby weather monitoring stations operated by Environment Canada
4. Grab samples of water quality parameters manually collected by the WRMD

Data Type 1 - Real time measurements obtained using the Hydrolab Datasonde Sensor

There are many different types of real time sensors currently on the market with sensors ranging in price, reliability, and measurement capabilities. The WRMD uses a companion pair of sensors known as the Hydrolab Multiprobe Series 4a Datasonde and the Hydrolab Minisonde for collecting real time water quality data (Figure 2.2). Part (a) of the figure shows the Minisonde on the left and the Datasonde on the right. Part (b) and (c) of the figure show a closer view of the Datasonde.

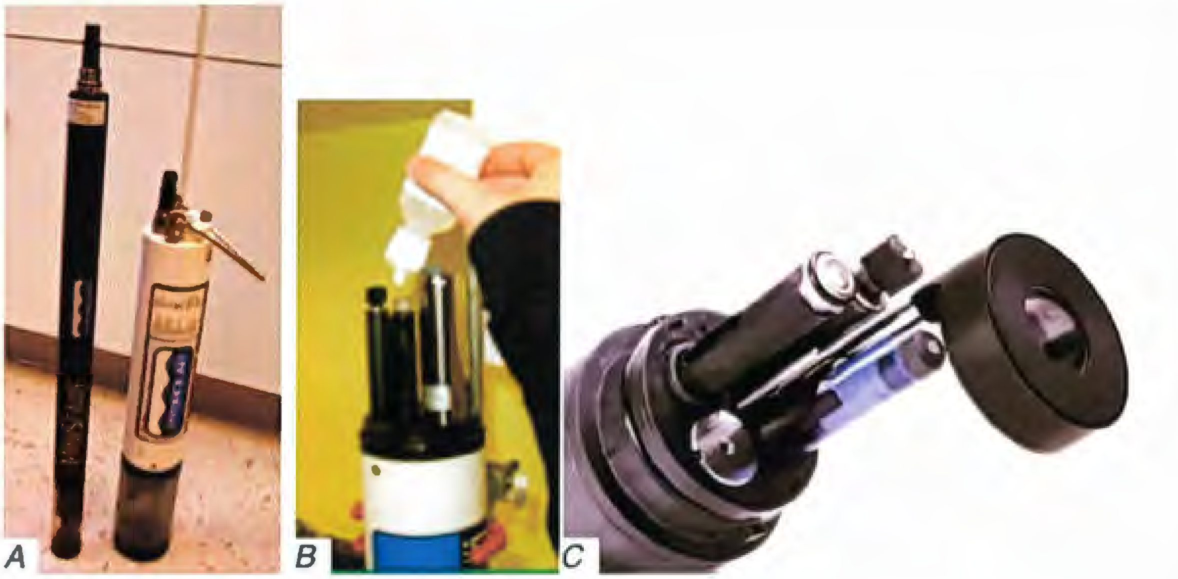


Figure 2.2 Hydrolab Multiprobe Datasonde Series 4a and Hydrolab Minisonde

The Hydrolab Datasonde is known to provide accurate readings of water quality indicators in real time. It is designed to be easily portable - with an outer diameter of 8.9 centimeters, a total length of 58.4 centimeters and a weight of 7.4 pounds. This makes it an ideal size for deployment in both large and small rivers and streams. It has a built in memory capacity of 120,000 measurements and can reliably operate in temperatures ranging from -5 to as high as +50 °C (Campbell Scientific, 2009). The version of the sensor used by the WRMD can record water temperature, pH, specific conductance, dissolved oxygen and turbidity.

The Datasonde can record water temperature from -5 to 50 °C with an accuracy of ± 0.10 °C at a resolution of 0.01 °C (Campbell Scientific, 2009). Water temperature is a measure of the amount of heat present in water and is one of the most important indicators of overall water quality. Not only does water temperature regulate the metabolism and growth rates of aquatic plants and animals but it also largely influences many chemical processes (Department of Environment and Conservation, 2009b). A number of previous studies have shown that when water temperatures go outside of a normal range in a river there will be consequences for the aquatic inhabitants. A study carried out by Hodgson and Quinn (2002) found that the spawning period of sockeye salmon in rivers in the North western United States are interrupted when water temperatures reach 19 °C as the salmon are forced to seek refuge from the higher than normal water temperatures. Lund et al. (2002) found that high water temperatures will induce a heat-shock response in juvenile salmonoids. Water temperature is known to be influenced by a number of factors - i.e. temperature of source water, industrial use of the water, and the heat exchange between the air and water interface.

The Datasonde can collect pH data in the range of 0 to 14 pH units with an accuracy of ± 0.2 units at a resolution of 0.01 units (Campbell Scientific, 2009). The pH of water is a measure of the hydrogen ion activity in a system and proper pH is essential for the survival of plant and animal species. The Canadian Council of Ministers of the Environment (CCME) specifies that pH should fall within the range of 6.5 to 9 pH units for the protection of aquatic life (Environment Canada, 2002). Not only can pH influence aquatic life but it can also change physical characteristics like water color, odor and taste. The pH in a river is known to be strongly influenced by daily biological and geological activities, which are in turn directly affected by the water temperature (Department of Environment and Conservation, 2009b).

The Datasonde can collect specific conductance data in the range of 0 to 100,000 µS/cm with an accuracy of ± 1 µS/cm at a four digit resolution (Campbell Scientific, 2009). Conductance is a measure of the ability of water to pass an electrical current. When this measure of conductivity is corrected to 25 °C it is referred to as specific conductance. Collecting specific conductance measurements is useful in that it can be used to provide an indirect measure of the amount of dissolved substances (salts) in an aquatic system. Pure water has a specific conductance of 0 to 200 µS/cm and bigger rivers tend to have values from 200 to 1000 µS/cm. Specific conductance levels above 1000 µS/cm represent quite saline conditions (Department of Environment and Conservation, 2008b). The WRMD use the specific conductance measurements collected by the sensor to obtain an estimate of the total dissolved solids (TDS) in the river or stream being monitored using the equation TDS (g/L) = specific conductance [µS/cm] * 0.00064. TDS is used a measure of the organic and inorganic solids in the water and is a general indicator of salinity. Water with large dissolved solids concentrations can produce scaly deposits and cause corrosion of pipes (Department of Environment and Conservation, 2008b).

The Datasonde can collect dissolved oxygen data in the range of 0 to 50 mg/L with an accuracy of ± 0.2 mg/L at a resolution of 0.01 mg/L (Campbell Scientific, 2009). Like water temperature, dissolved oxygen is considered to me one of the more important indicators of the health of aquatic ecosystems. Dissolved oxygen is a measure of the amount of oxygen dissolved in water and is controlled by a number of factors including oxygen consumption by aquatic organisms, the flow and depth of water and water temperature. The health of the system can be determined by considering three separate

DO levels – low (0-8 mg/L, high oxygen demand which can lead to fish kills), medium (8-12 mg/L, a healthy system) and high (12-20 mg/L which can lead to excessive algal growth). The WRMD uses the dissolved oxygen measurements collected by the sensor to obtain an estimate of the percent saturation of the river or stream being monitored. Percent saturation refers to the amount of dissolved oxygen contained in the water compared tot he amount that could potentially be there at the same temperature. Water can become supersaturated from excessive aeration (i.e. waterfalls) and have a percent saturation greater than 100%. Percent saturation levels below 60% or above 125% are undesirable (Department of Environment and Conservation, 2008b).

The Datasonde can also record turbidity levels in the water in real time. Turbidity readings for the stations in Newfoundland are known to be rather unreliable and cannot at this time be used to gain any reliable insight into actual turbidity levels in Newfoundland rivers and streams.

The Hydrolab Minisonde is similar to the Datasonde in that it can record water quality data in real time but it is not designed to be left in the water for long deployment periods and does not have the same capacity to record as many parameters as the Datasonde. The Minisonde is normally used by the WRMD for calibration purposes or as a temporary stand-in sensor for damaged Datasondes.

Historical records of real time measurements made by the real time sensors can be downloaded from the ADRS for every station in the network. The ADRS software is setup so that a user defines a period of interest and the desired record of real time data is then automatically retrieved. The system saves the records as Microsoft Excel files that contain the date and time of measurement, water temperature, pH, specific conductance, dissolved oxygen, and stage. The WRMD goes through a process of removing periods of time from the dataset where the sensor was not operating effectively and modifies the records to account for drift of the sensor measurements over time. The WRMD has come to find that measurements collected by the sensor tend to be less accurate the longer the sensor is left in the water. At the start of the deployment period the sensor will have been recently calibrated and will take accurate measurements of water quality parameters. Over the next few weeks this calibration will be lost and the measurements taken by the sensor will slowly drift away from the true value. The WRMD can gain an estimate of the size of this drift by sending personnel out to the sampling station with a recently calibrated Hydrolab Minisonde to take a companion set of measurements. The accurate readings taken by the Minisonde can be compared to the final Datasonde readings to see how they differ and changes can be made to the Datasonde readings to account for this difference. The product of making modifications for missing values and for sensor drift is referred to as a *Drift Corrected Historical Record* of sensor data for the station. It is the aim of the WRMD to keep these corrected records current but at this point in time, corrected records are not available past 2008.

Author's Note - Refer to Appendix A for a detailed description of how the ADRS records are modified to account for drift.

Data Type 2 - Real Time Measurements of Stage

Each of the RTWQ monitoring stations in the network are located in conjunction with nearby hydrometric monitoring stations operated by Environment Canada that record stage and streamflow in real time. Take for example real time station NF02YL0012 where the RTWQ sensor is installed across the river from hydrometric station 02YL003 (Figure 2.3). The real time stage data for each station is posted on the same Department of Environment and Conservation webpage as the RWTQ data. Historical records of the stage and streamflow can be obtained directly from Environment Canada.

At the time of carrying out this research it proved difficult to obtain complete records of streamflow for the RTWQ stations. It is unknown why records of stage level for the stations were complete but the corresponding streamflow in the same file for the first years of operation of the stations would be incomplete. Records from recent years of monitoring have improved and now usually include stage and streamflow.

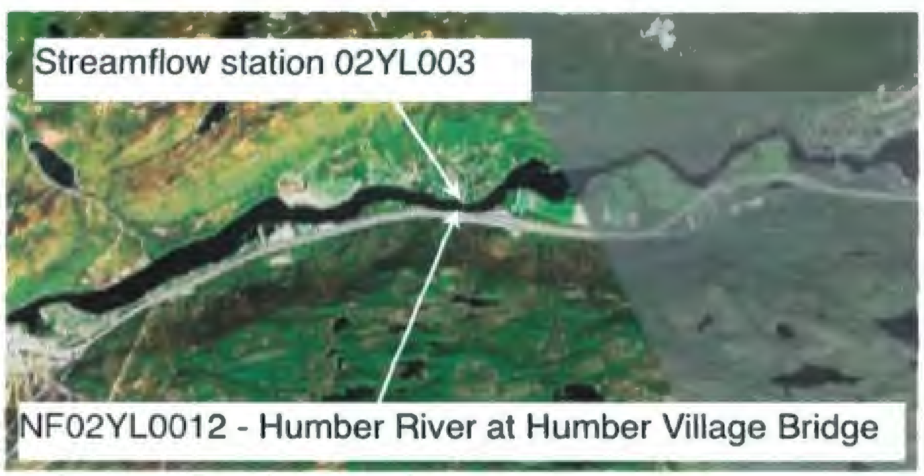


Figure 2.3 RTWQ Stations are Located in Conjunction With Hydrometric Stations

Data Type 3 - Nearby Measurements of Air Temperature

Although the RTWQ stations are not outfitted to record air temperature at this point in time, an estimate of air temperature at the station can be obtained using real time measurements made at nearby Environment Canada weather monitoring stations.

Data Type 4 - Grab Samples of Water Quality

The WRMD manually collects grab samples of water quality at the RTWQ stations at throughout the year. Once collected these samples are sent to Environment Canada laboratories for chemical analysis. Although grab samples only give an indication of water quality at the time of sample collection, they can provide information on the following water quality parameters that cannot be recorded by the RTWQ sensors:

- (1) Physical Properties, solids, and sediment values: alkalinity, color, conductivity, hardness, pH, total dissolved solids, total suspended solids, turbidity, and water temperature
- (2) Major ions and metals: boron, bromide, calcium, chloride, flouride, potassium, sodium, sulphate, ammonia, aluminum, antimony, arsenic, barium, cadmium, chromium, copper, iron, lead, magnesium, mercury, nickel, selenium, uranium and zinc.
- (3) Nutrient levels: dissolved organic carbon, nitrate, nitrite, kjeldahl nitrogen, and total phosphorus.

The historical record of grab sample data collected for each station in the network does tend to vary. Some of the older stations have the results of over 20 grab samples available for analysis while some of the newer stations have less than 10.

2.4 The RTWQ Stations Studied in This Thesis

The following four RTWQ stations are studied in this thesis:

- NF02YL0012 Humber River at Humber Village Bridge (west coast of Newfoundland),
- NF02YO0121 Peter's River near Botwood (central Newfoundland and no longer in operation),
- NF02ZM0178 Leary's Brook at Clinch Crescent (east coast of Newfoundland), and
- NF02ZM0009 Waterford River at Kilbride (east coast of Newfoundland).

All four of these RTWQ stations are part of the provincial network solely operated by the WRMD (Figure 2.4) and have the longest and most accurate monitoring records available for analysis. Although RTWQ records exist for the federal-provincial and industry network stations, the records available for analysis are and not as extensive as those available for the four provincial network stations and they have been excluded from analysis as a result.

Author's Note - as the records of the federal-provincial and industry RTWQ network stations are expanded over time it will be possible to include these stations into regression model development.

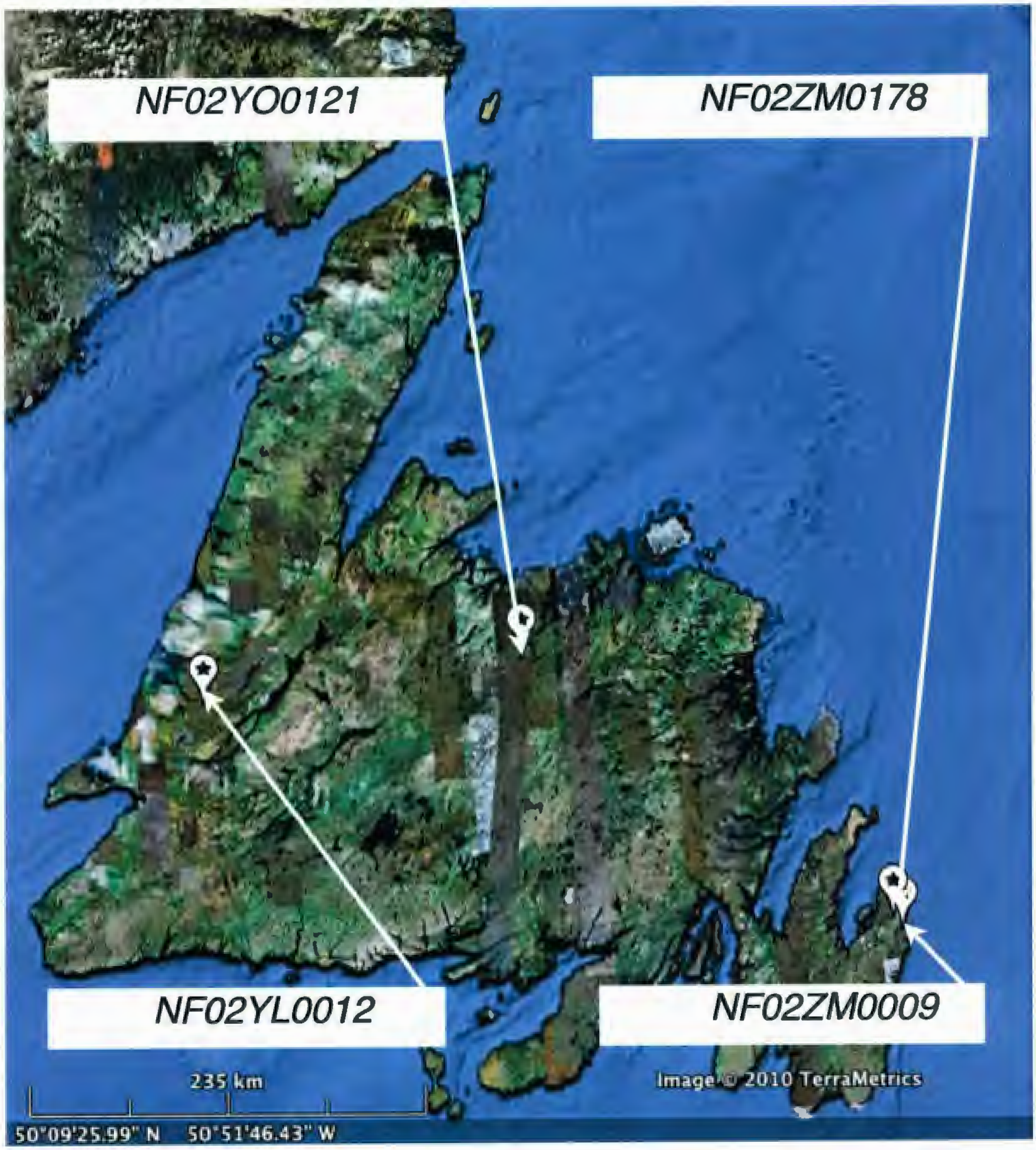


Figure 2.4 Google Earth Image of the RTWQ Stations Studied in this Thesis

NF02YL0012 - Humber River at Humber Village Bridge

Figure 2.5 RTWQ Station NF02YL0012 Humber River

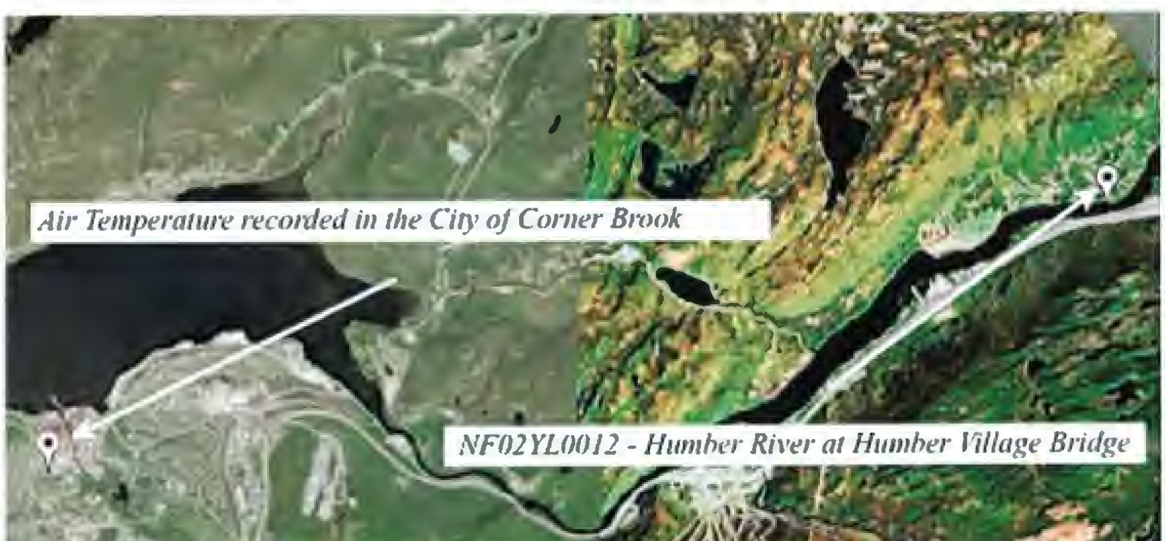


Figure 2.6 Proximity of Air Temperature Measurements to NF02YL0012

The Humber River is the second largest river system on the island of Newfoundland. The headwaters of the river flow all the way from the highlands of the Long Range Mountains on the West coast through the deep and heavily forested river valley into a wide marshy flood plane where the river drains the surrounding mountainous areas. The drainage area of the river is over 7000 square kilometers with forest comprising the largest chunk of this area (62.59%) followed by lakes (12.47%) wetlands (9.69%), barren land (7.23%), vegetation (6.88%) and other (1.13%). Monitoring the water quality of the Humber River is important as there is a a great deal of development pressure in the region. Not only are the waters of the river used for hydroelectric power and municipal consumption, but they are used for recreational purposes as well. There are two solid waste disposal sites and over fifty commercial farms in the area. Highways and access roads run along the river with and there are a number of bridges that cross the river (Department of Environment and Conservation, 2009c).

RTWQ station NF02YL0012 - Humber River at Humber Village Bridge (Figure 2.5) is located 12.5 kilometers from the outlet of the Humber River into the Bay of Islands. Corrected hourly measurements of water temperature, pH, specific conductance and dissolved oxygen are available for December 2003 to April 2008. Hourly measurements of real time stage were recorded at Environment Canada hydrometric station 02YL003. Hourly measurements of air temperature are recorded in the nearby city of Corner Brook (approximately 15 kilometers away - refer to Figure 2.6). 37 grab samples of water quality were collected at the station from May 2004 to August 2008.

NF02YO0121 - Peter's River near Botwood

Figure 2.7 WRMD Personnel Installing a Hydrolab Datasonde in Peter's River

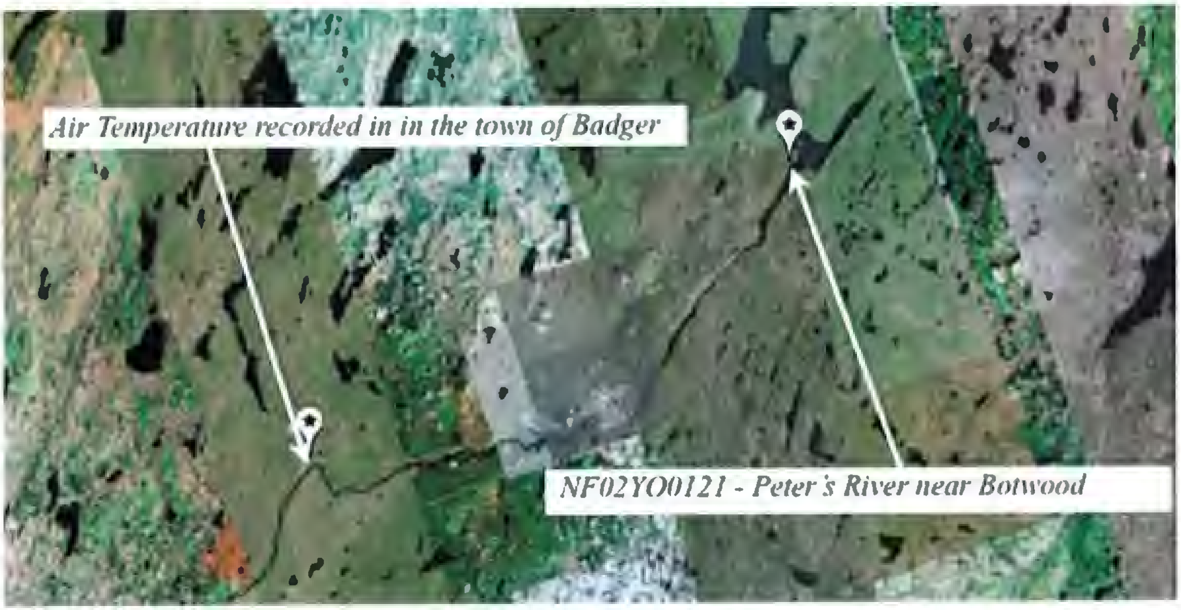


Figure 2.8 Proximity of Air Temperature Measurements to NF02YO0121

The Peter's River Basin is located in the central lowlands on the island of Newfoundland. The basin is approximately 80% forest, 14% lowlands, 3% lake and 3% barren land. The protected water supply area for the river is approximately 224 square kilometers (Acres International Limited, 2005).

A number of public roads, resource roads, abandoned railway lines and old trails allow access to almost all areas of the Peter's River Basin. The watershed area is currently used extensively for quarrying and recreational activity. A 1995 consultant's study found that all land use in this particular region of the province had the potential to deliver pollutants to the watercourse (Acres International Limited, 2005). The river used to supply drinking water for the nearby towns of Botwood and Peterview but this was no longer the case after 2006, when those towns began relying on another nearby source of water.

Real time monitoring station NF02YO0121 - Peter's River near Botwood (Figure 2.7) was first brought online in 2005 but was taken permanently offline three years later in 2008. Drift corrected hourly measurements of real time water temperature, pH, specific conductance and dissolved oxygen are available for the period June 2005 to February 2008. Hourly measurements of real time stage recorded at Environment Canada hydrometric station 02YO006 are available from June 2005 to February 2008. Hourly measurements of air temperature were recorded in the nearby town of Badger (approximately 50 kilometers away - Figure 2.8). 18 grab samples of water quality at the station were collected from June 2005 to February 2008.

NF02ZM0178 - Leary's Brook at Clinch Crescent in St. John's

Figure 2.9 RTWQ Station NF02ZM0178 Leary's Brook

Figure 2.10 Proximity of Air Temperature Measurements to NF02ZM0178

Leary's Brook is an urban water system that runs through a developed section in the capital city of St. John's. The total drainage area for Leary's Brook is 19.6 square kilometers. Forest makes up 74.36% of this total drainage area while barren land (12.31%), wetland (6.67%), vegetation (6.15%) and lakes (0.51%) make up the rest (Department of Environment and Conservation, 2009d).

Real time station NF02ZM0178 - Leary's Brook at Clinch Crescent in St. John's (Figure 2.9) is the main testing station for real time technology for the WRMD. It was the first station brought online in the real time network and real time data for the station dates back to 2002. Unfortunately, the equipment at the station is removed quite often for re-calibration and testing and the historical record for the station is quite erratic as a result.

Drift corrected measurements of real time water temperature, pH, specific conductance and dissolved oxygen collected at 15 minute intervals were available from September 2004 to December 2007. Measurements of real time stage were recorded at 15 minute intervals at Environment Canada hydrometric station 02ZM020. Hourly measurements of air temperature were recorded from September 2004 to December 2007 at the St. John's international airport (approximately 5 kilometers away - refer to Figure 2.10). 20 grab samples of water quality were collected at the station from April 2005 to September 2008.

NF02ZM0009 - Waterford River at Kilbride in St. John's

Figure 2.11 RTWQ Station NF02ZM0009 Waterford River

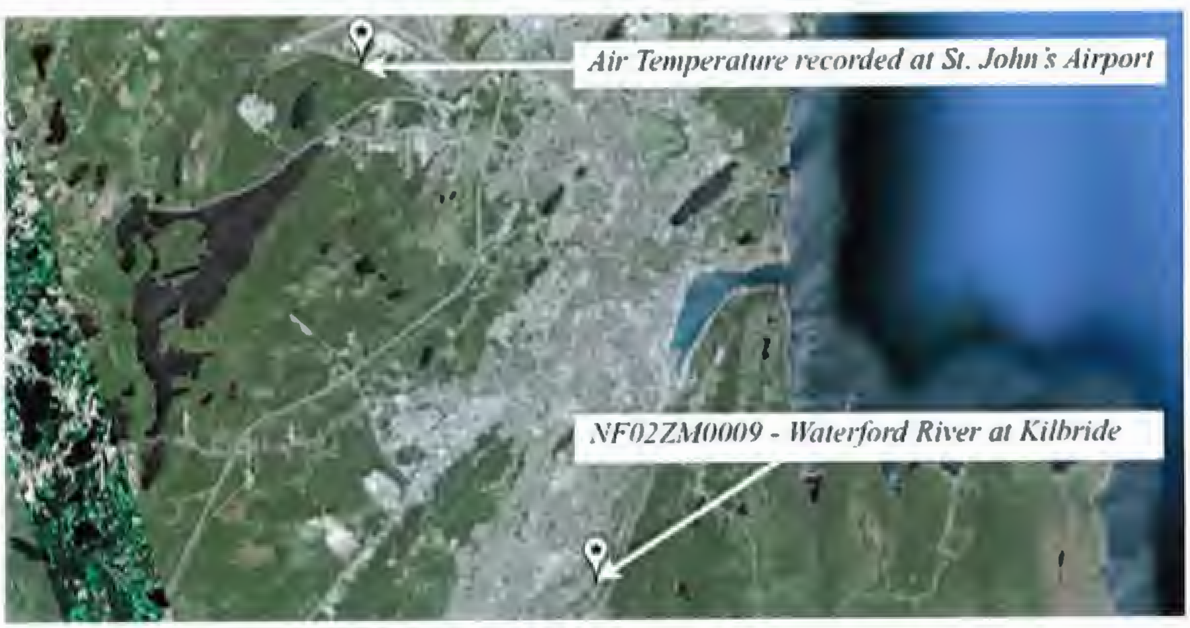


Figure 2.12 Proximity of Air Temperature Measurements to NF02ZM0009

The headwaters of the Waterford River are located in the town of Paradise, Newfoundland. From Paradise the river flows through to the capital city of St. John's until it reaches its outlet in the city harbor. The total drainage area for the river is roughly 52 square kilometers. Forest makes up 57.89% of this drainage area, barren land (19.36%), wetland (11.09%), vegetation (10.71%), lakes (0.75%) and other (0.19%) make up the remainder (Department of Environment and Conservation, 2009e).

Development pressure around the Waterford River is considered to be moderate with the majority of development in the basin located in two industrial parks. Highways, city streets and a number of access roads are dispersed throughout the basin. About 25% of the basin is dedicated to urban and sub-urban development (residential, commercial and industrial areas).

Real time station NF02ZM0009 - Waterford River at Kilbride (Figure 2.11) is located roughly thirteen kilometers from the headwaters of the river and was first brought on-line in 2005. Drift corrected hourly measurements of real time water temperature, pH, specific conductance and dissolved oxygen were available for July 2005 to March 2008. Hourly measurements of real time stage were recorded at Environment Canada hydrometric station 03OE003. Hourly measurements of air temperature were recorded from July 2005 to March 2008 at the St. John's international airport (approximately 10 kilometers away - refer to Figure 2.12). 20 grab samples of water quality were collected at the station from August 2005 to September 2008.

2.5 Summary of the Data Available for Analysis

A general statistical summary of the real time data collected at the stations is presented in Table 2.1. The Humber River station has the longest drift corrected record, the lowest overall mean water temperature, specific conductance and air temperature. It also has the highest overall mean dissolved oxygen and stage levels. Overall mean dissolved oxygen levels at the stations are all above the CCME minimum guideline for the protection of aquatic life (5.5 mg/L). Overall mean pH for each of the stations falls within the CCME guideline (6.5 to 9 pH units). In terms of stage levels, Leary's Brook and Waterford River are similar in size, with Peter's River being the medium sized station of the group. Overall mean specific conductance levels are much higher at the stations located in the capital city of St. John's (Leary's Brook and Waterford River) than they are at Peter's River and Humber River.

Table 2.1 - Overall Mean Values of Real Time Data Available for Analysis

Station Name	Date	Mean WT (°C)	Mean pH (pH unit)	Mean SC (µS/cm)	Mean DO (mg/L)	Mean Stage (m)	Mean AT (°C)
NF02YL0012 Humber River	Dec. 2003 - Apr. 2008	6.56	6.82	36.14	12.19	2.1	4.19
NF02YO0121 Peter's River	Jun. 2005 - Feb. 2008	8.53	6.65	43.51	10.91	1.15	4.55
NF02ZM0178 Leary's Brook	Sept. 2004 - Dec. 2007	7.91	6.67	432.95	11.26	0.76	5.67
NF02ZM0009 Waterford River	Jul. 2005 - Mar. 2008	8.16	6.69	515.46	10.92	0.56	6.15

Comparisons among the stations beyond the overall mean can be made by examining each of the parameters on a month by month basis. Figure 2.13 illustrates the monthly mean water temperature for each of the four stations. As is to be expected, water temperatures are the coldest in the fall and winter months and warmest in the spring and summer months. Water temperatures in summer tend to be the highest in Peter's River and temperatures in winter can get quite cold there as well. Note that water temperature recorded by the sensors never drop far below zero degrees - even in the coldest months of the air. Water temperatures in the Humber River do not tend to be as warm as the other stations in the summer months which is perhaps due to the larger size of the river (as larger rivers contain more water and take longer to heat up).

A comparison of the monthly mean air temperatures at each of the stations is made in Figure 2.14. Air temperatures at the stations is coldest in the fall and winter and warmest in the spring and summer. The coldest air temperatures are recorded at the Peter's River station. Monthly air temperatures at Leary's Brook and Waterford River are very similar which is to be expected as air temperature for both stations is recorded at the same Environment Canada weather station in St. John's. The highest monthly air temperatures can be found at the Humber River station. Note that monthly mean air temperatures in Newfoundland are not high - with daytime air temperatures in summer rarely going above 25 degrees Celsius.

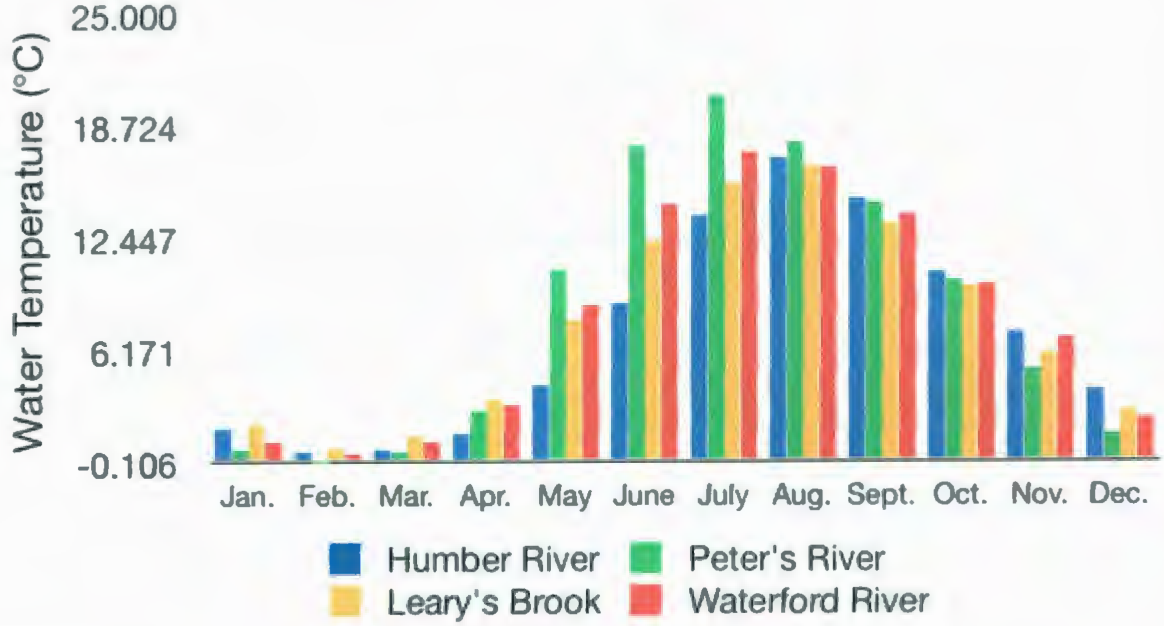


Figure 2.13 Monthly Mean Water Temperature at the RTWQ Stations

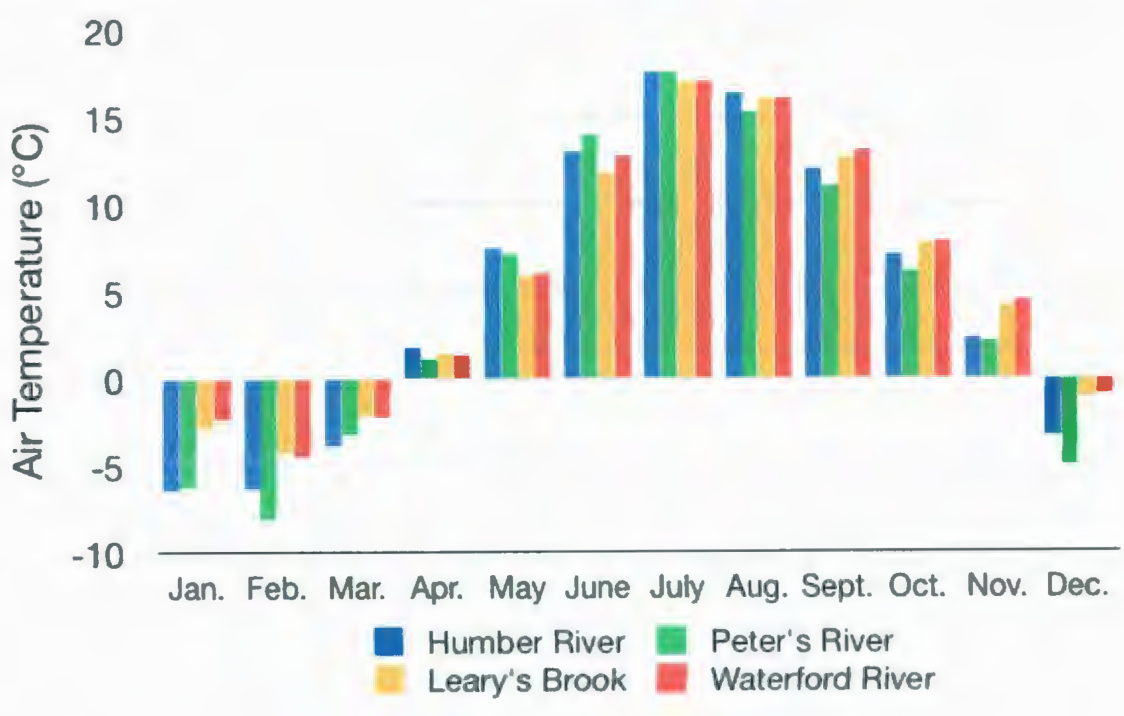


Figure 2.14 Monthly Mean Air Temperature at the RTWQ Stations

Figure 2.15 illustrates the monthly mean stage levels at the four stations in the network. This plot shows that stage levels throughout the year can be ranked from the highest to lowest as follows: Humber River, Peter's River, Leary's Brook and Waterford River. Stage levels at the Humber River station are the highest in the spring months (likely due to snowmelt) and lowest in the warmer summer months. Mean monthly stage levels at the Peter's River, Leary's Brook and Waterford River stations tend to be highest in the winter and spring.

Figure 2.16 compares the monthly mean pH levels at the four provincial stations. From the plot it can be noted that throughout the inter to spring months pH levels at the four stations are either close to or fall outside of the safe range of pH (6.5 to 9) as specified by the Canadian Council of Ministers of the Environment guidelines for the protection of aquatic life. The monthly mean pH levels tend to increase for all stations during the summer months and never go above 8 pH units.

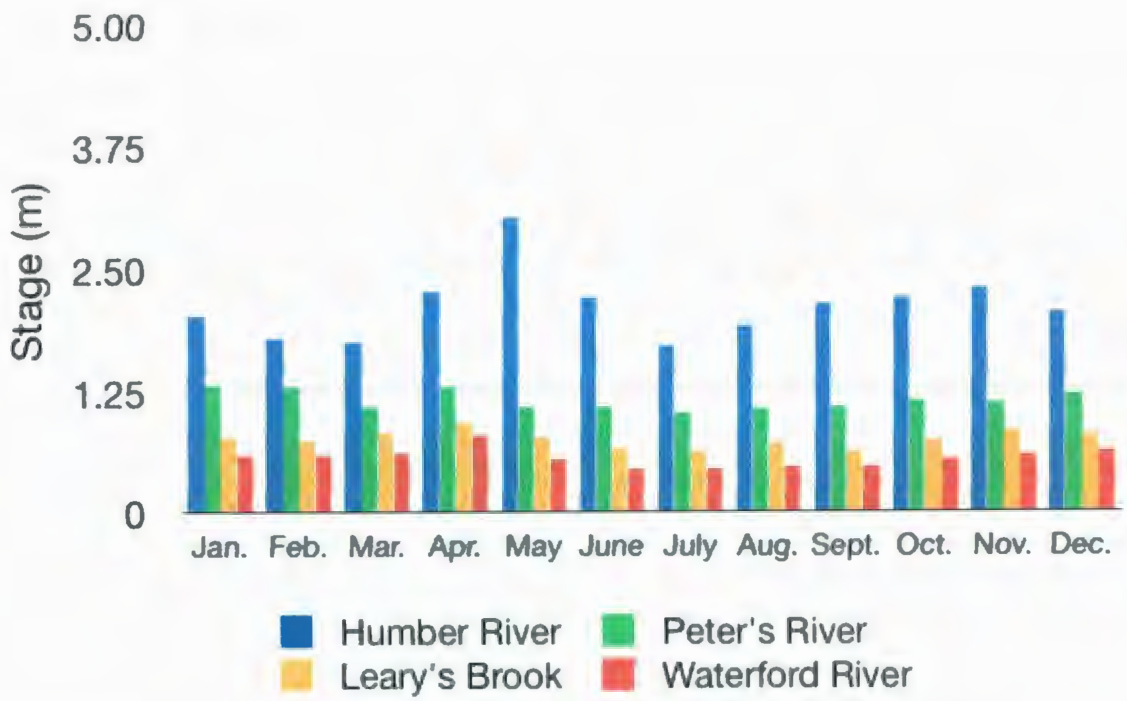


Figure 2.15 Monthly Mean Stage Levels at the RTWQ Stations

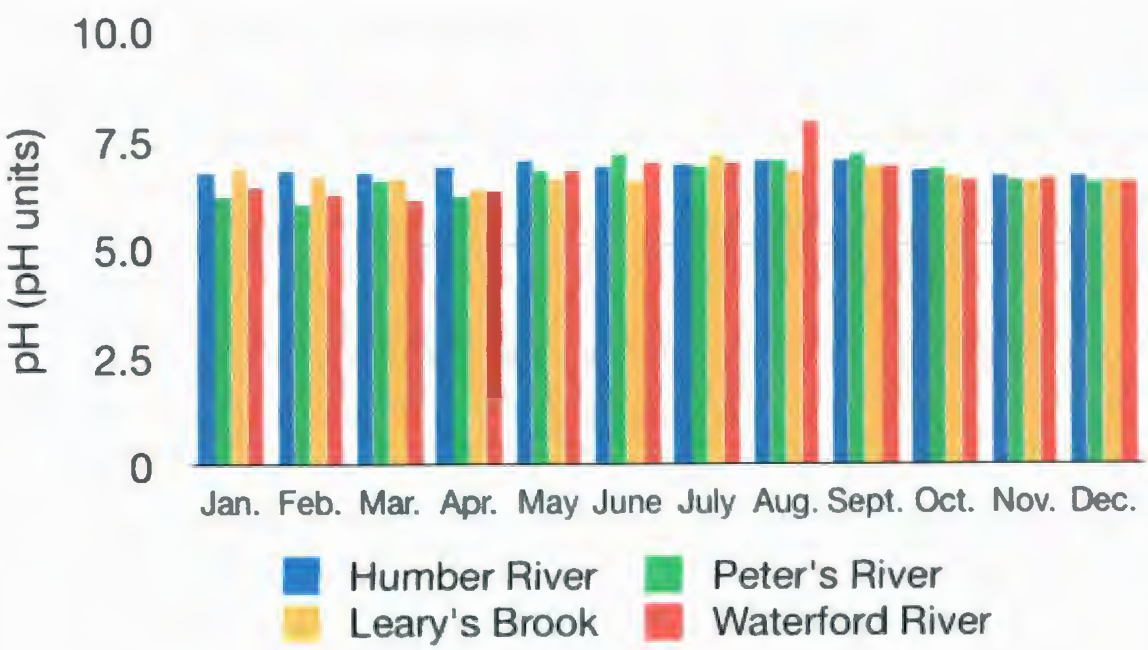


Figure 2.16 Monthly Mean pH Levels at the RTWQ stations

Figure 2.17 compares the monthly mean specific conductance at the four provincial stations. It can be noted from the plot that monthly mean specific conductance levels at the Leary's Brook and Waterford River stations can get rather high during winter and spring months (where specific conductance outside of 1000 μS/cm represents quite saline conditions). The high levels in winter are likely due to heavy road salting of the roads in the capital city of St. John's - when this salt washes off the roads and into the rivers the specific conductance levels will spike. Pure water tends to have specific conductance levels under 200 μS/cm so monthly mean levels at the Humber River and Peter's Station are quite low.

Figure 2.18 illustrates the monthly mean dissolved oxygen levels at the four stations. Monthly mean dissolved oxygen levels at the stations do tend to be lower in the warmer summer months - with levels at the Peter's River, Leary's Brook and Waterford River dropping below the 8 mg/L in July and August. When dissolved oxygen levels drop below 8 mg/L the health of the aquatic ecosystem can be threatened, but the monthly mean levels at the stations even at their lowest points are still rather close to the safe level.

A closer look at the variations in some of these parameters will be carried out in the next chapter when regression models are developed for linking the daily, weekly and monthly measurements of air temperature and stage to water temperature. In the fourth chapter a closer look at the variations in dissolved oxygen will be carried out when water temperature and stage are used to predict dissolved oxygen at the real time stations.

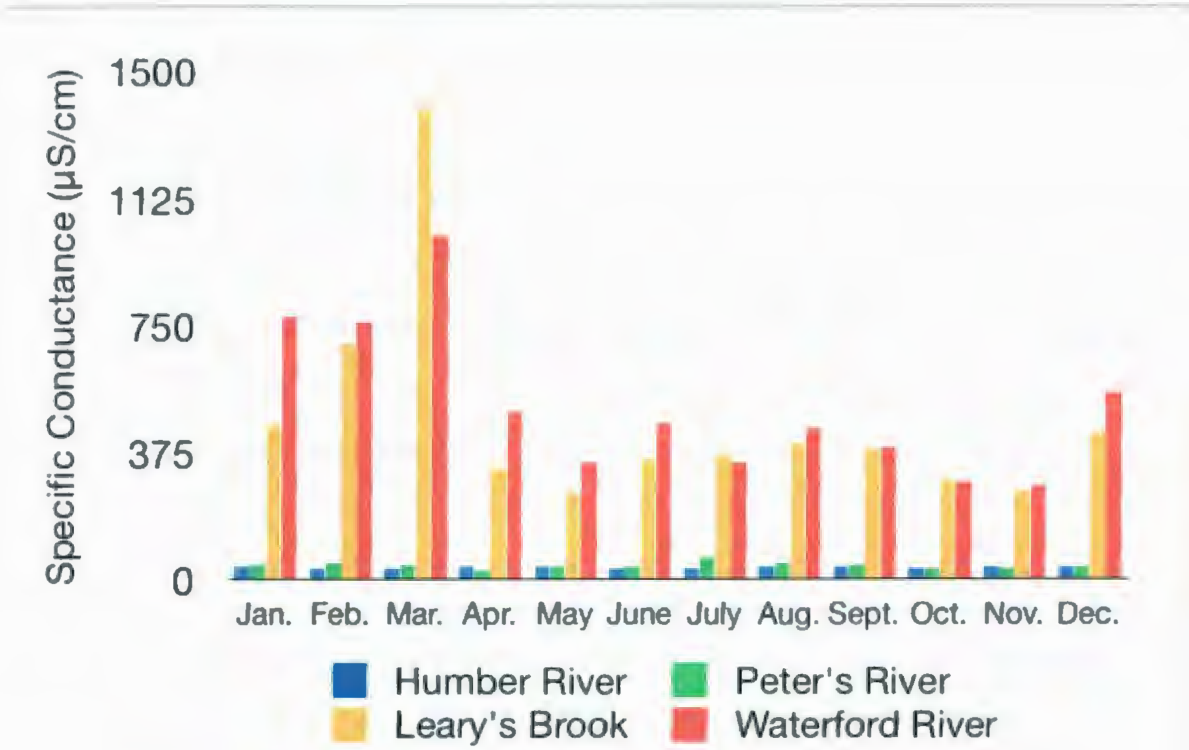


Figure 2.17 Monthly Mean Specific Conductance at RTWQ Stations

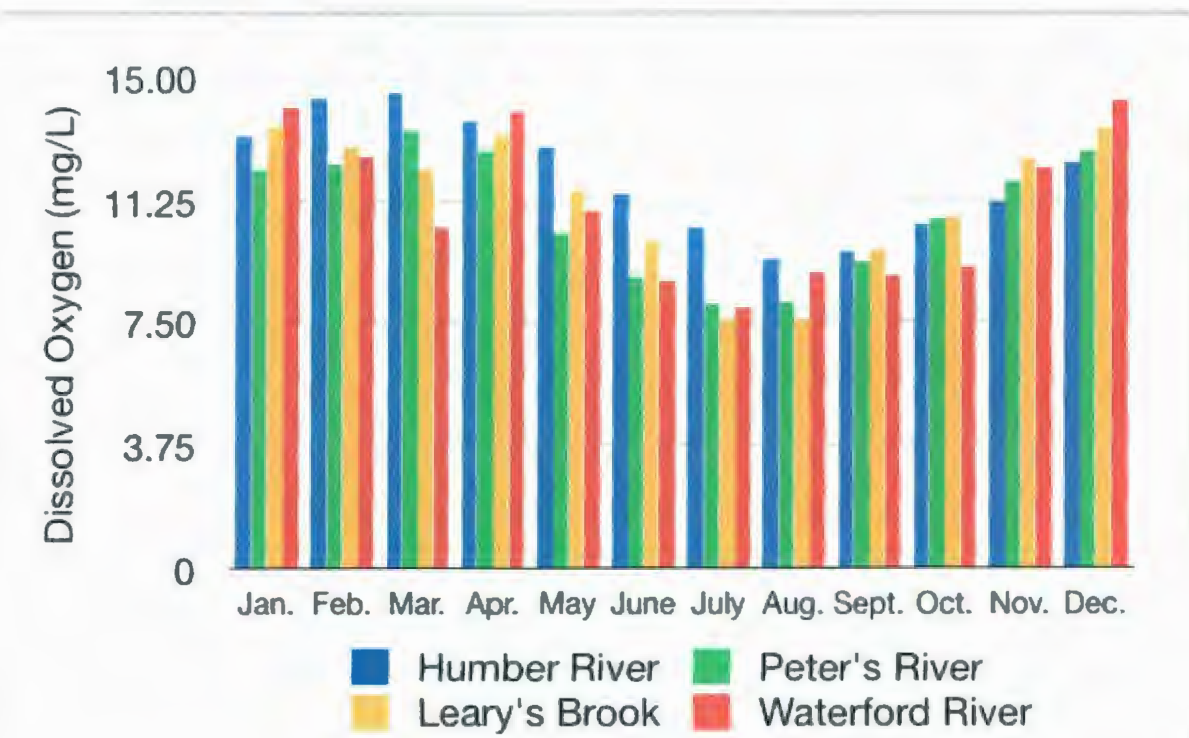


Figure 2.18 Monthly Mean Dissolved Oxygen Levels at the RTWQ Stations

Chapter Three

Development of Regression Models

for Water Temperature

3.1 Scope

Regression models for predicting water temperature at the real time stations are the main focus of this third chapter. A literature review of published works relevant to this kind of research is presented. A description of the methodology used in developing regression models follows. Models for predicting mean, maximum, and minimum water temperature at monthly, weekly and daily time scales are presented for each station. The models are tested using historical datasets specifically reserved for this purpose. A discussion of the results will be carried out at the end of the chapter.

3.2 Background Information and Literature Review

3.2.1 Background Information on Regression Modeling

Most often regression models are developed to learn something about the relationship that exists between variables of interest (i.e. water temperature and air temperature). The simplest type of regression model is referred to as the linear regression model, where just one explanatory variable is used to predict one response variable of interest. The linear regression model takes the form of:

$$y_i = \beta_0 + \beta_1 x_1 + \varepsilon_i \qquad \text{for } i = 0,1,2...,n \qquad \text{Equation 3.1}$$

where y_i is the i^{th} observation of the response variable, x_i is the i^{th} observation of the explanatory variable, β_0 is the intercept, β_1 is the slope, ε_i is the random error or residual for the i^{th} observation, and n is the size of the sample. The error around the linear model ε_i is a random variable with a mean of zero and a constant variance that does not depend on the value of the explanatory variable. Developing linear regression models between two variables that are at least somewhat linearly related can be quite easy these days as most pieces of statistical software will do most of the harder statistical calculations and work for the user. The goodness of fit and the appropriateness of the model can be determined using the statistical software output, but this will be discussed in more detail a little later in this chapter.

When dealing with water resources data it is quite often the case that more than one explanatory variable will need to be used for explaining variation the response variable. In these type of situations an extension of simple linear regression known as multiple regression can be used to develop a relationship that explains the relationship between the response variable and the two (or more) explanatory variables. This multiple regression model takes the form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \qquad \text{Equation 3.2}$$

where y is the response variable, β_0 is the intercept, β_1 is the slope coefficient for the first explanatory variable, β_2 is the slope coefficient for the second explanatory variable, β_K is the slope coefficient for the k^{th} explanatory variable, and ε is the remaining unexplained noise in the data. For this model, there are k explanatory variables, some of which may be correlated to each other. Sometimes picking the appropriate explanatory variables to include in the multiple regression model will be easy while at other times it can get quite difficult. Either way, a good model aims to explain as much of the variance of the response variable with the fewest number of explanatory variables (Helsel and Hirsch, 2002).

Sometimes water resources data will follow a linear relationship but it often nonlinear relationships exist between the explanatory variable(s) and the response variable. In those cases nonlinear regression models are developed for explaining the relationship between the explanatory variable (or variables) and the response variable of interest. The approach taken to developing these nonlinear models is the same as the linear models. Nonlinear models can get somewhat complex but fortunately the right types of statistical software can be used to make searching for the best fitting models easier.

3.2.2 Literature Review

A number of researchers have developed statistical regression models linking measurements of air temperature, streamflow and water temperature. Some of the earliest published work in this field was carried out by Johnson (1971) who studied six streams in New Zealand. In his research he found that monthly mean water temperatures and air temperatures at the streams could be described by a positive linear relationship. In a similar piece of research Song et al. (1973) studied streams in the state of Minnesota and found the relationship between monthly mean water temperature and air temperature to also be linear.

Smith (1979) went beyond mean measurements of water temperature and describes how linear regression models can also be fit to daily maximum and minimum water temperature and air temperature. Smith found that the regression models developed for daily minimum values tend to be more scattered and less reliable than those for daily mean or maximum values - due to the fact that the higher thermal capacity of the water will prevent development of the low night time minima characteristic of air temperature. In a later piece of work, Smith (1981) revisits the linear model and finds that the models are more accurate when working with data collected on larger time scales (i.e. months) rather than shorter time scales (i.e. days).

Stefan and Preud'homme (1993) developed linear regression models for predicting daily and weekly mean water temperature using data collected at 11 streams in the central United States. The air temperature data used in the study was obtained using weather stations that were from 0 to 144 miles away from the streams. The relationship between water and air temperature was less scattered when weekly means were studied and more scattered when daily means were studied. The authors found that the equations developed for shallower streams gave lower standard deviations than those developed for the larger and deeper streams.

Pilgrim and Stefan (1995) and Pilgrim et al. (1998) present linear regression models for predicting daily, weekly, monthly air temperature using water temperature data collected at 39 streams in the state of Minnesota in the United States. The air temperature data was obtained from weather stations on average 37.5 km away from the stream. The weekly and monthly models were more reliable for prediction purposes than the daily models. The authors attempted to lump the data from all streams in the area into one larger dataset to obtain a general linear regression model for the region but the resulting regression models was not as accurate as the models developed for individual streams.

The improved accuracy of models developed using longer time scales was also found in a study carried out by Webb and Nobilis (1997) where monthly mean models were more accurate than weekly mean models. The authors did find however that it was quite difficult to develop accurate models when working with an annual mean dataset (most likely due to the accuracy problems inherent in the dataset itself).

A number of authors have investigated the potential influence of streamflow on water temperature. Crisp and Howson (1982) developed a linear model for predicting 5-day and 7-day mean water temperatures from 8 streams in England using air temperature data recorded from monitoring stations up to 50 kilometers away. The authors found that a multiple regression model incorporating streamflow provided a negligible improvement of the linear model for prediction purposes. Hockey et al. (1982) developed regression models for predicting daily stream temperature using daily maximum air temperature and flow rate for a river in New Zealand. Their regression model showed that both flow rate and maximum air temperature influenced water temperatures in the river under natural flow conditions but there was a great deal of scatter in the relationship and more data would need to be collected if an accurate model was to be developed.

Webb (1987) used historical records of water temperature and air temperature to develop linear regression models for streams in the United Kingdom. In a similar piece of work Webb et al. (2003) develop linear regression models for streams in the United Kingdom (Figure 3.1) but also develop a multiple regression model that includes both air temperature and streamflow as explanatory variables for water temperature. Hourly, daily and weekly data collected from catchments in the United Kingdom were studied and streamflow was found to be significant only when the shorter time scales (i.e. hourly measurements) were studied. When larger time scales were of interest (i.e. daily and weekly means) streamflow was no longer of importance and a simple linear regression model was found to be appropriate. The strength of the linear relationship between water temperature and air temperature was found to increase as the time base being studied increased from hourly means to weekly means. The authors also studied the presence of hysteresis in the data and found that accounting for hysteresis in hourly data improved the fit of the regression models. The weekly mean data was grouped into 3 month periods (January-March, April-June, July-September, and October-December) but hysteresis was not found to be significant.

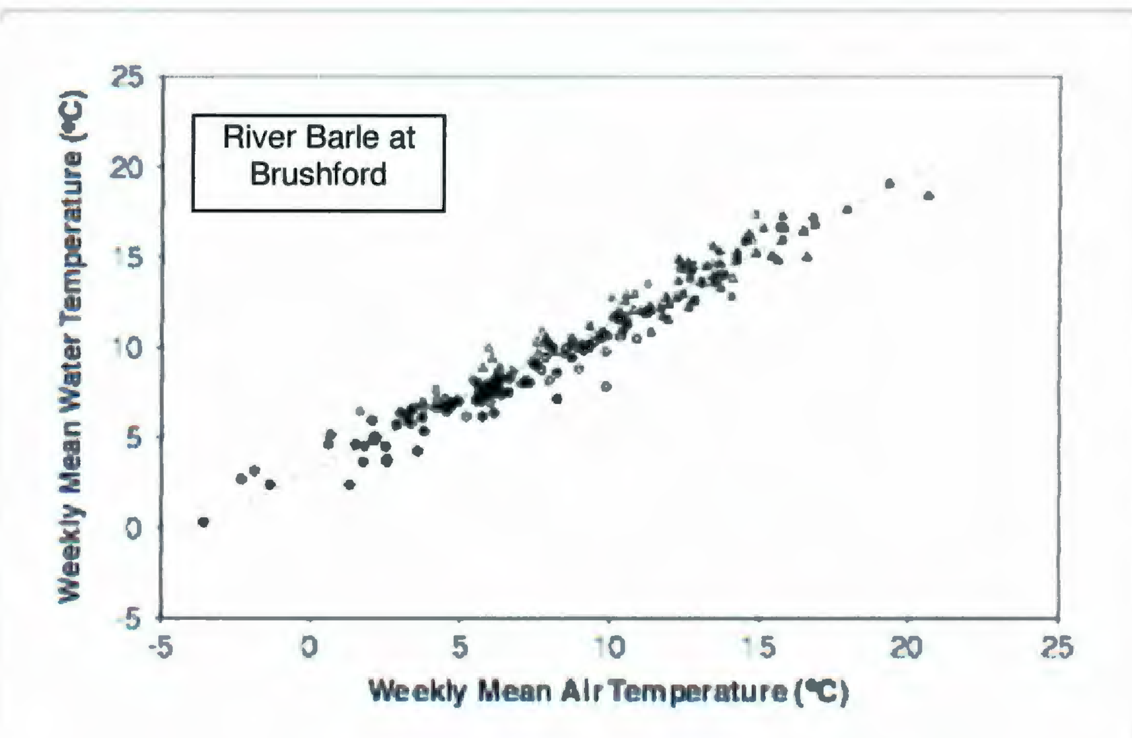


Figure 3.1 Linear Regression Model used by Webb et al. (2003)

Langan et al. (2001) studied a 30 year record of stream temperature collected at a stream in Scotland and found that the linear relationship between air and water temperature was stronger during summer than it was during the colder winter months.

The assumption that the relationship between water temperature and air temperature is best described by a linear equation was first brought into question when Mohseni et al. (1998) developed a nonlinear logistic function to describe the nonlinear S-shaped relationship that was observed in weekly maximum water temperature and air temperature data recorded at 584 streams in the United States. In this study air temperatures were obtained from weather stations that were 1.4 to 244 km away from the streams. The S-shaped relationship was found to be more accurate than a simple linear relationship for making predictions of weekly maximum stream temperatures. In the same study, the authors note that for some rivers the weekly and monthly stream temperature data showed signs of hysteresis due to snowmelt which kept water temperature close to zero in the spring even though air temperatures were rising. For rivers showing signs of hysteresis, one regression model was developed for the warming season and another for the cooling season to account for the heat storage effects (Figure 3.2).

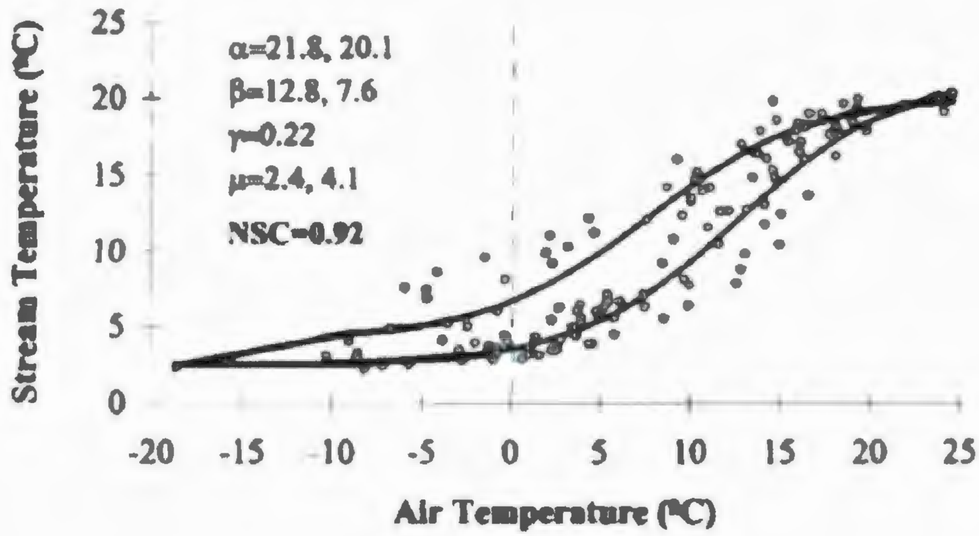


Figure 3.2 Mohseni Used Two S-shaped Models for Handling Hysteresis

(Image source - Mohseni et al., 1998)

Mohseni and Stefan (1999) discuss why the relationship between air temperature and water temperature should no longer be considered a linear one as the nonlinear S-shaped relationship better describes the physical relationship exists between the two parameters. Their research has shown that when air temperatures are high enough, the slope of the relationship between water and air temperature will level off despite further increases in air temperature. This leveling off above air temperature greater than 25 degrees Celsius is a result of evaporative cooling and back radiation from water surfaces. As air temperature continues to increase the capacity of the atmosphere to hold moisture increases and the rate of evaporative cooling increases as well. As the river or stream increasingly loses heat, the water temperature no longer increases linearly with the increase in air temperature and the relationship will taper off. When air temperatures drop below 0 °C the linear relationship also changes and the relationship levels off. Quite often groundwater inflow to a stream will directly affect the minimum stream temperature and even though temperatures continue to drop, the stream will reach some minimum temperature above zero degrees. The authors also note that streams in colder climatic zones may not always show a significant change in the slope of the relationship at higher air temperatures as air temperatures do not rise high enough to show the water temperature limiting the effect of evaporative cooling. Mohseni et al. (2002) revisit their work with the streams in the United States and add an additional parameter to their original S-shaped relationship so that boundaries for the relationship between weekly air temperature and stream temperature are put in place.

Caissie et al. (2001) modeled maximum daily stream temperature and maximum daily air temperature for a small stream in New Brunswick, Canada using the S-shaped regression model. The daily data was far too scattered to provide reliable models to logistic models were only developed using the weekly data. Stochastic models that broke the data into seasonal components were also developed. The simpler regression models compared quite favorably in terms of performance and were much easier to develop.

Neumann et al. (2003) compare the fit between a linear regression model and an S-shaped logistic model for predicting daily maximum stream temperatures using streamflow and air temperature data collected from 1993 to 1998 for the Truckee River in the state of Nevada. Linear models were found to work well for prediction purposes while the S-shaped model was also accurate but offered not major improvement over the linear model for that particular river. Morrill et al. (2005) compared the fit of the more commonly used linear regression model and the nonlinear S-shaped model developed by Mohseni et al. (1998) for evaluating the relationship between weekly mean values of air temperature and stream temperature values for 43 sites in the United States and internationally. The nonlinear S-shaped relationship was found to produce a better fit than the linear model.

Lagergaard Pedersen and Sand-Jensen (2007) used linear and nonlinear S-shaped regression models to study seasonal variations in daily water temperature for streams in Denmark. The nonlinear model was found to be more accurate than the linear one for the daily data and was then further used for examining the impact of a global warming scenario on streams in Denmark.

Not all of the published literature looking at the relationship between air temperature and water temperature has been devoted to developing statistical regression models as a number of authors have sought to describe the relationship between these two parameters using mathematic models that describe the physics of heat exchange between a river and the surrounding environment (Morin and Couillard, 1990; St. Hilaire et al., 2000). These other models require a great deal of input (meteorology, hydrology, stream geometry) if they are to be effective. For simplicity most authors and resource managers prefer to work with regression models for prediction in that are usually much easier to understand and use readily obtainable data. Benyaha et al. (2007) discusses the advantages and disadvantages of all the different modeling approaches.

3.3 Methodology Used for Developing Water Temperature Models

A four step methodology was used to develop regression models for water temperature at the real time stations: (1) get familiar with the commonly used models, (2) obtain the datasets necessary for regression, (3) use statistical software to determine the best fitting models, and (4) test out the best models for the purposes of predicting water temperature.

3.3.1 Step One - Get Familiar With the Most Commonly Used Models

The four most commonly used regression models for predicting water temperature found in the literature were: (1) linear regression using air temperature as an explanatory variable, (2) multiple regression using both air temperature and streamflow, (3) the first logistic S-shaped model proposed by Mohseni that uses air temperature, and (4) a follow-up S-shaped model proposed by Mohseni with an additional parameter for minimum water temperature. Table 3.1 presents the form of each of these models and lists some brief notes on each of them. *Author's Note - stage has been used instead of streamflow in this research as stage records are much more reliable for the RTWQ stations.*

The literature review has shown that the trend in recent years has been towards abandoning the traditional linear relationship between water temperature and air temperature in favor for the nonlinear logistic S-shaped relationship. However, not all authors have made the switch as some have found that the more complicated logistic model does not always offer up any major improvement. As for the multiple regression equation with stage, some authors have streamflow to be significant for some rivers but only at some time scales while others never find streamflow to be significant.

Table 3.1 The Most Commonly Used Regression Models for Water Temperature

Model	Form	Parameters
Linear Equation 3.3	$Tw = a_0 + a_1 Ta + \varepsilon$	- Tw is water temperature for a time period - Ta is air temperature for the same time period - a_0 and a_1 are the regression coefficients - ϵ is the error term.

- Strength of the relationship depends on the time interval being studied.
- Relationship is usually much stronger when examining monthly mean or maximum temperatures but is more scattered when daily means or maximums are used.
- Smith (1979) notes that regression models for minimum water temperatures tend to be more scattered than for mean and maximum water temperatures.
- Regression models are rarely developed for hourly water temperatures as it tends to be too scattered.

Model	Form	Parameters
Logistic 1 Equation 3.4	$Tw = \dfrac{\alpha}{1 + e^{\gamma(\beta - Ta)}}$	- Tw is the estimated water temperature - Ta is the measured air temperature - α represents the maximum stream temperature - γ measure of the steepest slope of the function - β is the air temperature at the inflection point.

- Original form of the non-linear logistic equation proposed by Mohseni and Stefan (1998).
- Some researchers have found this model is a great improvement over the linear model while others have found negligible improvement and tend to go with the linear model for simplicity.
- Hysteresis in the data may make it necessary to develop two models (one high and one low)

Model	Form	Parameters
Logistic 2 Equation 3.5	$Tw = \mu + \dfrac{\alpha - \mu}{1 + e^{\gamma(\beta - Ta)}}$	- additional parameter μ added to represent the estimated minimum stream temperature.

Notes: Modified form of the original Logistic 1 model proposed by Mohseni and Stefan (1998)

Model	Form	Parameters
MLR Equation 3.6	$Tw = a_0 + a_1 Ta + a_2 ST + \varepsilon$	- variable added for the influence of stage (ST)

- Other researchers have used streamflow as an additionally explanatory variable for water temperature.
- Stage has been used instead of streamflow in this research

3.3.2 Step Two - Develop the Necessary Datasets

The historical records of drift corrected Datasonde real time data, Environment Canada stage data and air temperature data were used to develop mean, maximum and minimum datasets at the daily, weekly and monthly time scales for the four real time stations.

Although the air temperature data being used in this research is collected at Environment Canada weather stations and not at the exact location of the RTWQ stations, it needs to be assumed that the Environment Canada data is a close approximation of the air temperature at the provincial network RTWQ stations. In this research the distances from the RTWQ stations to the weather stations are as follows: Humber River (approximately 15 kilometers), Peter's River (approximately 50 kilometers), Leary's Brook (approximately 5 kilometers), and Waterford River (approximately 10 kilometers). These distances are comparable to the work carried out by other authors: Crisp and Howson (1982) worked with air temperature data collected up to 50 kilometers away from their streams, Stefan and Preud'homme (1995) worked data collected from 0 to 144 miles away from their streams, and Pilgrim and Stefan (1995) worked with distances on average of 37.5 kilometers. Although it is possible that the meteorological data obtained at the stations may not match with what might be recorded at the RTWQ stations, there is currently no air temperature data collected at the RTWQ stations available for comparison. In the future the WRMD is looking to install air temperature equipment at the stations. With this new equipment in place it would be possible to quantitatively determine the impact of relying on the weather station air temperature when developing water temperature regression models. Until this equipment is installed, it will be necessary to assume air temperature data collected from reasonably close weather stations (i.e. 50 kilometers away or less) to be a good approximation of RTWQ station air temperature.

Most authors use streamflow as an additional explanatory variable for water temperature and not stage level (Crisp and Howson, 1982; Webb, 1987). In this research streamflow could not be used as it was not possible to obtain a complete record of streamflow for the RTWQ stations (determined after speaking with an Environment Canada representative in the early stages of research who noted that getting the missing hourly streamflow data would take a significant amount of effort, and the resulting dataset if developed would be in a rather unusable form). Although it would have been useful to compare models developed using streamflow and those using stage, that will not be possible until more extensive streamflow datasets are available in the future. Until more complete streamflow datasets are made available, stage can be considered as a substitute explanatory variable for streamflow as the two were found to be highly correlated using Minitab Statistical Software (Release 14) - with a correlation p-value of 0.00 obtained between stage and streamflow for each station.

These historical records contain over 25,000 rows of data and sorting through all these rows find the daily, weekly and monthly values would be a very time consuming process. Take for example finding daily mean values for the real time measurements collected at the Humber River station. Over the period of 2003 to 2008 there were over 30,000 measurements collected for each water quality parameter. Although it would entirely be possible to scroll down through 24 rows of data in Excel (each containing an hourly measurement) and then insert a row below this data so that an equation to find an average value for those 24 rows could be entered, this would get rather exhausting when done the required 1500 times.

In this thesis, macros were written for Minitab that sort through the rows of data and then automatically output the values a user was searching for (i.e. weekly mean values of water temperature for all of 2005). An overview of how these macros operate is given in Figure 3.3.

Author's note - the complete code for each of the macros can be found in Appendix B.

Macros written to search out the mean, maximum and minimum values
One macro written for handling Datasonde hourly RTWQ measurements
One macro written for handling Environment Canada stage level
One macro written for handling Environment Canada air temperature

After the drift corrected historical records are copied from Microsoft Excel into Minitab the macro can be run to find the values of interest (monthly macro shown)

Session

```
MTB > %Monthly
Executing from file: C:\Program Files\MINITAB 14\MACROS\Monthly.MAC
macro is used to get the monthly mean max and min values for real
time data
Author - Richard Harvey - February 2009
c1 has the date and time
c2 has the water temperature
c3 has the pH
c4 has the specific conductance
c5 has the dissolved solids
c6 has is the percent saturation
c7 has the dissolved oxygen
c8 has the turbidity (but leave this out of calculations)
Set desired year and months (i.e. 2004 1 12)
DATA> 2004 1 12
```

NF02YL0012 ***

↓	C1-D Date	C2 WT	C3 pH	C4 SC	C5 DS	C6 PS	C7 DO
1209	2/26/2004 1:50:00 AM	0.18	6.85	34.20	0.0220	98.01	14.18
1210	2/26/2004 2:50:00 AM	0.18	6.86	33.90	0.0218	97.52	14.10
1211	2/26/2004 3:50:00 AM	0.19	6.86	34.20	0.0220	97.93	14.18

Macro will then find and store the desired data

NF02YL0012 ***

↓	C15 Year	C16 Month	C17 Mean WT	C18 Max WT	C19 Min WT	C20 Mean pH	C21 Max pH	C22 Min pH	C2 Mean
1	2004	1	2.9353	4.01	2.22	6.76431	6.93	6.44	33.9
2	2004	2	0.4284	1.13	-0.10	6.95429	7.24	6.51	33.3
3	2004	3	0.3838	1.53	-0.10	6.82224	7.07	6.27	36.4
4	2004	4	1.2579	2.46	0.33	6.77868	6.91	6.50	37.2
5	2004	5	3.8283	6.65	1.86	6.74578	6.95	6.42	36.7
6	2004	6	7.3241	10.62	5.43	6.76471	6.93	6.65	35.9
7	2004	7	14.1083	17.51	10.21	6.68086	7.06	6.26	38.0
8	2004	8	17.8454	20.67	13.42	6.90165	7.31	6.58	38.5

Figure 3.3 Minitab Macros Seek Out the Mean, Maximum and Minimum Values

Once the mean, maximum and minimum values for the daily, weekly and monthly time scales were developed, each of the larger datasets were split into two separate datasets - a longer one for developing the regression models and another shorter one consisting of the last year of real time measurements in this historical record. Some stations have a rather short length of time covered by their drift corrected historical record and it was not always possible to have a full year of real time measurements available for model testing purposes without cutting significantly into the amount of data available for developing the regression models (i.e. Peter's River with a short drift corrected dataset ranging only from July 2005 to May 2007). For these shorter datasets the first priority was developing a longer dataset for regression modelling and then leaving some data left over for testing. For longer datasets like Humber River it was much easier to have multiple years for regression modelling and then at least a full year left over for testing purposes.

Table 3.2 presents a general statistical summary of the datasets for developing regression models using mean monthly, weekly and daily data. A more detailed statistical summary of the datasets can be found in Appendix C.

Initial investigations into the correlation between water temperature, air temperature and stage were carried out once the regression datasets were developed. The relationship between water temperature and air temperature was found to be strongly positive while the relationship between stage and water temperature was negative. Scatterplots of the mean, maximum and minimum air-water relationship showed the strong positive relationship between the two parameters. Figure 3.4 presents one of these scatterplots for the Peter's River station. Plots for the other stations can be found in Appendix D.

Table 3.2 Statistical Summary of the Regression Modeling Datasets

Time Periods Covered by the Datasets

For Developing Regression Models	*For testing Regression Models*
Humber River - Dec 2003 to Dec 2006	Humber River - Jan 2007 to Apr 2008
Peter's River - July 2005 to May 2007	Peter's River - July 2007 to Feb 2008
Leary's Brook - Sept 2004 to Dec 2006	Leary's Brook - May 2007 to Dec 2007
Waterford River - July 2005 to Mar 2007	Waterford River - Apr 2007 to Mar 2008

		Water Temperature (°C)			Air Temperature (°C)			Stage (m)			Dissolved Oxygen (mg/L)		
Dataset	Obs	Avg	Min	Max	Avg	Max	Min	Avg	Min	Max	Avg	Min	Max
Monthly Mean Models for Regression													
Humber River	37	7.07	0.38	17.85	4.89	-8.99	18.07	2.12	1.52	3.30	12.09	8.60	19.48
Peter's River	23	8.07	-0.10	20.58	4.28	-8.95	17.86	1.14	0.94	1.48	11.00	7.66	13.76
Leary's Brook	16	6.93	0.71	16.58	5.12	-3.65	16.98	0.79	0.63	0.95	11.64	6.45	13.87
Waterford River	21	8.09	0.22	17.71	5.94	-5.74	17.11	0.56	0.42	0.91	11.09	6.59	14.22
Weekly Mean Models for Regression													
Humber River	149	7.11	0.25	18.53	5.14	-12.94	20.51	2.13	1.39	3.66	12.05	8.56	19.61
Peter's River	91	8.34	-0.13	22.53	4.40	-13.76	20.82	1.16	0.93	1.84	10.95	6.99	14.26
Leary's Brook	57	6.92	0.27	17.50	4.98	-6.40	18.74	0.80	0.59	1.18	11.51	5.24	14.93
Waterford River	90	8.10	-0.17	18.73	5.92	-9.27	18.74	0.56	0.40	1.05	11.10	5.95	14.90
Daily Mean Models for Regression													
Humber River	986	7.25	0.02	20.12	5.39	-16.56	23.03	2.14	1.34	3.83	11.98	8.50	19.94
Peter's River	595	8.45	-0.29	27.88	4.61	-19.08	25.46	1.14	0.88	2.27	11.22	7.13	14.96
Leary's Brook	347	6.79	0.07	19.13	4.94	-10.87	21.22	0.81	0.57	1.35	10.82	8.74	13.14
Waterford River	587	8.12	-0.19	22.46	5.92	-13.46	21.22	0.56	0.39	1.43	11.07	5.23	15.53

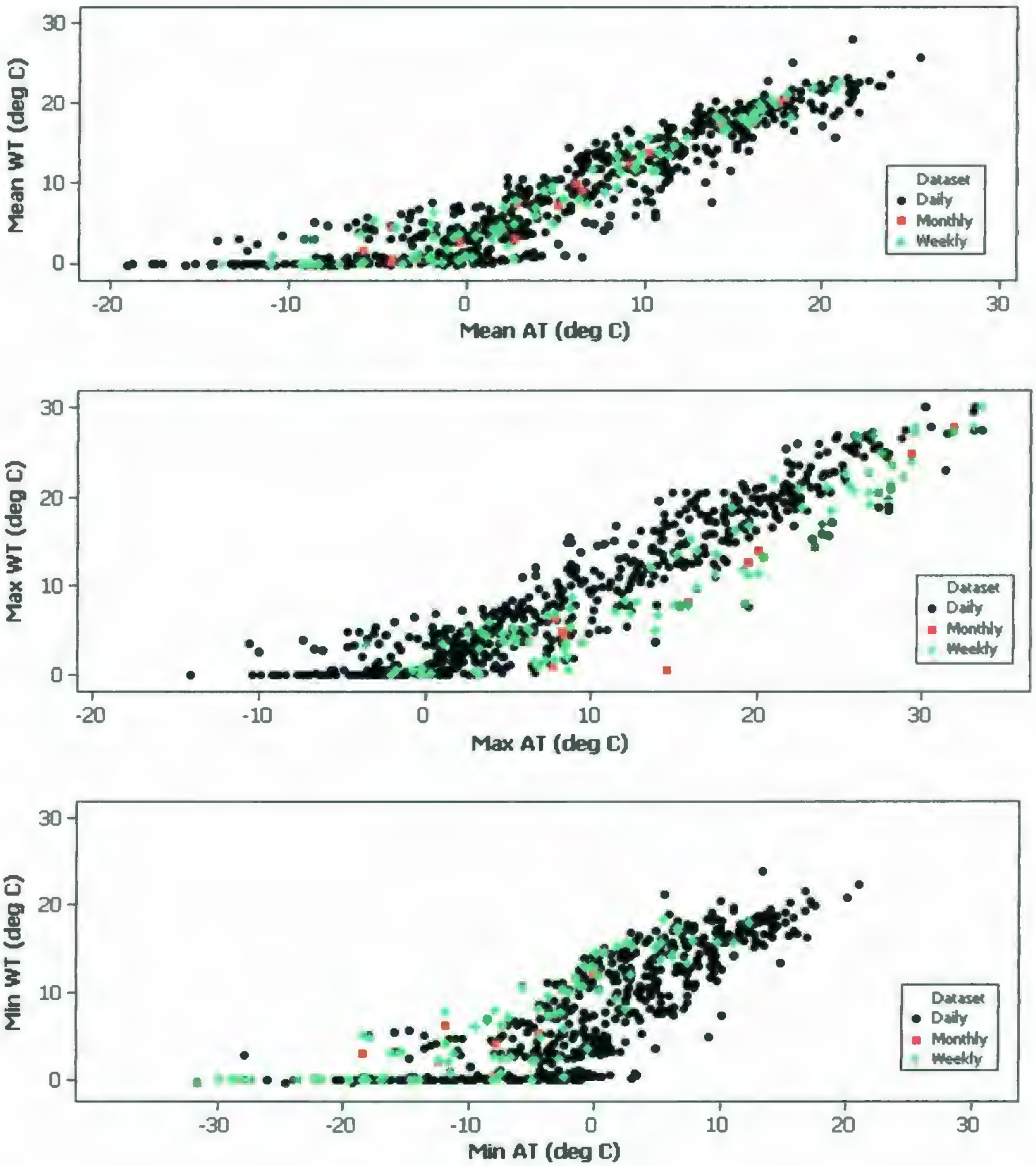


Figure 3.4 Investigations Showed Strong Water-Air temperature Relationship

(Peter's River Shown)

3.3.3 Step Three - Find the Best Fitting Models

Minitab Statistical Software (Release 14) and Datafit Curve Fitting Software (Release 8.0.32) were used together for fitting potential regression models and then determining the best fitting model overall. Minitab was used for quickly plotting the data, determining correlations between the observed data and for working with the mean, maximum and minimum datasets (i.e. finding unusual observations). Datafit was used for determining regression model parameters and determining the overall goodness of fit of the models.

The goodness of fit for the models was based on the following statistics: R^2, adjusted R^2, residual sum of squares (RSS - also referred to as SSE), and the standard error (the standard deviation of the residuals). The dimensionless measure, R^2 which is the fraction of the variance explained by regression, can be used as a dimensionless measure of fitting *y* on *x*. R^2 is calculated using equation 3.7 while RSS (or SSE) is calculated using Equation 3.8.

$$R^2 = 1 - \left(\frac{SSE}{SSy} \right) \qquad \text{Equation 3.7}$$

$$SSE = \sum_{i=1}^{n} \left[y_i - E(y_i) \right]^2 \qquad \text{Equation 3.8}$$

$$SSy = \sum_{i=1}^{n} \left(y_i - \bar{y} \right)^2 \qquad \text{Equation 3.9}$$

where $\bar{y}$ represents the mean of the response variable y and R^2 will range from 0 to 1.

Helsel and Hirsch (2002) note that the weakness in using R^2 as an indicator of the goodness of fit is that must increase, and the SSE decrease when any additional variable is added to the regression and this will happen no matter how little explanatory power that variable has. Another statistical measure available in the overall approach is the adjusted R^2 which is an R^2 value adjusted for the number of explanatory variables in the model. Adjusted R^2 is calculated as follows:

$$R_a^2 = 1 - \frac{(n-1)}{(n-p)}\frac{SSE}{SSy} = 1 - \frac{MSE}{(SSy/(n-1))}$$ Equation 3.10

The model with the highest R^2_a is identical to the one with the smallest standard error - or its square the MSE. MSE is calculated using:

$$MSE = \sqrt{\frac{\sum_{i=1}^{n}\left[y_i - E(y_i)\right]^2}{n-2}}$$ Equation 3.11

where y_i represents the value of the response variable at the ith data point, $E(y_i)$ represents the estimated value of the response variable at the ith data point, and *n* is the total number of samples being studied.

Either R^2_a should be maximized or MSE should be minimized as an overall measure of the quality of the model. Using software like Datafit to calculate all of these indicators of the goodness of fit allows the user to focus on which of the models is best for modeling the response variable being studied. Appendix E contains a brief summary note on points to remember when using curve fitting software like Datafit for seeking out the best fitting regression models. There is a tendency for some user to use the software blindly without ever checking to make sure the resulting models are both valid and meaningful.

Author's Note - one aspect of fitting regression models for measurements of water quality collected over time that is rarely mentioned in the literature is how to handle significant autocorrelation (the dependence or correlation of measurements in time) in the collected data. One of the main assumptions of regression is that the residuals are independent and any significant autocorrelation in the data will violate this assumption.

For the purposes of developing regression models for both water temperature and dissolved oxygen, the problem of high levels of autocorrelation was avoided by taking a random sample (without replacement) of the available data to interrupt any sequential time period in the data. Taking the Humber River station as an example, a random sample of the 64 monthly measurements (water temperature, air temperature, etc.) from 2003 to 2007 gives a new dataset of 194 observations in no particular time sequence.

3.4 A Note on Handling Hysteresis in the Data

Initial investigations into developing regression models for water temperature for the RTWQ stations proved to be successful for Peter's River, Leary's Brook and Waterford River but the models for Humber River were rather poor (with an adjusted R^2 values for the monthly mean first logistic model equal to 0.771).

Investigation into the Humber River data showed that a division should be made in the original dataset to account for hysteresis in the data (similar to the situation encountered in the work carried out by Mohseni et al., 1998). Water temperatures at the Humber River station tend to keep close to zero from February to July even though air temperatures during this time are rising. Figures 3.5 and 3.6 clearly shows the difference in mean water temperatures during the warming season (February to July - when water temperatures are lower) and the cooling season (August to January - when water temperatures are higher). Warming and cooling seasons were also observed in the maximum and minimum water temperature datasets.

Perhaps this hysteresis is due to snowmelt in the region which keeps water temperatures low but at this time it is not known for sure what the cause is in Humber River. To account for the hysteresis, separate models were developed for the warming season and the cooling season at the Humber River station.

The other three stations were investigated for the presence of hysteresis in the data but no clear division between a warming and cooling season could be identified for those sites and the regression datasets were kept whole as a result.

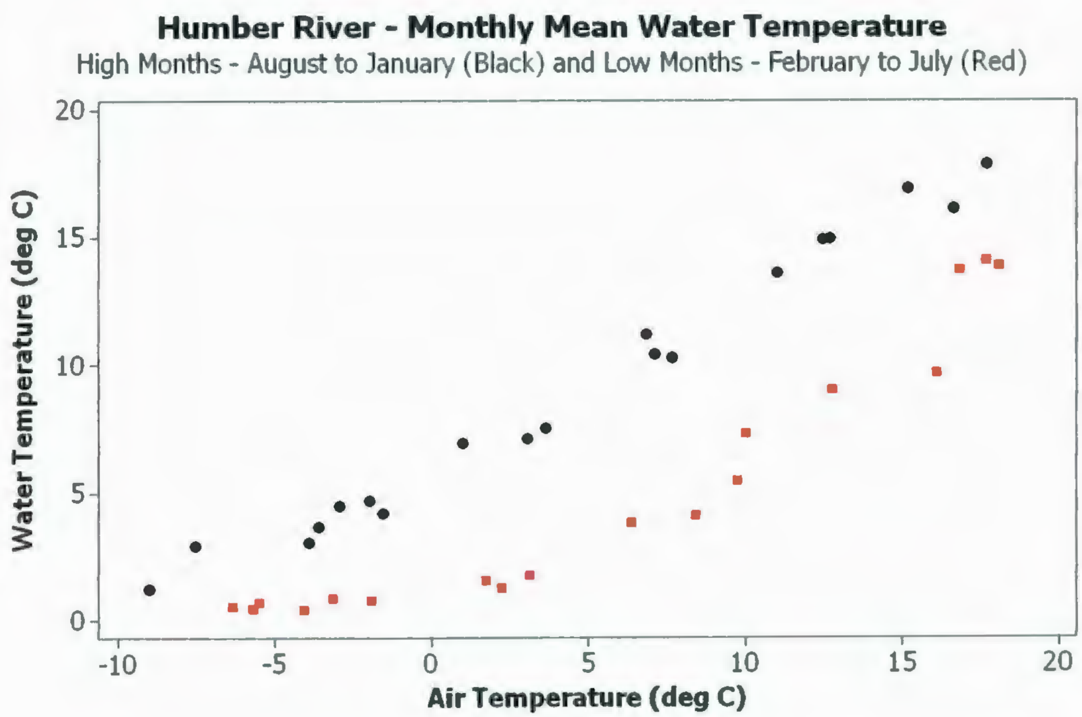


Figure 3.5 - Hysteresis in the Humber River Monthly Mean Dataset

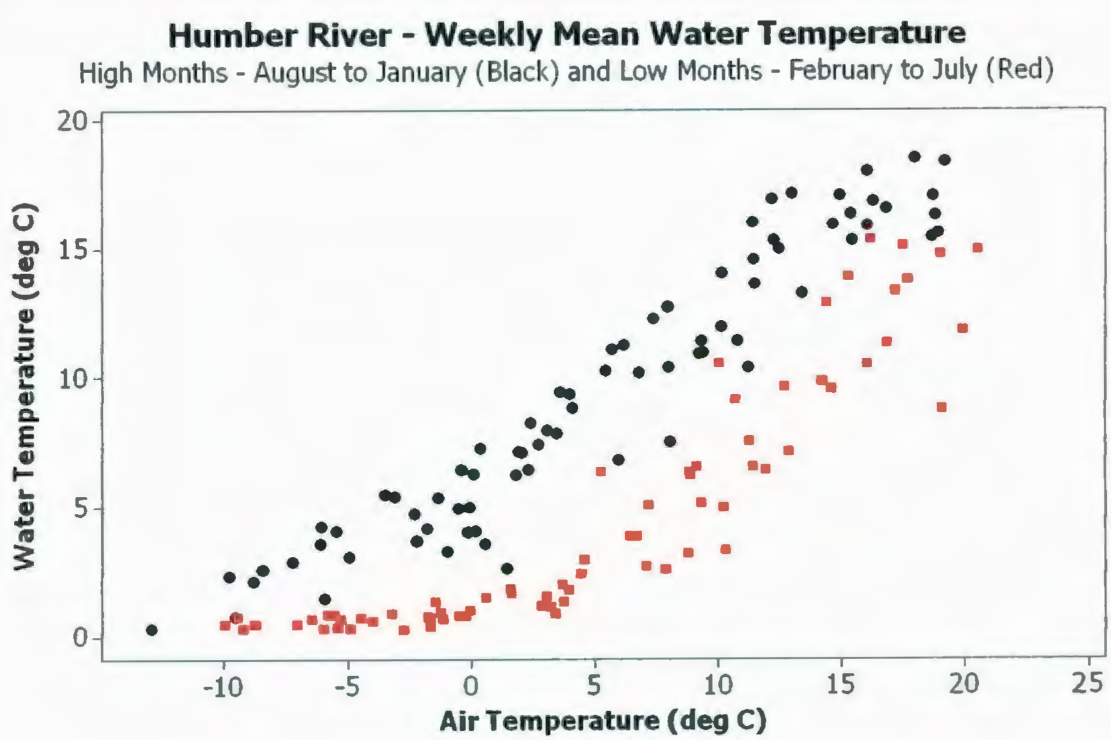


Figure 3.6 Hysteresis in the Humber River Weekly Mean Dataset

3.5 Regression Modeling Results for Water Temperature

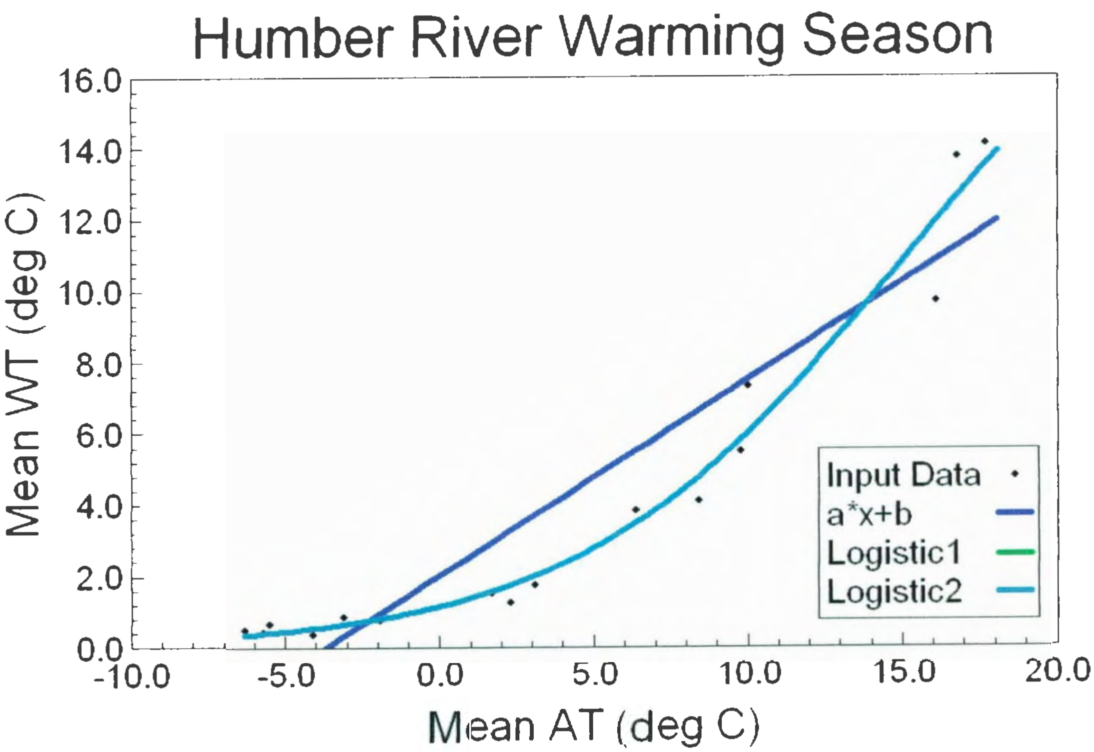
Table 3.3 presents the curve fitting results for the first logistic model. Figure 3.7 and 3.8 present the regression models for mean monthly water temperatures. The complete set of curve fitting results for the water temperature regression models (linear, logistic 1, logistic 2 and multiple regression with stage) are contained in Appendix F.

Table 3.3 First Logistic Water Temperature Regression Models

	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Humber River Cooling Season - High Water Temperature Months (August to January) *(19 Monthly Observations, 76 Weekly Observations and 500 Daily Observations)*									
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	21.41	20.31	20.50	27.87	27.54	22.06	37.65	18.26	18.85
b	0.15	0.15	0.13	0.15	0.11	0.11	0.11	0.16	0.15
c	7.20	6.56	6.78	21.25	19.74	11.57	11.11	-0.51	2.17
RSS	5.13	147.35	2150.71	69.37	361.24	2693.80	32.99	187.45	2349.29
R^2 Adj	0.99	0.93	0.84	0.88	0.84	0.81	0.90	0.89	0.82
Standard Error	0.57	1.42	2.08	2.08	2.22	2.33	1.44	1.60	2.17
Humber River Warming Season - Low Water Temperature Months (February to July) *(18 Monthly Observations, 73 Weekly Observations and 486 Daily Observations)*									
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	21.56	15.68	15.77	366.91	32.80	18.90	11.11	11.39	13.10
b	0.19	0.24	0.22	0.13	0.14	0.15	0.30	0.33	0.30
c	14.96	11.48	11.81	51.47	28.95	19.06	1.92	2.29	5.35
RSS	9.68	143.42	1911.98	12.75	384.49	2886.83	10.04	110.13	1756.17
R^2 Adj	0.97	0.91	0.83	0.98	0.81	0.77	0.95	0.90	0.83
Standard Error	0.80	1.43	1.99	0.92	2.34	2.44	0.82	1.25	1.91

Table 3.3 continued - First Logistic Water Temperature Regression Models

	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
	Peter's River								
	(23 Monthly Observations, 91 Weekly Observations, 595 Daily Observations)								
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	21.13	22.89	23.16	43.07	34.59	28.19	68.64	18.19	19.52
b	0.24	0.21	0.20	0.14	0.14	0.15	0.16	0.22	0.21
c	7.63	8.73	9.15	27.12	22.04	15.46	9.08	-1.56	3.06
RSS	31.03	237.02	3962.91	97.50	599.18	5386.46	52.19	612.91	6495.23
R^2 Adj	0.97	0.95	0.88	0.95	0.92	0.87	0.91	0.81	0.75
Standard Error	1.25	1.64	2.59	2.21	2.61	3.02	1.62	2.64	3.31
	Leary's Brook								
	(16 Monthly Observations, 57 Weekly Observations, 347 Daily Observations)								
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	18.28	18.41	19.06	1181.08	41.12	24.22	16.28	14.69	16.49
b	0.24	0.23	0.21	0.08	0.11	0.15	0.28	0.35	0.27
c	7.74	7.75	8.72	76.55	25.47	14.77	1.78	1.30	4.80
RSS	15.17	76.14	873.54	72.09	220.36	1523.39	17.57	108.29	1024.05
R^2 Adj	0.96	0.95	0.92	0.84	0.89	0.88	0.93	0.92	0.89
Standard Error	0.96	1.19	1.59	2.35	2.02	2.10	1.16	1.42	1.73
	Waterford River								
	(21 Monthly Observations, 90 Weekly Observations, 587 Daily Observations)								
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	18.47	18.95	19.56	70.24	32.34	22.82	9.90	16.34	17.44
b	0.25	0.25	0.23	0.11	0.14	0.18	2.44	0.25	0.26
c	7.41	7.77	8.36	32.13	19.94	12.24	-0.44	3.66	5.41
RSS	8.68	83.73	1596.62	77.11	433.75	2709.58	141.23	294.36	1705.31
R^2 Adj	0.99	0.98	0.94	0.93	0.92	0.91	0.95*	0.90	0.92
Standard Error	0.69	0.98	1.65	2.07	2.23	2.15	2.80	1.84	1.71



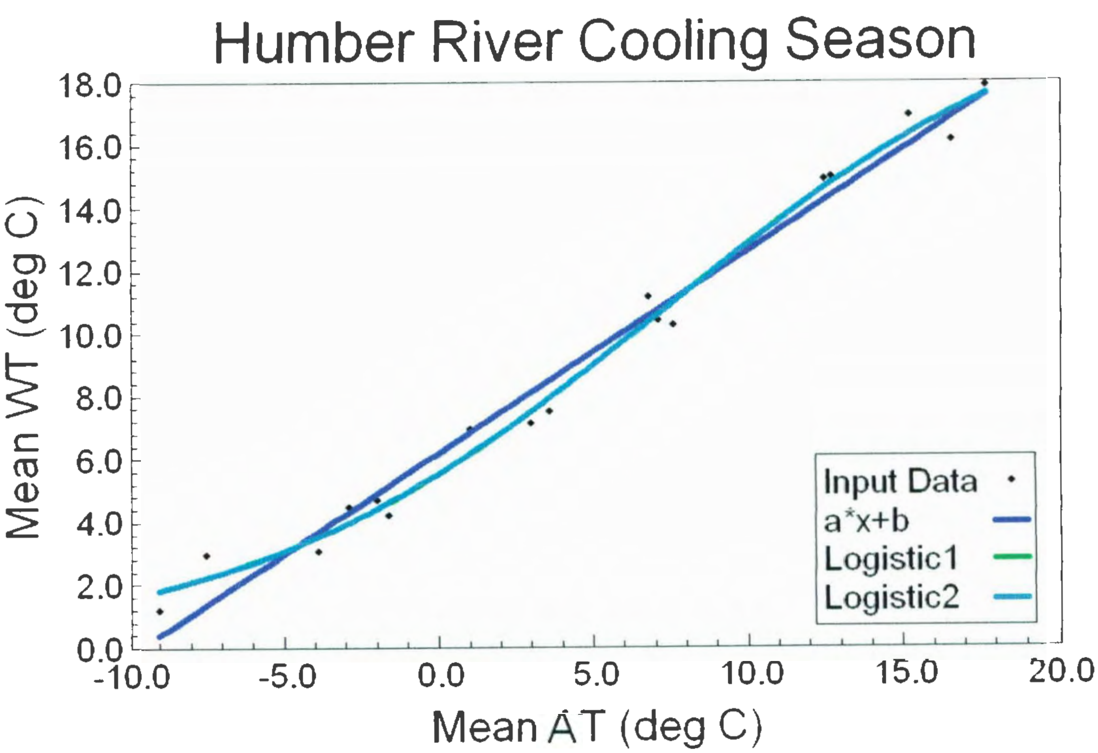


Figure 3.7 - Humber River Regression Models for Monthly Mean Water Temperature

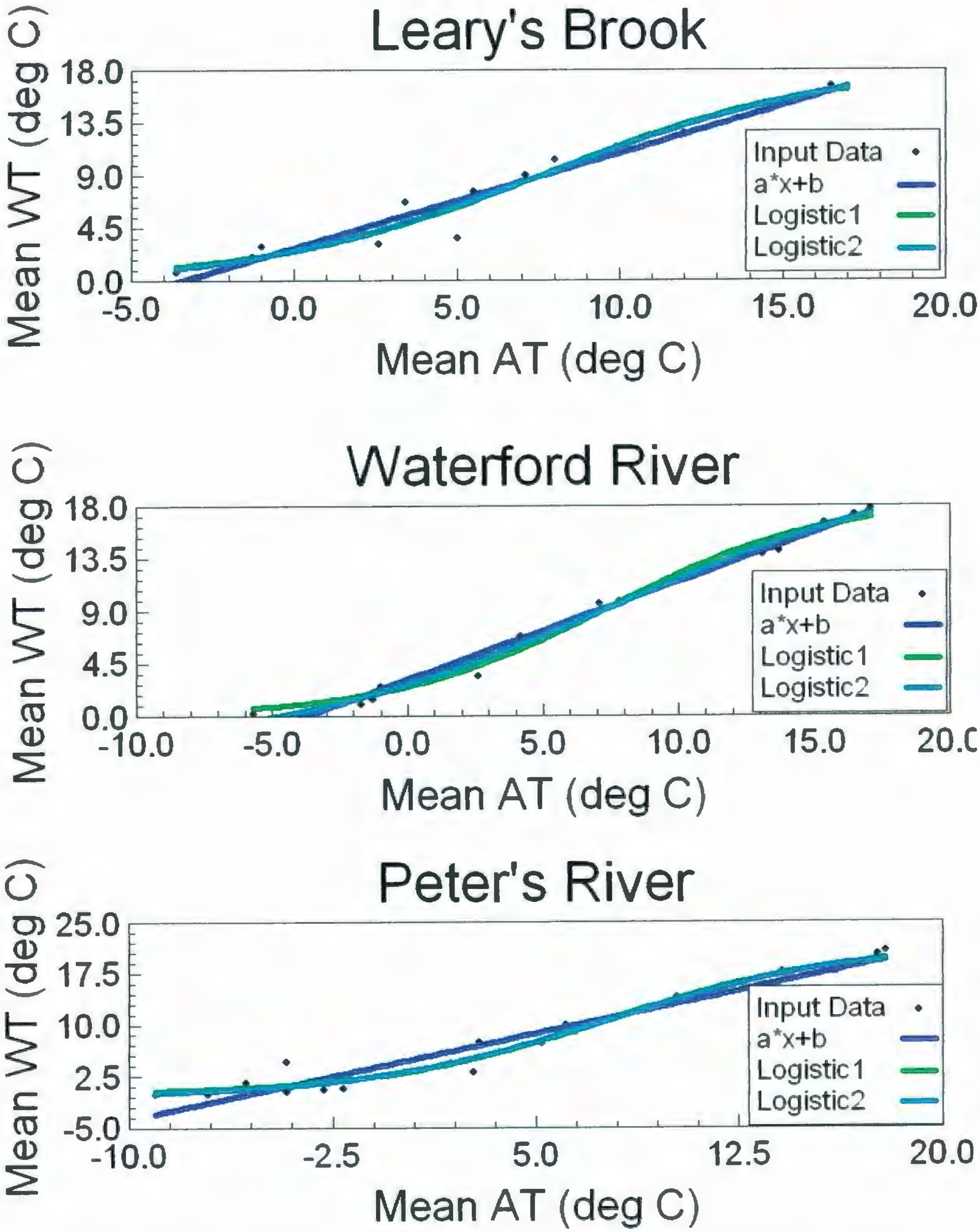


Figure 3.8 Leary's Brook, Waterford River and Peter's River Regression Models for Monthly Mean Water Temperature

3.6 Discussion

3.6.1 Linearity Versus Nonlinearity

The logistic models were found to best describe the S-shaped relationship between air temperature and water temperature at the stations. In all cases the level of explained variance of the first logistic model is equal to or higher than that of the linear model. For daily mean datasets, the first logistic function provided a significant increase in adjusted R^2 values - ranging from 3% for Waterford River to 9% for the Humber River warming season. For weekly mean datasets the increase in the level of explained variance ranges are in the range of 1 to 11%. The increase in explained variance at the monthly time scale is less significant - with no difference in the level of explained variance between the linear and first logistic model at the Waterford River station and a 1% increase at Leary's Brook and the Humber River cooling season. During the Humber River warming season though there is a significant increase in adjusted R-squared of 9%. It is interesting that when dealing with the mean water temperature datasets the logistic models often do not level off at higher water temperatures.

The first logistic model is a significant improvement over the linear model when dealing with the maximum and minimum water temperature datasets. Unlike the mean datasets, there is a bigger increase in adjusted R-squared at the monthly time scale for these datasets - where for maximum monthly water temperature the increase ranges from 5% at Peter's River to 15% for Humber River Warming season.

When working with the Waterford River monthly minimum water temperature data it was necessary to remove one outlier - September 2006 with the minimum water temperature of 0.01°C and minimum air temperature of 5.8°C (indicated in Table 3.3 with the asterisk). On the date of September 12, 2006 the water temperature in the historical drift corrected dataset drops from 13.92°C to 0.01°C and then within an hour rises back up to 12.09°C. There is no record of this drop on the maintenance forms for the station. It is highly likely that the drop is the result of communication problems with the monitoring equipment and it should not be considered an accurate measurement of water temperature. Once this outlier is removed from the dataset the adjusted R^2 for the first logistic model value for monthly minimum water temperature is 0.95 (up from 0.72 when the outlier was included in he original dataset).

From working with the data it was noted that the second logistic model never provides a substantial improvement in residual sum of squares or the amount of explained variation over the first logistic model. As a result, for the sake of prediction purposes for rivers in the provincial RTWQ network it is better to use the simpler first logistic model when modeling water temperature.

3.6.2 The Influence of Time-scale

The strength of the water temperature and air temperature relationship was strongest at all stations as the time scale was extended from daily mean observations to monthly mean observations. Figure 3.9 presents a comparison plot of the decrease in adjusted R-squared values as time scale is extended in the mean water temperature datasets.

Excessive scatter in the daily mean observations kept adjusted R-squared values lower and residual sum of square values for the daily first logistic models were rather high (from 873.54 for Leary's Brook to 3962.91 at Peter's River where at the monthly time scale it ranged from 5.13 for Humber River cooling season to 31.03 at Peter's River). Scatter at the weekly time scales is not as high as at the daily time scale and adjusted R-squared values for the first logistic models are still quite higher. Increased scatter in the smaller time scale datasets drives the residual sum of squares term quite high and also increases the standard error term (as shown in Figure 3.10). It is often easier to see the S-shaped relationships in the data at the weekly time scale than it is at the monthly time scale (where both models tend to fit equally well).

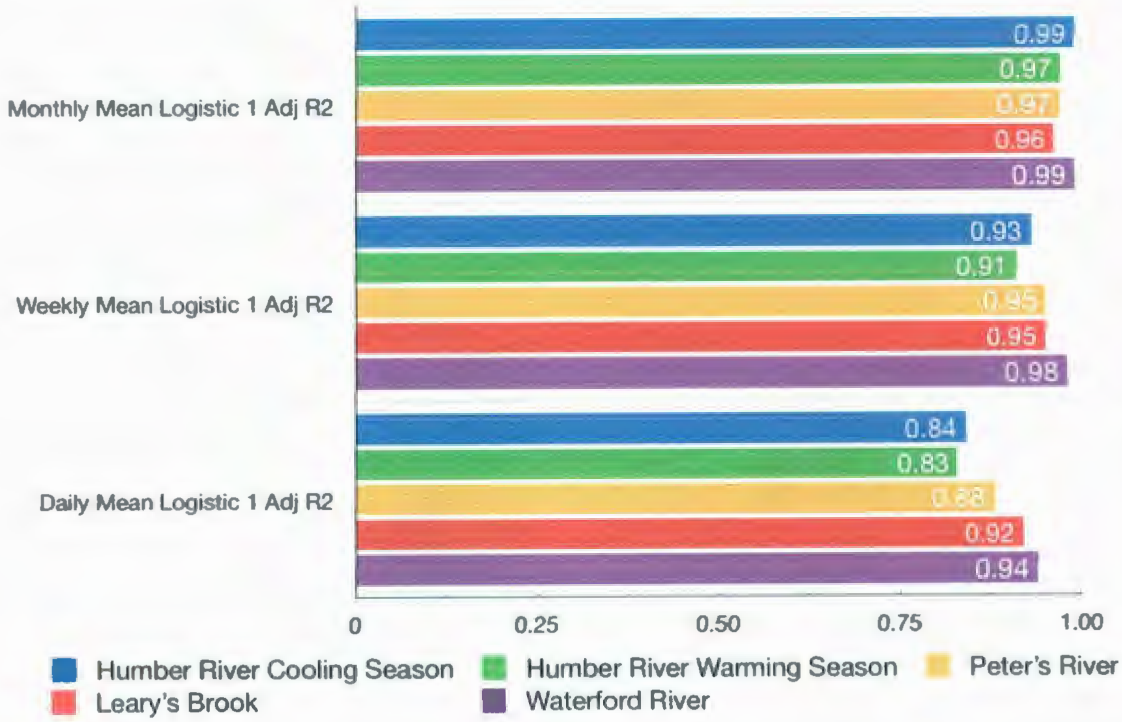


Figure 3.9 Comparing Adjusted R^2 for Mean Water Temperature Datasets

Figure 3.10 Standard Error Gets Larger as the Time Scale is Shortened

An important aspect of checking the validity of the models involves examining residual plots developed by Datafit (or Minitab). The residual plots for the monthly models presented in this thesis were all valid for checks of residual normality and constant variance. The residual plots of the weekly models were adequate for the normality assumption but an interesting pattern began to emerge on the variance plot - where at lower air temperatures there is smaller scatter in the residuals than at the higher air temperatures. For the most part, the lack of constant variance for the weekly models is slight but is even more visible in the residual plots developed for the daily models. There is no way to modify the daily models to get constant variance in these plots.

Residual scatter is smaller at the lower air temperatures as water temperature cannot drop far below 0 °C - i.e. scatter is more one sided. Water temperatures can take on a greater ranger of values at the higher air temperatures - i.e. scatter is two sided. For the purposes of developing regression models for prediction it is best that all the assumptions are valid. There is already a great deal of scatter in the daily models and the residual plots for the daily models are an indicator that the daily models will likely be not useful for accurately predicting water temperatures at the stations.

Figure 3.11 presents a comparison of the different residual plots for the Waterford River first logistic model for the warming season. It can be noted that at the daily timescale there is a distinct reverse funnel shape to the residual scatter, indicating variance in the data is larger at the higher air temperatures. Even though the daily models do violate the assumption of constant variance they have still been tested for prediction purposes later in this chapter.

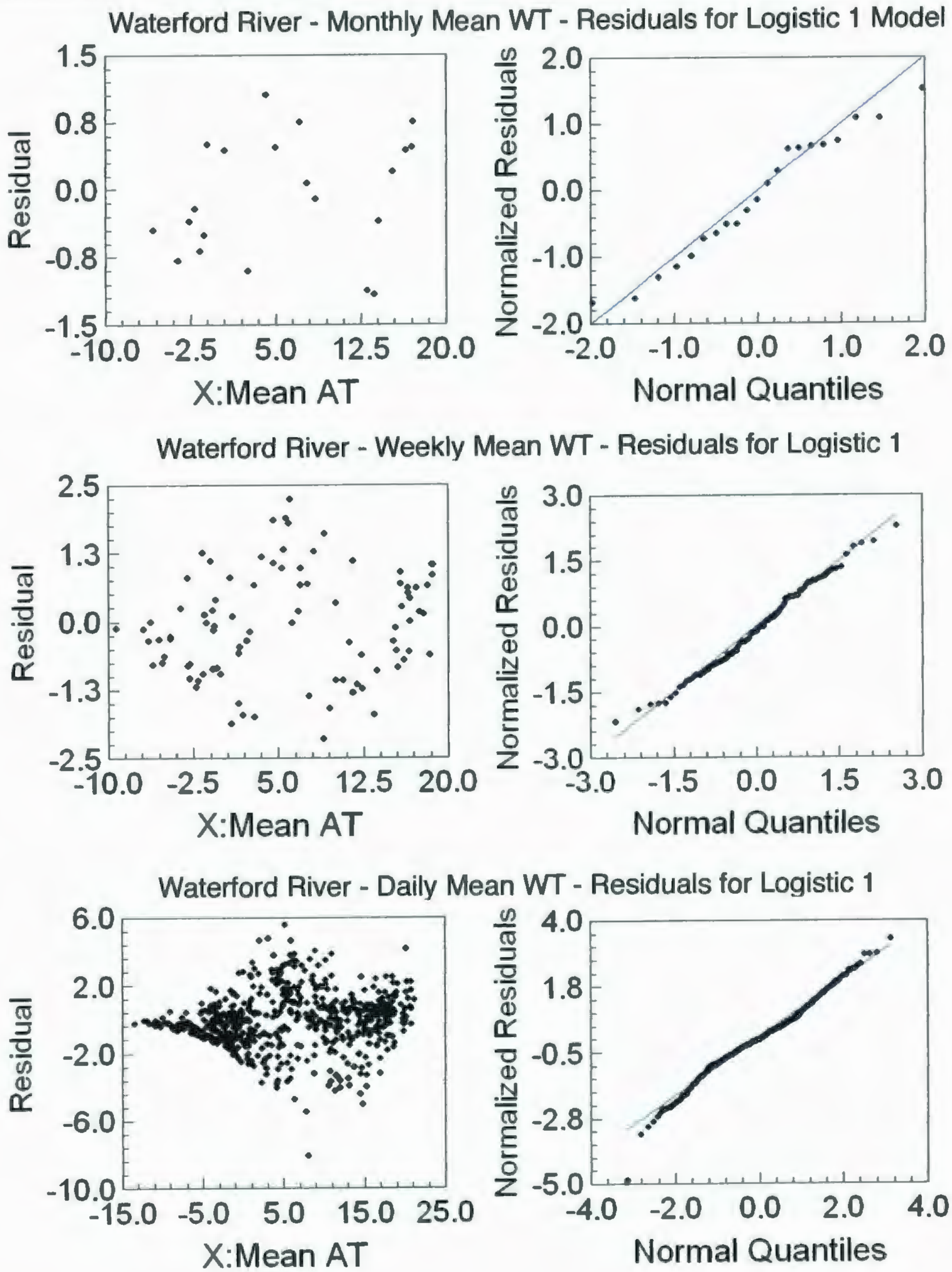


Figure 3.11 - Time Scale Residual Plots for the Waterford River Monthly, Weekly and Daily Mean Water Temperature Models

3.6.3 The Influence of Stage

The combined influence of stage and air temperature on water temperature was investigated through the use of multiple regression analysis. For the Humber River cooling season and Peter's River, stage was not a significant explanatory variable at the weekly and monthly time scales. Although stage is often a significant explanatory variable at the daily time scale, the multiple regression model never provides a significant improvement over the first logistic model. The loss of importance of stage at the extended time scales is a similar result as that found when other models developed models using streamflow at varying timescales (Crisp and Howson, 1982; Webb, 1987). Stage is never a significant explanatory variable at the Leary's Brook station.

Stage is an important explanatory variable at all time scales for the Humber River station during the warming season and at the Waterford River station. Adjusted R^2 values for the multiple regression for the mean water temperature datasets are quite high at the Waterford River station (ranging from 0.93 daily to 0.99 monthly) but the multiple regression model does not provide a better fit than the logistic models - refer to Table 3.4 for a comparison. Figure 3.12 presents the multiple regression model with stage and air temperature for daily mean water temperature for the Humber River warming and cooling seasons.

Table 3.4 - MLR Modeling Results for Water Temperature Models

	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Humber River Cooling Season - High Water Temperature Months (August to January)									
(19 Monthly Observations, 76 Weekly Observations and 500 Daily Observations)									
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	NS	NS	8.24	NS	NS	7.06	NS	NS	10.00
b			0.54			0.50			0.55
c			-0.76			-0.83			-0.80
RSS			2291.78			2755.01			2568.66
R^2 Adj			0.83			0.81			0.80
Humber River Warming Season - Low Water Temperature Months (February to July)									
(18 Monthly Observations, 73 Weekly Observations and 486 Daily Observations)									
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	7.88	6.38	5.98	0.75	3.75	4.84	9.13	8.86	7.42
b	0.57	0.53	0.50	0.72	0.54	0.46	0.33	0.39	0.47
c	-2.81	-2.02	-1.77	-2.57	-2.57	-1.96	-2.34	-1.92	-1.55
RSS	13.66	212.09	2340.63	44.84	374.05	2882.28	41.13	226.80	2709.59
R^2 Adj	0.96	0.87	0.80	0.92	0.81	0.77	0.80	0.80	0.74
Peter's River									
(23 Monthly Observations, 91 Weekly Observations, 595 Daily Observations)									
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	NS	NS	8.58	NS	NS	6.85	NS	21.23	12.78
b			0.70			0.74		0.45	0.54
c			-2.97			-3.63		-10.84	-4.24
RSS			5752.56			6167.02		849.88	8662.33
R^2 Adj			0.82			0.85		0.74	0.67

Table 3.4 continued - MLR Modeling Results for Water Temperature Models

	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Leary's Brook									
(16 Monthly Observations, 57 Weekly Observations, 347 Daily Observations)									
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	NS	NS	NS	NS	NS	NS	NS	NS	NS
b									
c									
RSS									
R^2 Adj									
Waterford River									
(21 Monthly Observations, 90 Weekly Observations, 587 Daily Observations)									
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	6.33	6.38	7.13	0.20	2.28	5.44	NS	10.46	9.26
b	0.79	0.77	0.73	0.92	0.82	0.73		0.60	0.67
c	-5.24	-5.06	-5.93	-2.80	-3.17	-4.76		-8.98	-7.75
RSS	5.04	108.07	1766.77	125.12	485.64	2498.05		391.20	2324.47
R^2 Adj	0.99	0.97	0.93	0.89	0.91	0.92		0.87	0.89

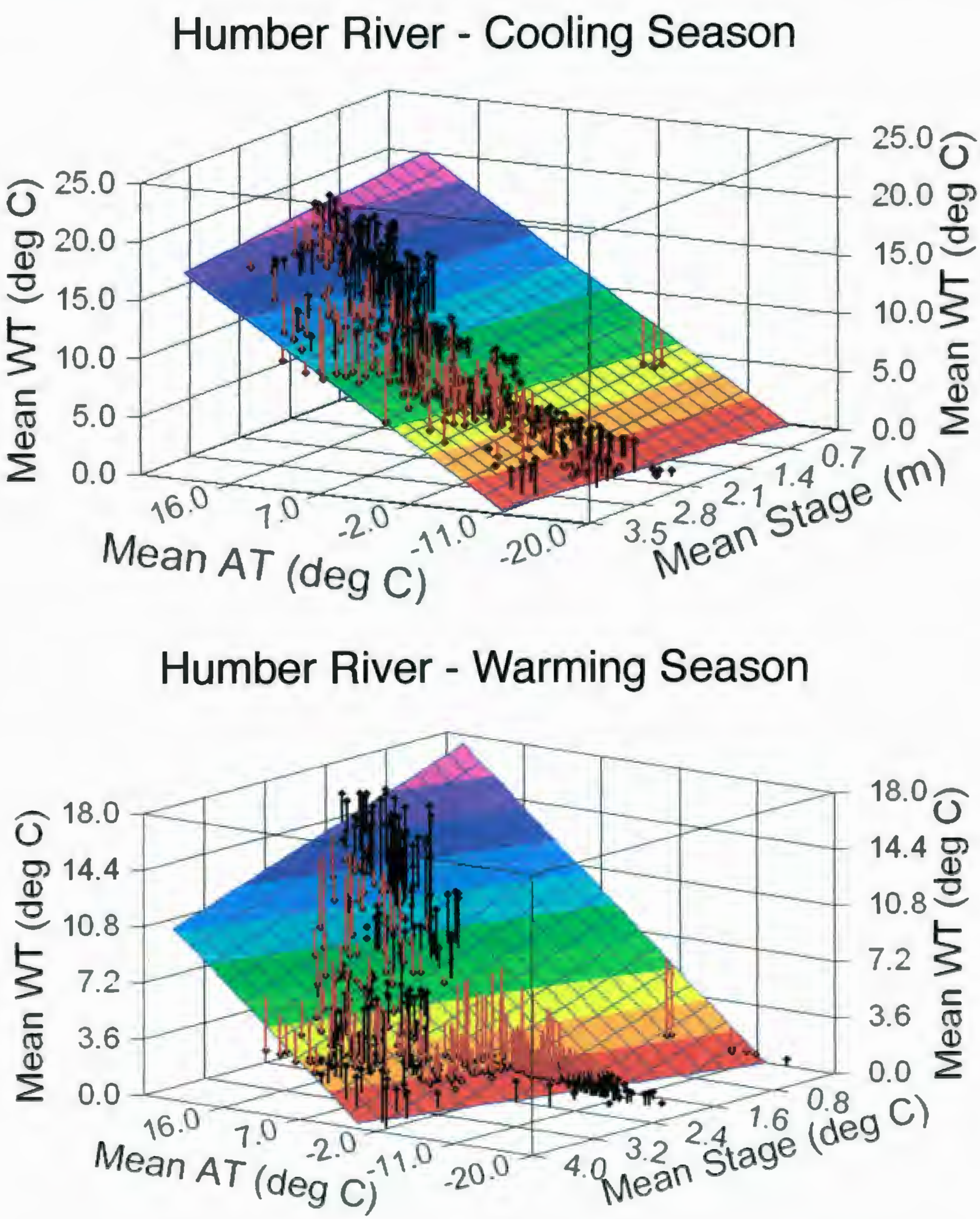


Figure 3.12 - Humber River Daily Mean Water Temperature Models With Stage and Air Temperature

3.6.4 An Alternative Approach to Handling Hysteresis in the Data

An alternative approach to developing separate regression models for the warming and cooling seasons at the Humber River station is to add an explanatory variable to the regression models that accounts for the time of the year the sample of water quality was taken. Curve fitting results for these alternative models were quite similar to the regression models developed for the warming and cooling seasons (with adjusted R^2 equal to 0.946, 0.904 and 0812 for monthly, weekly and daily mean water temperature modified logistic 1 models). The curve fitting results for this alternative approach can be found in Appendix G.

3.6.5 Using the Best Models for Prediction

The first logistic model was deemed to be the best option for modeling mean, maximum and minimum water temperature at the daily, weekly and monthly time scales. The first logistic model always has an adjusted R-squared value greater than or equal to that of the linear model and the residual sum of squares term of the first logistic model is generally lower than that of the linear model. The curve fitting results showed that there is little benefit in using the more complex second logistic model as the first logistic model can achieve comparable adjusted R-squared values. Stage is rarely a significant explanatory variable, and when it is significant the multiple regression model does not outperform the first logistic model. Although the first logistic model can be considered the best overall option for modeling, it is not advisable to develop one general logistic model for the stations as the models developed for each station are unique. A comparison between first logistic mean water temperature models is shown in Figure 3.13. This figure shows the two different curves for Humber River (warming and cooling), the steeper Peter's River curve, and the close similarity between the Waterford River and Leary's Brook (the two smallest rivers in the provincial network).

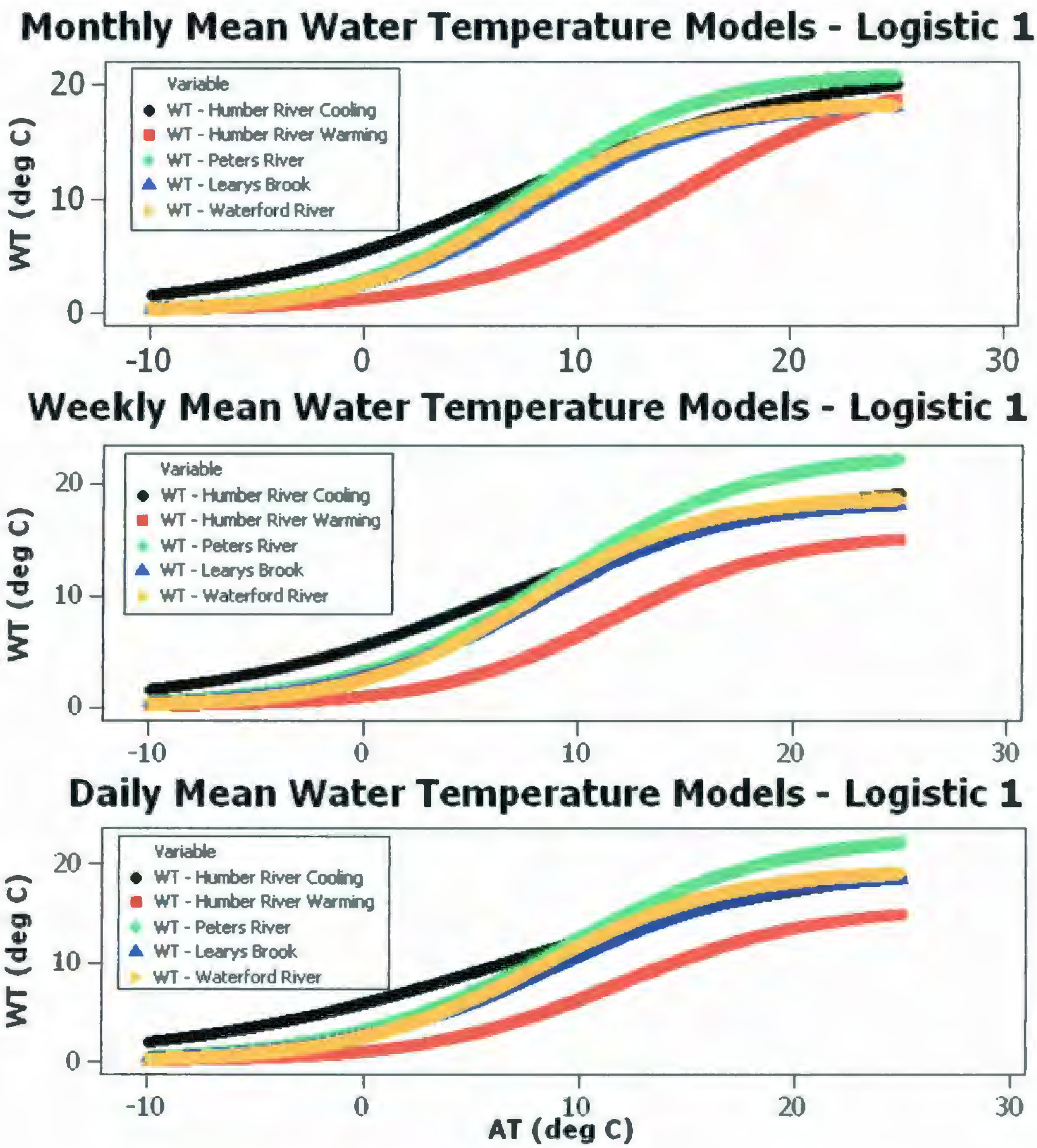


Figure 3.13 A Comparison of First Logistic Mean Water Temperature Models

The datasets reserved for prediction purposes were used to test the capability of the first logistic models for predicting water temperature at the real time stations. Model testing results for the mean water temperature models are shown in Table 3.5. When scatterplots of the observed versus predicted values were developed it was noted that there is less scatter in the monthly data and as a result the monthly models tend to perform better for predicting water temperature. One of these scatterplots is attached as Figure 3.14. Increased scatter at the weekly and daily time scales makes it more difficult to be as exact in prediction. It should be noted that there is a gap in the available observed the air temperature dataset for the Humber River - where there are no monthly mean measurements in the 5 to 13 °C range. This is solely the result of there being a limited dataset available for prediction purposes. As more air temperature and water temperature measurements are made available in the future it will be possible to test the accuracy of the developed water temperature models for the Humber River station for the full range of air temperatures that are experienced at the station.

The tables show both the absolute value of the difference between observed and predicted values and the absolute value of the percent error of the predictions. It should be noted that percent error has been calculated by subtracting the observed value from the predicted and then dividing by the predicted value. For low air temperatures the regression models will predict low water temperatures and sometimes these low water temperatures will inflate the size of the percent error - i.e. if the observed value of water temperature is 0.80 °C and the logistic model predicts a value of 0.10 °C then the percent error will be 700% even though the difference between the values is only 0.70 °C. *Author's Note - the complete testing results for the maximum and minimum water temperature models can be found in Appendix H.*

Table 3.5 Using the Logistic 1 Model for Predicting Water Temperature

		Absolute Value of Difference Abs[Pred - Obs]				**Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%**			
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**	**St Dev**
					Humber River				
		Cooling Season Monthly Mean WT = 21.41/(1+exp(0.15(7.20- Mean AT)))*							
0.989	5	0.85	1.61	0.29	0.67	19.9	59.8	1.86	22.91
		Warming Season Monthly Mean WT = 21.56/(1+exp(0.19(14.96- Mean AT)))*							
0.975	9	0.25	0.54	0.003	0.2	26.6	78.9	0.03	25.95
		Cooling Season Weekly Mean WT = 20.31/(1+exp(0.15(6.56- Mean AT)))*							
0.93	25	1.14	3.25	0.44	0.67	33.46	111.4	2.6	27.75
		Warming Season Weekly Mean WT = 15.68/(1+exp(0.24(11.48- Mean AT)))*							
0.91	38	0.634	4.66	0.011	0.8	86.24	557.92	1.73	122.31
		Cooling Season Daily Mean WT = 20.50(1+exp(0.13(6.79- Mean AT)))*							
0.84	130	1.786	5.53	0.012	1.35	37.92	126.3	0.34	28.59
		Warming Season Daily Mean WT = 15.78(1+exp(0.22(11.81 - Mean AT)))*							
0.83	268	1.04	7.77	3.82	1.43	79.39	837.4	0.02	125.87

Table 3.5 continued - Using the Logistic 1 Model for Predicting Water Temperature

		Absolute Value of Difference Abs[Pred - Obs]				Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%			
Adj R^2	Obs.	Mean	Max	Min	St Dev	Mean	Max	Min	St Dev
				Leary's Brook					
				Monthly Mean WT = 18.28/(1+exp(0.24(7.74- Mean AT)))*					
0.96	8	1.168	2.682	0.191	0.83	26.687	131.3	1.422	43.23
				Weekly Mean WT = 18.41/(1+exp(0.23(7.75- Mean AT)))*					
0.95	23	0.843	2.34	0.014	0.73	11.15	37.5	0.1	12.55
				Daily Mean WT = 19.06/(1+exp(0.21(8.72- Mean AT)))*					
0.92	136	1.639	5.01	0.01	1.19	25.54	140.3	0.14	28.31
				Waterford River					
				Monthly Mean WT = 18.47/(1+exp(0.25(7.41- Mean AT)))*					
0.99	12	0.497	1.244	0.026	0.36	14.45	39.62	0.191	12.03
				Weekly Mean WT = 18.95/(1+exp(0.25(7.77- Mean AT)))*					
0.98	48	0.803	2.49	0.03	0.61	22.09	138	0.4	26.09
				Daily Mean WT = 19.56/(1+exp(0.23(8.36- Mean AT)))*					
0.94	306	1.36	6.87	0.01	1.22	32.4	204.4	0.05	33.61
				Peter's River					
				Monthly Mean WT = 21.13/(1+exp(0.24(7.63- Mean AT)))*					
0.97	8	0.874	3.202	0.044	1.03	57.63	144	0.239	66.15
				Weekly Mean WT = 22.89/(1+exp(0.21(8.73- Mean AT)))*					
0.95	32	1.213	6.61	0.021	1.42	63.42	252.5	0.27	69.03
				Daily Mean WT = 23.16/(1+exp(0.20(9.15- Mean AT)))*					
0.88	196	1.923	10.8	0.024	1.82	78.41	662.94	0.12	98.42

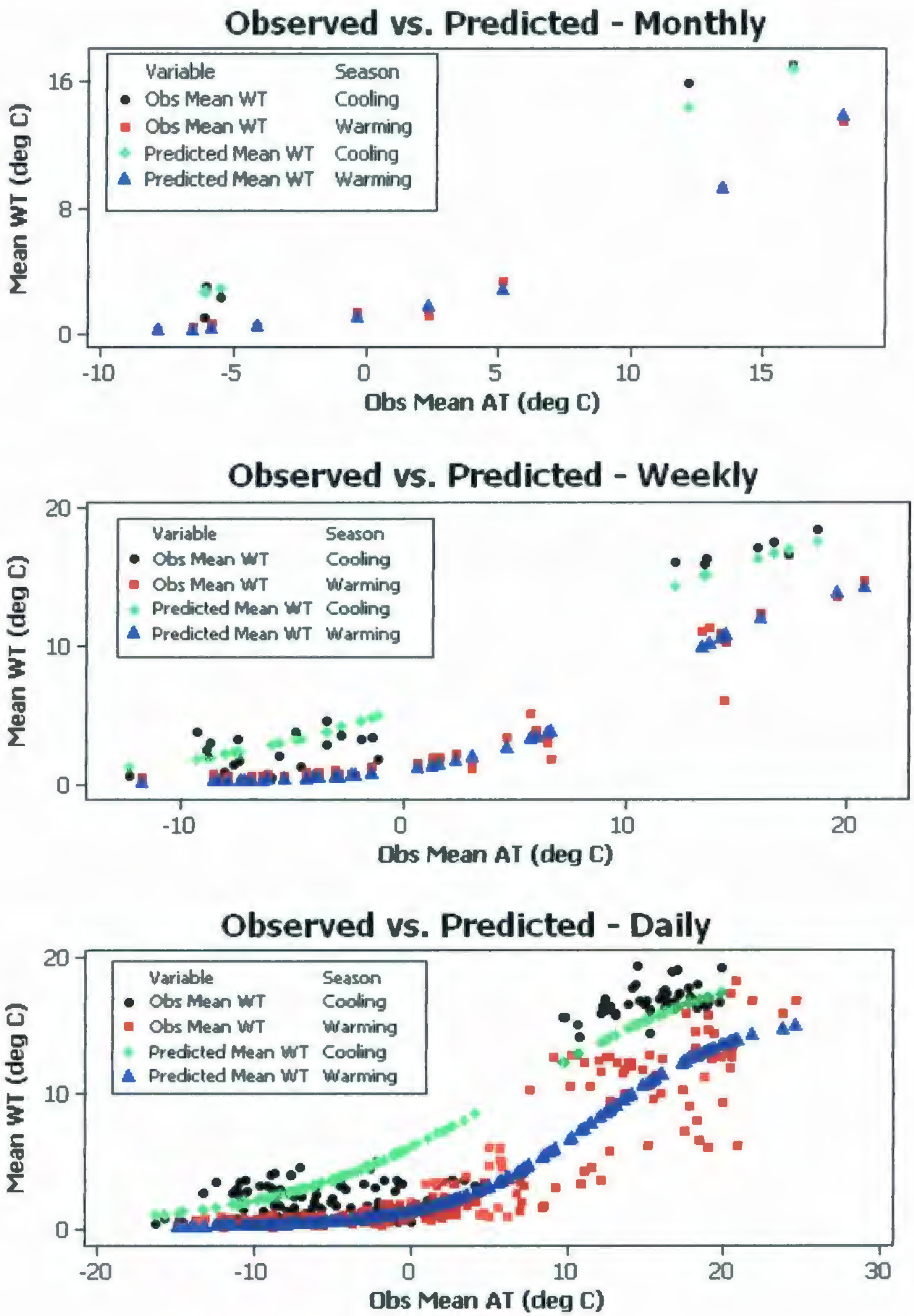


Figure 3.14 Scatterplot of Observed and Predicted Humber River Mean WT

Chapter Four

Development of Dissolved Oxygen Regression Models

4.1 Scope

Regression models for predicting dissolved oxygen at the real time stations are the main focus of this fourth chapter. A brief literature review of regression models for dissolved oxygen is first presented. The curve fitting results for dissolved oxygen models will be presented in detail. A unique curve plotting approach for relating air temperature to dissolved oxygen levels is presented. The chapter concludes with a discussion of results and a look at ways to make carrying out this kind of research in the future easier.

4.2 Literature Review

Dissolved oxygen levels in rivers and streams are known to be influenced by a number of factors such as oxygen consumption by aquatic organisms, depth of water and water temperature, where rising water temperatures cause dissolved oxygen levels to decrease. Most researchers have focused their regression modeling efforts for dissolved oxygen using measurements of water temperature.

Saffran and Anderson (1996) examined the linear relationship between minimum dissolved oxygen and maximum water temperatures for two monitoring sites along the Red Deer River in Alberta. From their research they found that during the summer months there was good correlation between dissolved oxygen and water temperature. There was good potential for developing regression models for predicting dissolved oxygen based on water temperature. Minimum dissolved oxygen levels in the river were negatively correlated with air temperature and water temperature, where the strongest correlation for dissolved oxygen was with maximum water temperature. Saffran and Anderson note that the relationship between minimum dissolved oxygen and maximum air temperature was linearly related but they lacked sufficient data at the time to fully examine the relationship and develop accurate regression models.

Although the focus in Morrill et al. (2005) was on developing regression models for predicting water temperature using air temperature measures, the authors also examined the impact of increasing water temperatures on dissolved oxygen levels in streams where dissolved oxygen levels were already close to critically low levels for many species Future dissolved oxygen levels are not predicted using regression models but were calculated by subtracting the monthly mean dissolved oxygen deficit in the streams from the mean saturated value. The high and low stream temperatures were then used to calculate the upper and lower saturated levels which were then used for calculating the stream dissolved oxygen levels.

Most of the published literature dealing with predicting dissolved oxygen levels goes outside of developing regression models into the realm of more statistically complex models. Rounds (2002) developed an artificial neural network model to predict dissolved oxygen concentrations in a river using air temperature, solar radiation, rainfall and streamflow as inputs. Gelda et al. (2001) develop a dynamic two-dimensional mass balance model for dissolved oxygen levels for rivers. Abdul-Aziz et al. (2007) use an extended stochastic harmonic analysis algorithm approach for predicting dissolved oxygen levels. In order for these more complex models to be effective in prediction they usually require a large number of inputs, and obtaining the necessary data for these inputs can be quite difficult. In this research, the aim was to determine if the available data being collected by the RTWQ network could be directly used for estimating dissolved oxygen without having to rely on a statistically complex model with a large number of inputs. In the same way that air temperature was used for predicting water temperature through regression modeling, it was hoped that the same regression modeling approach would work for predicting dissolved oxygen.

4.3 Methodology Used for Developing Dissolved Oxygen Models

The same methodology used for developing regression models for water temperature was used for developing models for dissolved oxygen.

4.3.1 Step One - Get Familiar With the Most Commonly Used Models

For this study three different regression models for dissolved oxygen were studied: simple linear regression (using water temperature - Equation 4.1) multiple regression (using water temperature and stage - Equation 4.2) and a nonlinear exponential decay model (using water temperature - Equation 4.3).

$$DO = a_0 + a_1 Tw + \varepsilon \qquad \text{Equation 4.1}$$

$$DO = \exp(a_0 + b_1 Tw) \qquad \text{Equation 4.2}$$

$$DO = a_0 + a_1 Tw + a_2 ST + \varepsilon \qquad \text{Equation 4.3}$$

4.3.2 Step Two - Develop the Necessary Datasets

The Minitab macros used for obtaining datasets for the water temperature regression models were written to also find mean, maximum and minimum dissolved oxygen values at the monthly, weekly and daily time scales. The same periods of time used for defining the model development and model prediction datasets for the water temperature datasets were used for the dissolved oxygen datasets. Table 3.2 in the previous chapter presented a general statistical summary of the datasets for developing dissolved oxygen regression models using mean monthly, weekly and daily data. A more detailed statistical summary of the datasets can be found in Appendix I.

Initial investigations into the correlation between dissolved oxygen, water temperature and stage were carried out once the regression datasets were developed. The relationship between dissolved oxygen and water temperature was found to be strongly negative (i.e. as water temperatures increased the dissolved oxygen levels decreased). The relationship between dissolved oxygen and stage was found to be positive (i.e. dissolved oxygen levels tended to be higher when the stage level was higher). Due to the negative correlation between water temperature and dissolved oxygen and the positive correlation between stage and dissolved oxygen the following datasets were investigated: (1) mean dissolved oxygen, stage and water temperature, (2) minimum dissolved oxygen, minimum stage and maximum water temperature and (3) maximum dissolved oxygen, maximum stage and minimum water temperature. Scatterplots of water temperature and dissolved oxygen showed the strong negative relationship between the two parameters. Figure 4.1 presents one of these scatterplots for the Humber River station. Plots for the other stations can be found in Appendix J.

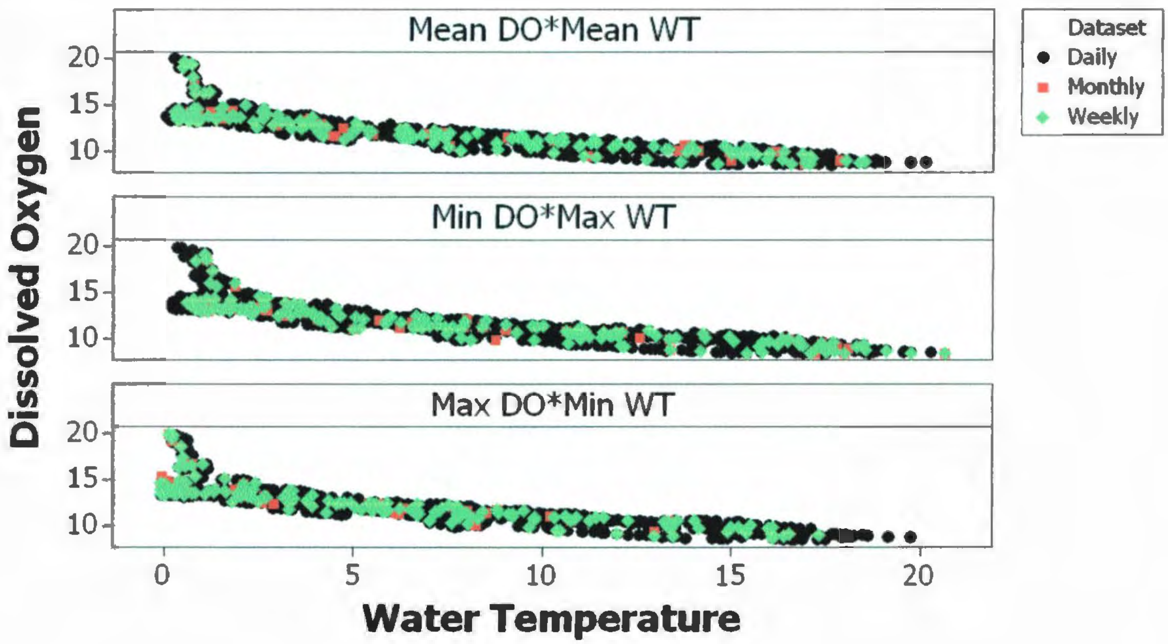


Figure 4.1 Humber River Dissolved Oxygen-Water Temperature Relationship Before the Removal of Unusually High Observations

Correcting Issues with the Dissolved Oxygen Datasets

The Hydrolab Datasonde sensor has the capability to detect dissolved oxygen levels in the range of 0 to 50 mg/L but analysis of the developed dissolved oxygen datasets showed that measurements recorded by the sensor do go outside of this range. There might be a number of reasons as to why dissolved oxygen values might go outside the measurement range - including, but not limited to, malfunction of the dissolved oxygen sensor probe, communication errors between the sensor and the data logger, and calibration error. Values of dissolved oxygen outside of the 0 to 50 mg/L range were removed from the dataset and were not considered in analysis. The same problem was not observed for the other measured parameters during the same problematic dissolved oxygen periods. As a result it was not necessary to remove these other parameters from the dataset used for analysis.

Initial explorations with the Humber River dissolved oxygen dataset showed that at lower water temperatures there was a wide range of recorded daily mean dissolved oxygen level (refer to Figure 4.2). The scatterplot shows that when air temperatures drop below 2.5 °C the dissolved oxygen levels will range from 12.5 to 20 mg/L. Although those levels are not high enough to interfere with aquatic health, the variation in the data posed a problem for regression modeling.

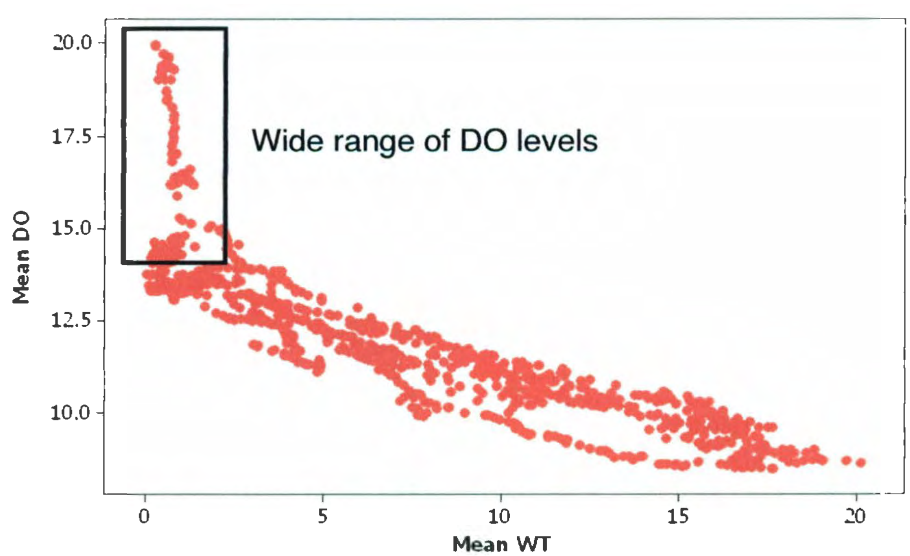


Figure 4.2 Unusually High DO Levels at Low WT in the Humber River Daily Mean Dataset

When this range of values was investigated it was determined that the majority of the higher observations were recorded during a stretch of Datasonde readings dating from February 16, 2006 to April 11, 2006. The drift corrected hourly observations for the period showed that until January 6, dissolved oxygen levels were around 13 mg/L. The sensor was taken offline until February 16 and when it came back online dissolved oxygen levels were some of the highest they had been since the sensor was first installed at the station. The sensor was taken offline again on March 17 when the levels were still high (Figure 4.3). Deployment records for the station showed that the high levels were not due to changes in the physical conditions in the river but were due to sensor malfunction. During this period in time the default time delay for the Datasonde was not allowing the sensor to warm-up enough to accurately read the dissolved oxygen concentrations. A field visit on April 11 allowed WRMD personnel to reset the time delay and after that period dissolved oxygen values return to normal. Once the high measurements were removed from the Dataset, the relationship between water temperature and dissolved oxygen station looked to be more reasonable (Figure 4.4). At the daily time scale there still was a considerable amount of scatter in the range of the dissolved oxygen measurements.

There were no sets of unusually large dissolved oxygen measurements recorded at the other stations and only those measurements outside of the 0 to 50 mg/L range were removed for regression modeling. Although there does tend to be a fair amount of scatter in the dissolved oxygen datasets (i.e. at a mean water temperature of 15 °C dissolved oxygen might be anywhere from 8 to 13 mg/L) no warming and cooling seasons could be determined and regression was carried out using datasets without seasonal division.

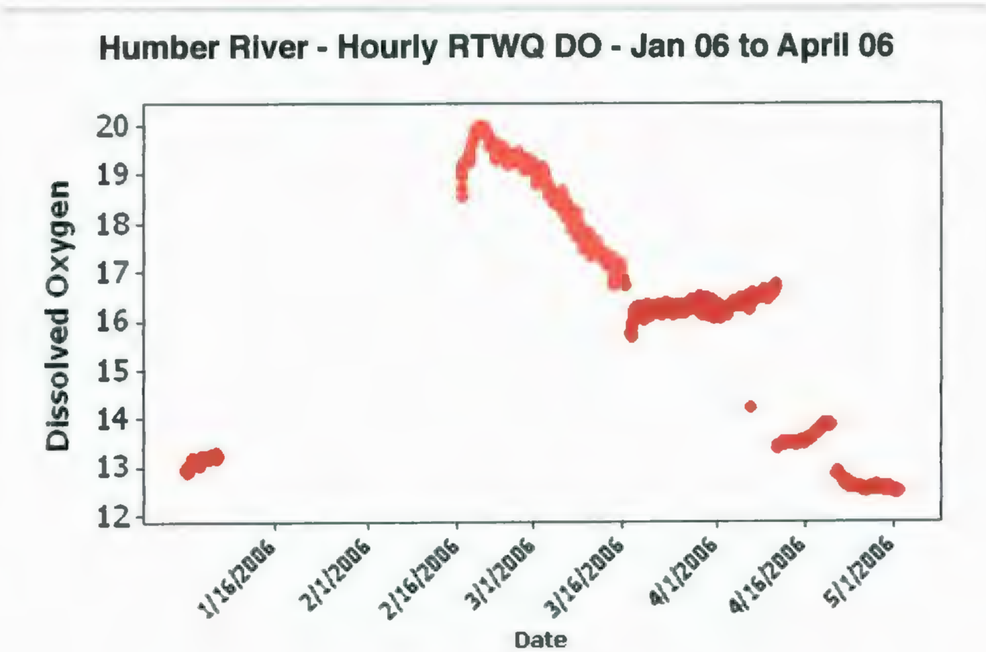


Figure 4.3 Hourly Observations of DO Collected January to April 2006 - Humber River

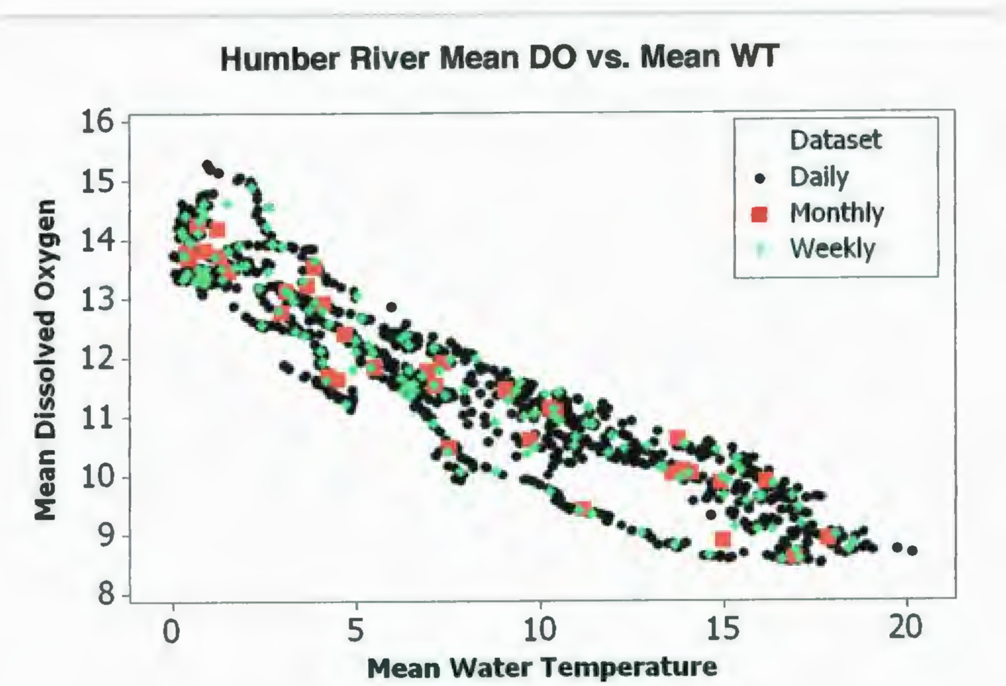


Figure 4.4 Relationship is More Reasonable Once the Higher DO Points are Removed

4.3.3 Step Three - Find the Best Fitting Models

Minitab and Datafit were used to find the best fitting dissolved oxygen models. Like the water temperature datasets in Chapter Three, the dissolved oxygen datasets were randomized to remove correlation.

4.4 Regression Modeling Results for Dissolved Oxygen

Table 4.1 presents the curve fitting results for the linear and exponential models and Figure 4.5 presents the regression models for mean monthly dissolved oxygen. The complete set of curve fitting results for the dissolved oxygen regression models (linear, exponential and multiple regression with stage are contained in Appendix K.

4.5 Discussion

4.5.1 Linearity Versus Nonlinearity

The goodness of fit of the linear and exponential decay models were found to be quite similar where in most cases there is very little difference in the shape of the two models. When dealing with mean dissolved oxygen, both models have high adjusted R^2 values at all time scales - with the exception of Leary's Brook where the models are only good for monthly and weekly observations and at the daily time scale the exponential model drops off to an adjusted R-squared of 0.68 and the linear model has an adjusted R-squared of 0.71. There is a considerably large amount of scatter in the dissolved oxygen values at the lowest and highest water temperatures in the Leary's Brook dataset (Figure 4.6). It is unknown at this time why this scatter is so high - potentially it is the result of this being a smaller stream so changes in air temperature are quick to influence dissolved oxygen.

Whatever the cause might be, the daily model for this station is quite unreliable and can only be used to gain a general idea of what daily dissolved oxygen might be at that station.

4.5.2 The Effect of Time Scale

Similar to the water temperature models, the goodness of fit of the dissolved oxygen models are better at the monthly time scale than they are at the weekly and daily time scale. As the time scale is shortened, the adjusted R^2 values tend to decrease, although when dealing with the mean dissolved oxygen data this decrease is usually less than 5%. High amounts of scatter in the daily observations force the residual sum of squares to be quite high for the daily models. The residual plots for the models were checked to ensure the assumptions of regression modeling were not violated. There was only a slight difference in the residual plots for the different time scales for the dissolved oxygen models. Figure 4.7 presents a plot of the residual plots for the Humber River monthly mean dissolved oxygen models. For this particular station there is slightly larger variation in the dissolved oxygen levels at the lower water temperature than there is at higher water temperatures at the daily time scale.

Table 4.1 Linear and Exponential Dissolved Oxygen Regression Models

	Mean Datasets Mean Dissolved Oxygen Mean Water Temperature Mean Stage			Minimum Datasets Minimum Dissolved Oxygen Maximum Water Temperature Minimum Stage			Maximum Datasets Maximum Dissolved Oxygen Minimum Water Temperature Maximum Stage		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Humber River (37 Monthly Observations, 149 Weekly Observations and 986 Daily Observations)									
*Linear Model: DO = a*Tw + b*									
a	-0.291	-0.291	-0.288	-0.258	-0.277	-0.277	-0.347	-0.305	-0.290
b	13.871	13.900	13.857	13.489	13.740	13.740	14.268	13.992	13.867
RSS	9.05	47.07	342.93	9.16	50.50	50.50	10.44	48.44	341.22
R^2 Adj	0.90	0.88	0.87	0.90	0.88	0.88	0.88	0.87	0.87
*Exponential Decay Model: DO = exp(a + b*Tw)*									
a	2.643	2.644	2.642	2.620	2.636	2.636	2.668	2.649	2.641
b	-0.026	-0.026	-0.026	-0.024	-0.025	-0.025	-0.029	-0.026	-0.026
RSS	8.26	43.61	317.68	8.29	46.45	46.45	9.42	45.38	316.88
R^2 Adj	0.91	0.89	0.88	0.91	0.89	0.89	0.90	0.88	0.88
Peter's River (23 Monthly Observations, 91 Weekly Observations, 595 Daily Observations)									
*Linear Model: DO = a*Tw + b*									
a	-0.271	-0.270	-0.270	-0.216	-0.230	-0.245	-0.330	-0.303	-0.286
b	13.189	13.208	13.214	12.498	12.811	13.030	13.887	13.615	13.346
RSS	4.92	27.51	220.58	12.23	46.61	365.03	5.96	27.01	203.54
R^2 Adj	0.94	0.93	0.91	0.88	0.89	0.87	0.92	0.92	0.91
*Exponential Decay Model: DO = exp(a + b*Tw)*									
a	2.588	2.591	2.591	2.545	2.566	2.581	2.634	2.618	2.599
b	-0.025	-0.026	-0.026	-0.023	-0.023	-0.024	-0.028	-0.027	-0.026
RSS	5.42	28.56	224.11	13.34	52.68	374.09	5.76	25.25	198.11
R^2 Adj	0.94	0.92	0.91	0.87	0.88	0.87	0.92	0.92	0.92

Table 4.1 continued - Linear and Exponential Dissolved Oxygen Regression Models

	Mean Datasets Mean Dissolved Oxygen Mean Water Temperature Mean Stage			Minimum Datasets Minimum Dissolved Oxygen Maximum Water Temperature Minimum Stage			Maximum Datasets Maximum Dissolved Oxygen Minimum Water Temperature Maximum Stage		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Leary's Brook (16 Monthly Observations, 57 Weekly Observations, 347 Daily Observations)									
*Linear Model: DO = a*Tw + b*									
a	-0.347	-0.407	-0.398	-0.341	-0.355	-0.351	-0.442	-0.418	-0.402
b	14.207	14.328	14.199	12.626	12.757	13.401	15.895	15.302	14.689
RSS	15.05	63.15	687.79	118.71	238.54	1093.68	11.26	48.52	2130.05
R^2 Adj	0.83	0.81	0.71	0.41	0.51	0.59	0.82	0.84	0.42
*Exponential Decay Model: DO = exp(a + b*Tw)*									
a	2.666	2.673	2.662	2.567	2.557	2.607	2.772	2.734	2.691
b	-0.030	-0.036	-0.035	-0.037	-0.035	-0.033	-0.034	-0.033	-0.033
RSS	17.28	76.56	770.17	126.67	265.69	1180.39	10.93	51.53	2209.88
R^2 Adj	0.80	0.77	0.68	0.37	0.46	0.55	0.82	0.83	0.40
Waterford River (21 Monthly Observations, 90 Weekly Observations, 587 Daily Observations)									
*Linear Model: DO = a*Tw + b*									
a	-0.378	-0.371	-0.373	-0.335	-0.330	-0.349	-0.409	-0.397	-0.383
b	14.148	14.097	14.098	12.649	13.412	13.828	14.791	14.496	14.249
RSS	15.13	120.07	886.03	102.32	229.78	1132.24	15.77	104.83	851.31
R^2 Adj	0.88	0.81	0.80	0.55	0.71	0.77	0.84	0.81	0.79
*Exponential Decay Model: DO = exp(a + b*Tw)*									
a	2.667	2.663	2.663	2.617	2.631	2.651	2.697	2.680	2.668
b	-0.035	-0.035	-0.035	-0.043	-0.035	-0.035	-0.033	-0.030	-0.035
RSS	14.73	116.89	861.44	96.34	226.17	1090.66	15.33	101.64	825.25
R^2 Adj	0.89	0.82	0.81	0.58	0.72	0.78	0.85	0.82	0.80

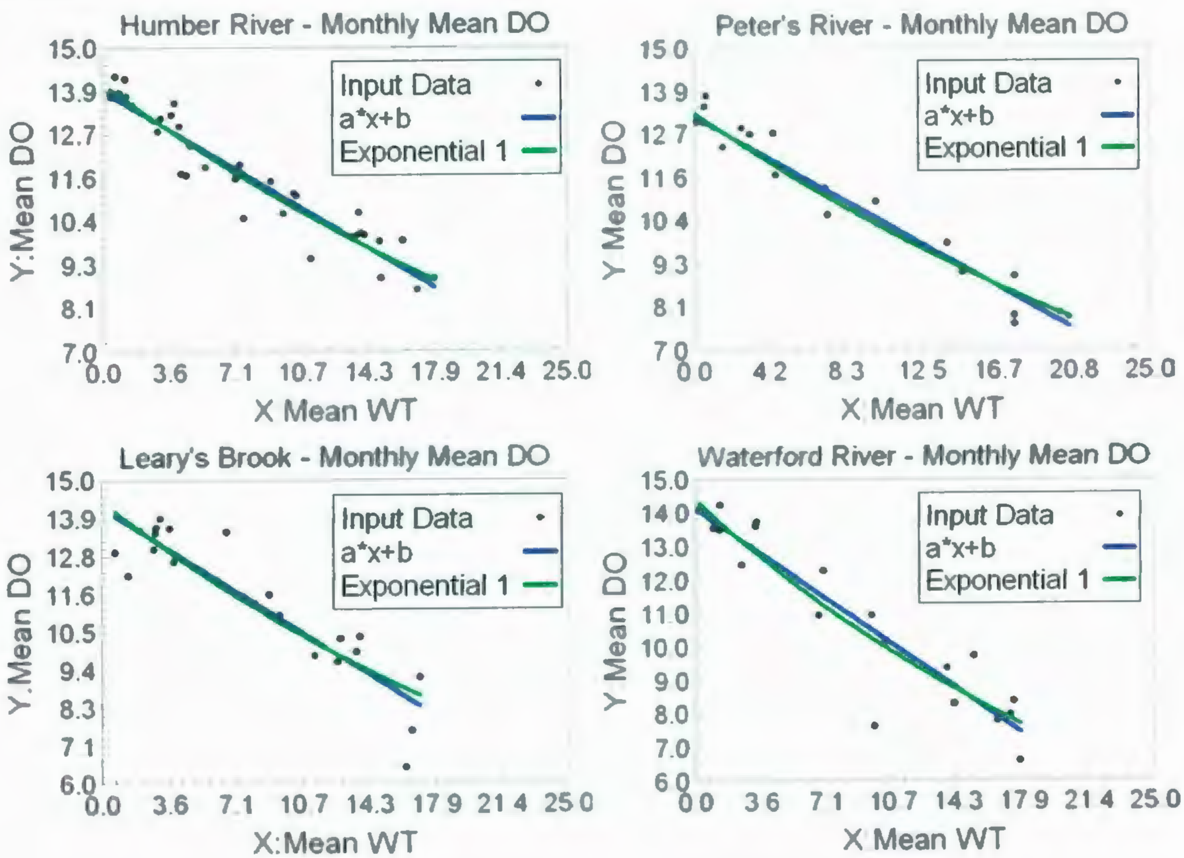


Figure 4.5 Regression Models for Monthly Mean Dissolved Oxygen

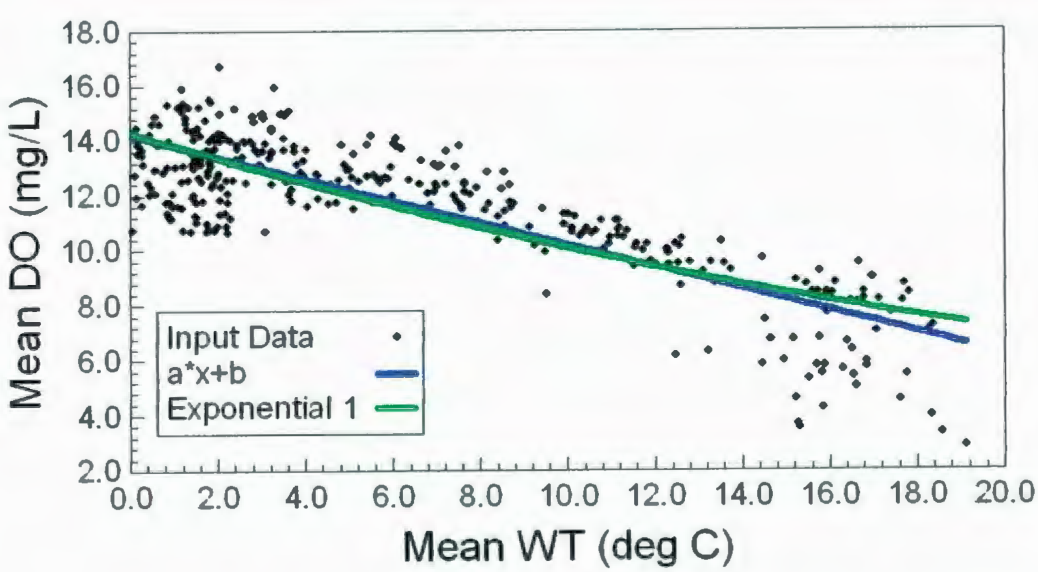


Figure 4.6 Scatter in the Leary's Brook Daily Mean Dissolved Oxygen Dataset

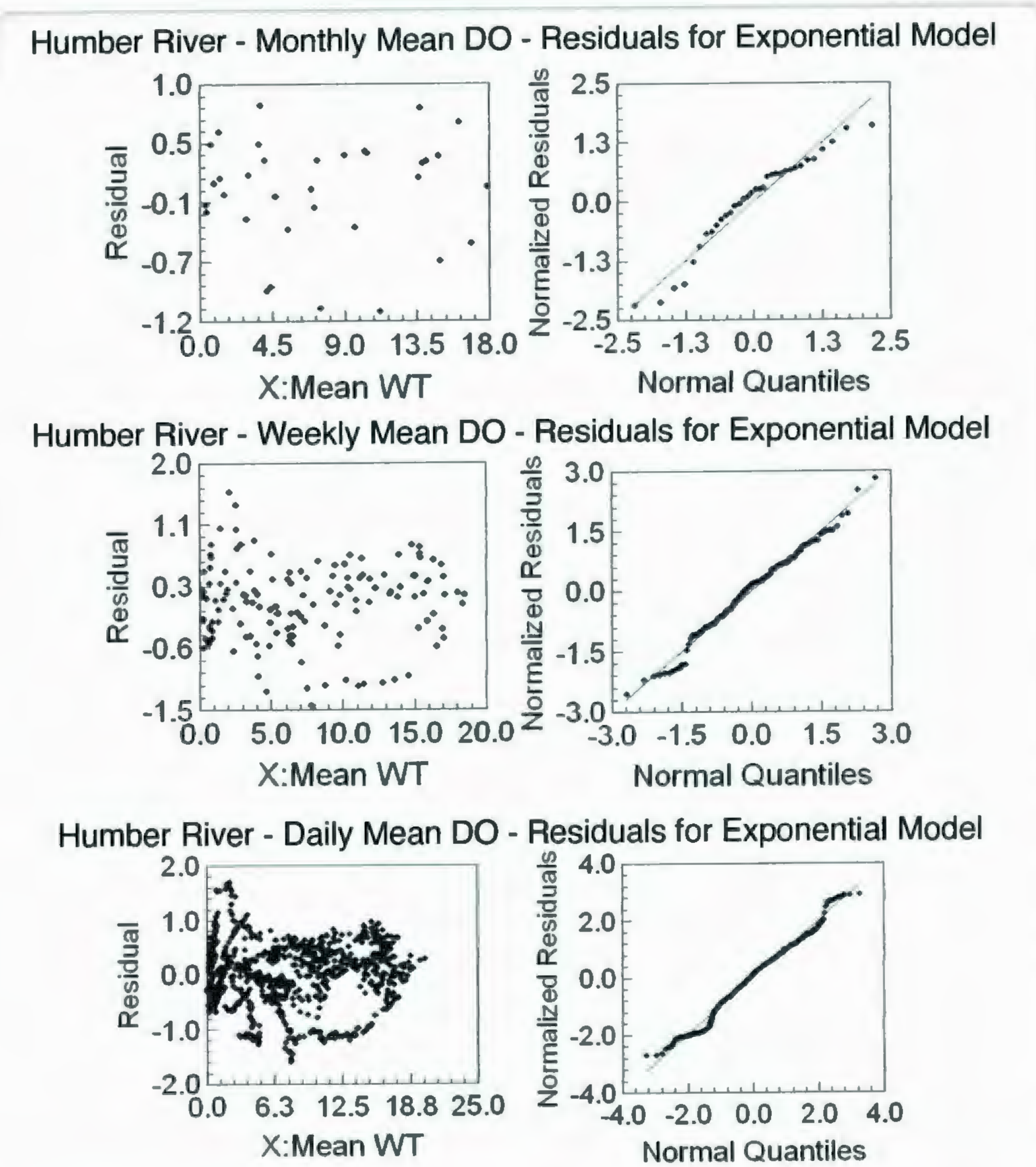


Figure 4.7 A Comparison of Residual Plots for the Humber River Dissolved Oxygen Models

4.5.3 Maximum and Minimum Datasets

For the two smallest rivers in the network, the fit of the models to minimum dissolved oxygen data is rather poor. For monthly minimum dissolved oxygen the adjusted R^2 value for the exponential model is quite low for both (0.37 for Leary's Brook and 0.58 for Waterford River). These models are of little use to modeling monthly values. The models remain poor for Leary's Brook at the weekly and daily time scales, but rather surprisingly the models improve slightly at the Waterford River station as the time scale is shortened (adjusted R^2 of 0.72 and 0.78 for weekly and daily minimum dissolved oxygen). For the two larger rivers in the network, the fit of the models to the minimum dissolved oxygen data is good, with adjusted R^2 values for Humber River and Peter's River equal to 0.90 and 0.88. Overall the models tend to fit better to minimum dissolved oxygen data collected at the larger rivers than at the smaller rivers. The models fit quite well to the maximum dissolved oxygen datasets, with adjusted R^2 values above 0.80 - except for Leary's Brook where the adjusted R^2 for the daily maximum dissolved oxygen drops to 0.40 (likely due to high levels of scatter at this station that was previously discussed).

4.5.4 The Influence of Stage

Stage was only a significant explanatory variable for dissolved oxygen at the daily time scale and was as a result the curve fitting results for the multiple regression model were excluded from Table 4.2 for this reason. Although the multiple regression model at the daily time scale had high adjusted R^2 values, the model is not a better choice than the linear and exponential models.

4.5.5 Similarities in the Exponential Models

Figure 4.8 presents a graphical summary of all the exponential decay models developed for modeling dissolved oxygen at the station. For the two larger rivers - Humber River and Peter's River, there is very little difference between the mean, maximum and minimum dissolved oxygen models. The only noticeable difference comes at the lower water temperatures, where the maximum dissolved oxygen models would tend to predict higher dissolved oxygen concentrations which is to be expected.

The exponential models Leary's Brook and Waterford River are only slightly different depending on the dataset being used, with minimum dissolved oxygen models predicting lower dissolved oxygen values and maximum dissolved oxygen models predicting higher values. It should be noted that the monthly and weekly minimum models for these two stations show a concern for low dissolved oxygen (i.e. less than 6.0 mg/L) when dealing with higher water temperatures (i.e. greater than 20° C). The daily minimum dissolved oxygen models do not show such a significant drop in dissolved oxygen at these same levels - but levels do still drop below 7.0 mg/L at those high water temperatures. High water temperatures appear to have a more significant impact on the minimum recorded dissolved oxygen levels at these two smaller water bodies. With this being said, there is a great need to carefully monitor dissolved oxygen levels at the Leary's Brook and Waterford River stations during the summer months to ensure levels do not stay low for long periods of time.

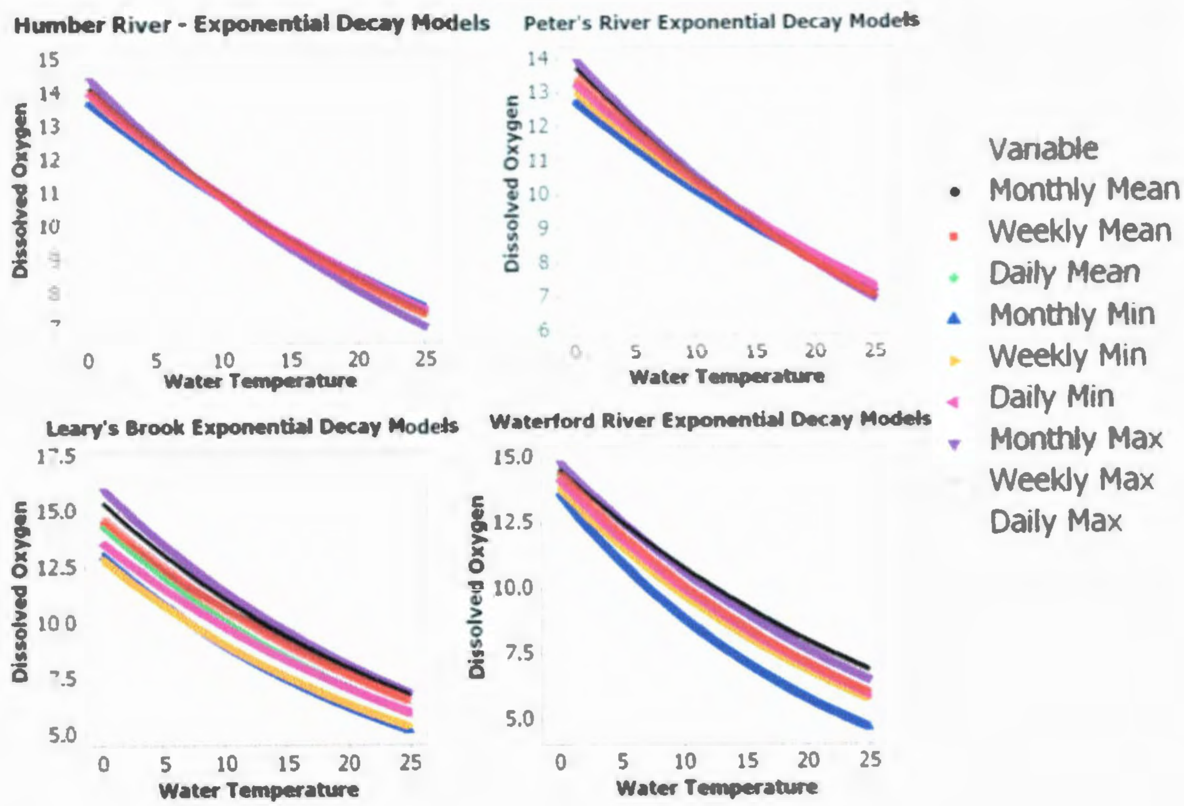


Figure 4.8 Similarities Between the Exponential Dissolved Oxygen Models

4.5.6 Using the Best Models for Prediction

All the developed regression models were tested using the dissolved oxygen datasets reserved for prediction purposes. The exponential decay model was deemed the best overall choice for modeling dissolved oxygen levels at the stations as it tended to perform better than the linear model for handling lower dissolved oxygen levels recorded at higher water temperatures. A summary of the model testing results for the mean dissolved oxygen models are shown in Table 4.2. A scatterplot of the Humber River monthly, weekly and daily observed versus predicted mean dissolved oxygen levels is shown in

Figure 4.9. Figures 4.10 and 4.11 contain a comparison of the mean and standard deviation of the difference between observed and predicted values for the models developed for all the stations. The complete testing results for the maximum and minimum dissolved oxygen models can be found in Appendix L.

Author's note - like the datasets that were reserved for testing the water temperature regression models, the gap in the gap in the available observed dissolved oxygen dataset is the result of having a limited dataset available for testing the models developed for this station. As more data is made available it will be possible to test the monthly and weekly models for accuracy over the 5 to 7 °C range. The daily mean values available for testing do give an idea of the accuracy of the daily model over this water temperature range for this station.

Table 4.2 - Using the Logistic 1 Model for Predicting Dissolved Oxygen

		Absolute Value of Difference Abs[Pred - Obs]				**Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%**			
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**	**St Dev**
				Humber River					
				Monthly Mean DO = exp(2.643 - 0.0258 Mean WT)					
0.91	14	0.76	1.77	0.04	0.53	6.51	16.11	0.3	5.02
				*Weekly Mean DO = exp(2.644 - 0.0258 * Mean WT)*					
0.89	63	0.74	2.9	0.01	0.59	6.18	21.48	0.04	5.05
				*Daily Mean DO = exp(2.642 - 0.0256 * Mean WT)*					
0.88	398	0.76	2.97	0	0.59	6.36	22.21	0	5

Table 4.2 continued - Using the Logistic 1 Model for Predicting Dissolved Oxygen

		Absolute Value of Difference Abs[Pred - Obs]				Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%			
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**	**St Dev**
				Peter's River					
				Monthly Mean DO = exp(2.588 - 0.0255 Mean WT)					
0.940	8.000	0.720	1.420	0.020	0.500	6.690	10.770	0.140	4.140
				Weekly Mean DO = exp(2.591 - 0.0256 Mean WT)					
0.92	32	0.68	1.94	0.02	0.58	6.38	21.87	0.17	5.75
				Daily Mean DO = exp(2.591 - 0.0256 Mean WT)					
0.91	196	0.72	2.32	0	0.59	6.78	24.7	0.02	5.82
				Leary's Brook					
				Monthly Mean DO = exp(2.666 - 0.0302 Mean WT)					
0.8	8	0.37	0.86	0.1	0.26	3.78	9.03	0.82	2.91
				*Weekly Mean DO = exp(2.673 - 0.0357 * Mean WT)*					
0.77	23	0.84	1.71	0.21	0.39	8.85	19.44	1.58	4.92
				*Daily Mean DO = exp(2.662 - 0.0351 * Mean WT)*					
0.68	132	0.88	1.83	0.05	0.41	9.51	22.09	0.37	5.2
				Waterford River					
				Monthly Mean DO = exp(2.667 - 0.0353 Mean WT)					
0.89	10	1.49	3.09	0.05	1.17	13.88	36.8	0.5	11
				Weekly Mean DO = exp(2.663 - 0.0347 Mean WT)					
0.82	42	1.71	4.09	0.07	1.15	16.3	41.81	0.65	11.22
				Daily Mean DO = exp(2.663 - 0.0351 Mean WT)					
0.81	272	1.7	8.14	0	1.25	16.28	74	0.01	12.35

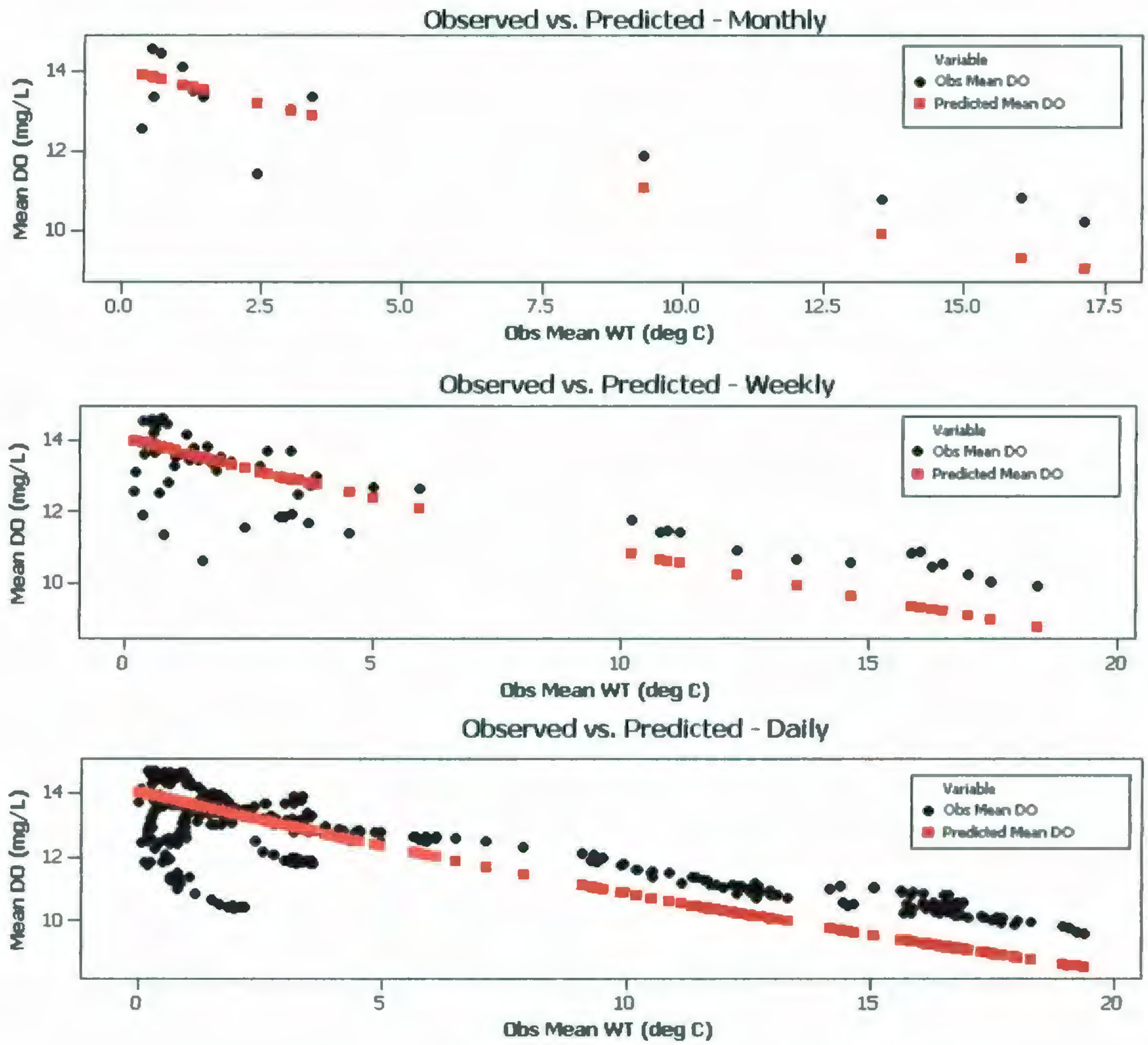


Figure 4.9 - Scatterplot of Humber River Observed Versus Predicted Dissolved Oxygen

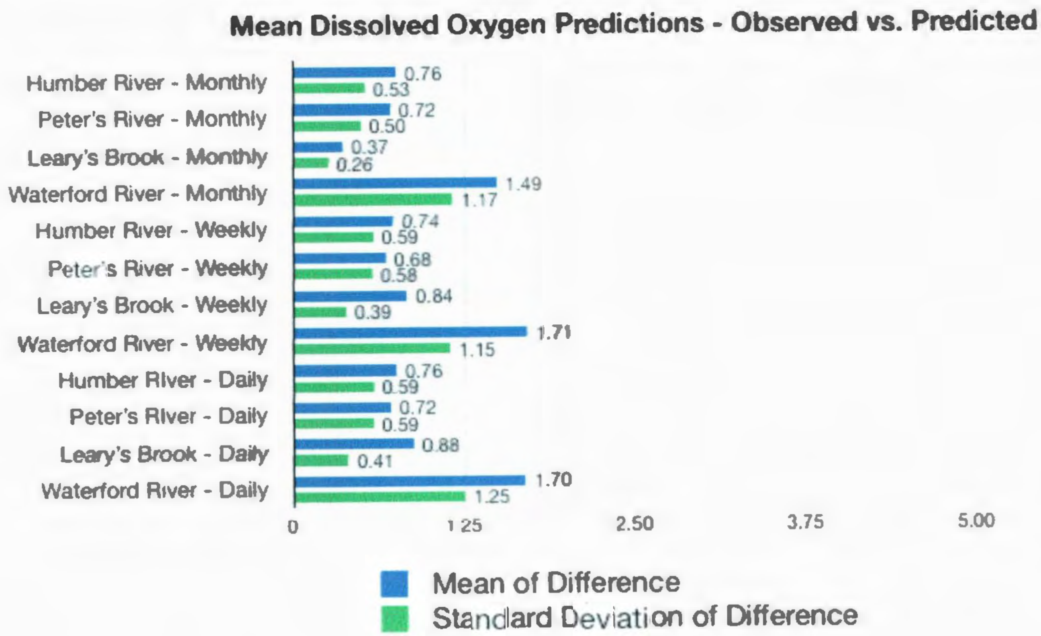


Figure 4.10 Mean and St. Dev. of Difference for Exponential Model Predicted DO

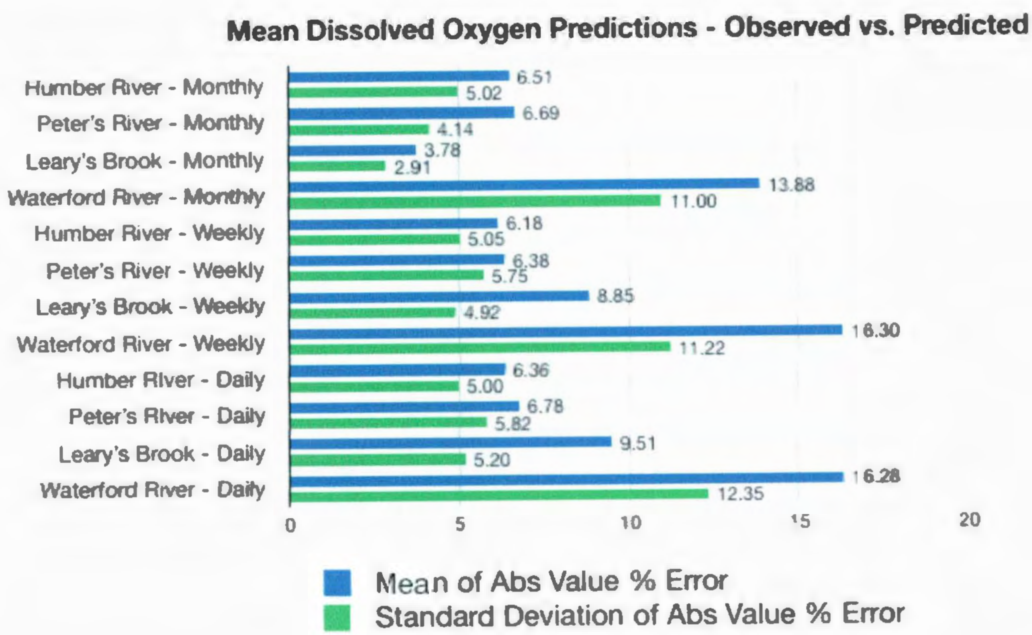


Figure 4.11 Mean and St. Dev. of % Error for Exponential Model Predicted DO

4.5.7 Visually Linking Air Temperature to Dissolved Oxygen

Regression models were developed in Chapter 3 to link nearby measurement of air temperature to water temperature at the RTWQ stations. In this chapter regression models were developed to link water temperature to dissolved oxygen levels at the RTWQ stations. A Minitab macro was written to let the user visually link these two regression models. The macro is designed so that a user can define their own logistic model for air-water temperature and their own exponential decay model for water temperature-dissolved oxygen. Once both models are defined the user then enters a value for air temperature that will be visually displayed on the plot as a reference point. The complete code for the macro can be found in Appendix M. Figures 4.12 to 4.16 present the plots developed for monthly mean models for each of the real time stations.

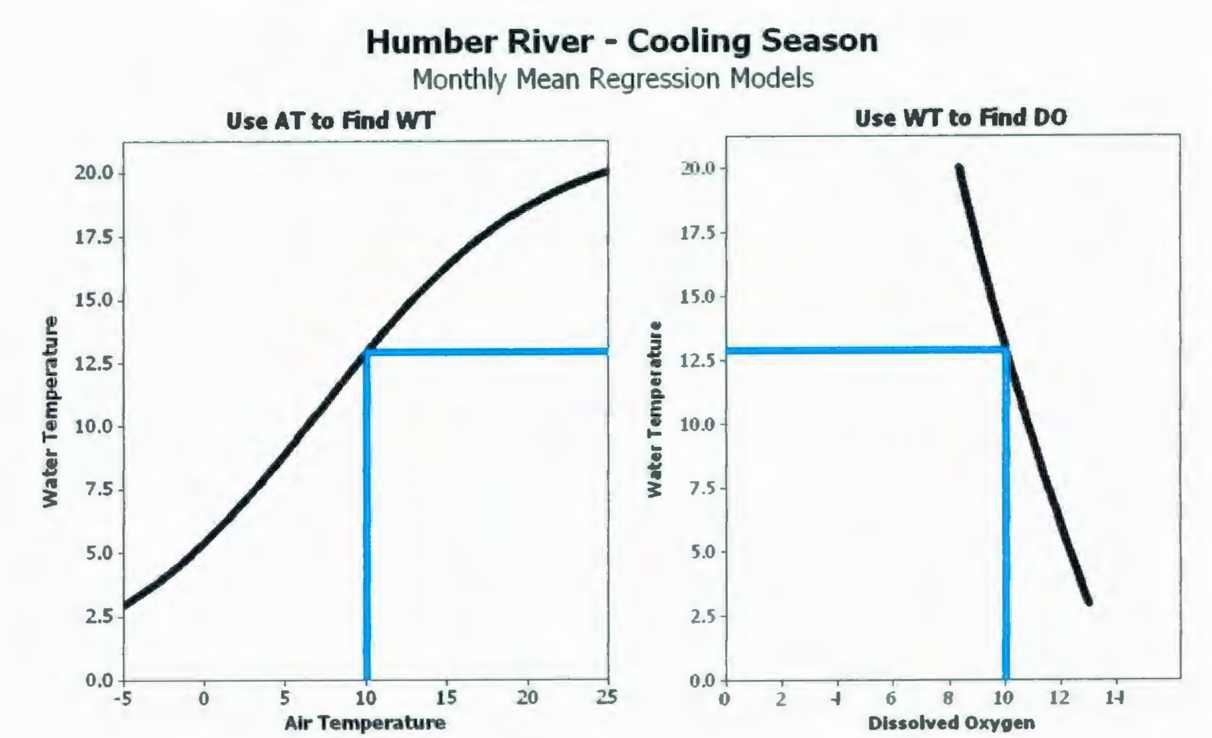


Figure 4.12 Humber River Cooling Season - WT, AT and DO Plot

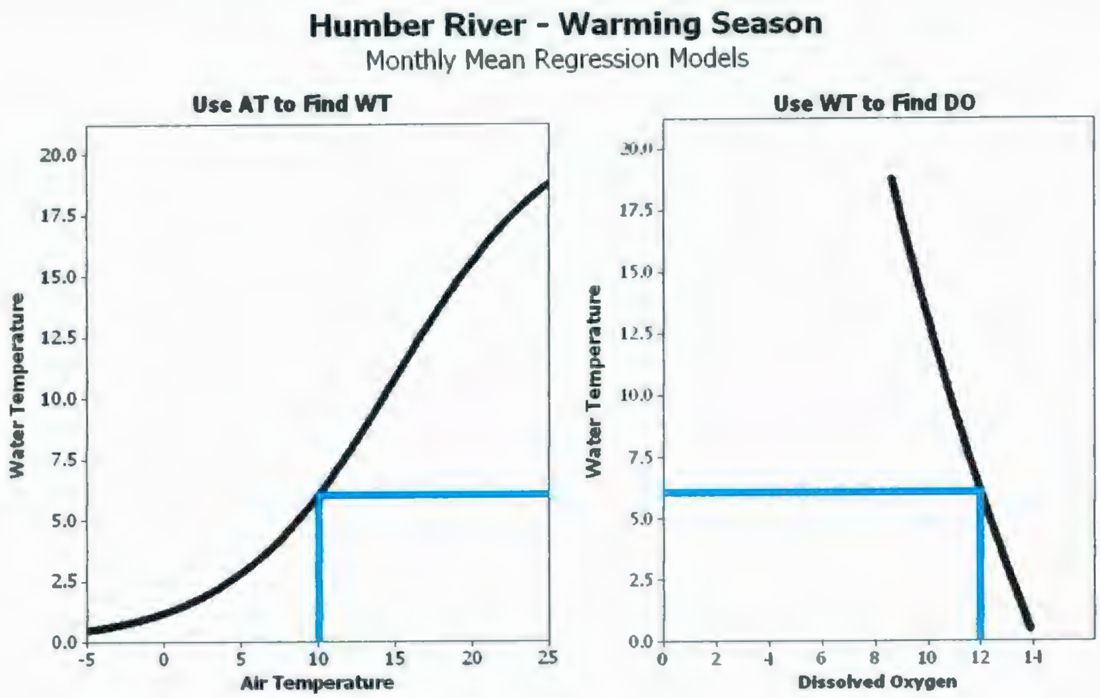


Figure 4.13 Humber River Warming Season - WT, AT and DO Plot

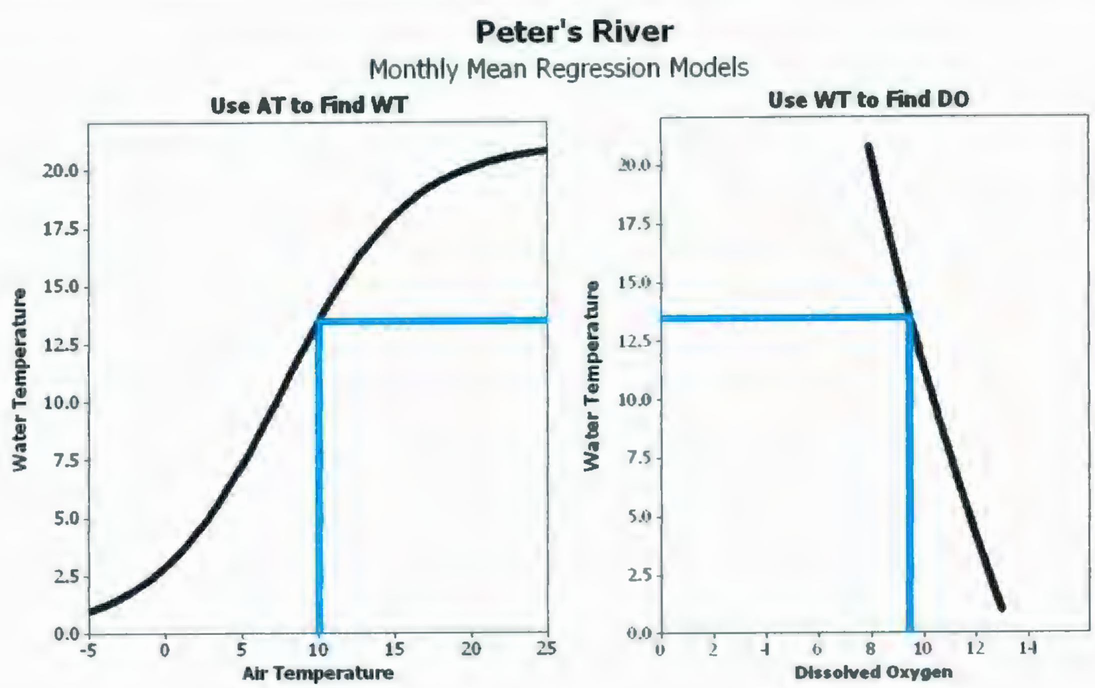


Figure 4.14 Peter's River - WT, AT and DO Plot

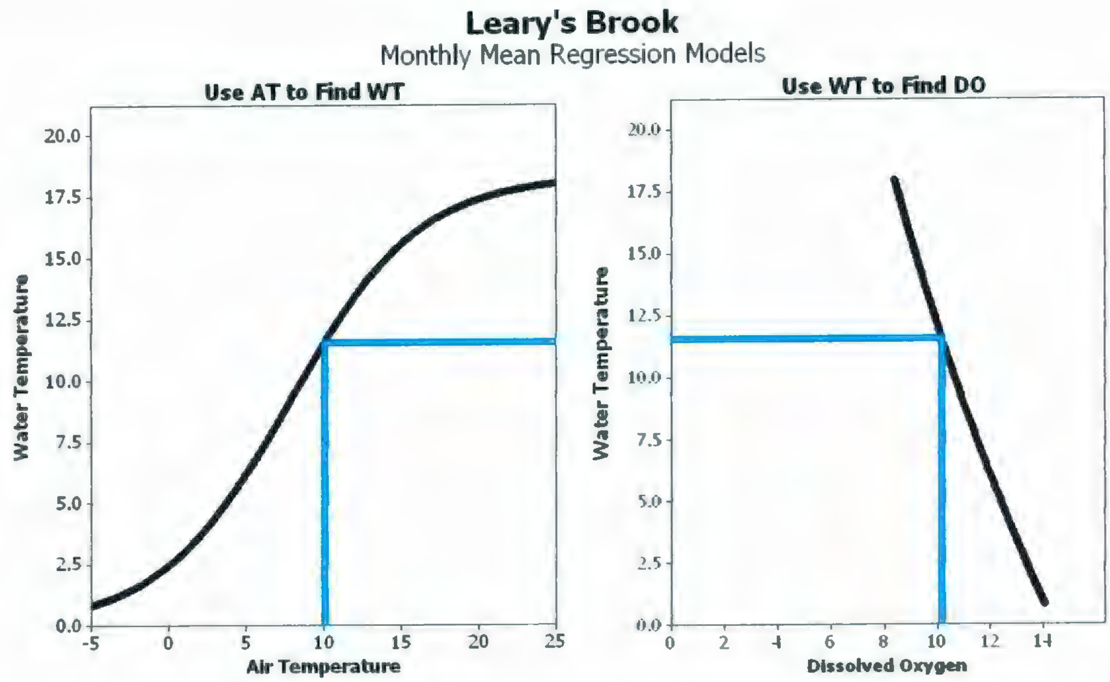


Figure 4.15 Leary's Brook - WT, AT and DO Plot

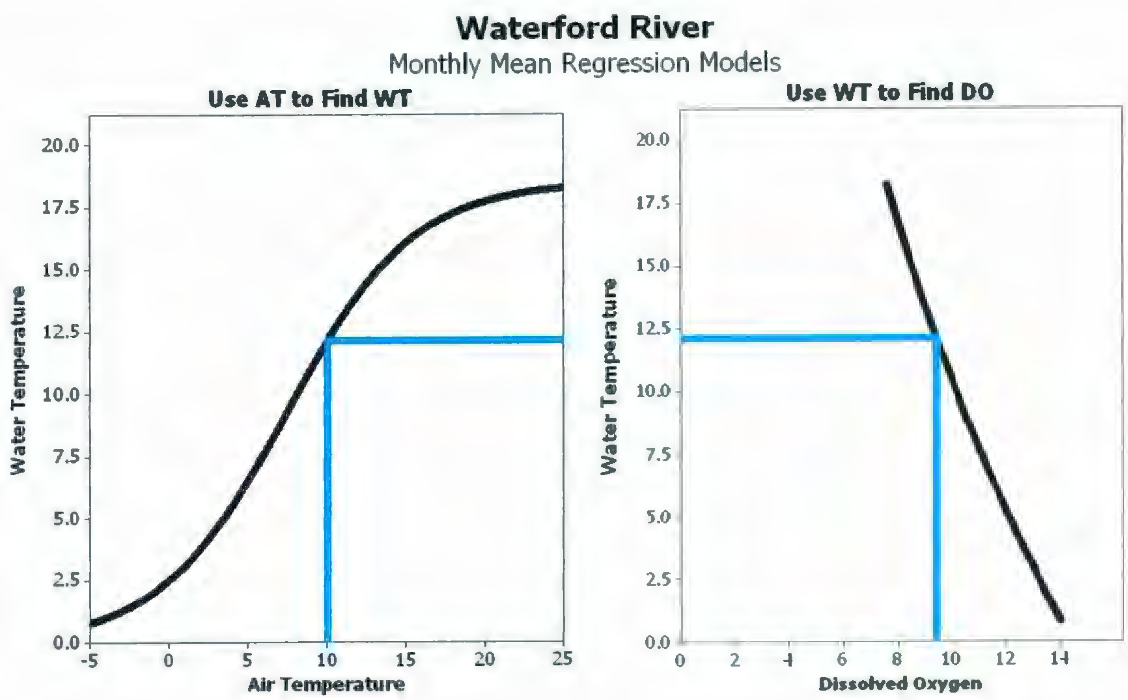


Figure 4.16 Waterford River - WT, AT and DO Plot

Chapter Five

Development of Regression Models

for Grab Samples

5.1 Scope

This chapter presents an in-depth look at a process used for developing regression equations that link real time measurements of water quality to manually collected grab samples of water quality. The chapter begins with a review of relevant literature. This is followed by a summary of the approach the USGS has taken for modeling grab sample data using real time measurements. The models developed for each of the four provincial stations using an approach similar to that of the USGS are presented and future directions for this research are discussed.

5.2 Literature Review

Clifton and Gilliom (1989) developed relationships for predicting dissolved solids and selenium using data collected in the San Joaquin River in California from 1985 to 1987. The authors found that both dissolved solids and selenium levels at the site could be estimated using continuously recorded streamflow and specific conductance measurements. Hill and Gilliom (1993) followed up this work with an investigation of an expanded 1985 to 1988 San Joaquin River dataset. They found that dissolved solids, boron and selenium concentrations in the river were positively correlated with each other and negatively related to streamflow.

The United States Geological Survey can be considered to be at the forefront of using regression models to link real time and grab sample measurements of water quality. Christensen et al. (2000) first used real time water quality monitoring and grab samples collected at two USGS stations in Kansas from 1995 to 1998 to develop regression equations for estimating alkalinity, dissolved solids, total suspended solids, chloride, sulfate, atrazine and fecal coliform bacteria concentrations. 2 years of grab sample data (35 to 55 samples) were found to be sufficient enough to develop a good relation between the surrogate (real time) and constituent (grab sample chemical). The developed equations were tested using data collected in 1999 at the sites and were found to give good estimates of the grab sample observations.

Christensen (2001) discusses the approach used by the USGS for developing regression models for estimating grab sample measurements for Rattlesnake Creek near Zenith, Kansas - in the Quivira National Wildlife Refuge. A real time sensor was installed at the USGS streamflow station on Rattlesnake Creek in 1998 and real time measurements of specific conductance, pH, water temperature, dissolved oxygen and turbidity were collected at hourly intervals from December 1998 to June 2001. USGS personnel made visits to this station every two weeks to keep the sensor calibrated and ensure the station was properly maintained. Along with the collected real time data, the USGS also manually collected water quality samples at the station. Four quarterly samples, five event samples and one quality-assurance sample were collected every year at the station and analyzed for physical properties, solids, sediment, major ions, nutrients, and bacteria. The first step in the USGS methodology for developing regression models for estimating daily and annual mass loads of chemicals of concern at the station was to plot each of the possible explanatory variables against the response variable to identify patterns in the data. Once the data has been plotted, a stepwise procedure and an overall method were used to identify the explanatory variable(s) to include in the regression model for the chemicals of concern. The stepwise procedure involves adding each of the explanatory variables (pH, specific conductance, water temperature, dissolved oxygen and turbidity) to the regression equation one at a time to determine if there was a statistically significant correlation. Explanatory variables were considered to be of significance if the probability value was less than 0.05. If several models showed themselves to be acceptable, Christensen selected the one with the lowest PRESS statistics, where the Prediction Sum of Squares (PRESS) is the sum of the squares of the prediction error. By minimizing PRESS, the model with the least error in the prediction of future observations in selected. Additionally, only explanatory variables with a physical basis for their inclusion into the regression models were considered in modeling. Christensen goes beyond the PRESS statistics and used an additional four diagnostic statistics to evaluate potential regression models - the mean square error (MSE), the coefficient of determination (R^2), the relative mean absolute error (RMAE) and the relative percentage difference (RPD). RMAE is expressed as a percentage and is calculated using the equation:

$$RMAE = \frac{\frac{1}{n}\sum_{i-1}^{n}|A - B|}{M_B} \times 100 \qquad \text{Equation 5.1}$$

where A is the estimated concentration, B is the measured concentration, and M_B is the average of all the measured concentration.

RPD is the relative percentage differences between the measured and estimated chemical concentrations and was calculated using the equation:

$$RPD = \left[|B - A| / (A) \right] \times 100 \qquad \text{Equation 5.2}$$

where A is the measured chemical concentration, and B is the estimated chemical concentration.

The USGS used graphical plots to examine the linearity of the developed relation between the explanatory and the response variables and it is sometimes necessary to transform certain variables to eliminate curvature in the data and convert the models to linear equations. Christensen (2001) used a graphical approach to identify outliers in the data but did not remove any outliers in the datasets used in the report. In the end, regression equations for estimating alkalinity, dissolved solids, total suspended solids, sediment, chloride, flouride, sulfate, nitrate, total organic nitrogen, total phosphorus and fecal coliform bacteria were developed for the river. The number of samples used to develop the regression equations was small. ranging from only nine to a maximum of twenty samples. A summary of the models developed in Christensen (2001) is presented in Table 5.1.

Table 5.1 - Regression Models for Grab Samples in Christensen (2001)

Chemical or Property	Regression Equation	n	Concentration	R^2
Alkalinity	$\log_{10} ALK = -0.000368Q - 0.000148WT^2 + 2.36$	18	ALK 91-224 Q 3.6-840 WT 3.4-31.5	0.71
Dissolved Solids	$DS = 0.549SC + 14.3$	18	DS 264-5460 SC 453-9930	0.999
Total Suspended Solids	$\log_{10} TSS = 0.818 \log_{10} NTU + 0.348$	18	TSS 14-270 NTU 5-270	0.825
Suspended Sediment	$\log_{10} SSC = 0.926 \log_{10} NTU + 0.438$	9	SSC 14.3-1820 NTU 5-480	0.926
Sodium	$Na = 0.203SC + 0.0938Q - 117$	18	Na 50-1880 SC 453-9930 Q 3.6-840	0.998
Chloride	$Cl = 0.319SC + 0.113Q - 172$	18	Cl 67-3000 SC 453-9930 Q 3.6-840	0.999
Flouride	$\log_{10} F = -0.000255Q + 0.162 \log_{10} SC - 0.892$	18	F 0.2-0.6 Q 3.6-840 SC 453-9930	0.826
Sulfate	$SO_4 = 0.0268SC + 13.17$	18	SO_4 12-269 SC 453-9930	0.983
Nitrate	$\log_{10} NO_3 = -0.000442SC + 2.60 \log_{10} SC - 0.000998WT^2 - 7.37$	20	NO_3 0.014-2.13 SC 453-9930	0.829
Total Organic Nitrogen	$TN = 0.00317NTU + 0.0234WT - 0.0000655SC + 0.469$	20	TN 0.050-2.5 NTU 5-480 WT 3.4-3.15 SC 453-9930	0.806
Total Phosphorus	$TP = 0.00103NTU - 0.227 \log_{10} SC + 0.0057WT + 0.776$	20	TP 0.025-0.755 NTU 5-480 SC 453-9930 WT 3.4 - 31.5	0.96
Fecal coliform bacteria	$\log_{10} FCB = -3.40 \log_{10} WT + 0.432 \log_{10} NTU + 6.53$	18	FCB 90-20000 WT 9.3-32.2 NTU 5-480	0.661

5.3 Methodology Used for Grab Sample Regression Models

The USGS approach to developing regression models for grab samples was used as a framework for the approach taken in this research to developing similar kinds of models for the Newfoundland provincial real time network.

5.3.1 Develop Real Time-Grab Sample Datasets

The WRMD routinely collects grab samples of water quality and then send these samples to a laboratory for analysis. Once the samples are back from the laboratory, the results are entered into a historical record of samples that have been collected over the years. These historical records contain the grab sample number, the date of collection, and the physical and chemical properties of the samples. Unfortunately, the historical records of grab samples maintained by the WRMD currently do not contain the exact hour the samples were collected and only contain the day.

The time of day of grab sample collection is an essential piece of information for linking the grab sample data to the real time pH, specific conductance, dissolved oxygen, water temperature, turbidity and stage level data recorded by the real time sensors. The only way to identify the grab sample collection time is to refer to maintenance forms for the sites. Every time a grab sample is collected, WRMD personnel fill out a corresponding maintenance form that contains the deployment period of the sensor, the Hydrolab minisonde calibration readings for the sensor, the name of the sampler, and most importantly for this research, the time of grab sample collection and whether or note the sample was taken upon removal or reinstallation of the real time sensor. Once the time of collection for the grab sample is determined it is then possible to refer to the historical records of real time water quality to find Hydrolab sensor measurements taken at the same time.

Under ideal circumstances, models would be developed by linking grab samples taken upon reinstallation of the Hydrolab datasonde to the accurate real time measurements of water quality made by the re-calibrated sensor at that time. Unfortunately, the grab sample datasets developed for the stations in the network are rather small and usually consist of measurements taken both when the sensor is reinstalled and when the sensor is removed (i.e. when the sensor measurements have drifted over time away from the actual value). It has always been WRMD standard practice to only collect grab samples upon reinstallation, but over the years removal grab samples have worked their way into the historical records.

In this research regression models have been developed using two datasets for each station: (1) only reinstallation samples and (2) removal and reinstallation samples. Grab samples without matching real time data (i.e. the sensor was for whatever reason not recording or transmitting measurements at that point in time) have not been included in the datasets.

5.3.2 Use Statistical Software to Identify the Best Models

The USGS used a combination of plotting the variables, implementing a stepwise variable selection approach, and implementing an overall approach for model selection using PRESS, MSE, and R^2 to identify the best overall model. For this thesis, relationships between explanatory variable(s) and the response variable of interest were first investigated using graphical scatter plots of the variables in Minitab. These plots provided an idea of what variables might warrant consideration for inclusion in the developed model.

After these plots were investigated, an *All Subsets* approach in the Datafit curve fitting software was used for selecting the best model from the available explanatory variables. The all subsets procedure is an exhaustive examination of every possible combination of independent variables being in and out of the model. Datafit computes all the possible combinations for one independent variable, then moves on to two independent variables, then three, etc. Once the computations are finished, the software returns the best models for each possible number of independent variables based on the following statistics: R^2, residual sum of squares, standard error (the standard deviation of the residuals), and Mallow's Cp. Mallow's Cp is a statistic designed to achieve a good compromise between the need to explain as much of the variation in the response variable as possible by including all the relevant variables while also minimizing the variance of the resulting estimates (minimizing the standard error) by keeping the number of coefficients small (Helsel and Hirsch, 2002). Mallow's Cp is defined by the following equation:

$$Cp = p + \frac{(n-p)\times\left(s_p^2 - \hat{\sigma}^2\right)}{\hat{\sigma}^2} \qquad \text{Equation 5.3}$$

where n is the number of observations, p is the number of explanatory variables plus 1, s_p^2 is the mean square error of this p-coefficient model, and $\hat{\sigma}^2$ is the best estimate of the true error (usually taken to be the minimum MSE among all the possible models). The best model will be the one with the lowest Cp value.

Often times search procedures like the *All Subsets* procedure in Datafit will recommend a number of different models if the explanatory variables shows signs of multicollinearity (the existence of linear relationships between the variables). Multicollinearity can result in inaccurate estimates of the regression coefficients, deflation of the probability values for the regression coefficients and can make it difficult to draw appropriate conclusions from the search results presented by the software. There are a number of different sources of multicollinearity - including explanatory variables which are inherently related to each other like water temperature and dissolved oxygen, extreme outliers and including additional variables that are generated from existing ones like dissolved oxygen and percent saturation. The Datafit software gives two different sets of statistics to help identify multicollinearity: the correlation matrix and variance inflation factors (VIF).

The correlation matrix is an array of the correlation coefficients that are calculated from all the possible pairings between the explanatory and response variables. The correlation matrix lets the user identify which variables correlate with each of the other variables. A perfect correlation of 1.0 would indicate a perfect linear relationship while a correlation of 0.0 would indicate no relationship was present.

The second measure for detecting multicollinearity in Datafit is the variance inflation factor (VIF) which indicates how well each independent variable can be predicted from all the other independent variables. VIF is calculated using the following equation:

$$VIF = \frac{1.0}{1.0 - R_i^2}$$ Equation 5.4

where R_i^2 represents the individual R^2 and is not the same as the overall R^2 of the regression model. It is better for the overall R^2 of the model to be high and the individual R^2 to be low (meaning that there was low collinearity between variables). If an individual R^2 is high then the VIF will end up being greater than 1.0 while a low individual R^2 will result in a VIF that approaches 1.0. When looking at the VIF values in the software output it should be kept in mind that (1) if the VIF is high for one or more variables (greater than 10.0) than multicollinearity can be assumed to be a problem, (2) if the VIF is greater than 4.0 for one or more variables than multicollinearity may be a problem and (3) if VIF is less than 4.0, multicollinearity is likely not a problem.

A sample *All Subsets* parameter selection output in Datafit is presented in Figure 5.1. Note that X1 is the water temperature, X2 is pH, X3 is specific conductance, X4 is dissolved oxygen, X5 is stage level and Y is alkalinity recorded in the grab samples. The real time measurements of turbidity has not been included as an explanatory variable as the sensors tend to provide an unreliable estimate of the turbidity levels at the station. The correlation matrix shows that water temperature is highly correlated with dissolved oxygen and as a result probably only one of them should be used in regression. All the VIF values are less than 4.0 so in this case multicollinearity should not be a problem. The all subsets results shows that the best model would only include variable X5 (stage) as an explanatory variable but the goodness of fit of the model is quite low (R2 of only 19% which is very low).

Correlation matrix

	X1	X2	X3	X4	X5	Y
X1	1	0.1965696097	-0.04687666038	-0.7956738421	-0.3459202428	-0.03019338939
X2	0.1965696097	1	0.1280900274	-0.08537055605	0.02868819246	0.170626995
X3	-0.04687666038	0.1280900274	1	-0.0370833616	0.1146312569	0.257846719
X4	-0.7956738421	-0.08537055605	-0.0370833616	1	0.1329922959	0.09270088559
X5	-0.3459202428	0.02868819246	0.1146312569	0.1329922959	1	0.43646887
Y	-0.03019338939	0.170626995	0.257846719	0.09270088559	0.43646887	1

Variance Inflation Factors

Variable	R2	VIF
X1	0.7153340431	3.512889321
X2	0.09151226172	1.100730321
X3	0.04461618913	1.046699754
X4	0.6690331893	3.021450997
X5	0.1936711843	1.240188842

All Subsets Results

Index	**Var Count**	**R2**	**RSS**	**Std. Error**	**Cp**	**Variables**
1	1	0.1905050745	71.7536302	1.766273776	1.665652652	X5
2	2	0.2342666682	67.87460253	1.756476361	2.440339125	X3,X5
3	3	0.262063447	65.41069606	1.764878197	3.662036367	X1,X4,X5
4	4	0.3172518632	60.51879485	1.739522849	4.116774707	X1,X3,X4,X5
5	5	0.3214224262	60.14911614	1.779253482	6	X1,X2,X3,X4,X5

Figure 5.1 Datafit All Subsets Output for Humber River Alkalinity

In this particular case although there are no problems with high levels of multicollinearity, there is likely going to be no model that will be useful for predicting alkalinity at the Humber River station. This lack of a useful model was a situation that was encountered many times for most of the Humber River grab sample water quality parameters and for many of the grab sample parameters for the other stations - but more on this will be mentioned later in this chapter after the modeling results are presented.

Once the *All Subsets* results in Datafit have been analyzed its then possible to use Datafit for identifying the best fitting regression models in Datafit. Unlike the regression models for water temperature and dissolved oxygen where specific models were being solved (i.e. logistic with air temperature), in this case the form of the model is unknown. Datafit can solve either every possible model for the explanatory variables or solving groups of models. When Datafit is used to solve every possible model often times the best fitting models will be tenth and ninth order polynomials which are rather useless for WRMD modeling purposes.

A better approach for finding more reasonable regression models is to use Datafit to solve smaller groups of models (i.e. single term intercept, polynomial, inverse polynomial, user defined models, etc.) to find more reasonable regression models.

The model fitting results presented by Datafit will present statistics like residual sum of squares, the standard error of the estimate, the probability values for the independent variables, R^2 and adjusted R^2. Unlike the USGS approach, in this research the adjusted R^2 has been used instead of R^2 for identifying the best fitting models.

5.4 Taking a First Look at the Grab Sample Datasets

Table 5.2 presents a general statistical summary of the grab sample datasets developed for each of the real time stations. The largest historical record of grab samples with matching real time data belonged to the Humber River station (31 samples), while Peter's River only had 18 samples with matching real time data. There is rarely a balance in the number of samples available for analysis from each year of collection - where for some years there might be 8 samples with real time data while the next there might only be three.

Table 5.3 on the following page presents the range of physical properties, major ions, elements, metals and nutrients recorded in the grab samples. The last column in the table presents the water quality guidelines established by the CCME for those parameters.

Table 5.2 Overview of the Grab Sample Datasets

Dataset	# of Samples with matching RTWQ data	# per year	Parameters Not Detected			
Humber River 12 special 19 reinstallation (original)	31	6 in 2004 5 in 2005 20 in 2006	bromide flouride potassium	antimony arsenic cadmium	copper lead mercury	nickel selenium uranium zinc
Peter's River 15 reinstallation 3 removal	18	6 in 2005 3 in 2006 8 in 2007 1 in 2008	bromide	antimony arsenic cadmium	lead mercury	nickel selenium
Leary's Brook 8 reinstallation 12 removal	20	4 in 2005 9 in 2006 3 in 2007 4 in 2008	bromide		mercury	nickel selenium
Waterford River 10 reinstallation 10 removal	20	2 in 2005 9 in 2006 5 in 2007 4 in 2008		antimony arsenic cadmium	mercury	nickel selenium uranium

Table 5.3 Range of the Grab Sample Measurements and CCME Guidelines

Parameter	Humber River (31)	Peter's River (18)	Leary's Brook (20)	Waterford River (20)	CCME Guideline (mg/L)
		Real Time Measurements			
WT (ºC)	0.63-18.8	-0.2-28.7	0.6-17.3	0.5-20.4	--
pH (pH units)	6.7 - 7.6	4.9-8.1	5.3-14.0	5.8-11.2	6.5-9.0
SC (µS/cm)	24.0-42.9	34 - 84.9	167-1329	235-1060	--
DO (mg/L)	8.6-19.3	7.5-14.3	9.3-16.3	7.1-24.7	> 5.5 mg/L
Stage (m)	1.5-3.5	0.9-1.5	0.6-0.9	0.4-1.2	--
		Grab Sample Measurements			
Alkalinity (mg/L $CaCO_3$)	10-20	8-34	0-13.0	6-21	--
Color (TCU)	22-112	15-74	0-24.0	8-26	--
Cond. (uS/cm)	39-56	41-89	210-2400	219-1200	--
Hardness (mg/L CaCO3)	7-17	16-33	10.0-53.0	17-52	--
pH (pH units)	6.6-7.6	6.5-7.6	6.1-7.1	6.6-7.4	6.5-9.0
TDS (mg/L)	25-36	25-58	107-959	142-625	--
TSS (mg/L)		Not recorded in the grab samples			--
Turbidity (NTU)	0.4-4.2	0.4-0.7	0.3-19.2	0.5-3.8	--
Boron (mg/L)	0-0.03	0-0.02	0-0.1	0-0.03	--
Bromide (mg/L)	ND	ND	ND	0-1.1	--
Calcium (mg/L)	3-5	4.8-10	4-18	5-17	--
Chloride (mg/L)	3-5	2-6	50-510	51-360	--
Flouride (mg/L)	0-0.11	0-0.1	0-0.1	0-0.5	--
Potassium(mg/L)	ND	0-0.3	0-5	1-2.6	--
Sodium (mg/L)	0-3	0-3.5	32-390	33-210	--
Sulphate (mg/L)	3-4	0-4	7-27	7-18	--

Table 5.3 continued - Range of the Grab Sample Measurements and CCME Guidelines

Parameter	Humber River (31)	Peter's River (18)	Leary's Brook (20)	Waterford River (20)	CCME Guideline (mg/L)
Ammonia (mg/L)	0-0.24	0-0.1	0-0.3	0-0.2	Total 10.3 *pH = 7.0* *WT = 10.3*
DOC	0.8-14.0	3.8-11	1.4-5.6	2.2-7.8	--
Nitrate(ite) (mg/L)	0-0.13	0-1.4	0.2-0.6	0.5-1.2	2.9 nitrate
Kjeldahl N (mg/L)	0-0.37	0.1-0.4	0.1-0.4	0-0.6	--
Total Phosphorus (mg/L)	0-0.09	0-0.1	0-0.1	0-0.3	--
Aluminum (mg/L)	0.05-0.17	0-0.1	0-0.5	0-0.17	0.005 - 0.1
Antimony (mg/L)	ND	ND	ND	ND	--
Arsenic (mg/L)	ND	ND	ND	ND	0.005
Barium (mg/L)	0-0.01	0-0.012	0-0.1	0-0.036	--
Cadmium (mg/L)	ND	ND	ND	ND	0.000017
Chrom. (mg/L)	0-0.001	0-0.005	ND	0-0.017	0.0089
Copper (mg/L)	ND	0-0.003	0-0.006	0-0.004	0.002-0.004
Iron (mg/L)	0.04-0.13	0.1-0.3	0.14-1.3	0.1-0.4	0
Lead (mg/L)	ND	ND	0-0.0093	0-0.0008	0.001-0.007
Magnes.(mg/L)	0-1.0	1-2	0-2	1-2.5	--
Mangan.(mg/L)	0-0.03	0-0.023	0-0.3	0-0.2	--
Mercury (mg/L)	ND	ND	ND	ND	0.00026
Nickel (mg/L)	ND	ND	ND	ND	0.025-0.15
Selenium (mg/L)	ND	ND	ND	ND	0.001
Uranium (mg/L)	ND	ND	0-0.0001	ND	--
Zinc (mg/L)	ND	0-0.009	0-0.086	0-0.03	0.03

ND - parameter not detected in the grab sample measurements with matching real time data
CCME guidelines are for the protection of aquatic life - freshwater

5.4.1 The Humber River Grab Sample Dataset

The historical record of grab samples collected at the Humber River real time station contained 44 grab samples of water quality collected from May 19, 2004 to August 11, 2008. 23 of the samples either had no matching maintenance forms to use as a reference n for collection time or the maintenance form for that sample failed to indicate the time of day the sample was collected. The WRMD was contacted about the missing forms and they were able to locate the log books of the staff members who had collected the samples at the station. With the help of the log books, the time of collection was determined for 11 of those samples while 12 of the samples were identified as being part of a special sample collection program at the station carried out in 2006. Unlike normal grab samples of water quality collected upon the reinstallation of the sensor, these special samples were collected while the sensor was still in the water. After consultation with the WRMD it was determined that these special samples could be included in the regression analysis dataset - leaving a total of 37 samples available for analysis.

31 of the samples collected from May 2004 to December 2006 were originally used to develop regression models for grab sample water quality at the station while 6 of the samples were reserved for model testing. Poor model fitting results called for follow up attempts at modeling with new datasets - one containing all 37 samples and another containing only the 25 samples taken upon reinstallation of the sensor. Of the available five real time measurements of water quality for use as potential explanatory variables, water temperature was highly correlated with dissolved oxygen (a correlation p-value of 0.000 for

the 31 samples) and dissolved oxygen was removed from the set of potential surrogates for the grab sample data to simplify modeling efforts. Water temperature and stage were also closely correlated for the 31 samples (p-value 0.035) but both measurements were left in the set of potential surrogates. 13 of the of the 38 measured physical and chemical properties were never detected in the grab samples collected at the Humber River station - bromide, flouride, potassium, antimony, arsenic, cadmium, copper, lead, mercury, nickel, selenium, uranium and zinc.

5.4.2 The Peter's River Grab Sample Dataset

The original historical record of grab samples collected at the Peter's River station showed 25 samples were collected from December 2004 to February 2008. 18 of these samples were paired with real time measurements - 3 of these samples were taken upon removal of the sensor. The removal samples appeared to be unreliable and were not included in the regression datasets - leaving only 15 samples available for regression. Due to the small number of samples both reinstallation and removal samples were combined for use in regression modeling and no samples were reserved for model testing purposes. The Peter's River station is now offline so its likely that no more grab samples will be collected for this station. Nine of the measured physical and chemical properties were never detected in the Peter's River grab samples - bromide, antimony, arsenic, cadmium, lead, mercury, nickel, selenium, and uranium.

5.4.3 The Leary's Brook Grab Sample Dataset

The historical record of grab samples collected at the Leary's Brook station contained 33 samples collected from May 2004 to September 2008 - with almost half of these samples collected in 2006. Only 20 of these samples could be matched to real time data and 12 of these samples were taken upon removal of the sensor. Both removal and reinstallation samples were combined into one regression modeling dataset. Due to the small number of samples it was not possible to set any samples aside for model testing purposes. The collection of grab samples is an ongoing process at the real time stations and it is expected that as new samples are collected they can either be used to test the models or can be added to the existing datasets to determine if better models can be identified.

Although only four of the measured physical and chemical properties were never detected at the station - bromide, mercury, nickel and selenium, there were a number of other chemicals whose measured levels were quite low - boron, flouride, total phosphorus, antimony, arsenic, barium, cadmium, chromium, copper, lead, uranium and zinc.

5.4.4 The Waterford River Grab Sample Dataset

The historical record of grab samples for the Waterford River station contained 22 samples were collected from August 2005 to September 2008. 20 of these samples could be matched to real time data but half of these samples were taken upon removal of the sensor. Again, small sample size forced the combination of the removal and reinstallation samples into one dataset - with no samples kept aside for model testing. Antimony, arsenic, cadmium, mercury, nickel, selenium and uranium were never detected in the samples collected at the station and levels of boron, total phosphorus, barium, chromium, copper and lead were very low. *Author's Note - Appendix N contains a more detailed statistical overview of the grab sample datasets for the real time stations.*

5.5 Regression Modeling Results for Grab Samples

The developed grab sample datasets were used to develop regression models for the following categories of grab sample measurements:

- *Physical Properties, Solids and Sediment:* alkalinity, conductivity, hardness, pH, total dissolved solids, turbidity and water temperature).
- *Major Ions, Elements and Metals*: boron, bromide, calcium, chloride, flouride, potassium, sodium, sulphate, ammonia, aluminum, antimony, arsenic, barium, cadmium, chromium, copper, iron, lead, magnesium, mercury, nickel, selenium, uranium, and zinc.
- *Nutrients*: dissolved organic carbon (DOC), nitrate(ite), kjeldahl nitrogen, and total phosphorus.

Table 5.4 (physical properties, solids and sediment), Table 5.5 (major ions, elements and metals) and Table 5.6 (nutrients) contain the regression modeling results for these categories of grab sample measurements. Only those models with an adjusted R^2 value greater than 0.40 have been included in the tables - unless the measured grab sample parameter is of particular interest. A discussion of the developed grab sample regression models follows directly after the tables.

Table 5.4 Models for Grab Sample Physical Properties, Solids and Sediment

Station	Parameter	Regression Model	n	Range	Adj R^2
Peter's	Alkalinity	ALK = 0.313SC + 5.33	15	ALK 8-34	0.514
Waterford	Alkalinity	Log(ALK) = 3.8 – 0.001SC – 1.44ST	20	ALK 6-21	0.799
No significant equations: Humber (ALK 10-20) and Leary's (ALK 0-13)					
Leary's	Color	Color = 16.12 + 0.49WT – 0.01SC	19	Color 0-24	0.635
No significant equations: Humber (color 22-112), Peter's (color 15-74) and Leary's (color 0-24)					
Peter's	Conductivity	Conductivity = 0.78SC + 19.78	15	Cond 41-89	0.809
Leary's	Conductivity	Conductivity = 1.23SC – 66.80	19	Cond 210-2100	0.904
Waterford	Conductivity	Conductivity = 1.07SC – 21.0	20	Cond 219-1200	0.95
No significant equations with real time specific conductance: Humber (cond. 39-56)					
Peter's	Hardness	Hardness = 35.3 + 0.2WT -636.7SC	15	Hard. 16-33	0.664
Leary's	Hardness	Hardness = 4.18 + 0.61WT + 0.03SC	19	Hard. 10-53	0.848
Waterford	Hardness	Hardness = 12.35 + 0.03SC – 12.11ST	20	Hard. 17-52	0.799
No significant equations: Humber (hardness 7-17)					
Peter's	pH grab	pH Grab = 0.26pH +5.42	15	pHgrab 6.5-7.7	0.403
Waterford	pH grab	pH Grab = 0.10pH + 6.29	20	pHgrab 6.6-7.4	0.9
No significant equations with real time pH: Humber (pH grab 6.6-7.6) and Leary's (pH grab 6.1-7.1)					
Humber	TDS	TDS = 26.28 -0.26WT + 1.98ST	31	TDS 25-36	0.55
Peter's	TDS	TDS = 0.56SC + 9.20	15	TDS 25-58	0.831
Leary's	TDS	TDS = 0.66SC – 7.17	19	TDS 107-959	0.848
Waterford	TDS	TDS = 0.56SC + 10.34	20	TDS 142-625	0.901
Peter's	Turbidity	Turb = 1.26 -0.01WT - 3.94pH	15	Turb 0.4-0.7	0.392
No significant equations: Humber (turb 0.4-4.2), Leary's (turb 0.3-19.2) and Waterford (turb 05-3.8)					
Humber	Grab WT	Grab water temperature = 0.99WT + 0.31	16	GrabWT 0.8-16	0.999
Grab sample water temperature recorded only 6 times in Peter's and never in Leary's and Waterford					

Table 5.5 Regression Models for Grab Sample Major Ions and Metals

Station	Parameter	Regression Model	n	Range	Adj R^2
Humber	Calcium (mg/L)	Ca =-0.05WT + 4.51	31	Ca 3.0-5.0	0.235
Peter's	Calcium (mg/L)	Ca = 0.09SC + 2.86	15	Ca 4.8-10.0	0.62
Leary's	Calcium (mg/L)	Ca = 0.10SC + 3.45	19	Ca 4.0-18.0	0.804
Waterford	Calcium (mg/L)	Ca = 0.30 + 0.01SC + 2.51ST	20	Ca 5.0 - 17.0	0.87
Humber	Chloride (mg/L)	Cl^- = 5.3 - 0.04WT - 37.54SC	31	Cl^- 3.0-5.0	0.291
Peter's	Chloride (mg/L)	Cl^- = 0.04SC + 1.93	15	Cl^- 2.0-6.0	0.245
Leary's	Chloride (mg/L)	Cl^- = 0.35SC -28.01	19	Cl^- 50 – 510	0.954
Waterford	Chloride (mg/L)	Cl^- = 0.33SC – 36.00	20	Cl^- 51 - 360	0.903
Leary's	Potassium (mg/L)	K = 0.003SC - 0.045	19	K 0.0-5.0	0.736
Waterford	Potassium (mg/L)	K = 0.002SC + 0.66	20	K 1.0 – 2.6	0.487
No significant models: Humber (K 0-0.4) and Peter's (K 0-0.3)					
Leary's	Sodium (mg/L)	Na = 0.23SC – 20.78	19	Na 32.0-390.0	0.895
Waterford	Sodium (mg/L)	Na = 0.227SC – 26.213 (One removed Feb 8/06, Na = 161)	19	Na 33.0-210.0	0.975
No significant models: Humber (Na 0-3.2) and Peter's (Na 0-3.5)					
Leary's	Sulphate (mg/L)	SO_4^{2-} = 0.014SC + 4.37	19	SO_4^{2-} 7.0 – 27.0	0.835
Waterford	Sulphate (mg/L)	SO_4^{2-} = 0.01SC + 6.59	20	SO_4^{2-} 7 – 18	0.747
No significant models: Humber (SO_4^{2-} 0-4.0) and Peter's (SO_4^{2-} 0-4.0)					
Waterford	Aluminum (mg/L)	Al =0.15 +0.11 logST	20	Al 0.03 – 0.17	0.62
No significant models: Humber (Al 0.05-0.17) and Peter's (Al 0.04-0.11), and Leary's (Al 0.04-0.48)					
Leary's	Barium (mg/L)	Ba = 3.21SC + 0.0005	19	Ba 0.0– 0.05	0.817
Waterford	Barium (mg/L)	Ba = 3.12SC - 0.003	20	Ba 0 – 0.036	0.664
No significant models : Humber (Ba 0-0.01) and Peter's (0-0.012)					

Table 5.5 continued - Regression Models for RTWQ Station Grab Sample Major Elements and Ions

Peter's	Iron (mg/L)	Fe = - 0.003SC + 0.28	15	Fe 0.05-0.26	0.535

No significant models: Humber (Fe 0.04-0.13)), Leary's (Fe 0.1-1.3) and Waterford (Fe 0.08-0.39)

Peter's	Magnesium (mg/L)	Mg = 0.019SC + 0.42	15	Mg 1.0-2.0	0.506
Leary's	Magnesium (mg/L)	Mg = 0.001SC + 0.31	19	Mg 0 – 2.0	0.576
Waterford	Magnesium (mg/L)	Mg = 1.77 – 251.5/SC + 0.38/ST	20	Mg 1.0-2.5	0.587

No significant models: Humber (Mg 0 -1.0)

Waterford	Manganese (mg/L)	Mn = 0.05 - 0.003WT + 0.0001SC	20	Mn 0-0.20	0.680

No significant models : Humber (Mn 0-0.03), Peter's (Mn 0-0.023) and Leary's (Mn 0.03-0.31)

Leary's	Zinc (mg/L)	Zn = 4.60SC + 0.0002	19	Zn 0-0.09	0.684
Waterford	Zinc (mg/L)	Zn = -0.07 + 0.02log(SC) – 0.02/ST	20	Zn 0 – 0.03	0.612

No significant models : Humber (Zn not detected) and Peter's (Zn 0-0.009)

Never detected at the stations - bromide, antimony, arsenic, cadmium, mercury, nickel and selenium. No significant models (with adjusted R^2 above 0.40) could be developed for - boron, flouride, ammonia, chromium, copper, lead, and uranium.

Table 5.6 Models for Grab Sample Nutrients

Station	**Parameter**	**Regression Model**	**n**	**Range**	**Adj R^2**
Humber	Nitrate(ite) mg/L	NO_3^- = -0.005WT + 0.08	31	NO_3^- 0-0.13	0.249
Peter's	Nitrate(ite) mg/L	NO_3^- = 0.03 – 0.01WT + 0.005SC (Remove one outlier NO_3^- = 1.4)	14	NO_3^-: 0 – 0.4	0.696
Waterford	Nitrate(ite) mg/L	NO_3^- = -0.02WT + 1.0	20	NO_3^- 0.5 – 1.2	0.402

No significant models: Leary's (NO_3^- 0.19-0.59)

Peter's	DOC	DOC =25.85 -4.91*log(SC)	15	DOC 3.8-11	0.54

No significant models: Humber (DOC 0.8-6.6), Leary's (DOC 1.4-5.6) and Waterford (2.2-7.8)

No significant models (with adjusted R^2 above 0.40) could be developed for kjeldahl nitrogen and total phosphorus recorded at any of the stations.

5.6 Discussion

5.6.1 Models for Grab Sample Physical Properties, Solids and Sediment

Regression models were investigated for grab sample physical properties, solids and sediment - these being alkalinity, conductivity, hardness, pH, total dissolved solids, turbidity and water temperature.

Alkalinity

Alkalinity is the capacity for solutes in water to react with and neutralize acid. It is an important indicator of water quality as it represents the capacity for a body of water to neutralize acidic pollution from rainfall or wastewater. When rivers and streams have low alkalinity they can be adversely affected by acidic inputs and the corresponding drop in pH of the water can harm acid-intolerant forms of aquatic life - where fish are particularly susceptible to harm from low pH (U.S. Environmental Protection Agency, 1999). Alkalinity at the stations are for the most part in the 0 to 25 mg/L $CaCO_3$ range - with levels at Peter's River being the highest and levels at Leary's Brook the lowest (Figure 5.2).

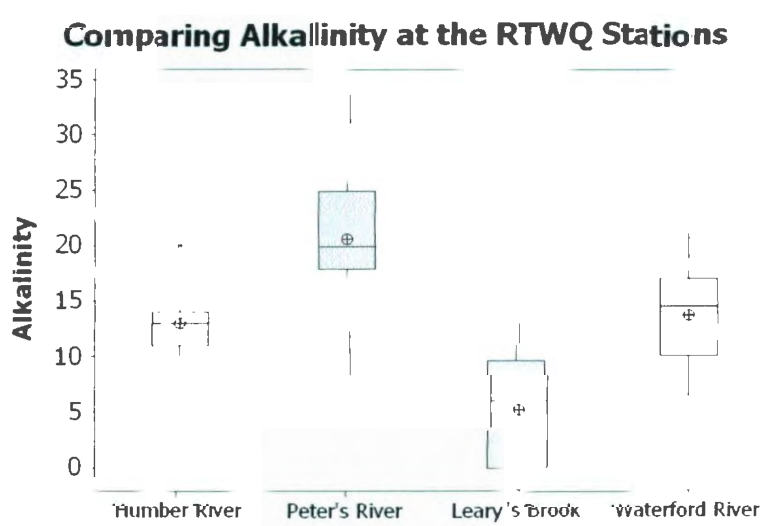


Figure 5.2 Comparing Alkalinity at the RTWQ Stations

No statistically significant equations could be developed using the available datasets for predicting alkalinity at the Humber River and Leary's Brook stations. At Peter's River, specific conductance was used as a surrogate to define a linear relationship to alkalinity - but the model has a low adjusted R^2 of 0.514. At the Waterford River station, 20 samples of specific conductance and stage were used as surrogates for alkalinity and adjusted R^2 was equal to 0.799. The goodness of fit of this model is similar to that developed by Christensen et al. (2002) - where streamflow and water temperature were used as surrogates for logarithmically transformed alkalinity with an R^2 of 0.710.

Water Color

No statistically significant relationships could be developed linking the real time measurements and grab sample measurements of water color. A relationship was developed for Leary's Brook using water temperature and specific conductance, but with an adjusted R^2 of 0.635 the model would not be very useful for making predictions.

pH Level

Regression models using real time measurements of pH as a surrogate for grab sample pH were investigated primarily for comparing how real time measurements and grab sample measurements of the same parameter match up together. Real time and grab sample pH measurements were not correlated at the Humber River and Leary's Brook stations, and were only slightly correlated at the Peter's River station. This is not surprising, as field experience has shown that often times when the sample is removed from the water the pH tends to change from what the true value would be. A linear relationship was developed for the Waterford River station (adjusted R^2 of 0.90).

Specific Conductance

Regression models using real time sensor specific conductance as a surrogate for grab sample conductivity were investigated. Except for Humber River, close linear relationships were developed between the two measurements (with adjusted R^2 being above 0.80). Scatterplots of real time specific conductance and grab sample conductance are shown in Figure 5.3. There is a complete lack of a relationship between the Humber River real time and grab sample conductance. Perhaps if the range of values in Humber River was larger it would be easier to define a relationship but the current grab sample measurements cannot be paired with real time specific conductance. The linear relationship between the two parameters is strongest for the two smaller rivers - where conductance levels recorded by the grab samples are much higher than the levels recorded at Humber River and Peter's River.

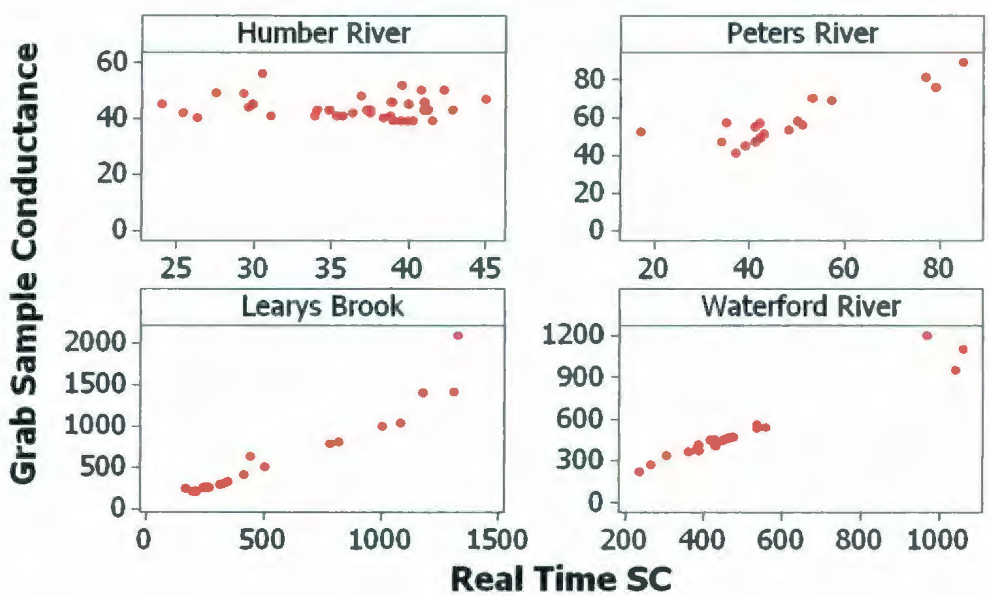


Figure 5.3 Grab Sample Conductance Versus Real Time Specific Conductance

Water Hardness

According to Hem (1992), its possible to use parameters like specific conductance and stage to establish a relationship with water hardness (which is alkalinity as $CaCO_3$). Specific conductance, water temperature and stage were used as surrogates for hardness at the two smaller stations in the network - where water temperature and specific conductance were used for Leary's Brook (adjusted R^2 of 0.848) and specific conductance and stage were used for Waterford River (adjusted R^2 of 0.799). No statistically significant regression model could be developed for Humber River water hardness. For Peter's River, water temperature and specific conductance were used as surrogates for hardness but the adjusted R^2 was only 0.664. The range of hardness measurements in the grab samples collected at Leary's Brook and Waterford River is much wider than at Peter's River and Humber River (refer to Figure 5.4). The variation in the Humber River data in particular is rather small - with 18 of the 31 measurements having a hardness of 10.

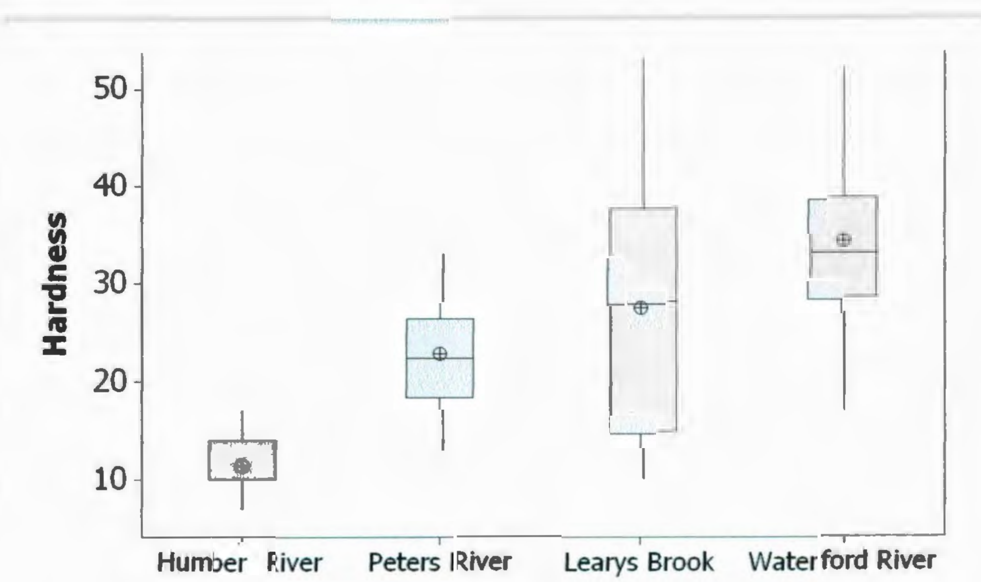


Figure 5.4 - Comparing Water Hardness at the RTWQ Stations

Total Dissolved Solids

Dissolved solids are the organic and inorganic material dissolved in a sample of water (Bates and Jackson, 1984). Water with high concentrations of dissolved solids tends to be salty while water with low concentrations of dissolved solids tends to be fresh. Dissolved solid levels in a water body can fluctuate as a result of pollution - i.e. wastewater discharges high in salts, irrigation, the clearing of land near a stream or the spreading of road salt during icy winter conditions (Texas State, 2009). Total dissolved solids levels in the grab samples collected at Leary's Brook and Waterford River are much higher than they are at Peter's River and Humber River. Large spikes in total dissolved solids occur during the winter months at the two urban streams - likely the result of road salt being washed into the rivers (Figure 5.5).

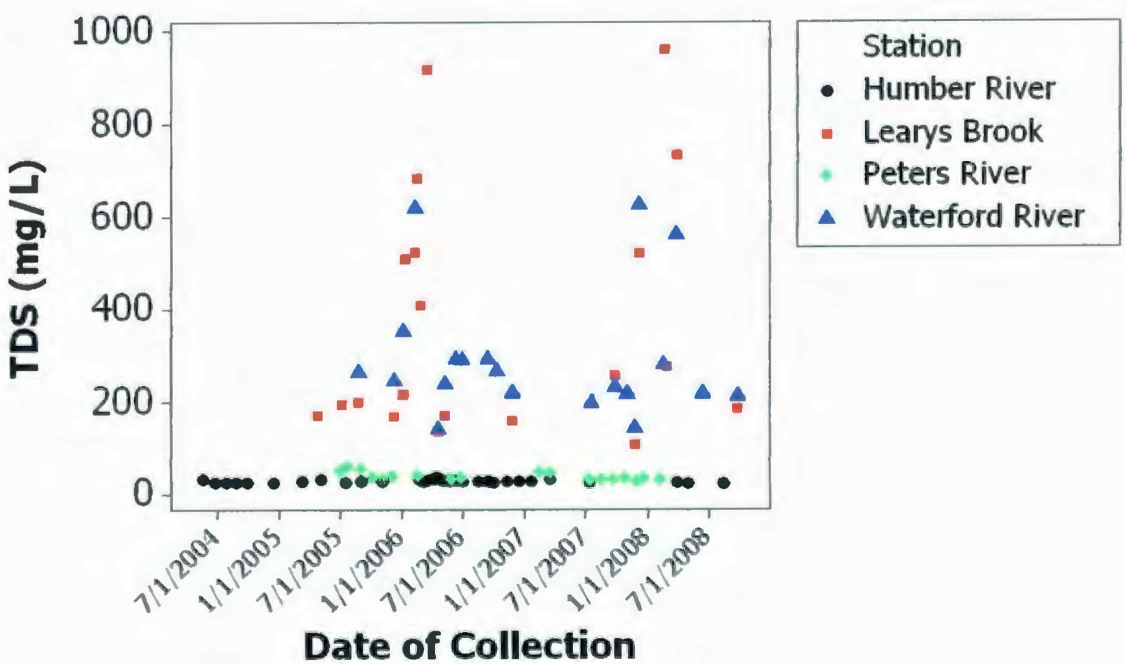


Figure 5.5 Grab Sample TDS by Date of Grab Sample Collection

According to Hem (1992) dissolved solid levels are usually related to the specific conductance of water but the strength of the relationship will depend on the ions present in the water. Some water bodies can be characterized by strong relationships between dissolved solids and specific conductance and at these locations chemical constituents (i.e. chloride) can usually be predicted with great accuracy - although during periods of high streamflow where a large portion of the streamflow is the result of large amounts of rain, the relationship can change. Rainfall was not investigated as a potential surrogate in this research so its influence on the development of total dissolved solids models at this point in time is unknown. The use of stage as a surrogate did not result in the development of a significant total dissolved solids model for any of the stations.

Grab sample measurements of total dissolved solids at the Humber River station were poorly related to the real time data. Specific conductance and total dissolved solids at the station were not significantly related to each other (refer to Figure 5.6). The best fitting model for the station did not use specific conductance but used water temperature and stage as surrogates for a model with an adjusted R^2 of 0.550.

The total dissolved solids models for the other three stations are more useful, with specific conductance acting as a surrogate in a linear regression model for each station (Figure 5.6). Adjusted R^2 values for the models are all above 0.80.

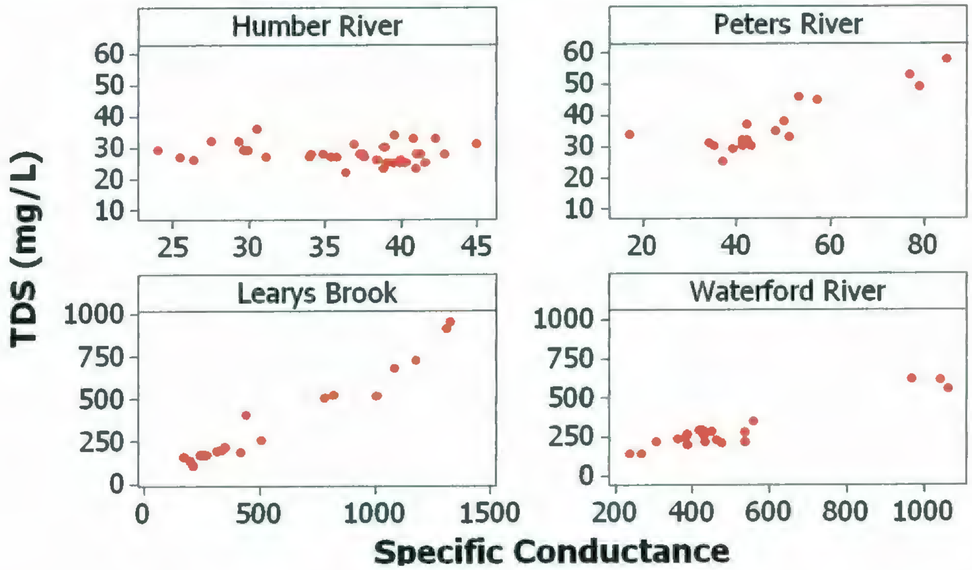


Figure 5.6 Relationship Between Real Time SC and Grab Sample TDS

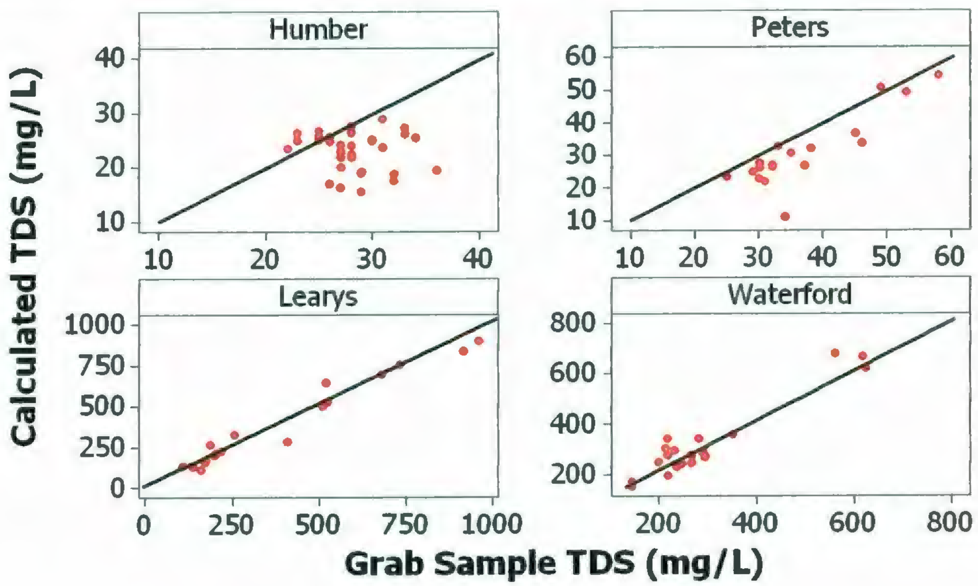


Figure 5.7 Comparing WRMD Calculated TDS and Grab Sample TDS

The WRMD currently has developed an equation for using specific conductance recorded by the real time sensors to estimate total dissolved solids:

$$TDS(g/L) = \text{Specific Conductance}(\mu S/cm) \times 0.00064 \qquad \text{Equation 5.5}$$

A comparison can be made between the total dissolved solids levels predicted by these equations (converted to mg/L) and the grab sample total dissolved solids - Figure 5.7 on the previous page. The scatterplot shows that the prediction equation is fairly accurate for Leary's Brook and Waterford River (with values falling along the 45° line). The equation tends to overestimate TDS at the Peter's River station while there is little correspondence between calculated and grab sample TDS at the Humber River station.

Turbidity

Regression models for grab sample turbidity were investigated for the stations but no significant models could be developed using the real time measurements as a surrogate. It would be useful if the turbidity levels recorded by the real time sensor could be brought on-line so that grab sample turbidity and real time turbidity could be compared.

Grab Sample Water Temperature

Regression models using real time sensor water temperature as a surrogate for grab sample water temperature were investigated for the Humber River station. A near perfect linear relationship was found for the two parameters (adjusted R^2 of 0.999). Measurements of grab sample water temperature are not recorded for the Leary's Brook and Waterford River stations and were only recorded six times a the Peter's River station.

5.6.2 Models for Grab Sample Major Elements, Ions and Metals

Statistically significant regression models were developed for the following major elements, ions and metals - calcium, chloride, potassium, sodium, sulphate, magnesium, manganese, aluminum, barium, iron and zinc. No statistically significant regression models could be developed for boron, flouride, ammonia, chromium, copper, lead, and uranium. Bromide, antimony, arsenic, cadmium, mercury, nickel and selenium were never detected at the stations.

Calcium

When the concentration of a charged ionic species increases the conductivity of a solution will also increase. For this reason, specific conductance should be close correlated to ionic species at the stations (Hem, 1992). Statistically significant regression models using specific conductance as a surrogate for calcium were developed for Leary's Brook and Waterford River where the highest calcium levels were recorded (4 - 18 mg/L). No signifiant equations could be developed for Peter's River and Humber River (Figure 5.8).

Calcium is one of the most important contributors to water hardness (where hardness in the grab samples is measured in mg/L $CaCO_3$). The close relationship between the two parameters at the Leary's Brook station is shown in Figure 5.9. Calcium levels at the Humber River station remained rather constant being primarily in the 4.0 - 5.0 mg/L range, and no significant regression model could be found relating specific conductance to calcium. Being such a large river it is likely that outside inputs into the river are highly diluted and do little to change the natural levels of the river - unlike a smaller water body like Leary's Brook where inputs into the stream are not diluted to the same extent and cause variation in the natural levels of the water. Without any large variation in the levels of calcium the lack of a model is not of concern - as the current dataset shows that calcium levels at the station will always be in the 3.0 to 5.0 mg/L.

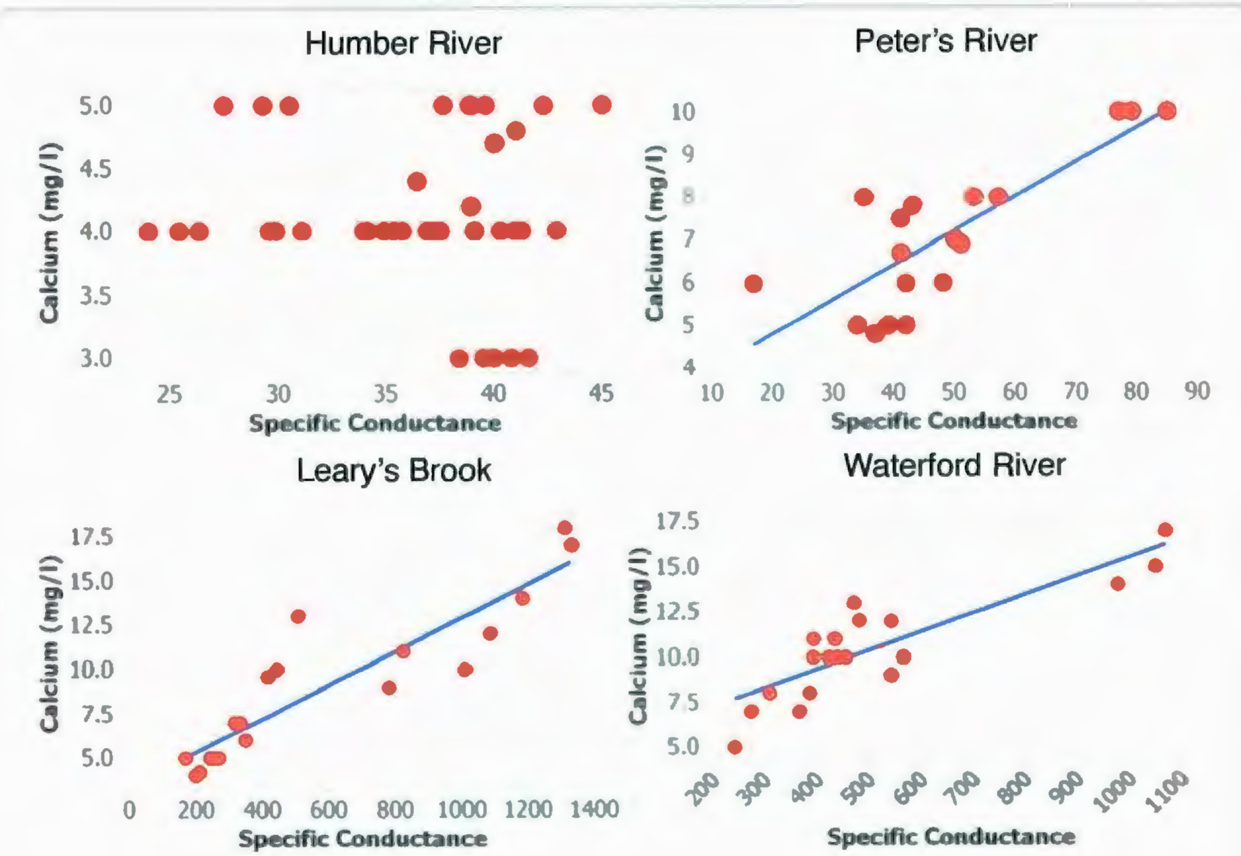


Figure 5.8 Relationship Between Specific Conductance and Calcium

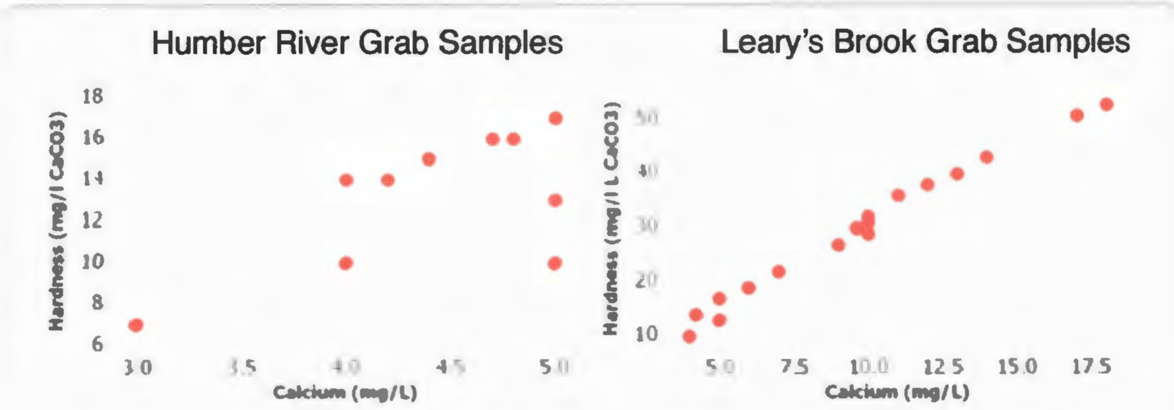


Figure 5.9 - Relationship Between Grab Sample Calcium and Water Hardness

Chloride

A close linearly relationship was found to exist between chloride and real time specific conductance measurements at the Leary's Brook and Waterford River stations. No statistically significant relationship could be defined for Peter's River and Humber River but the models for those two stations are included in the Table 5.5 for comparison. Like the calcium data, chloride measurements at the Humber River station were all in the 3.0 - 5.0 mg/L range.

Sodium and Sulphate

Grab sample measurements of charged ionic species like sodium and sulphate are lower and less spread out in the larger rivers (Humber River and Peter's River) than at Leary's Brook and Waterford River. Take for example the spread of the sodium measurements shown in Figure 5.10. At Humber River and Waterford River sodium is in the 0 - 3.5 mg/L and no significant relationship can be defined. The range of sodium levels at the Leary's Brook and Waterford River stations is from 32.0 -390.0 mg/L and it was much easier to define a clear linear relationship with specific conductance measurements (Figure 5.11).

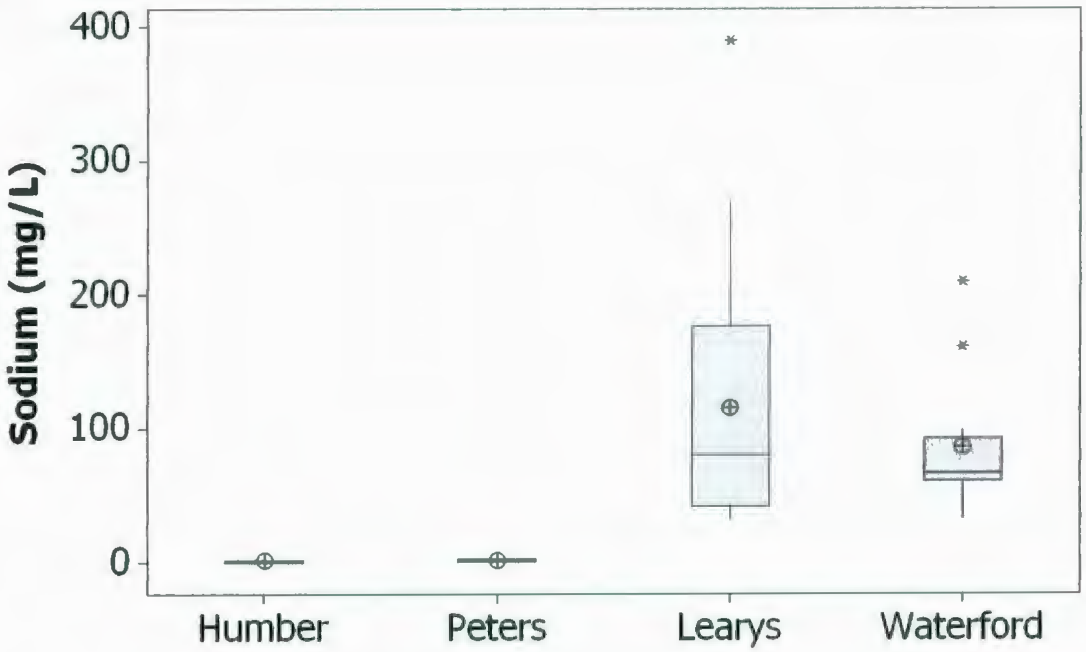


Figure 5.10 Comparing Sodium at the RTWQ Stations

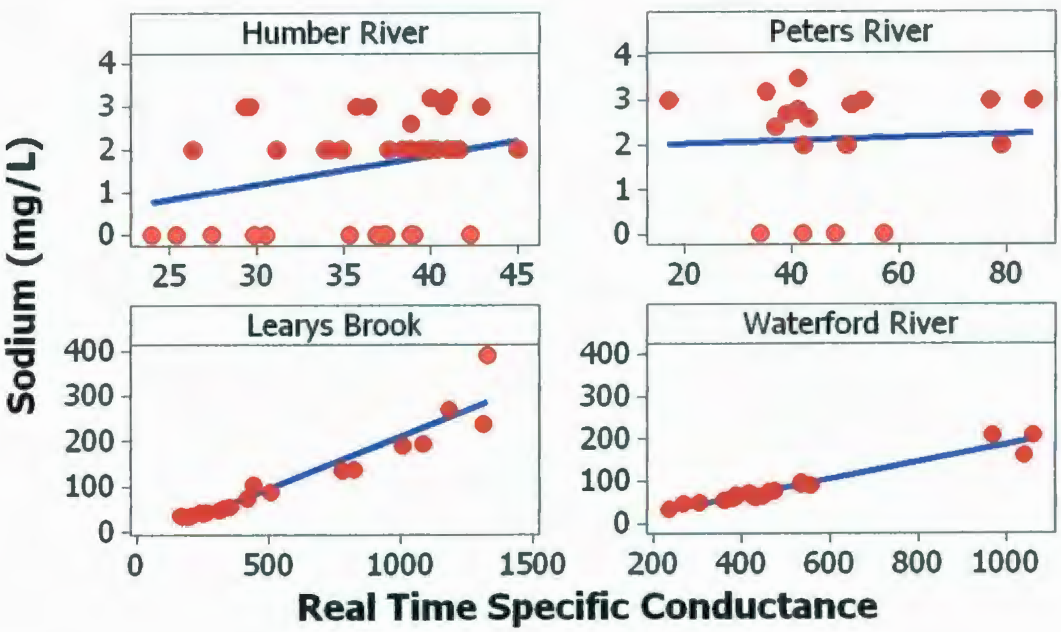


Figure 5.11 Relationship Between Specific Conductance and Sodium

Magnesium, Manganese, Iron and Barium

Although most of the regression models developed for magnesium, manganese, iron, and barium had adjusted R^2 values less than 0.70, one regression model using specific conductance as a surrogate for Leary's Brook grab sample Barium had an adjusted R^2 of 0.817 (Figure 5.12). It should be noted though although Barium levels at this station are the highest of the four stations, these barium levels should not pose an ecological threat. Here in Canada there is no CCME guideline for barium, but in the United States a 1.0 mg/L allowable limit for barium has been set for any freshwater that is to be used for domestic supply.

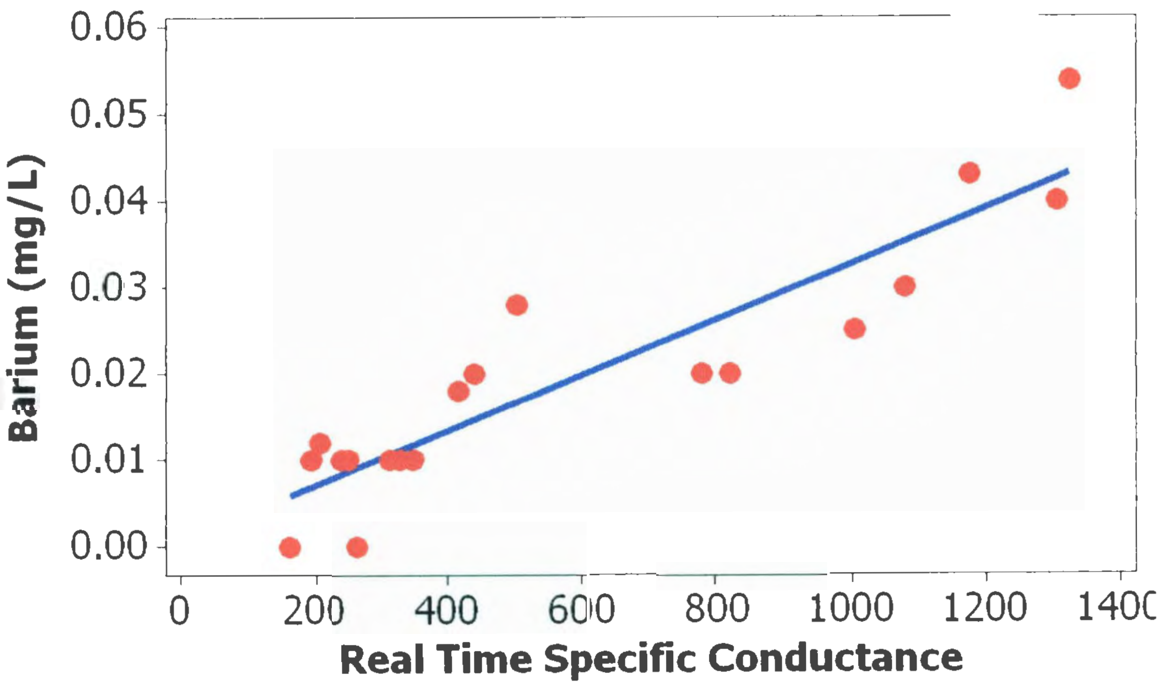


Figure 5.12 Regression Model for Leary's Brook grab sample barium

Zinc

Zinc is known to be very toxic to microscopic organisms in aquatic environments and the CCME has established a maximum allowable zinc concentration of 0.03 mg/L. Although the grab samples collected at Humber River, Peter's River and Waterford River never go outside of this limit, five of the Leary's Brook grab samples are above the limit (Zn 0.036, 0.04, 0.05, 0.079, and 0.086 mg/L). All five of these measurements occur during the winter months of February and March - Figure 5.13. During the winter months specific conductance at the station is also very high.

Unfortunately there is a fair amount of scatter in the grab sample zinc measurements (Figure 5.14) and the best regression model that could be developed for the Leary's Brook dataset has a low adjusted R^2 of 0.684. It is recommended that more grab samples of water quality be collected during these winter months to gain a better idea of how often zinc levels at the station are above the CCME and to hopefully improve the fit of the regression model.

Author's Note - the 0.036 grab sample measurement cannot be paired with a specific conductance measurement from the same day due to problem with the sensor. As a result the regression model was developed using only 19 of the 20 grab samples.

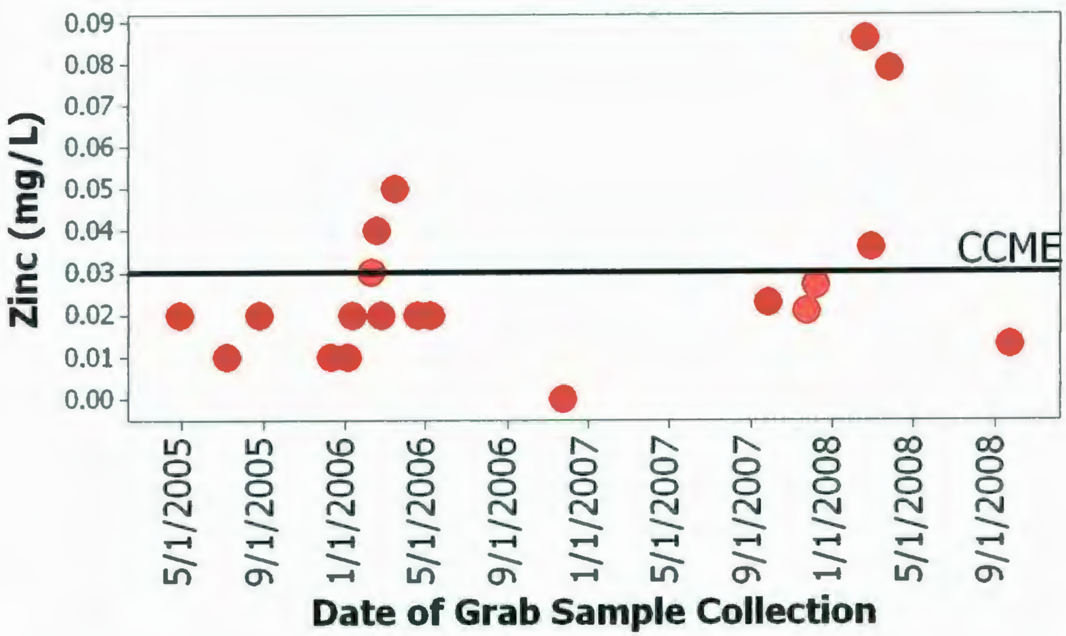


Figure 5.13 Five Leary's Brook Samples of Zinc are Above the CCME Guidelines

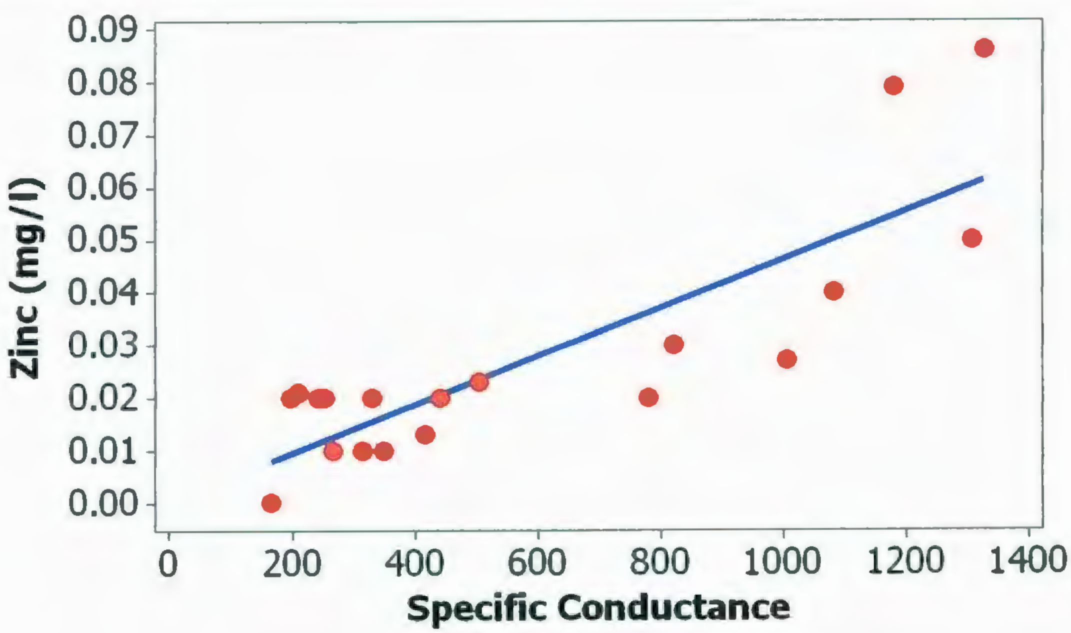


Figure 5.14 Leary's Brook Zinc Regression Model

Aluminum

Aluminum is one of the most abundant elements in the earth's crust but the presence of aluminum ions in water is usually the result of industrial waste. Aluminum is a concern for water quality as high concentrations of aluminum can become toxic to aquatic life if the pH is lowered (Kentucky Watershed Watch, 2009). A previous study of water quality carried out in Newfoundland noted that aluminum levels in most rivers in the province are the result of natural sources and should not pose a threat (CCME, 2004), four observations at the Waterford River station go above the CCME limit of 0.1 mg/L four times (0.14, 0.14 , 0.16 and 0.17 where these values come from both winter and summer grab sample measurements). Although all RTWQ data was examined as being a potential surrogate, the only reasonably well-fitting model was developed using logarithmically transformed stage - with an adjusted R^2 of 0.62 (Figure 5.15). Aluminum levels tends to be highest during the highest stage levels at the station.

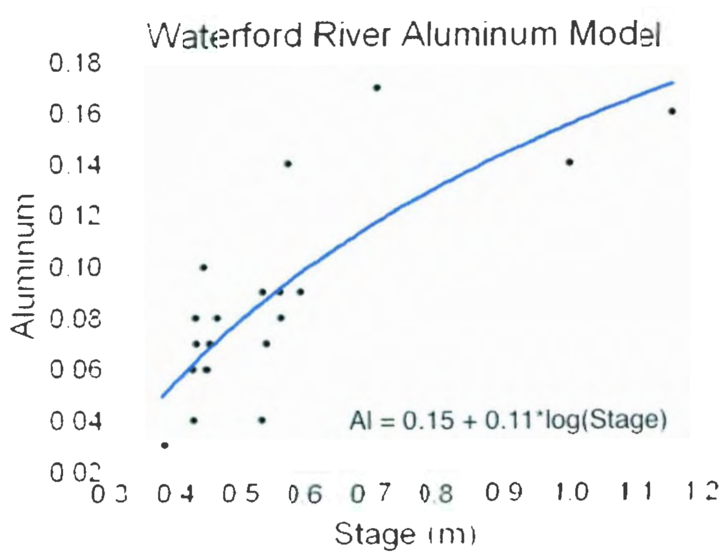


Figure 5.15 Waterford River Aluminum Model

5.6.3 Models for Grab Sample Nutrients

Its important to monitor nutrient levels at the stations as excessive nutrient levels can be harmful to aquatic organisms (Mueller and Helsel, 1996). The potential for developing regression models for nitrate(ite), kjeldahl nitrogen, dissolved organic carbon and total phosphorus was examined. There was very little success in finding real time surrogates for nutrient levels at the stations. No models could be developed for kjeldahl nitrogen and total phosphorus with adjusted R^2 values above 0.40. There is a small chance that models for nitrate(ite) might be possible, but at this point no models could be developed with adjusted R^2 values above 0.70.

Nitrogen and Kjeldahl Nitrogen

Nitrogen can take several forms in rivers and streams - elemental nitrogen, ammonia, nitrate and nitrite. The grab sample results for the real time stations contain one combined measurement for nitrate(ite). The nitrate(ite) measurements at the stations are not outside of the CCME guideline limit of 2.9 mg/L for nitrate. The USGS had some success using specific conductance and water temperature as surrogates in their research (adjusted R^2 of 0.829 for twenty measurements - Christensen et al., 2002) but the levels of nitrate in that stream were in the 0.014-2.13 mg/L range while the levels at the RTWQ stations are only in the 0-1.2 mg/L range.

No regression model could be developed for Leary's Brook nitrate(ite) with an adjusted R^2 above 0.10 and regression models developed for Humber River and Waterford River using water temperature as a surrogate for nitrate(ite) were poor (adjusted R^2 of

0.249 and 0.402, respectively). Water temperature and specific conductance were used as surrogates for Peter's River nitrate(ite) but adjusted R^2 was only 0.696. One outlier of nitrate(ite) 1.4 mg/L was removed from the Peter's River dataset for this model.

Kjeldahl nitrogen is calculated by taking the sum of the organic nitrogen, ammonia and ammonia levels in the grab sample. No statistically significant regression models could be developed for kjeldahl nitrogen at the stations.

Phosphorus

Phosphorus is a key element necessary for the overall health of aquatic ecosystems but in its elemental form it can be quite toxic to aquatic organisms. There are no CCME guidelines for phosphorus in freshwater but the United States Environmental Protection Agency suggests phosphorus levels should be no more than 0.1 mg/L for streams that do not empty into reservoirs (USEPA, 1986). Total phosphorus levels at Humber River, Peter's River, and Leary's Brook were all below this cutoff, but two grab samples collected at Waterford River violate this limit (0.11 and 0.31 mg/L). No surrogates could be determined for modeling phosphorus using the available datasets for the real time stations.

The USGS used turbidity, specific conductance and water temperature as surrogates for phosphorus measured in the 0.025-0.755 mg/L range. The RTWQ grab sample datasets cannot be paired with real time measurements of turbidity (as these measurements are currently unreliable) so it is unknown if turbidity is the missing link in developing these regression models.

Dissolved Organic Carbon

Dissolved organic carbon (DOC) can be used as an indictor of organic loadings in rivers and streams. The range of DOC grab sample measurements at Humber River, Leary's Brook and Waterford River was quite small and no statistically regression model could be developed for these stations. The range of DOC values at Peter's River were broader and a first order logarithmic model was developed for the station using specific conductance as a surrogate. Unfortunately the fit of this model was poor (adjusted R^2 of 0.54).

5.6.4 Further Investigations into the Humber River Grab Sample Models

There was great difficulty in developing regression models for the grab samples collected at the Humber River station. Major ion levels recorded in the grab samples collected at the station tend to be low and show little variation. It is quite likely that being the largest river in the provincial RTWQ network, all inputs into the river are diluted. When inputs are diluted they will not change the natural levels in the river. With that being said, there is little need to develop models for the Humber River station if the levels remain as constantly low as the available grab sample datasets indicate.

A comparison in the levels of grab sample measurements between a large river like the Humber River and a smaller stream like Leary's Brook is shown in Figure 5.11. The variation in grab sample sodium levels at the Humber River station was very small - specific conductance could range anywhere from 25 to 35 µS/cm and sodium would take on a value of 0, 2 or 3 mg/L). Sodium levels at Leary's Brook were spread over a much broader range (32.0-390 mg/L) for specific conductance anywhere from 200 to 1400 µS/

cm. The higher sodium levels at the Leary's Brook station occur during winter when large amounts of road salt being washed into the water after winter storms and during periods of heavy rainfall and snowmelt. Although these investigations into grab sample regression model development did not consider rainfall and snowmelt as potential surrogates, their inclusion for future model development would be quite useful for grab sample prediction.

If possible a large portion of the future Humber River grab sampling effort should be focused around collecting samples near significant events (i.e. sever rainfall, road salting during winter, etc.) to try to determine if ion, metal and nutrients concentrations at the station go outside of the ranges indicated by the current historical grab sample dataset. If higher observations of major elements and ions like sodium and chloride can be added to the historical dataset then there should be a greater chance of developing models for these stations.

Chapter Six

Investigations into the Use of Control Charts for Handling Real Time Water Quality Data

6.1 Scope

This chapter takes an investigative look into the use of statistical process control charts for monitoring water quality data collected by the Newfoundland RTWQ network. This chapter will be of most interest to those looking to take the control chart traditionally used for monitoring industrial manufacturing processes and use it for monitoring data that is highly autocorrelated in time. A literature review of the origins of statistical control charts and how they have been used in the field of environmental engineering is presented. This is followed by a look into the different approaches for implementing control charts for the RTWQ network that were investigated in this research.

6.2 Literature Review

The origins of the field of statistical quality control date back to the early 1920s and the work of Dr. Walter A. Shewhart who at the time was employed as a member of the technical staff at Bell Labs in the United States. Managers at the company relied on written reports for determining the quality of products being manufactured by the company and within these reports were charts showing the month to month performance of a quality characteristic. Most managers at Bell Labs found the charts difficult to interpret and as a result the managers were not able to easily distinguish variations in quality of the manufactured products that were due to chance (as most processes have some natural inherent variability) from variations that were the result of some actual change in the performance of the manufacturing equipment. In 1924, Shewhart drew limit lines around the historical average performance of the quality characteristic shown on these charts -

developing what is now called the Shewhart Control Chart (Figure 6.1). Shewhart's lines were set up so that points plotted outside of the lines had a low probability of occurring solely as a result of chance – i.e. points outside of the lines could only occur due chance 5% of the time. With these limit lines added to the charts, managers quickly see that points outside of the lines had 20 to 1 odds of being the result of some real change in the process. When a number of points in a row plotted outside these lines it was likely that the process had reached an out of control state and action would need to be taken to correct the problem.

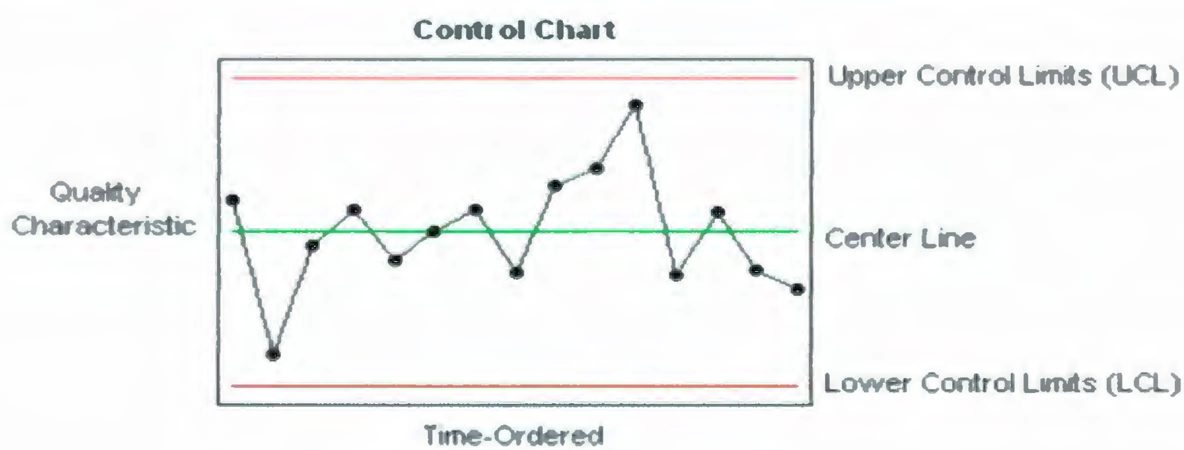


Figure 6.1 The Traditional Shewhart Control Chart

Shewhart knew from experience that even the best designed and well maintained processes would show some natural variability or background noise. If this background noise was small then there was no reason for inspection managers to be concerned. Occaisionally there are other kinds of variability present that will be much larger then the usual background noise on its own and are the result of some assignable cause. When a process operates in the presence of one of these assignable causes the process enters into an out of control state. It is not an unusual situation for a process to operate in a state of statistical control for a period of time but then an assignable cause will shift the process to an out of control state. Shewhart's control charts were designed so that managers could quickly detect the presence of these assignable causes and make the required change in the process to fix the problem.

Unfortunately Shewhart's control charts didn't catch on with the inspection managers working at Bell Labs - they were more concerned with sticking to production schedules than dealing with inferior products. It took about three decades after the charts were first developed before they were adapted by other manufacturing companies (Juran, 1997). Their popularity grew during the 1950s to the point that the Shewhart control chart (and variations of the original chart) is now recognized as being one of the more essential tools for handling variations in manufacturing and processing industry data.

Typically industrial and manufacturing measurements are independent over time with a constant mean and variance. Wardell et al. (1992) noted that quite often industrial process data is not independent but is actually autocorrelated - meaning that the value at one point in time is influenced by either the previous or following values. In these situations, using a traditionally designed control chart for monitoring the correlated data will lead to big problems. Vander Wiel (1996) examined the applicability of Shewhart individuals charts, cumulative sum charts, exponentially weighted moving averages and likelihood ratio schemes for monitoring data that is correlated over time. Signaling probabilities and average run lengths were used in his research to show that the cumulative sum chart can usually be designed to perform better than the other charts. The traditional Shewhart control chart in particular was found to be a poor choice for monitoring this correlated data. One of the more popular approaches for using control charts for handling correlated data is to first model the collected data using a Box and Jenkins Autoregressive Integrated Moving Average (ARIMA) time series model (Box and Jenkins, 1976) and then use control charts on the residuals of the data.

The idea of fitting a time series model to environmental data for statistical process control was first explored in the late 1970s and early 1980s. Berthouex, Hunter and Pallesen (1976, 1978) studied effluent from two sewage treatment plants and found that the standard Shewhart control chart could not effectively be used for monitoring the environmental data as the assumptions of constant mean, normal distribution and independence were violated. They took the approach of fitting an ARIMA time series model to the daily samples of the sewage treatment plant data and then used control charts on the residuals of the data so that the assumptions for the control charts would not be violated. In their research, the authors remind their readers that quite often the state of a particular monitoring operation may not require such high levels of statistical sophistication. Fitting ARIMA models can be quite difficult and it takes a great deal of experience to know what kinds of models will work best for the data being collected. In some scenarios it would likely be better to avoid spending unnecessary time modeling the data and just plot the raw data and use a simple control chart without any limit lines. The human eye is remarkably sensitive to changes in these kinds of plots and usually resource managers will be served well enough with the simpler plots.

Yourstone and Montgomery (1989) note how the majority of the time series approaches for environmental data at the time only focused on using a smaller sample of all available data to determine if a process was out of control in the past. In their work they propose a real time approach that allows the user to develop ARIMA models for the data being studied in real time and then use the residuals of this model fitting to determine if the process is in control. At that point in their research their work was still in the theoretical stage and only simulated data is used for evaluating the data and no real world results are discussed.

Alwan and Roberts (1988) modeled the systematic nonrandom behavior in data using ARIMA models proposed by Box and Jenkins (1976) and then developed two charts instead of one standard control chart for monitoring the data. The Common-Cause Chart - a chart of the fitted values from the ARIMA models was used to gain a better understanding of the process being studied, while the Special-Cause Chart - a chart of the residuals from the ARIMA models was used to detect any special causes. Although they do not explicitly work with environmental data, their approach is useful in that it can deal with correlation in the data and the SCC chart can be used to determine when a process goes out of control.

Wardell et al. (1992) found that traditional Shewhart control charts couldn't adequately handle autocorrelation in data but another control chart, the exponential weighted moving average control chart was adequate for detecting small shifts in the data if autocorrelation was not excessively high. The authors also investigated the Alwan and Roberts (1988) approach for developing control charts for autocorrelated data. The Alwan and Roberts approach was found to be a better option for reliably being able to detect large shifts in the data. Lu and Reynolds (2001) looked at using cumulative sum control charts for monitoring a process that could be modeled by an AR(1) time series. When autocorrelation was high the CUSUM charts worked better on the residuals from the time series model than they did on the original data.

MacNally and Hart (1997) were two of the first authors to publish research using control charts on actual water quality data. They studied the usefulness of cumulative sum (CUSUM) control charts for monitoring water quality trends within large storages like water reservoirs. The CUSUM approach was found to be effective for detecting changes in nutrient levels that were simulated for a water reservoir but the effectiveness of the approach hinged upon three important assumptions. First, the variance of the water quality parameter being studied had to remain constant over time and it had to be easily estimated. Second, there could be no serial autocorrelation within the time sequence of the data. Finally, there could not be any strong seasonal variation in the data and the samples being used had to be randomly selected. The authors spend a considerable amount of time in their paper discussing the importance of satisfying the assumption of no autocorrelation in the data - as CUSUM charts used on strongly correlated data resulted in an unacceptably high probability of making a Type-1 error (a so called false alarm where what looks to be a statistically significant change or problem on the control chart is not really a problem at all). A note is made in their paper that CUSUM charts will likely not be useful for handling seasonal or pulsed patterns in the data.

Smeti et al. (2007) found that typical SPC charts were of no use for dealing with highly autocorrelated daily toxicity data collected from treated-water tanks in Greece. They found that the approach of Alwan and Roberts (1988) was able to eliminate the autocorrelation in the daily toxicity data.

Manly (1994) proposed an adaptation of the CUSUM method for detecting systematic changes in one or more monitored variables at more than one site. All other previous research at that point had dealt with data collected only at one site but Manly looked into the situation of monitoring being preformed at a number of independent sites with data being collected at regular time intervals. His method of developing CUSUM charts was illustrated using water quality data collected from 48 Norwegian lakes over a 4 year period that was first presented in Mohn and Volden (1985). There was no autocorrelation or spatial correlation in the data. A procedure was developed for obtaining independent observations of water quality at each site by randomizing the observations so that there were no underlying systematic changes present in the data. This randomized data can then be compared to the observed data so that conclusions can be drawn regarding the presence of a systematic change. The CUSUM plots can then be used to determine the types of changes that might have occurred. Manly (1997) describes these randomization tests in greater detail.

Manly and MacKenzie (2000) build on the first work carried out in Manly (1994). Their approach is illustrated using a dataset consisting of dissolved reactive phosphorus measurements collected in 25 New Zealand rivers once every December from 1989 to 1996. Their CUSUM approach was effective for detecting changes in the distribution of a water quality variable but only for low levels of serial autocorrelation. When high levels of autocorrelation were present in the data for the individual sites the method needed to be modified. Manly (2002) developed a free piece of software known as the CUSUM Analysis Tool (or CAT 2.2) for implementing the CUSUM method proposed in Manly and MacKenzie (2000). The software is setup so that a user can enter in water quality data for a number of sites each year and then develop CUSUM charts for those years.

Manly and MacKenzie (2003) build on the CUSUM method described in Manly and MacKenzie (2000). Their original method is modified so that it can handle serial correlation at individual sites and moderate spatial correlation between sites. An algorithm is proposed that can be used to find a set of sites within a dataset that have negligible spatial correlation. Three examples are used to illustrate their new method. The first uses the dissolved reactive phosphorus dataset used in their previous work. Using the algorithm they reduce the dataset to include only 15 out of the 25 rivers. A follow up example uses data from Mohn and Volden (1985) - calcium and nitrate measurements taken from 48 lakes in southern Norway in 1976, 1977, 1978 and 1981. When analyzing calcium spatial correlation was high and only 3 out of the 48 lakes on the edges of the study area could be used for analysis. This limited dataset isn't very useful for monitoring purposes and their CUSUM method lost the majority of its effectiveness. For the nitrate measurements spatial correlation was lower and all 48 lakes could be used for analysis. Their CUSUM method was more effective for this larger dataset.

6.3 An Overview of the Most Commonly Used Types of Control Charts

Minitab can be used to develop a variety of statistical process control charts, from the simple Shewhart chart to the more complex exponential weighted moving average (EWMA) chart. These charts either plot individual observations or show subgroups of the data (i.e. combining 24 hourly measurements into one daily observation). Four of the most commonly used types of control charts were investigated in this research for monitoring the RTWQ data - (1) the traditional shewhart chart, (2) the cumulative sum chart, (3) the moving average chart, and (4) the exponentially weight moving average chart.

6.3.1 The Shewhart Chart (X-bar chart)

The Shewhart chart, also known as the X-bar chart in Minitab, is the simplest type of control chart for detecting a change in the level of a process. Berthouex and Brown (2002) note that the Shewhart chart does not indicate a change in the variability of a process but can be combined with a range chart so that the precision of the observations can be checked. Using these two charts together allows a user to track the process level and the process variation at the same time and detect the presence of special causes. The points plotted on the Shewhart chart at each recording interval is an average of the subgroup of n observations made at time t to calculate:

$$V_t = \bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_i$$ Equation 6.1

The acceptable amount of variation in the process level and the precision are defined on the charts by control limits that bind a specified percentage of all the results expected as long as the process remains in control (i.e. 99.7% of the values would be inside the limits if the process was in control). The control limits are only valid when the variation is randomly distributed above and below the average level of the process. The equations for the X-bar and Range control chart limits are defined as follows:

$$\begin{aligned} &\bar{X} \text{ chart} \quad \text{Central Line} = \bar{\bar{X}} \\ &\qquad\qquad \text{Control Limits} = \bar{\bar{X}} \pm k_1 \bar{R} \end{aligned}$$ Equation 6.2

$$\begin{aligned} &R \text{ chart} \quad \text{Central Line} = \bar{R} \\ &\qquad\qquad \text{Upper Control Limit} = k_2 \bar{R} \end{aligned}$$ Equation 6.3

where $\bar{\bar{X}}$ is the overall mean for the sample means (the average of the $\bar{X}$ used to make the chart), $\bar{R}$ is the mean sample range (the average of the ranges used to make the chart), and *n* is the number of replicates used to compute the average and the range at each sampling interval. *R* is the absolute difference between the largest and smallest values in the subset of *n* measured values at a defined sampling interval. The coefficients k_1 and k_2 depend on the size of the subsample used to calculate the overall mean and the mean sample range. If instead of subgroups of data the user decides to use only one observation at time *t* then:

$$V_t = y_t$$ Equation 6.4

and this is referred to as an individual observation chart or X-chart. Although an X-chart can be interpreted much like a Shewhart chart, using only one observation at each sampling time reduces the power of the chart to detected shifts in the performance (Berthouex and Hunter, 2002) and these charts are only useful for detecting large shifts in the process mean.

6.3.2 The Cumulative Sum Chart (CUSUM)

The CUSUM chart was first proposed by Page (1954) as an effective alternative to the Shewhart control chart. One of the biggest disadvantages of the Shewhart control chart is its inability to detect small shifts in the process and the CUSUM chart was designed to quickly detect small departures from the mean level (i.e. in fewer sampling intervals than the Shewhart chart). The CUSUM is known to be one of the best available charts for monitoring changes in the process level (Berthouex and Brown, 2002). The basic idea behind the chart is to plot the cumulative deviations from T, the mean or target level. The deviation at time t is $y_t - T$ while at time t-1 the deviation is $y_{t-1} - T$ and so on. All of these deviations are summed from time $t=1$ to the current time t to give the cumulative sum:

$$V_t = \sum_{t=1}^{t} (y_t - T)$$ Equation 6.5

Processes that are stable will show deviations that randomly vary around zero and the sum of the deviations from the target level will average zero. If the mean process performance shifts upwards over time then the deviations will tend to show more positive values and the cumulative sum values plotted on the chart will show an upwards trends, while the reverse is true for mean process performance that shifts downward. Unlike the Shewhart chart, control lines on the CUSUM chart are not parallel. One of the methods of adding these lines to the CUSUM chart is to use a procedure referred to as V-mask - where if all points fall within the arms of the V-mask then the process is in a state of statistical control.

6.3.3 The Moving Average Chart (MA)

The MA chart is a useful control chart for situations when single observations are used instead of subgroups of observations. A moving average gives equal weight to a sequence of the past values, where the weight will depend on how many of the past values are to be remembered. The time over which a moving average is to be calculated can be adjusted by the user in Minitab to represent the memory of the environmental system being studied as it responds to pollutants. The moving average is calculated by taking an average of the *k* most recent data points, or:

$$V_t = \frac{1}{k} \sum_{t-(k-1)}^{t} y_t \qquad \text{Equation 6.6}$$

Thus a daily moving average would use the latest 24 hourly observations made at the observation station. The moving average is useful for smoothing out random fluctuations in the data and can help the user focus on trends in the data.

6.3.4 The Exponential Weighted Moving Average Chart (EWMA)

The exponentially weighted moving average places more weight on the most recent observations than it does on older observations. The EWMA is calculated as:

$$V_t = (1-\lambda)\sum_{i=0} \lambda^i y_{t-i} \qquad \text{Equation 6.7}$$

where λ is a constant between 0 and 1 that determines the length of the EWMA memory. As λ increases from 0 to 1 the smoothing of the observations increases and trends in the data tend to stand out more clearly. When λ is kept small the memory of the EWMA is short and the weight given to older observations shrinks towards zero. When λ is larger, the EWMA has a long memory, but usually the EWMA is still dominated by the last four to six observations.

6.3.5 A Quick Summary of the Charts

The Shewhart chart ($\overline{X}$ chart) plots the average of a subsample and gives equal weight to all previous observations. It is a useful chart for checking shifts in the process that are relatively large compared with the variability in the process. This chart has become so popular as it is a direct plot of the data that lets the user visually inspect the observations. The Individuals (X chart) is similar to the Shewhart chart but plots individual samples instead of subsamples. It also tends to be insensitive to small shifts in the process.

The cumulative sum (CUSUM) chart gives equal weight to all previous observations but is much quicker at detecting small departures from the mean level than the Shewhart chart. The CUSUM chart serves the same purpose as the Shewhart chart but by plotting the cumulative change from the target level it is much quicker at detecting small changes in the observations and is a powerful improvement over the Shewhart chart. A potential downside of the CUSUM chart though is that it does not show the actual values of the observations.

The moving average (MA) chart gives equal weight to the *k* most recent observations and gives no weight to every other observation. Unlike the Shewhart chart which uses subsamples, the MA chart can be used for studying individual observations. The exponentially weighted moving average (EWMA) chart gives the most weight to the most recent observations. It is a particularly useful chart for taking into account any serial correlation and drift in data being studied.

6.4 What Charts Might Work Best for Water Quality Data?

A common approach for using control charts for process monitoring is to first develop a Shewhart chart for the data and then if necessary use additional charts (i.e. CUSUM, EWMA) to learn more about the process. By first plotting the observations on the Shewhart chart the user gets to visually inspect the data collected over time and identify any changes in the process. However, the Shewhart chart will not always provide useful information to the user (i.e. changes in the process might be too small for the chart to detect anything). In those cases follow-up charts let the user make further investigations into the collected data. Control charts for process monitoring will only provide the user with useful insights into the process if the underlying statistical conditions of the charts are satisfied - independence of the observations, constant variance and normally distributed variations). Water quality observations rarely satisfy the set of control chart conditions as observations of parameters like dissolved oxygen recorded over time will show signs of serial correlation and seasonality.

When water quality data is collected sequentially over time, there is a high tendency for those samples taken close together to be more similar than those taken farther apart. For example, dissolved oxygen might change a great deal over the course of a month but measurements made one hour apart will usually be very similar. This tendency for neighboring observations to be related to each other is referred to as autocorrelation. The autocorrelation function is the fundamental tool used by statisticians for diagnosing the structure of a time series and determining the amount of autocorrelation in a set of data.

The correlation of two variables (x and y) is defined by the correlation coefficient:

$$r(x,y)=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum(x_i-\bar{x})^2\sum(y_i-\bar{y})^2}} \qquad \text{Equation 6.8}$$

where the denominator in the equation is used to scale the correlation coefficient so that $-1\le r(x,y)\le 1$. Adjacent and nearby time series observations are correlated - i.e. z_t and z_{t-k} are correlated (where *k* is the lag distance, which is measured as the number of sampling intervals between the observations). The sample autocorrelation at lag of *k* is:

$$r_k=\frac{\sum_{t=k+1}^{n}(z_t-\bar{z})(z_{t-k}-\bar{z})}{\sum_{t=1}^{n}(z_t-\bar{z})^2} \qquad \text{Equation 6.9}$$

where *n* is the total number of observations recorded in the time series. This sample autocorrelation (r_k) is used to estimated the autocorrelation of the population (ρ_k). The denominator is used to scale the correlation coefficient so that $-1\le r_k\le 1$ (Berthouex and Brown, 2002). The inherent problems with using a control chart to investigate observations of real time data taken over a longer period of time (i.e. one deployment period of the sensor) is shown in Figure 6.2.

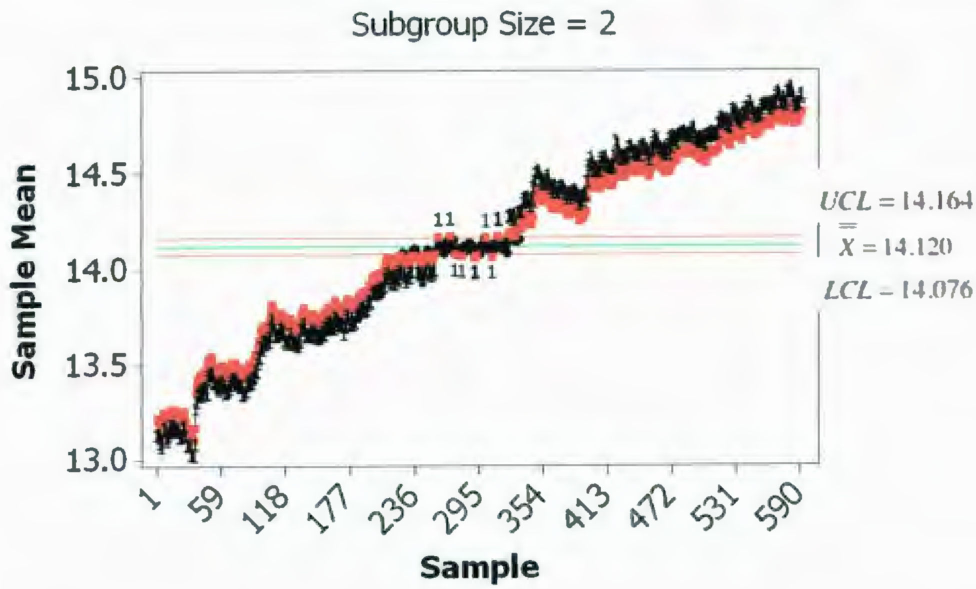


Figure 6.2 Shewhart Chart for Dissolved Oxygen in the Humber River (January to March 2005)

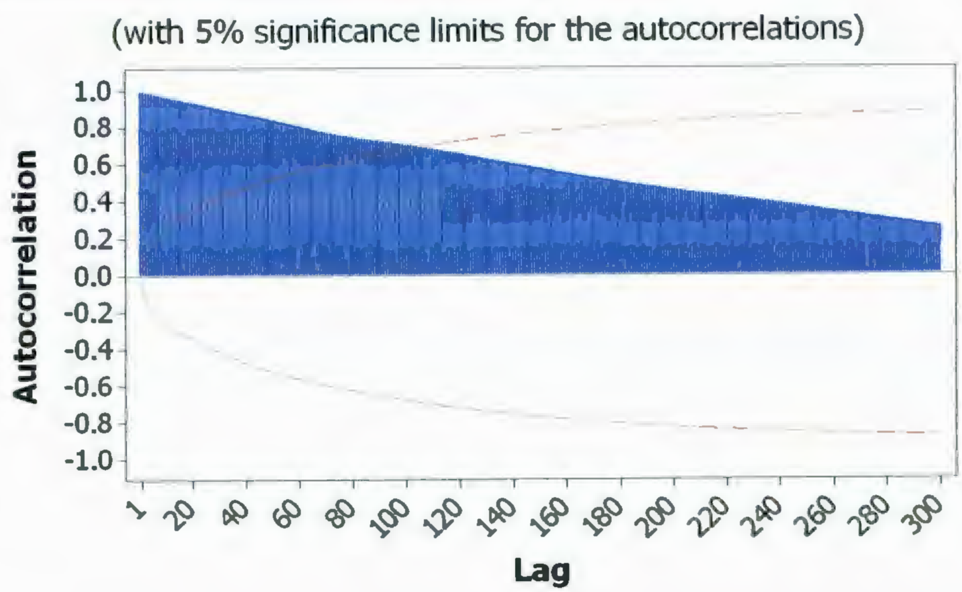


Figure 6.3 - Autocorrelation Function for Dissolved Oxygen Measurements in the Humber River (January to March 2005)

Figure 6.2 presents a Shewhart control chart (subgroup size = 2) developed for hourly measurements of dissolved oxygen collected at the Humber River station from January 14 to March 10, 2005 (one randomly selected deployment period of the sensor). Although the chart is useful in that it shows how the measurements of dissolved oxygen in the month increase from 13 to 15 mg/L over the course of the deployment period, the control lines (developed using three standard deviations from the mean) are so close to the overall process mean that virtually every observation is *out of control*. At first glance, the Shewhart chart makes it seem that there are major problems with dissolved oxygen at the station, but in terms of water quality, dissolved oxygen is still quite safe (never drops below the CCME guideline of 5.5 mg/L). The real problem here is that measurements of dissolved oxygen, unlike measurements of the weight of an auto part or a box of light bulbs, will vary over time (due to changes in water temperature and stage) and each sample of dissolved oxygen taken close together tends to be highly related. The autocorrelation function (ACF) shown in Figure 6.3 is the collection of the sample autocorrelation $r_k's$ for $lag = 0,1,2,...,(n/4)$ developed for the hourly dissolved oxygen observations. The two red dash lines represent confidence intervals for an independent series of data. Data with no autocorrelation would have coefficients that plot inside of the confidence intervals. This collection of hourly dissolved oxygen observations plots outside of these confidence bands so we know that autocorrelation cannot be ignored. It is not until a lag time of about one hundred hours (just over four days) that observations start to lose this high correlation.

6.5 Implementing Control Charts for the RTWQ Network - Results

Fortunately there are ways to implement control charts even when autocorrelation is high in the observations. Seven different approaches were investigated in this research:

- (1) Use control charts for detecting large shifts in the process over short periods of time

- (2) Develop charts using larger subgroups

- (3) Use control charts for monitoring monthly mean values

- (4) Use control charts to study the uncorrelated residuals from ARIMA models

- (5) Use control charts to study uncorrelated residuals from harmonic analysis of the data

- (6) Use the Manly and Mackenzie (2003) approach to compare measurements of water quality collected at different stations in the RTWQ Network.

- (7) Modify control chart limits to represent useful water quality limits

6.5.1 Approach 1 - Use Charts to Detect Large Shifts Over Short Periods of Time

Autocorrelation in short RTWQ datasets (i.e. 24 hourly measurements) tends to be much less than it is in longer RTWQ datasets (i.e. a complete set of hourly measurements for one month) and control charts like the EWMA chart can be useful for detecting large shifts in the real time measurements. Take for example one randomly selected day of real time observations in the Humber River dataset - September 30, 2006 (Figure 6.4).

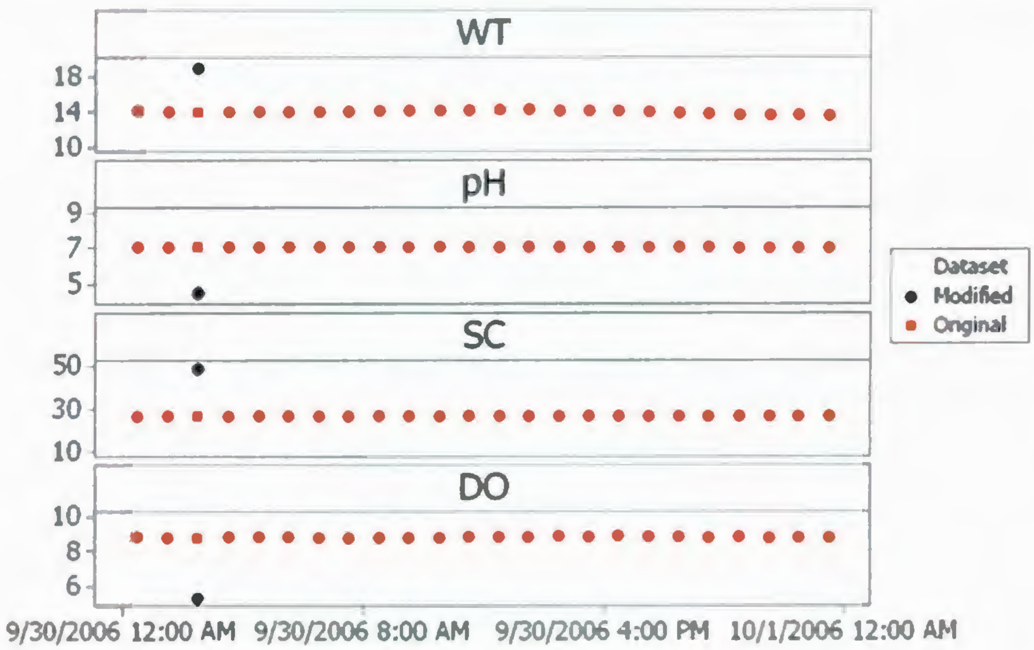


Figure 6.4 - Humber River RTWQ Measurements (September 30, 2006)

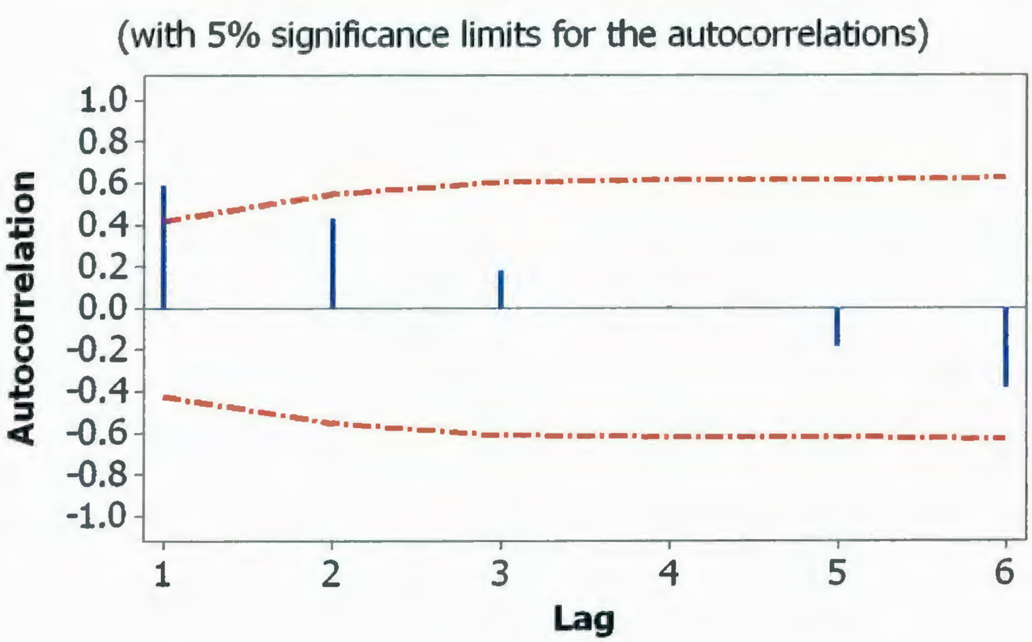


Figure 6.5 - Autocorrelation in the Dissolved Oxygen Values

All the 24 hourly measurements of water quality collected that day fall within safe levels and tend to remain fairly constant. The amount of autocorrelation in the dissolved oxygen values not as high as it was for the larger (but lag one still shows significant autocorrelation). An EWMA chart (with a subgroup size of 1 for each observation and a weight of 0.95 so that the EWMA has a longer memory) will show all the points as being in a state of statistical control - Figure 6.6.

A CUSUM chart (with a subgroup size of 2) developed for the same hourly observations will show all the observations as being out of control (Figure 6.7 presents the CUSUM chart for dissolved oxygen). The CUSUM chart tends to be too sensitive to the small variations in the hour to hour measurements recorded by the sensors.

The original dataset was modified to include one large shift in the levels of the real time observations (shown in Figure 6.4 as the modified datapoint). The same EWMA chart (subgroup = 1 and EWMA weighting of 0.95) will easily capture this out of control datapoint for all the real time parameters (Figures 6.8 - presents the plots for modified water temperature, pH, dissolved oxygen and pH).

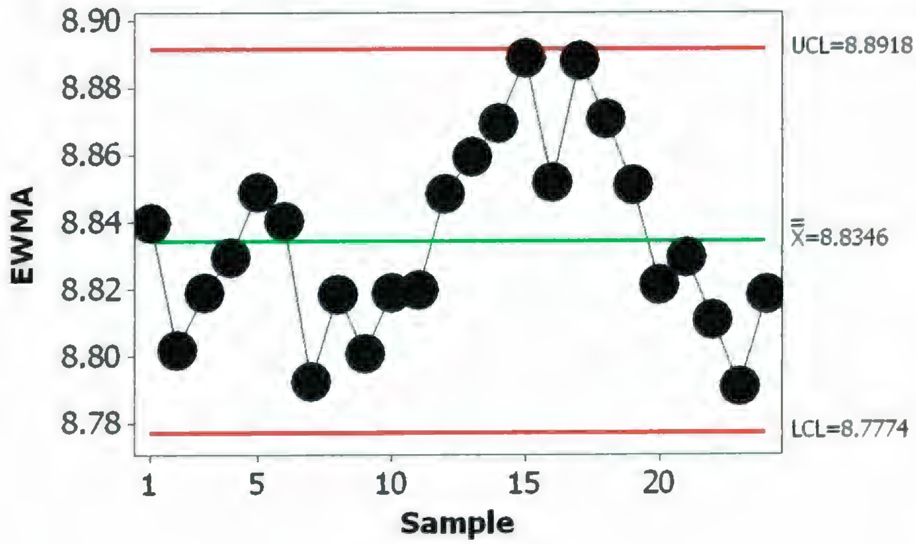


Figure 6.6 EWMA Chart (Subgroup = 1, EWMA = 0.95) for Original Dissolved Oxygen

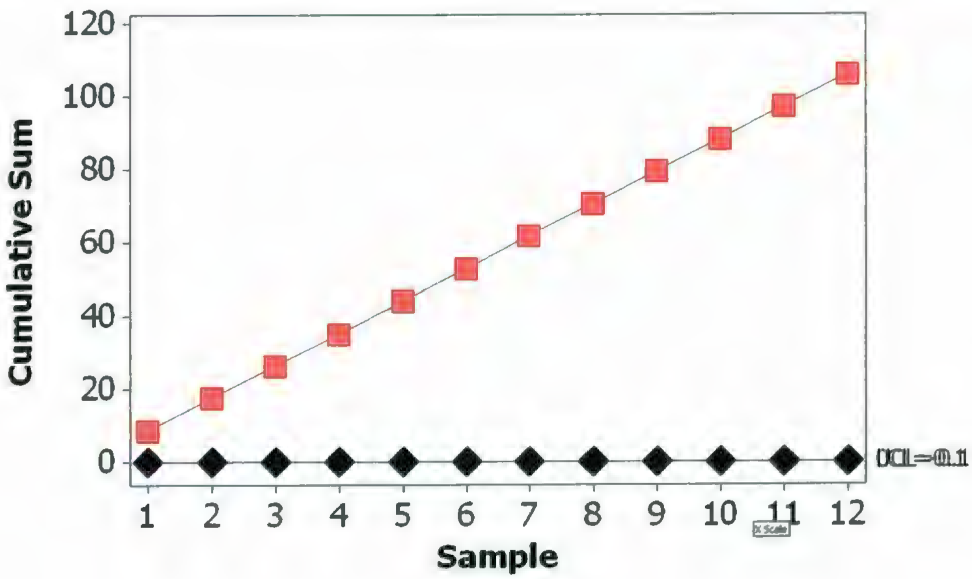


Figure 6.7 CUSUM Chart (Subgroup = 2) for Original Dissolved Oxygen

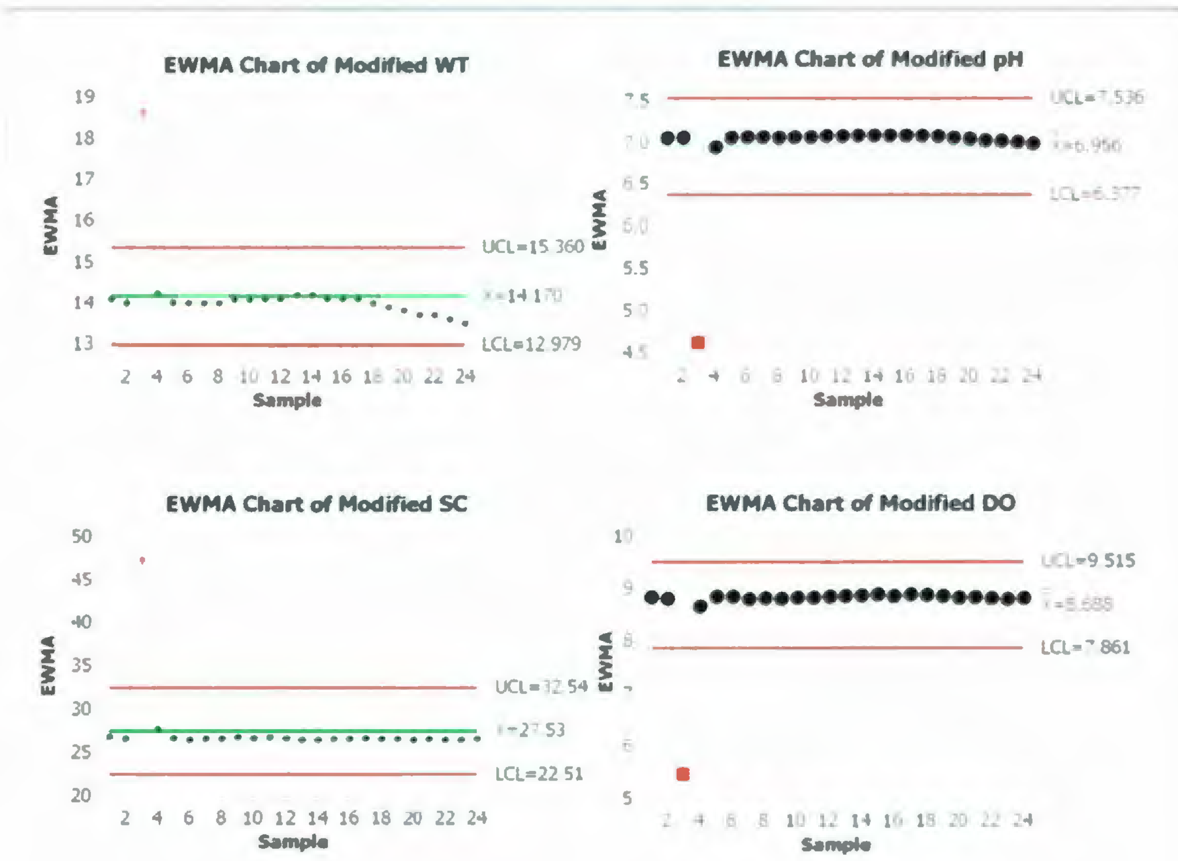


Figure 6.8 EWMA Chart (Subgroup = 1, EWMA = 0.95) for Modified Data

Experience in using EWMA charts for studying short time periods of RTWQ data has shown that the charts work best for identifying large shifts in RTWQ levels over short periods of time. If the charts are developed for studying a larger number of observations (i.e. one month of hourly observations), autocorrelation in the data tends to be higher and the observations tend to show trends (i.e. dissolved oxygen levels in July will start to drop as water temperature over the month rise). As a result the terms of the control chart are violated and an unrepresentative number of observations will show up as being out of control, even though these observations might only represent small shifts away from the overall mean of the process.

6.5.2 Approach 2 - Develop Charts Using Larger Subgroups

Increasing the size of the subgroups used in the control charts can increase the power of the chart to detect shifts in performance. A second approach to implementing the control charts sought to increase the size of the subgroups used in the control charts. The EWMA (subgroup = 1, EWMA = 0.95) developed for the 24 hourly measurements collected at the Humber River station was useful in that it could pick up on large shifts in the levels of the RTWQ parameters. However for one week and one month of hourly measurements, the same EWMA tended to show both large shifts and a number of smaller shifts as being out of control.

Figure 6.9 presents a EWMA (subgroup = 1, EWMA = 0.95) for one week of hourly measurements of dissolved oxygen collected at the Humber River station from September 30 to October 6, 2006. The chart flags 76 observations as being out of control (approximately 45% of the total number of hourly observations collected over the week) as being out of control. Increasing the subgroup size to 12 (Figure 6.10) is a little easier to work with but still roughly half of the observations are shown as being out of control. The increased autocorrelation and variation in the dissolved oxygen measurements over the course of the week have reduced the usefulness of the plot for monitoring the hourly measurements.

The benefit of increasing the subgroup size is only truly apparent when there is a large shift in the observations at the station. The same week of hourly measurements recorded at the Humber River station was modified to include a dissolved oxygen reading significantly lower then the rest (the original range of the data was in the 8.7 to 9.1 mg/L range and a value of 6.50 mg/L was added). Figure 6.11 presents an EWMA chart with a subgroup size of 1 - the large shift can be identified but there are a number of other points that are being flagged by the chart as being out of control. Figure 6.12 presents an EWMA chart with a subgroup size of 12 - it would be much easier to work with this chart to note that somewhere in the first subgroup there is an observation that is unusually low for the week and it should be investigated further. Although increasing the size of the subgroups also provides the advantage of decreasing the amount of clutter on the plots there is a downside to this design approach - with larger subgroup sizes, the points plotted on the charts out of control will be detected as a subgroup which may contain individual observations that are still in control. This mixing of in and out of control observations into large subgroups will result in difficulty when picking up individual abnormal observations of water quality and will require extra monitoring effort on behalf of the user to look through the raw data to further investigate the subgroup observations.

Investigations were made into using larger subgroups for studying monthly sets of hourly observations and even full deployments periods of hourly observations, but control charts did not work well for handling the variation in the water quality measurements over these longer periods of time (with over 75% of the observations showing as being out of control). Although traditional control charts could potentially be used to identify large shifts in the data, any chart developed for an ordinary dataset (with no large shifts in the levels and only day to day variations in the levels) will show

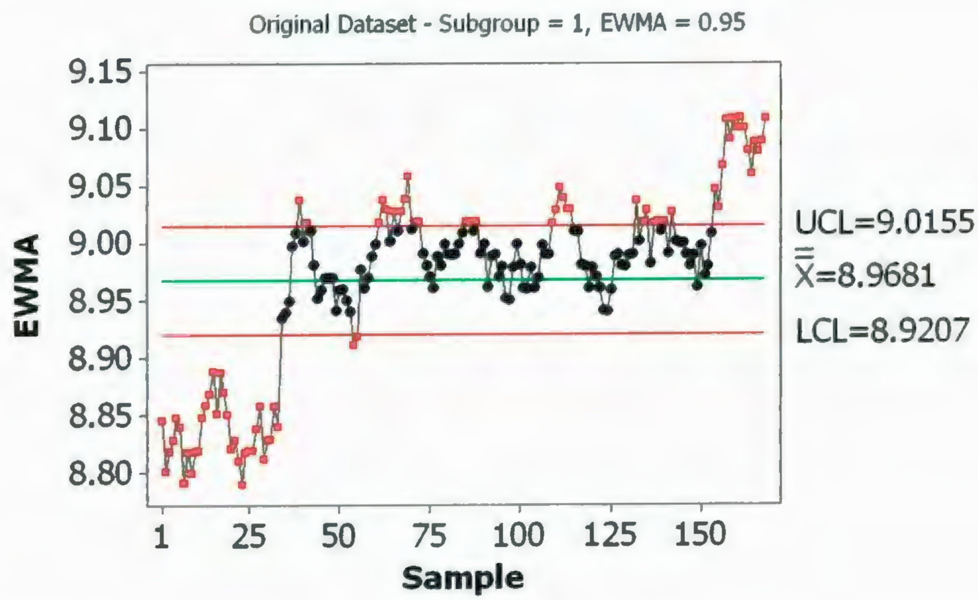


Figure 6.9 - EWMA Chart (Subgroup = 1, EWMA = 0.95) for One Week of Hourly Measurements

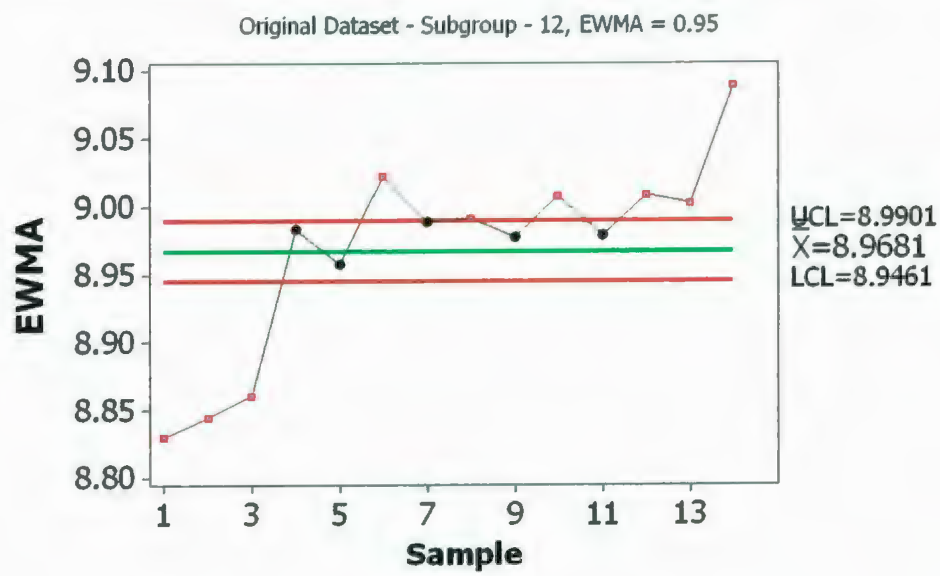


Figure 6.10 - EWMA Chart (Subgroup = 12, EWMA = 0.95) for One Week of Hourly Measurements

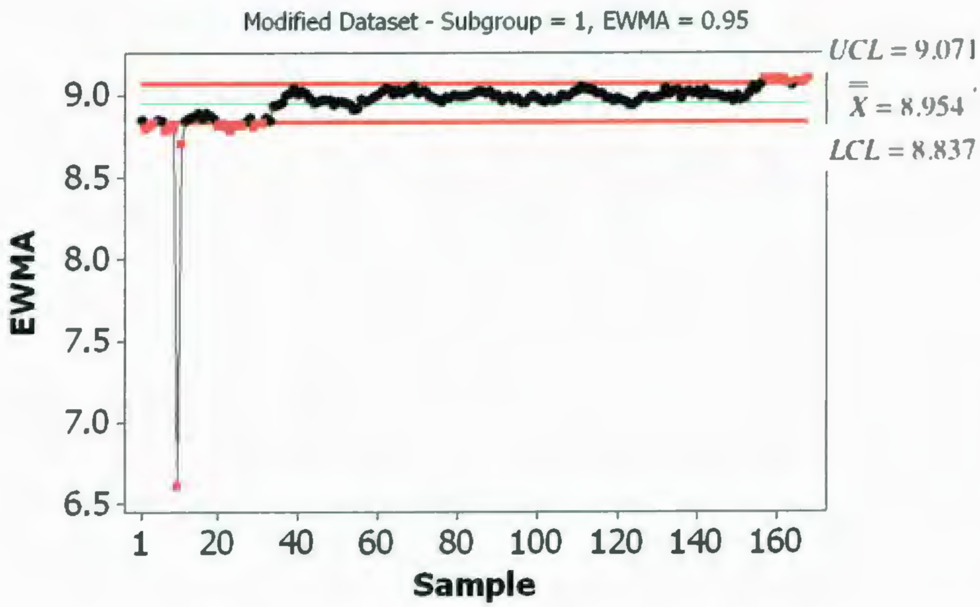


Figure 6.11 - EMWA Chart (Subgroup = 1, EWMA = 0.95) For Modified Week of Hourly Measurements

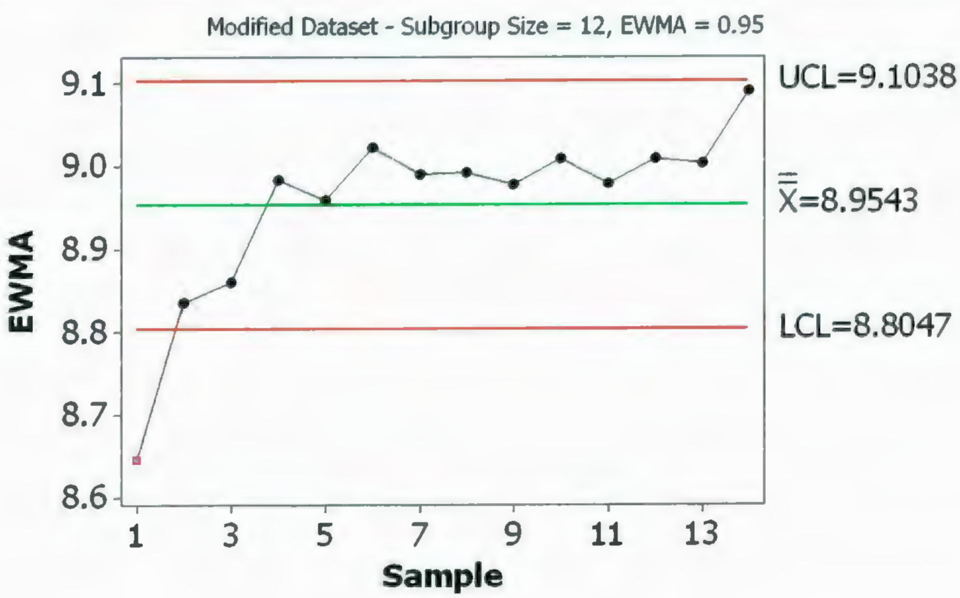


Figure 6.12 - EMWA Chart (Subgroup = 12, EWMA = 0.95) For Modified Week of Hourly Measurements

6.5.3 Approach 3 - Use Charts to Study Uncorrelated Monthly Mean Values

One approach to using the control charts to study longer stretches of real time data is to increase the time scale of the observations being studied - i.e. instead of using the charts for highly correlated hourly observations use them for studying monthly mean values that are much less autocorrelated.

Investigations into daily mean observations of real time data showed that autocorrelation is still significant enough to pose a major problem for the control charts. It is only somewhat significant for weekly mean datasets and only slightly significant for monthly mean datasets. A comparison of the autocorrelation functions for mean weekly and monthly dissolved oxygen recorded at Humber River is shown in Figure 6.13 - where for the weekly mean data, autocorrelation is no longer significant after a lag of seven weeks while for the monthly mean data autocorrelation is no longer significant after a one month. Shewhart control charts can be developed for the weekly and monthly observations - Figure 6.14 shows the monthly Shewhart chart (subgroup size of 2 months) developed for a dataset consisting of monthly mean dissolved oxygen values recorded at the Humber River station during 2006.

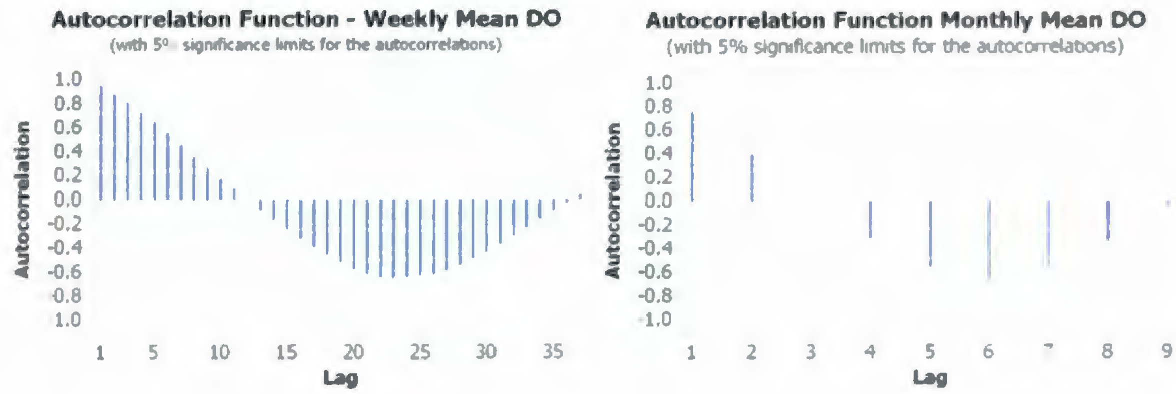


Figure 6.13 Autocorrelation Functions for Weekly and Monthly Mean DO

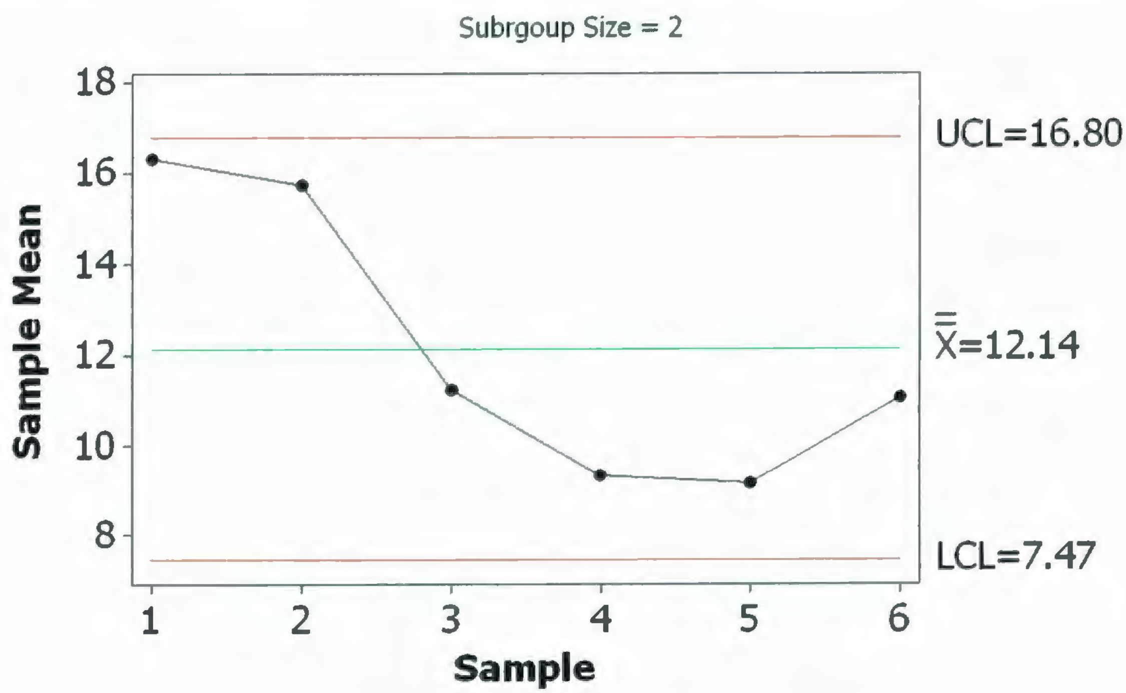


Figure 6.14 Shewhart Chart (Subgroup = 2) for Monthly Mean DO (2006)

The Shewhart chart shows none of the monthly mean dissolved oxygen levels over the year were out of statistical control.

A CUSUM chart for the data tends to be too sensitive to the variations in the dissolved oxygen levels (Figure 6.15) and also provides the challenge of needing to define a target level for the measured observations. For water quality parameters that change as the seasons pass, CUSUM charts developed over long periods of time will be of little use for monitoring purposes and will tend to show the majority of observations as being out of control. An EWMA chart developed for the same monthly data with a subgroup size of 2 and an EWMA of 0.95 (Figure 6.16) is similar to the Shewhart chart in that all observations are shown as being in control. picks up on five of the monthly means as being out of control. When the EWMA is lowered to a range of 0.2 to 0.3 more observations are likely to show up as being out of control as the memory of the chart is quite short. It does need to be noted though that the chart is being developed for a dataset consisting of observations that vary throughout the year and are not constant - so technically the terms of that make the control chart valid are being violated.

Control charts can also be developed for studying the weekly mean observations for Humber River in 2006 (Figure 6.17). The problem with any charts developed for the weekly data though is that autocorrelation in the dataset is more significant than it was for the monthly data and it is likely that the control limits being plotted on the charts are being affected by the autocorrelation - an unrepresentative number of points might be showing up as being out of control.

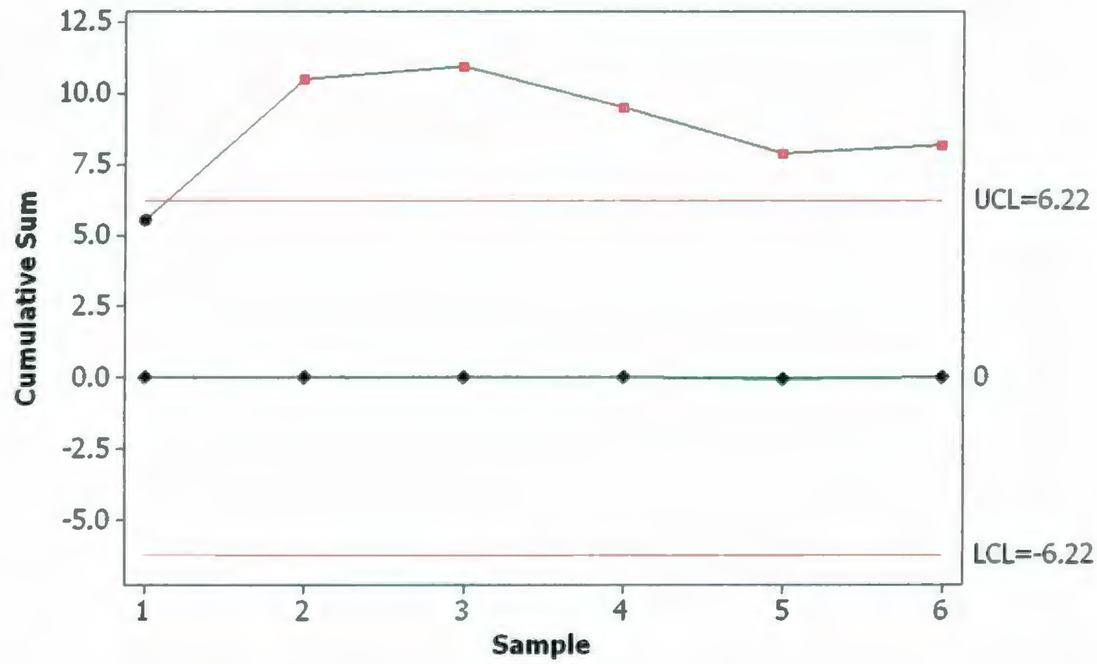


Figure 6.15 - CUSUM Chart (Subgroup = 2, Target = 10.0) for Monthly Mean DO (2006)

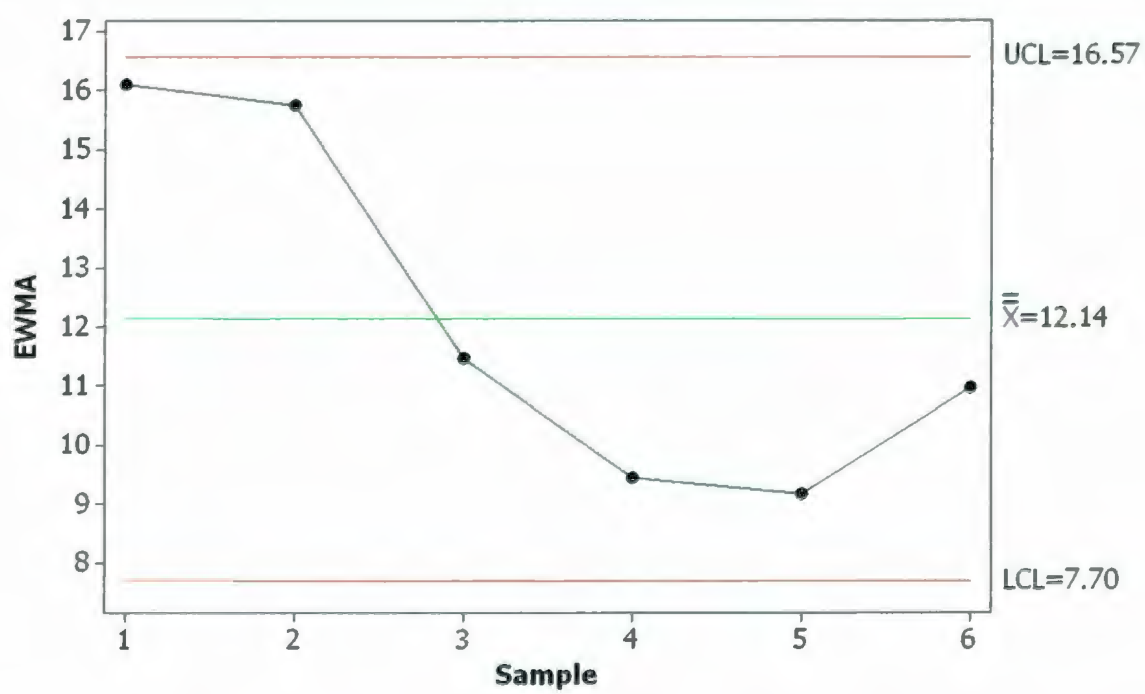


Figure 6.16 - EWMA Chart (Subgroup = 2, EWMA = 0.95) for Monthly Mean DO (2006)

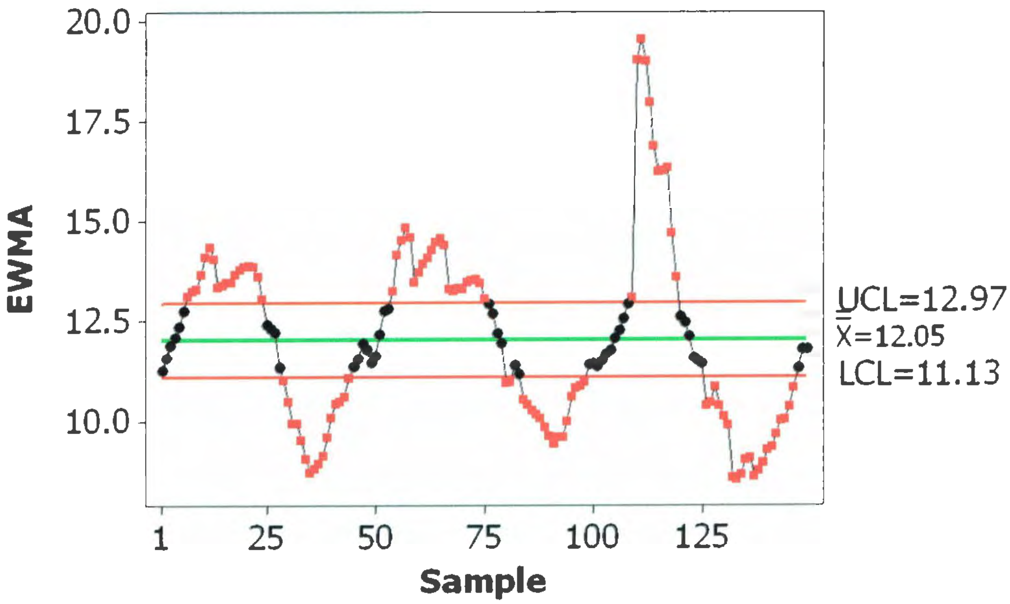


Figure 6.17 - EWMA Chart (Subgroup = 1, EWMA = 0.95) for Weekly Mean DO

6.5.4 Approach 4 - Use Charts to Study Uncorrelated Residuals from ARIMA

The Alwan and Roberts (1988) approach for using control charts for highly autocorrelated data involves first fitting an autoregressive integrated moving average (ARIMA) time series model to the observations and then using control charts to study the residuals from the fit of the model. ARIMA is used to describe a family of models that can tend to be quite complicated. The simplest of the models use the autoregressive (AR) part to describe a stationary time series whose values fluctuate about a fixed level. Other simple models use only the moving average (MA) part to describe a non-stationary process that drifts over time and that does not have a fixed mean. The more complicated models can end up using both the AR and MA parts together and can include additional features that account for seasonality, drift and trends in the data.

The Box-Jenkins methodology is used to fit ARIMA models to environmental data - where the type of model to be used is first identified, the parameters are estimated, and residuals from the fitting are checked for normality, constant variance and independence (Box et al., 1994). Although the steps of the approach appear to be rather straightforward, the analysis of time series data tends to be a frustrating and mysterious subject that most textbooks tend to avoid. The fitting of appropriate ARIMA models takes a considerable amount of experience working with time series data. This is not to say that ARIMA models cannot be developed for the RTWQ network data as when levels of autocorrelation are low (i.e. for monthly mean observations) it is possible to use Minitab to fit models and then use control charts on the residuals to identify points of interest. When autocorrelation is higher (i.e. for daily means or longer stretches of hourly observations) it becomes extremely difficult to fit models to the data - in fact entire research papers are often dedicated fitting ARIMA models to just one highly autocorrelated dataset. This is perhaps the biggest downside of the Alwan and Roberts approach - for every set of data to be studied using this approach it is necessary to go through the process of developing an ARIMA model and then analyzing the residuals to find unusual points. As a result this approach is probably of more use to the WRMD for research purposes then it is for day to day monitoring of the stations. An example of how to use the approach for studying slightly autocorrelated data is given below.

Fitting ARIMA Models to Monthly Observations - An Example

A dataset consisting of monthly mean pH levels recorded at the Humber River station is shown in Figure 6.18. The first step in fitting an ARIMA model to this data is to use Minitab to look for any linear trend in the observations over time (Figure 6.19). The plot shows that there is a slight upward trends at the station but in this case it is not necessary to account for the trend in the model. Minitab next needs to be used to check for normality of the data (Figure 6.20). In this case the monthly mean pH values are normal as the probability value is greater than 0.05. Autocorrelation in the data is checked using the autocorrelation function and partial autocorrelation function plots in Minitab (Figure 6.21). The plots tells us that autocorrelation at lag one is important, while lag one and lag four is important for partial autocorrelation.

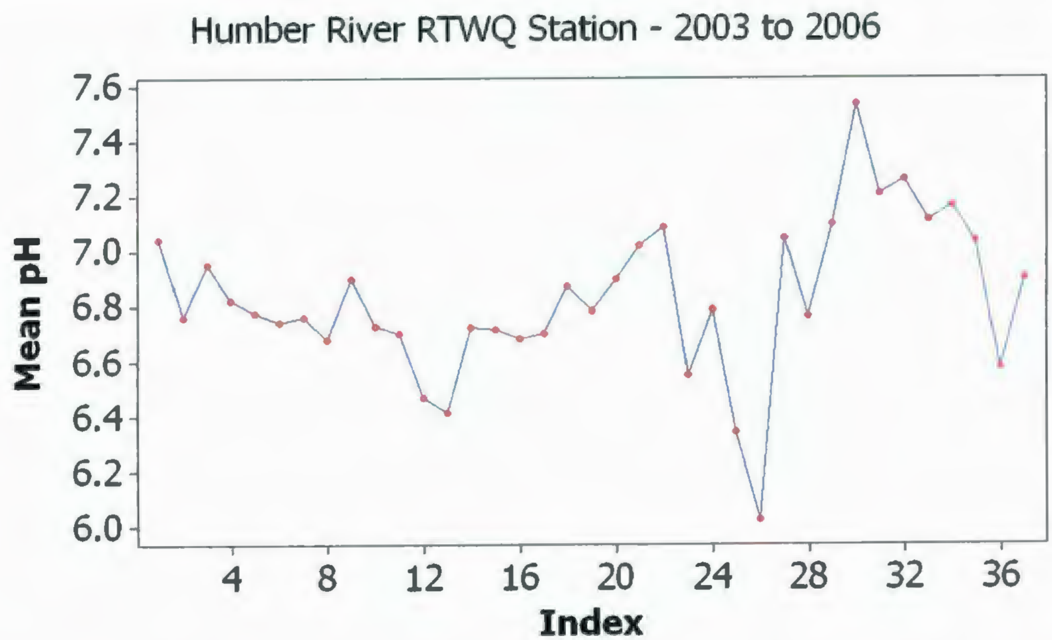


Figure 6.18 Time Series Plot of Monthly Mean pH Levels at Humber River - 2003 to 2006

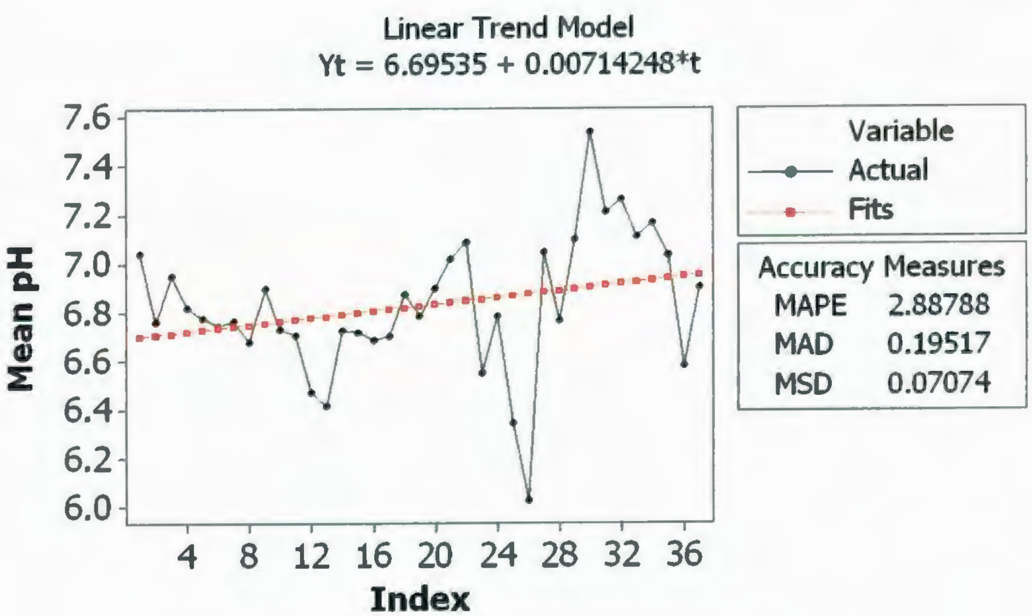


Figure 6.19 Linear Trend Analysis for Monthly Mean pH Levels

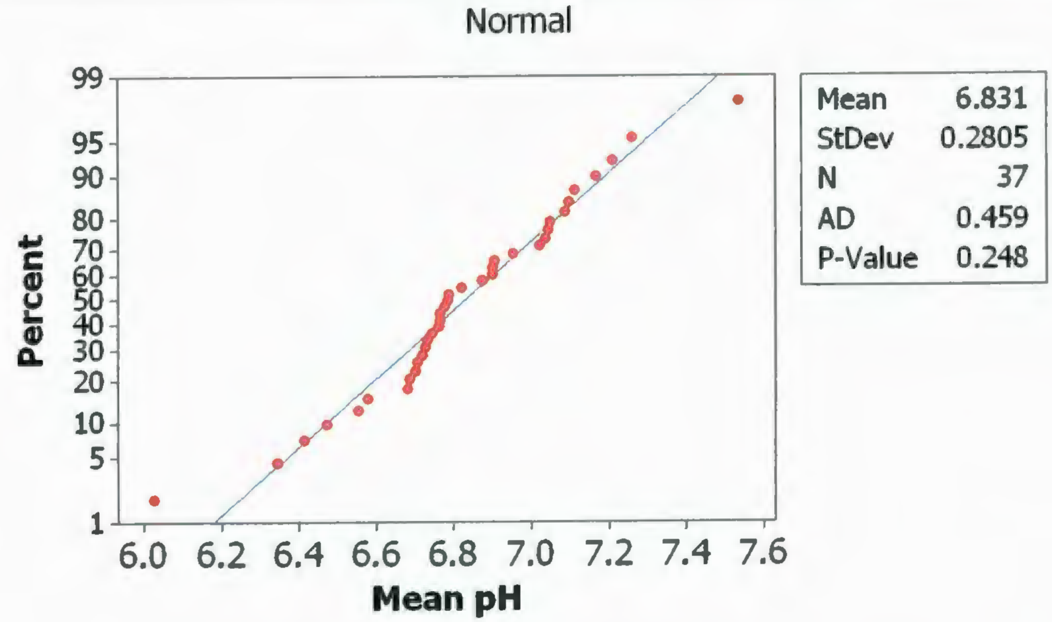


Figure 6.20 Normal Probability Plot of Monthly Mean pH Levels

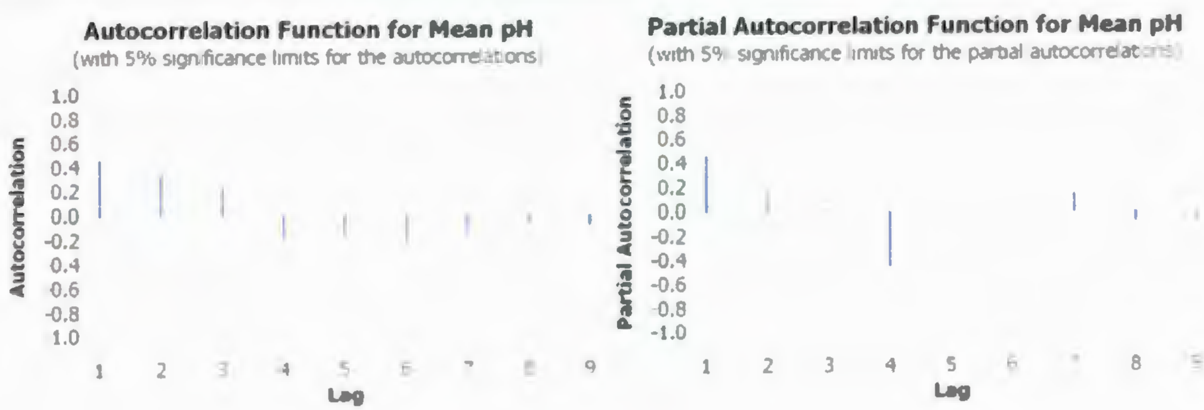


Figure 6.21 ACF and PACF for the Monthly Mean pH Levels

With linear trend, normality, autocorrelation and partial autocorrelation examined its then possible to start fitting ARIMA models to the data in Minitab (Figure 6.22). In this particular case a first order autoregressive model, AR(1), is the best choice for the monthly mean pH observations. *Note - in some cases there may be a number of models that work well and the model with the fewest number of parameters should be used (principle of parsimony).*

```
Final Estimates of Parameters

Type          Coef  SE Coef      T      P
AR   1      0.4719   0.1490   3.17  0.003
Constant  3.61111  0.04129  87.47  0.000
Mean      6.83771  0.07817


Number of observations:  37
Residuals:     SS =  2.20631 (backforecasts excluded)
               MS =  0.06304  DF = 35
```

Figure 6.22 Minitab Output for AR(1) Model Parameters

Once the parameters for the model are determined then the residuals need to be checked for normality, constant variance, and independence (Figure 6.23) autocorrelation and partial autocorrelation (Figure 6.24). The plots show that the diagnostics for the fitting are all fine and that there is no major autocorrelation in the residuals.

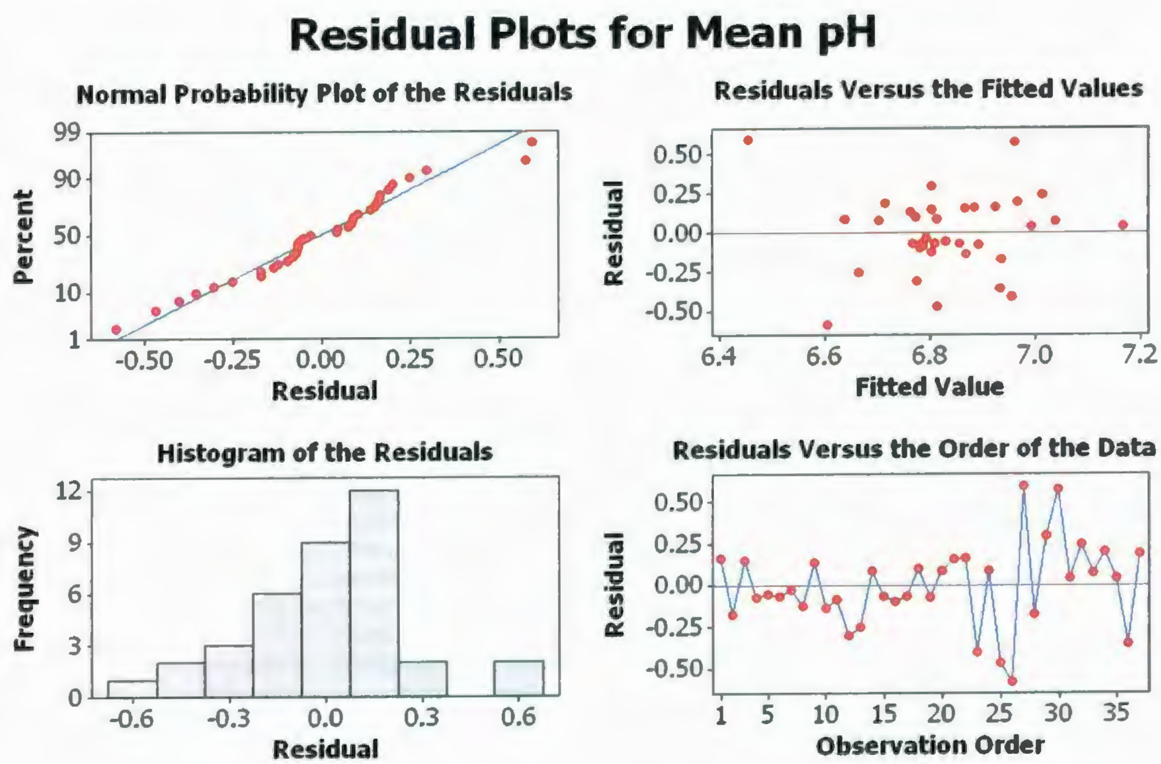


Figure 6.23 Checking Residuals for the AR(1) Model for Monthly Mean pH

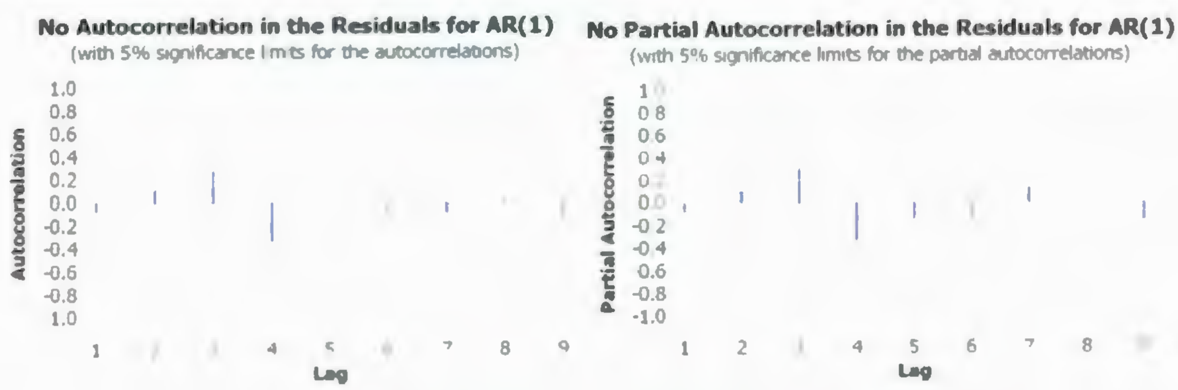


Figure 6.24 ACF and PACF for the Residuals of the AR(1) Model

It is now finally possible to go ahead and fit a control chart to the data. Figure 6.25 presents a Shewhart chart with a subsample size of 2 for the residuals. The control chart shows that subsample 13 is out of control. Looking back at the original data this subsample represents an average monthly mean in December 2005 and January 2006 of 6.03 pH units.

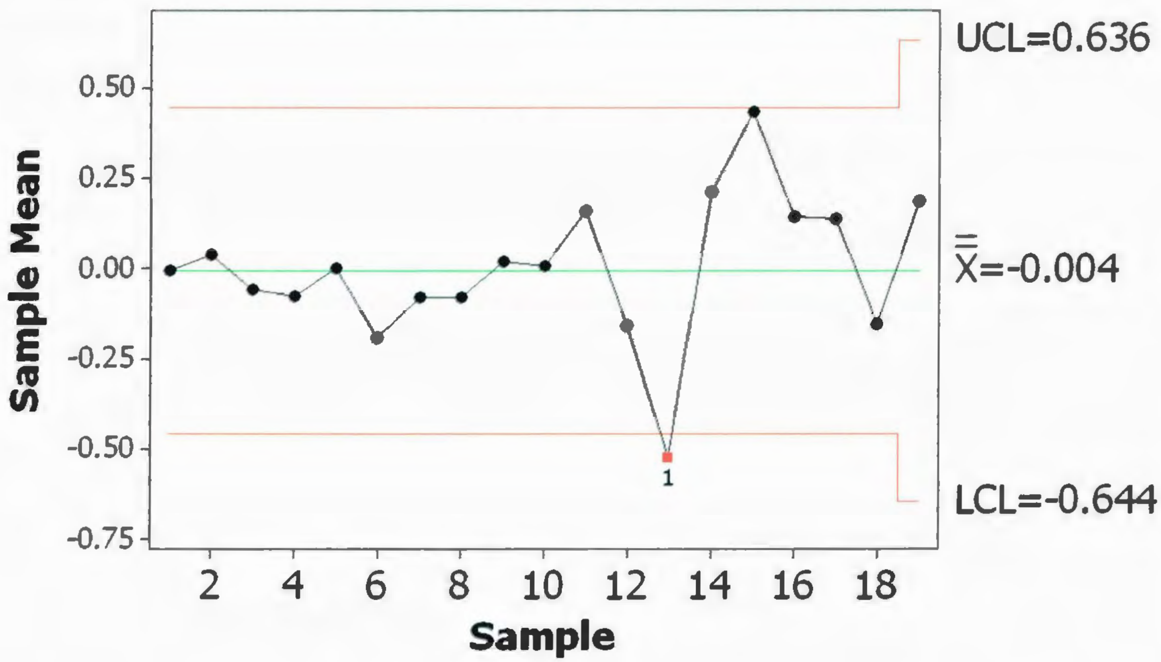


Figure 6.25 Shewhart Chart for the Residuals of AR(1) Model

It is also possible to use the ARIMA approach for control charts for handling observations with seasonality - i.e. monthly observations of dissolved oxygen. Appendix Z includes the results of fitting an AR(1) model with a seasonality component to Humber River monthly mean dissolved oxygen.

The Difficulties in Fitting ARIMA Models to Daily Observations - A Brief Example

The difficulties in fitting an ARIMA model to observations with high amounts of autocorrelation periods can be demonstrated by quickly looking at a set of mean daily pH levels at the Humber River station collected during one sensor deployment period (Figure 6.26). Analysis of the data in Minitab showed that these daily mean observations show a significant downward linear trend, the data is significantly not normal (with a p-value in the normal plot less than 0.005), autocorrelation is significant up to a lag four and partial autocorrelation is significant at lag one. Fitting a model to this data proved much too difficult for the purposes of the WRMD (i.e. the simple AR or MA models will no longer work and something more complex needs to be developed). Eliminating autocorrelation in the residuals proved to be just as difficult. Trying to fit models that would show high autocorrelation along with seasonal or daily variation (i.e. water temperature) would be next to impossible.

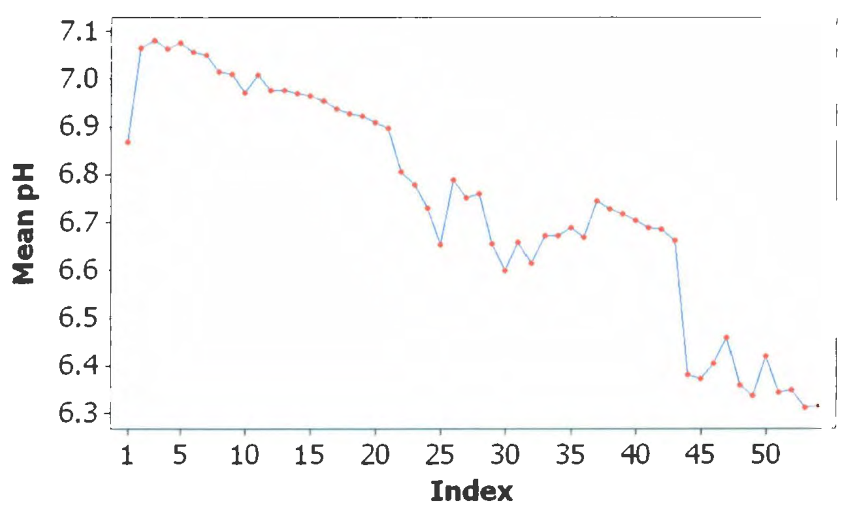


Figure 6.26 Fitting ARIMA Models to the Autocorrelated Daily Mean pH is Difficult

6.5.5 Approach 5 - Use Charts to Study Residuals from Harmonic Analysis

Harmonic analysis can be used in a similar manner to the ARIMA approach for finding uncorrelated residuals that can be examined using control charts. Like the ARIMA approach, spectral analysis of time series data is a rather complex subject. This approach proved to be useful for studying observations with low levels of autocorrelation but was not useful for studying hourly and daily observations collected over longer deployment periods with high levels of autocorrelation.

An Example of Using Harmonic for Low Levels of Autocorrelation

The time series plot of monthly mean water temperature presented in Figure 6.27 shows a smooth periodic function that repeats itself every year with no major trend over time. From everyday experience, its already known that the time period for water temperature will follow a 12 month (yearly) cycle. This periodicity in the data is due to the revolution of the Earth around the sun and is quite common for many hydrological parameters.

Author's Note - for those cases when the time period for a set of observations is not already known, it is possible to use a technique known as spectral analysis to identify the time period of a set of data. Appendix P illustrates how Minitab can be used to carry out spectral analysis to find the time period for a dataset.

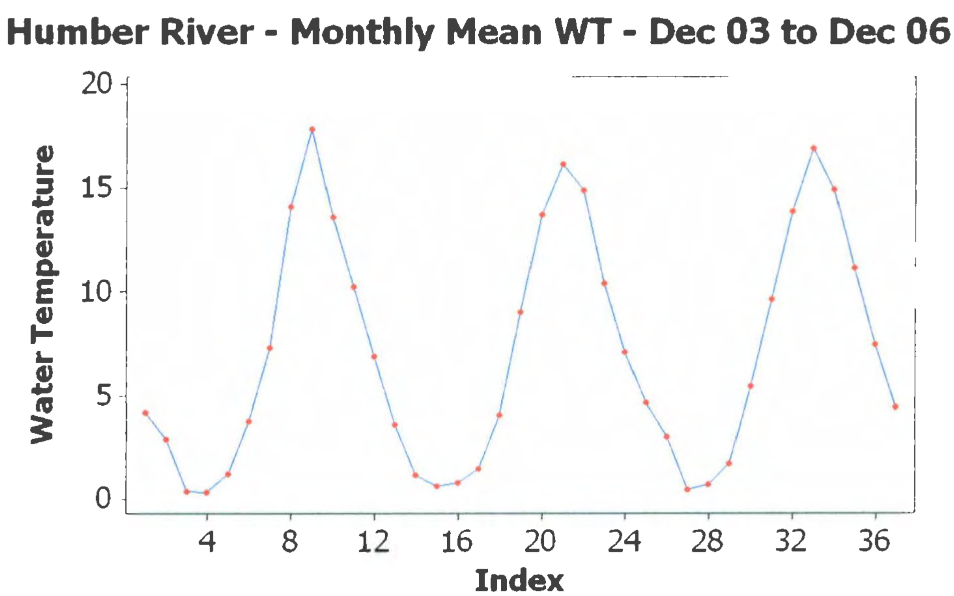


Figure 6.27 - Humber River Monthly Mean Water Temperature December 2003 to 2006

These monthly means can be represented by a parametric function with sine and cosine terms, a technique known as Fourier Harmonic Analysis. A periodic time series z_t with no trends over time can be represented by:

$$z_t = \alpha + \beta_0 \sin(2\pi ft) + \beta_1 \cos(2\pi ft) + \beta_2 \sin(4\pi ft) + \beta_3 \cos(4\pi ft) + \ldots \qquad \text{Equation 6.10}$$

where 2πf is a constant that depends on the frequency of the observations being studied. In this case the frequency of the data is 1/12 therefore 2πf is equal to 0.5236. The number of pairs of sine and cosine terms that need to be used to find a good overall fit to the observations can be determined using Minitab. A Minitab worksheet can be setup with the time of collection in the first column and water temperature (Z_t) in the second column of the worksheet. A constant term k1=2π/12=0.523 is defined and the pairs of cosine and sine can be calculated. A Minitab screenshot is attached in Figure 6.28 for reference.

Once the sine and cosine pairs have been calculated it is possible to perform a regression of Z_t versus the sine and cosine terms. The fit between the original values and fitted values with the sine and cosine terms is quite close (Figure 6.29 - where PFIT1 are the fitted observations for a model with one pair of sine and cosine terms, PFIT2 has two pairs, and PFIT3 has the first pair of sine and cosine and the second cosine term). Regression analysis showed that the best fit could be found using sin1, cos1 and cos2 (with an adjusted R^2 of 98.5%):

$$z_t = 7.15 + 1.38\cos 1 - 7.63\sin 1 - 1.57\cos 2 \qquad \text{Equation 6.11}$$

The residuals from the fit of the model to the observed values can be stored and analyzed for autocorrelation (Figure 6.30). The autocorrelation function shows that no autocorrelation in the residuals is of significance. Once autocorrelation in the dataset has been removed it is then possible to use a control chart to study the dataset. The CUSUM chart was used to detect small shifts in the process level (i.e. shifts away from the time series model), but no points were flagged as being of significance in this case (Figure 6.31).

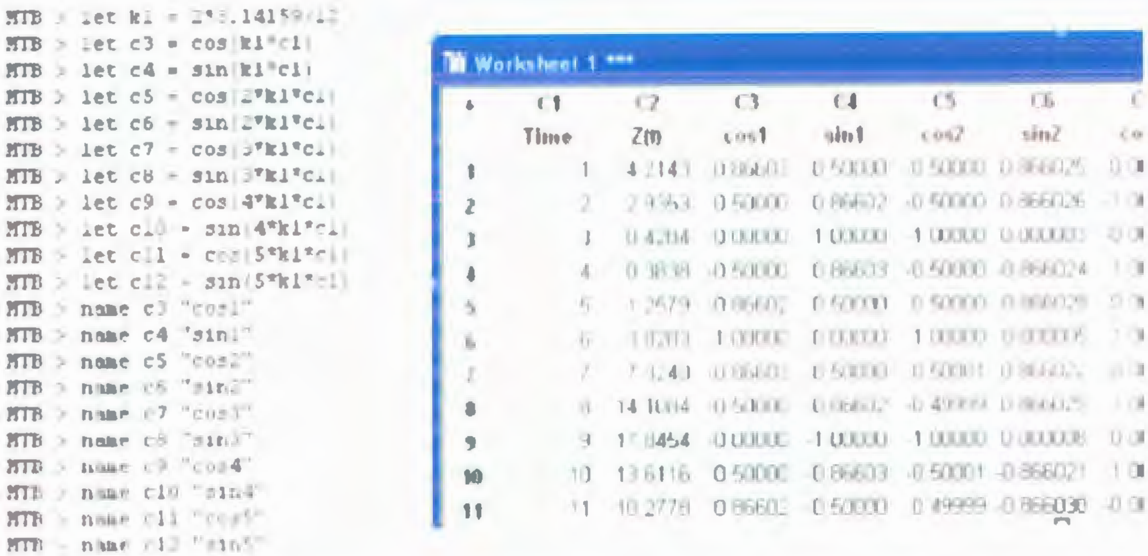


Figure 6.28 - Screenshot of Using Minitab to Find Cosine and Sine Terms

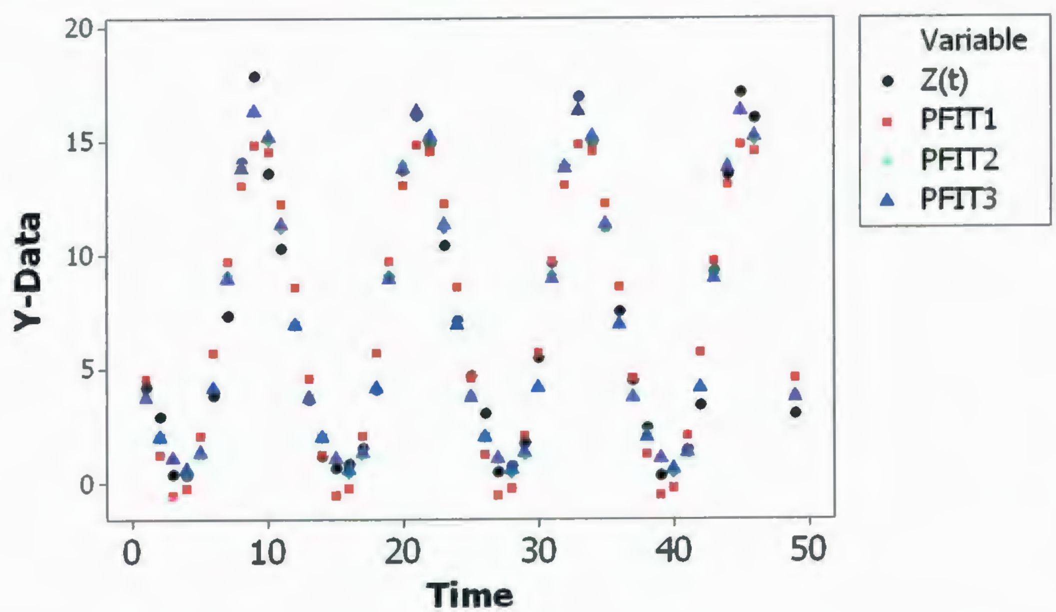


Figure 6.29 - Comparing Observed and Fitted Water Temperature Values

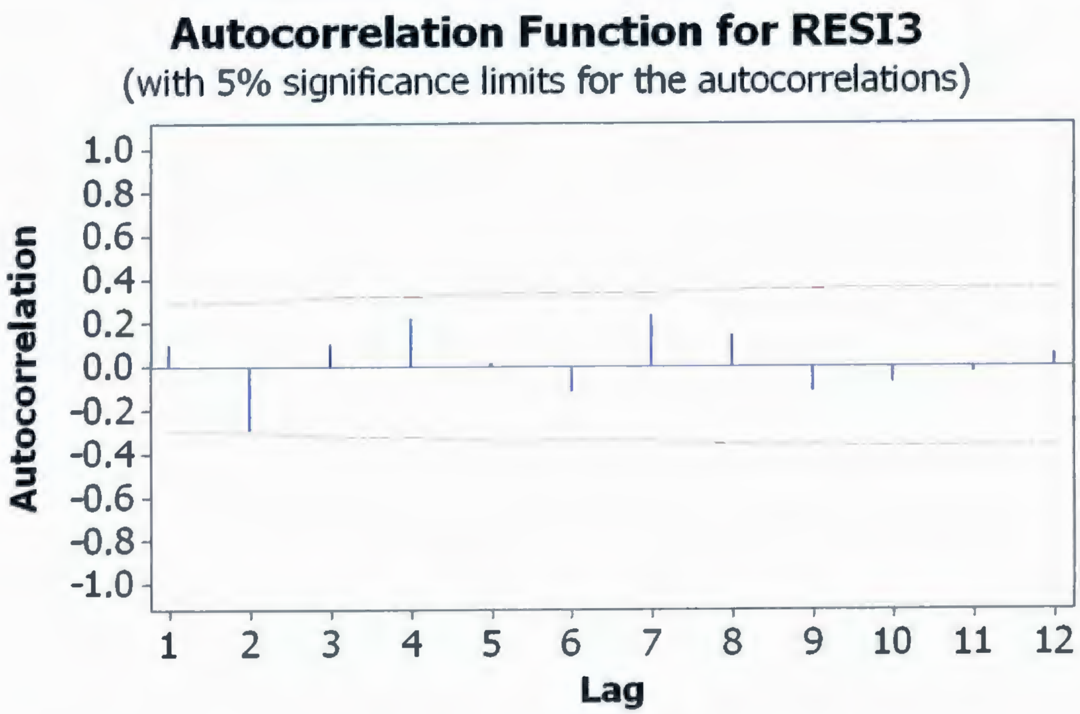


Figure 6.30 No Autocorrelation in the Residuals for the Harmonic Model

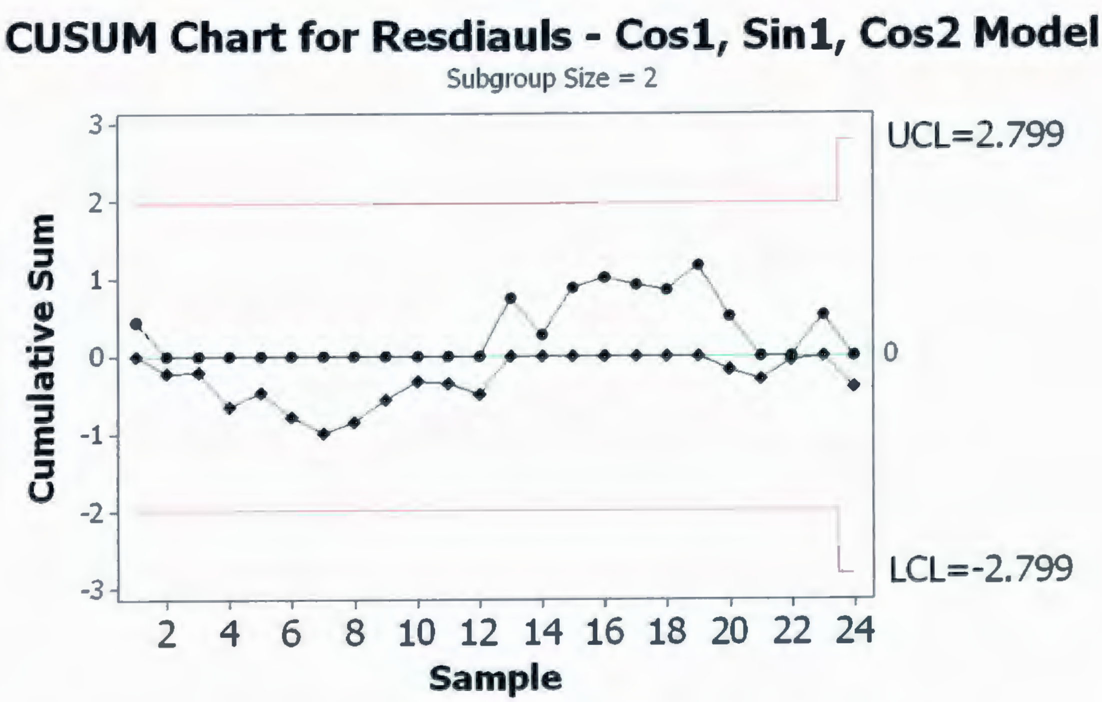


Figure 6.31 CUSUM Chart Shows No Points of Concern

6.5.6 Approach 6 - The Manly and MacKenzie Approach

Manly and MacKenzie (2000) used a CUSUM control chart to compare levels of dissolved reactive phosphorus recorded at 25 rivers in New Zealand every December from 1989 to 1996. Manly developed a free piece of software (the CUSUM Analysis Tool) to help other researchers carry out similar analysis on their own data. Investigations were made into using this software to compare monthly measurements of water quality collected at the four provincial network RTWQ stations. For example, the potential for the software to identify differences in monthly mean specific conductance collected every February at the four stations from 2004 to 2007. Unfortunately the limited size of the dataset posed a problem for the software and no meaningful control charts could be developed.

Although the Manly and MacKenzie approach is different from the usual approach of using control charts to study observations made at one station, it could prove to be a useful technique for the WRMD once the size of the historical water quality datasets is bigger - comparing monthly mean specific conductance recorded every February at every station in the network over five consecutive years.

6.5.7 Approach 7 - Modify Control Charts to Show More Useful Limits

The previous approaches have demonstrated the potential for using control charts in Minitab to monitor the real time data but the developed charts do not always provide information that would be useful to resource managers. When Minitab is used to develop these charts for the data (either individual observations, observations made at larger time scales, or subsets of observations) the limit lines on the chart are drawn to show when the process is out of statistical control. Take the Shewhart Chart for example, where limits lines are set three standard deviations away from the process mean. Points on the chart are flagged as being out of control when they plot outside these limits. A more useful control chart for resource managers would define limit lines based on water quality guidelines for the parameter being studied. A series of modified control charts were developed that allow the user to work with hourly observations of RTWQ data.

The Modified One-sided Control Chart for Dissolved Oxygen

The CCME recommended minimum dissolved oxygen level for the protection of aquatic life is 5.5 mg/L. A minitab macro was written that would plot each hourly observation of dissolved oxygen for a time period and flag points that were below this minimum. Two of these one-sided control charts are shown in Figure 6.32 (for the hourly measurements at Humber River recorded from January to March 2006) and Figure 6.33 (the same dataset randomly modified to include unsafe dissolved oxygen levels). The low dissolved oxygen levels shown in Figure 6.33 are not actual observations, but were manually inserted in the raw data to represent some imaginary threat to water quality at the station.

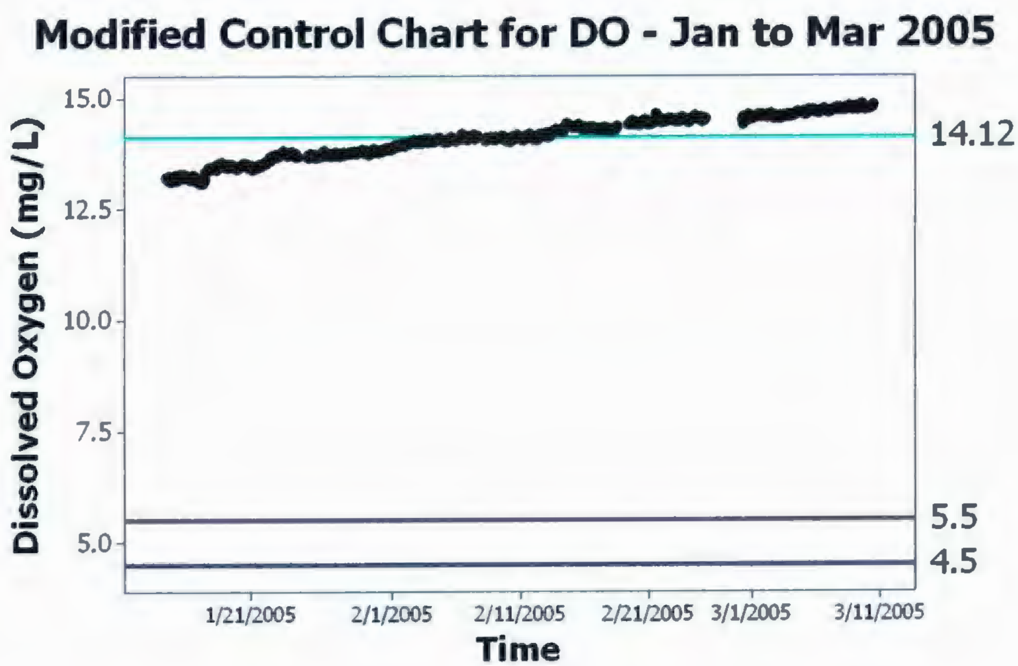


Figure 6.32 Modified One-sided Control Chart for Hourly Dissolved Oxygen

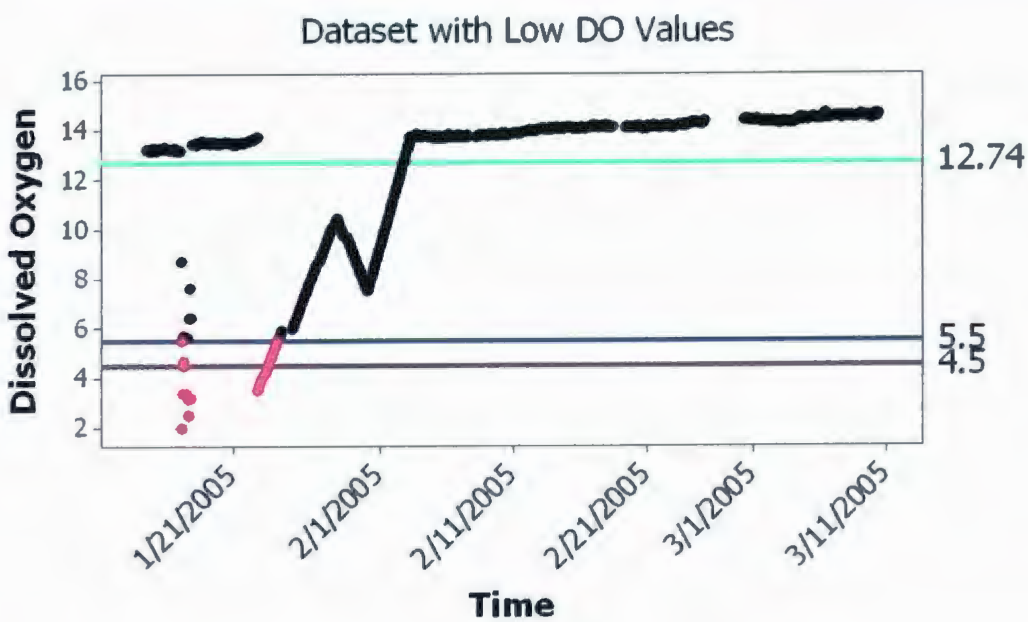


Figure 6.33 Modified One-sided Control Chart for a Modified Dissolved Oxygen Dataset

Figure 6.32 and 6.33 show how particularly useful this type of modified control chart is for identifying dissolved oxygen levels that pose a threat over long periods of RTWQ observations (i.e. a full deployment period of the sensor). Whenever values are above the CCME guideline of 5.5 mg/L the points plot in black and whenever they drop below the specified minimum they are flagged in red.

The Modified One-sided Control Chart for Water Temperature

Water temperature at the real time stations will never drop far below zero degrees and although cold air temperatures in Newfoundland tend to keep the temperatures in the rivers low throughout the year, it is useful to have a modified control chart that can flag points that pass an upper threshold (i.e. 18 °C). A modified one sided control chart was developed that would flag points above a user specified upper threshold (Figure 6.34). This upper one-sided control chart would also be useful for studying specific conductance levels at the stations.

The Modified Two-Sided Control Chart for pH Level

The CCME guideline for pH for the protection of aquatic life is from 6.5 to 9. A modified control chart with these two sided safe pH limits was developed - as shown in Figure 6.35. The chart is designed to flag any point outside of these thresholds.

Author's Note - a copy of the macros used for developing these control charts is included in Appendix Q.

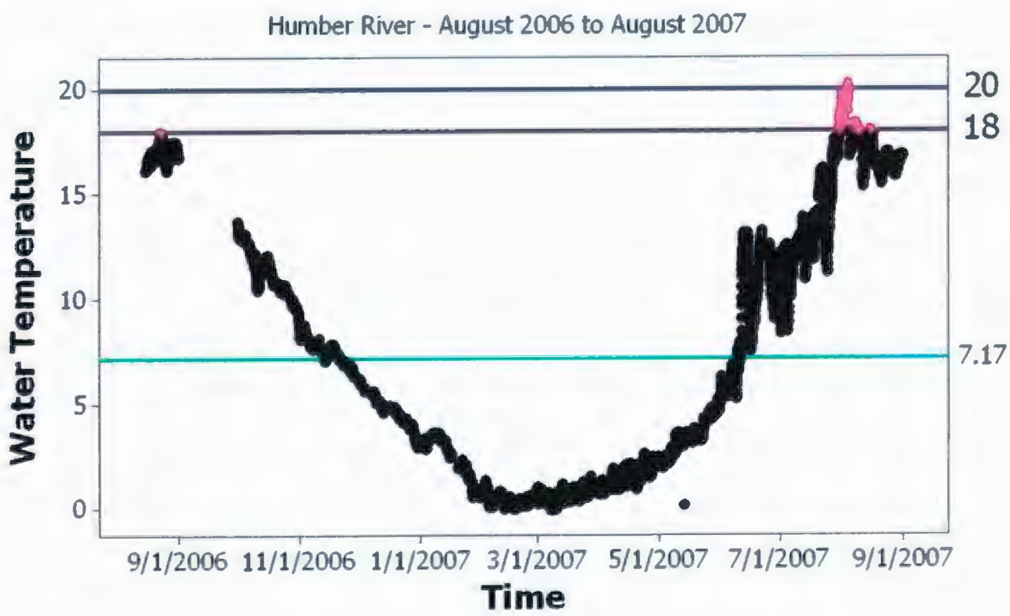


Figure 6.34 Modified One-sided Control Chart for Water Temperature

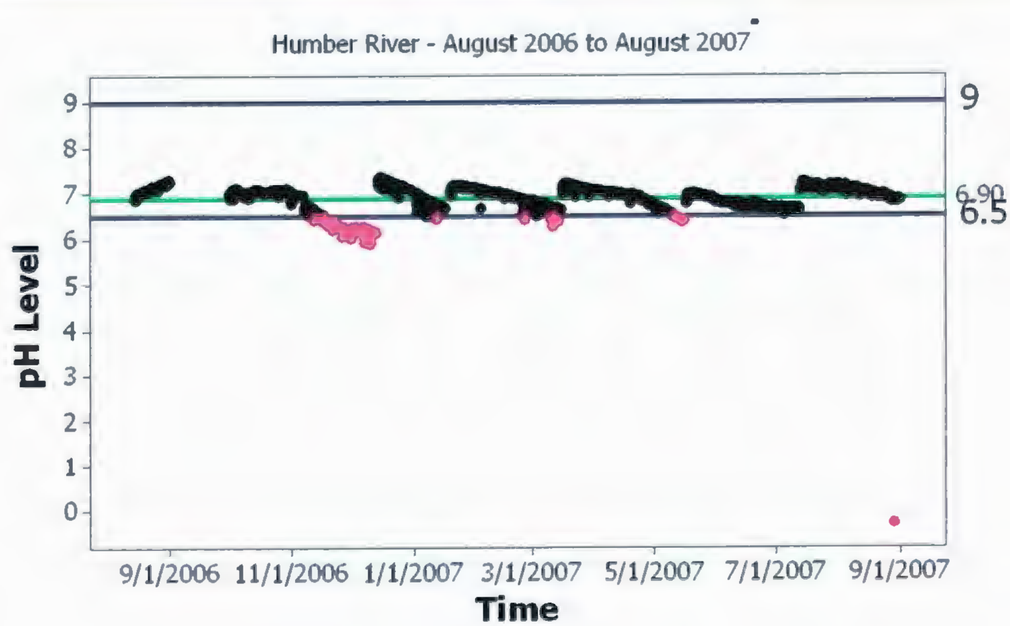


Figure 6.35 Modified Two-sided Control Chart for pH

6.6 Discussion

Investigations into the use of control charts for monitoring RTWQ data have shown that there is potential for implementing these charts but only in certain ways. The biggest obstacle that needs to be overcome in using the charts is finding a way to deal with autocorrelation in the observations.

The first approach (using the charts to detect large shifts over short periods of time) showed potential for using the EWMA chart to flag large shifts in performance in datasets consisting of hourly observations taken over the course of 24 hours. Autocorrelation in the datasets tended to increase when longer sets of data were analyzed and the charts became less useful for identifying shifts. The second approach (use of larger subgroups for the control charts) worked well for detecting large shifts in the data over longer periods of time then the first approach. However, if no large shift in the data was present then autocorrelation in the longer datasets tended to make the control chart limits less useful.

The third approach investigated the use of the charts for uncorrelated monthly means and found that as long as autocorrelation is low, charts like the Shewhart chart and EWMA chart will work well for identifying unusual shifts in the RTWQ levels. The biggest downfall of this approach is that it will only work for means taken on longer time scales and will not work as well for studying daily and weekly means. Furthermore the WRMD would not be able to use the charts to study individual observations of water quality or observations close to each other in time.

The fourth and fifth approaches worked with the residuals from ARIMA and harmonic models. Fitting these models to water quality data takes a considerable amount of previous experience working with time series datasets - even for datasets with low levels of autocorrelation. Fitting ARIMA models to water quality data with high levels of autocorrelation like that recorded by the real time sensors tends to be incredibly complex and mystifying. With this being said, it was still possible to fit ARIMA and harmonic models to monthly mean observations of water quality collected by the sensors and then use control charts to study the residuals and identify points of interest. It is not feasible to use the same approach for observations with high levels of autocorrelation (hourly, daily mean). Overall this approach is very statistically complex and is likely not the best option out of the seven approaches for the WRMD.

The Manly and MacKenzie approach for comparing observations between stations might eventually be useful for the WRMD but at this point the historical datasets of the stations are not long enough for the CUSUM analysis software to work properly.

The seventh approach of using Minitab to develop modified control charts would provide resource managers with a tool to identify observations that threaten water quality. Although these modified control charts are only time series plots with horizontal reference lines they are quite easy to develop (with the aid of the macros) and they avoid the inherent problems with autocorrelation and variance in the data that plagued the other attempts. Resource managers can use these charts to take a quick look at the data being collected and identify the timing of any problems at the stations through visual inspection alone.

Chapter Seven

Summary of Results

7.1 Scope

This chapter provides a summarizing discussion of the observations and conclusions that were presented in the previous chapters. This includes the results from regression modeling of water temperature, dissolved oxygen, and grab samples as well as the investigations into developing control charts for monitoring RTWQ data. This chapter also presents a brief discussion of ways the WRMD can modify historical records of RTWQ data so that future work in these research areas can be carried out successfully.

7.2 Summary of Results

7.2.1 Regression Models for Water Temperature

Regression models were developed for modeling daily, weekly and monthly mean, maximum and minimum water temperatures at four RTWQ monitoring stations in Newfoundland. The curve fitting results presented in Chapter 3 have shown that the relationship between water temperature and air temperature at the RTWQ stations is more S-shaped than it is linear. Both logistic models were fit to the data at the stations but the first logistic model is a better choice in terms of simplicity. It was noted that the S-shaped relationship often did not level off as much at higher temperatures as was expected. Perhaps this is because air temperatures in Newfoundland never reach warm enough levels - Mohseni and Stefan (1999) note the logistic model will behave in this manner for rivers in colder climates.

Although adjusted R^2 values were usually higher for monthly mean models than they were for weekly mean models, the S-shaped relationship was easier to see at the weekly time scale. This is likely the result of the differences between the high levels of water temperature contained in the two datasets - where the weekly mean datasets contain all the highest water temperatures during the summer months, while the monthly mean dataset sometimes misses out on the higher values. For example, if water temperature during four weeks in August are 19, 22, 25, 20 °C the average value for the month will only be 21.5°C. By using the weekly water temperatures, higher mean values are included in the dataset and the logistic model gets the opportunity to level off. At the daily time scale there is a considerable amount of scatter in the data and the models are less suited for prediction purposes. There is a considerable difference in the amount of variation in the observations of water temperature at low air temperatures than there is in water temperature at higher air temperatures.

Each of the stations were investigated for signs of hysteresis in the data but only the Humber River station (the largest river in the network) needed to be subdivided to account for warming and cooling seasons. There are two different approaches to handling hysteresis. The first involves splitting the dataset into two seasons and developing separate regression models for the datasets. The second involves adding an additional explanatory variable model to the existing regression models to account for the time of year the sample was collected. Both approaches resulted in good models, but the first approach makes it easier to compare the shape and fit of the models to those developed for other stations.

Stage level at the stations is usually only significant at the daily time scale and loses significance as the time scale is extended. Even when stage is significant the goodness of fit of the multiple regression model is never better than the logistic model. It would have been useful to obtain real time streamflow observations for the stations as most of the published work in this area of research deals with streamflow and not stage.

Overall curve fitting results for water temperature at the RTWQ stations can be considered a success. With successful models developed linking air temperature to water temperature it was expected that similar success could be found linking water temperature to dissolved oxygen levels at the stations.

7.2.2 Regression Models for Dissolved Oxygen

When modeling dissolved oxygen at the real time stations in the provincial network the relationship between dissolved oxygen and water temperature follows an exponential decay model more so than a linear model. The advantage in using the exponential decay model over the linear model is more apparent at higher water temperatures than it is at lower water temperatures. As water temperature increases the exponential model begins to level off while the linear model keeps dropping. Water temperatures at the stations are rarely high enough to force mean dissolved oxygen levels on the monthly and weekly time scale below 5 mg/L, but if water temperatures could reach these levels I believe the exponential model would be better at identifying the lower dissolved oxygen level.

The exponential regression model tends to perform better for handling mean dissolved oxygen then it does for maximum and minimum dissolved oxygen. The maximum and minimum datasets showed a considerable amount more variation in observations. The mean datasets are better at avoiding major outliers (i.e. mean weekly measurement might be 12 mg/L while the maximum value recorded that week might be 26 mg/L). There can be a great deal of variation in daily measurements, where for either physical or sensor-related reasons the dissolved oxygen level will often randomly jump for one particular measurement. These jumps will be captured in the maximum and minimum datasets but will be smoothed out by the mean datasets. There is a considerable amount of scatter in the daily mean observations of dissolved oxygen. This scatter makes it difficult to reliably predict daily dissolved oxygen levels at the stations.

7.2.3 Regression Models for Grab Samples

Investigations into regression models for grab samples have shown that there is potential for relating real time measurements to *select* grab sample measurements at *select* real time stations. An approach similar to that taken by the USGS works well for identifying for developing models for grab samples of major ions at urban stations like Leary's Brook and Waterford River. It tends to be easier to use real time parameters like specific conductance as surrogates for grab sample data when the grab sample measurements of ions, elements, metals and nutrients vary over the course of a year. The models presented in this thesis are not definitive in that they have not yet been tested for validity using new observations. In order for accurate regression models to be developed for the stations it will be necessary to expand the datasets with grab sample observations that can be paired with real time data and are taken upon reinstallation of the real time measurements.

Unfortunately, the datasets available for regression modeling in this research were small and it was not possible to set aside any of the grab samples for model testing. This model testing will be a necessary component of the model development process in the coming years.

There was considerable difficulty developing statistically significant regression models for the Humber River monitoring station. Levels of major ions, metals and nutrients at the station tend to be lower than the other stations and this may be one reason why there was such difficulty in modeling. Comparisons between the grab sample data and real time data have shown that measurements of water quality that should match do not always match at this station. Grab sample conductance in particular is not correlated to real time conductance and this is suspicious. Further investigations with new grab samples collected at the station should provide further insight into the problem.

Once the grab sample datasets are expanded and new models are developed these regression models should be quite useful for resource managers for optimizing the number of costly site visits and for gaining a better understanding of concentrations of major ions in the real time network rivers. At this point, the framework for modeling at this has been explored and expectations are high for reliable models to be developed for the RTWQ network.

7.2.4 Control Charts for Monitoring RTWQ Data

Seven different approaches for using control charts to monitor RTWQ data were investigated in Chapter 6. Although it is possible to use control charts on certain smaller groups of data (i.e. 24 hourly measurements), autocorrelation in the data can become significant for larger periods of time and the underlying statistical conditions for the chart s will be violated. Although it is possible to model the RTWQ data using ARIMA models and then use control charts to study the uncorrelated residuals, this approach is only useful when working with small amount of autocorrelation and is not possible for studying a full deployment period of hourly RTWQ measurements. Furthermore the developed control charts will only show when the observations go out of a state of statistical control and will specifically flag points that threaten water quality.

A modified form of a control chart was developed to sidestep the problems of autocorrelation in the data. These modified charts are nothing more than simple time-series plots of the data with reference lines added to represent healthy water quality limits. Although these limits are not defined in any statistical manner they will likely prove to be more useful to the resource managers at the WRMD.

Chapter Eight

Recommendations

8.1 Scope

This chapter presents a brief list of the different ways the WRMD can modify historical records of RTWQ data so that future work in the research areas examined in this thesis can be carried out successfully. Some points to consider when designing an effective control chart for the RTWQ network are are also presented.

8.2 Recommendations

8.2.1 Regression Modeling of WT and DO

- *Develop Models for Other Stations in the Network:* Regression models for dissolved oxygen and water temperature have only been developed for the RTWQ provincial network stations in this thesis. There are a number of other stations in operation in the province (i.e. those belonging to the Federal-Provincial network) and the potential exists for developing regression models similar to those developed in this thesis.

- *Use the Models to Study the Influence of Global Warming*: There is potential to use the water temperature and dissolved oxygen models to gain a further understanding of the potential impact of climate change on water quality. Most climate change models developed in recent years make predictions of how air temperature will change in the future. Regression models that use air temperature to predict water temperature can be used to determine the impact of rising air temperature on water quality. Those climate change induced changes in water temperature can then be related to dissolved oxygen levels in the rivers. For example, Lagergaard, Pedersen and Sand-Jensen (2007) used linear and nonlinear S-shaped regression models to study seasonal variations in daily water temperature for streams in Denmark and then used the S-shaped model to examine the impact of a global warming scenario on the streams. The Minitab macro developed for plotting both the water temperature and the dissolved oxygen regression models on the one plot should prove to be useful for this research.

- *Improve the Historical Records:* There are a few changes that should be made to the historical records if future regression models for water temperature and dissolved oxygen are to be developed for other stations or if the models developed in this research are to be modified as larger datasets become available with time.

- *Keep the Historical Datasets Maintained and Up to Date:* Working with the datasets retrieved from the ADRS will take a fair amount of preprocessing time before they can be used for analysis (removing blanks, changing the order of parameters in the columns, matching column headings). If future work is to be carried out analyzing the historical data it would be useful to make improvements to the records to reduce this preprocessing time. This would entail keeping a record on file at the WRMD that contains an easy to use record of the measured parameters at the station - with notes added to the records to explain any unusual measurements of water quality. The drift corrected datasets used in this thesis were only available up to early 2008. All real time observations collected after that point had not been corrected for drift and could not be used for regression analysis as a result. In speaking with personnel at the WRMD it was determined that it would be useful to have these drift corrected datasets updated after every deployment period of the sensor is complete. It is the tendency for a number of deployment periods to pass by before the datasets are updated. In some cases a member of the WRMD staff has updated the datasets but this member was not always familiar with the station being monitored or what happened over the deployment period. In those cases there is the good chance that data will be added to the historical record of drift corrected observations that should not be there (i.e. high dissolved oxygen values at the Humber River station that were obtained when the sensor was not operating correctly).

- *Add complete records of Streamflow to the Historical Datasets*: It is not always easy to obtain a historical record of streamflow at the real time stations. It took a considerable amount of time in this research to track down records of stage level recorded at the Environment Canada hydrometric stations. Quite often the records of stage would have some measurements of streamflow but these were usually incomplete. As a result streamflow could not be used as a potential explanatory variable for regression. For future work in this area it would be useful to obtain a complete record of stage and streamflow recorded at the stations on file at the WRMD.

- *Collect Air Temperature Measurements at the Stations*: In this thesis regression models were developed linking real time water temperature measurements to air temperature measurements made at nearby weather monitoring stations. This approach of using nearby recorded air temperatures when air temperature is not recorded at the station is used by the USGS in their work. It would be better for developing regression models if air temperature was recorded at the station in real time.

8.2.2 Regression Modeling of Grab Samples

The collection of grab samples of water quality is an ongoing process at the real time stations. The ability to develop statistically significant regression models that are known to be reliable will increase with the addition of these new collected grab samples. There are a few changes that can be made to the way grab samples are collected and the way historical records are maintained to make future investigations in this field of research straightforward.

- *Collect grab samples taken upon reinstallation of the sensor and not upon removal:* If new grab samples are to be added to the historical records used for developing regression models they should only be collected upon reinstallation of the re-calibrated real time sensor. In this research, smaller datasets have forced regression models to be developed using both removal and reinstallation grab samples (knowing that the removal samples are not as accurate). At this point its unknown how much the use of these removal samples is affecting the regression modeling results. It also advisable to replace the older removal samples in the datasets with new reinstallation samples as more samples are collected in the coming years.

- *Collect grab samples that can be easily matched to real time data*: Grab samples collected with the purpose of being used for developing regression models should be easily matched with real time sensor data. At the start of this research, the historical records of grab samples for each station was quite large, but once those samples with no matching real time data were removed from the dataset, the number of useable samples for regression decreased significantly (i.e. an initial 26 Peter's River samples was cut down to only 18 samples). It should be the aim of the WRMD to have at collect at least four samples a year that can be paired to real time data. It would also be useful to have at least one sample per season. Currently the number of grab samples can vary significantly each year in the historical records. An approach of taking four periodic samples along with collecting samples after significant events (i.e. rainstorms, heavy snowfall) giving an average of about 5-7 samples a year would be useful for regression modeling.

- *Bring real time measurements of turbidity on-line* The current real time sensors tend to provide rather unreliable measurements of turbidity and turbidity real time measurements could not be used for regression modeling. Real time turbidity was used by the USGS as a surrogate for grab sample measurements for total suspended solids, total organic nitrogen and phosphorus. There was little success in developing regression models for phosphorus in the Newfoundland rivers, but perhaps this might change if reliable real time turbidity measurements can be brought on-line and paired to the grab sample data.

- *Carry out testing of the models once more samples become available*: Once the grab sample datasets are large enough (i.e. over 25 samples) set aside some of the newer measurements to test the developed regression models. In this research there were not enough samples available for testing the models. This is an essential component of developing regression models that needs to be incorporated in future work so that reliable models are developed.

- *Ensure accurate information is recorded on the maintenance forms:* The maintenance form used by the WRMD for calibration has changed a fair amount over the years. The maintenance forms have recently been modified so that they are now much more accurate and there is less chance for them to be misinterpreted. For instance, the older maintenance forms did not specify how date was to be recorded and quite often a sample would appear to be made in February (i.e. 2/7/2005) but would be recorded as being collected some other month (i.e. July - 7/2/2005). The incorrect sample date would then be stored in the historical records of grab samples kept on file for the station. There were a number of times in this research where a considerable amount of time was spent tracking down the correct time of sample collection. It is to be hoped that new changes to the forms will make this problem a thing of the past.

- *Include the time of collection on the historical records of grab samples:* The historical records of grab samples currently kept by the WRMD do not contain the time of day the sample was taken but only only the day of collections. The files should be modified to include a column for entering the time of collection.

8.2.3 Using Control Charts for RTWQ Data

The following points should be remembered If traditional statistical control charts are to be used for monitoring the data collected at the stations:

- *What Chart to Use?* The Shewhart, CUSUM, MA and EWMA charts can be used for determining when the RTWQ data goes out of a state of statistical control but all of these charts are sensitive to autocorrelation in the data. If small amounts of autocorrelation are present in the data then it is recommended to use the Shewhart and EWMA charts to study the data as the CUSUM chart tends to be too sensitive to small shifts in the data and will not be overly useful for studying observations that vary over time due to seasonal changes.

- *How to Deal With Autocorrelation:* if there are low levels of autocorrelation in the data then it is possible to use control charts to study the residuals from ARIMA models or harmonic analysis. If high amounts of autocorrelation are present in the data then the statistical terms of the traditional control charts will be violated and the charts themselves will be useful. It is not practical to try to model highly autocorrelated RTWQ data with the ARIMA and harmonic models. If autocorrelation levels are high, avoid using the traditional forms of the control chart and use the modified control chart that plots the observations and uses control chart limits that are meaningful for aquatic health. These modified charts are the only control charts that allow the user to ignore autocorrelation on the data and focus solely on identifying events that threaten water quality.

Bibliography

Abdul-Aziz, O., Wilson, B. and Gulliver, J., 2007, *An Extended stochastic harmonic analysis algorithm: Application for dissolved oxygen,* Water Resources Research, Volume 43, 1 - 19.

Acres International Limited, 1995, *Integrated Watershed Management Plan for Peters River Basin*, Environment Canada and Newfoundland Department of Environment and Lands.

Alwan, L. and Roberts, H., 1988, *Time Series Modeling for Statistical Process Control*, Journal of Business and Economic Statistics, Volume 6(1), pp. 87-95.

Bartram, J. and Ballance, R., 1996, *Water Quality Monitoring - A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes*, United Nations Environment Program and the World Health Organization.

Bates, R. and Jackson, J., 1984, *Dictionary of geological terms*, Doubleday, New York.

Benyaha, L., Caissie, D., St. Hilaire, A., Ouarda, T., and Bobee, B, 2007, *A Review of Statistical Water Temperature Models*, Canadian Water Resources Journal, Volume 32(3), pp. 179-192.

Berthouex, P. and Brown, L., 2002, Statistics for Environmental Engineers - 2nd Edition, Lewis Publishers, USA.

Berthouex, P., Hunter, W., and Pallesen, L., 1976, *The use of stochastic models in the interpretation of historical data from sewage treatment plants*, Journal of Water Research, 10, pp. 689-698.

Berthouex, P., Hunter, W., and Pallesen, L., 1978, *Monitoring sewage treatment plants: some quality control aspects*, Journal of Quality Technology, Volume 10(4), pp. 139-149.

Box, G. and Jenkins, G., 1963, *Further Contributions to Adaptive Quality Control: Simultaneous Estimation of Dynamics: Non-zero Costs*, Bulletin of the International Statistics Institution, 34th Session.

Box, G. and Jenkins, G., 1976, *Time Series Analysis, Forecasting and Control* - 2nd Edition, Holden Day, San Francisco.

Box, G., Jenkins, G., and Reinsel, G., 1994, *Time series analysis, forecasting and control,* 3rd Edition, Englewood Cliffs, New Jersey, Prentice-Hall.

Caissie, D., El-Jabi, N., and Satish, M., 2001, *Modeling of maximum daily water temperatures in a small stream using air temperatures*, Journal of Hydrology, 251, pp. 14-28.

Campbell Scientific, 2009, *Hydrolab Series 4a Product Information*, Retrieved January 25, 2009 from http://www.hydrolab.com/products/DataSonde4a.asp

Canadian Council of Ministers of the Environment, 2004, *Application and Testing of the Water Quality Index in Atlantic Canada - Report Summary,* Retrieved February 20, 2010 from http://www.ccme.ca/assets/pdf/awi_en_posting.pdf

Christensen, V., 2001, *Characterization of Surface-Water Quality Based on Real-Time Monitoring and Regression Analysis, Quivira National Wildlife Refuge, South-Central Kansas, December 1998 Through June 2001*, Water Resources Investigations Report. United States Geological Survey 4248.

Christensen, V., Jian, X. and Ziegler, A., 2000, *Regression analysis and real-time water-quality monitoring to estimate constituent concentrations, loads and yields in the Little Arkansas River, south-central Kansas, 1995-1999*, United States Geological Survey, Water Resources Investigations Report 2000-4126.

Clifton, D. and Gilliom, R., 1989, *Sources and concentrations of dissolved solids and selenium in the San Joaquin River and its tributaries, California, October 1985 to March 1987*, United States Geological Survey Water-Resources Investigations Report 88–4217, Retrieved February 20, 2009 from http://pubs.er.usgs.gov/pubs/wri/wri884217

Crisp, D. and Howson, G., 1982, *Effect of Air Temperature upon Mean Water Temperature in Streams in the North Pennines and English Lake District*, Journal of Freshwater Biology, 12, pp. 359-367.

Department of Environment and Conservation, 2003, *Water Quality Trends in Selected Water Bodies of Newfoundland and Labrador*, Water Resources Management Division, St. John's, Newfoundland, Canada.

Department of Environment and Conservation, 2009a, Canada-Newfoundland Water Quality Monitoring Agreement, Retrieved January 15, 2009 from http://www.env.gov.nl.ca/env/Env/waterres/WQMA/WQMA.asp

Department of Environment and Conservation, 2009b, *Real Time Water Quality Information - Glossary*, Retrieved February 20, 2009 from http://www.env.gov.nl.ca/wrmd/RTWQ/Glossary.asp

Department of Environment and Conservation, 2009c, *Water Quality Station Profile - NF02YL0012*, Retrieved October 21, 2009 from http://www.canal.gov.nl.ca/canal/root/main/station_details_e.asp?envirodat=NF02YL0012

Department of Environment and Conservation, 2009d, *Water Quality Station Profile - NF02ZM0178*, Retrieved October 21, 2009 from http://www.canal.gov.nl.ca/canal/root/main/station_details_e.asp?envirodat=NF02ZM0178

Department of Environment and Conservation, 2009e, *Water Quality Station Profile - NF02ZM0009*, Retrieved October 21, 2009 from http://www.canal.gov.nl.ca/canal/root/main/station_details_e.asp?envirodat=NF02ZM0009

Environment Canada, 2002, *Application and testing of the CCME Water Quality Index for Select Water bodies in Atlantic Canada*, Retrieved September 10, 2009 from http://www.ec.gc.ca/soer-ree/English/resource_network/status_report_e.pdf

Gelda, R., Effler, W. and Owens, E., 2001, *River dissolved oxygen model with zebra mussel oxygen demand*, Journal of Environmental Engineering, Volume 127(9), pp. 790-801.

Glasgow, H., Burkholder, J., Reed, R., Lewitus, A., and Kleinman, J., 2004, *Real-time remote monitoring of water quality: a review of current application, and advancements in sensor, telemetry, and computing technologies*, Journal of Experimental Marine Biology and Ecology, 300, pp. 409 - 448.

Hem, J., 1992, *Study and interpretation of chemical characteristics of natural water - 3rd Edition*, United States Geological Survey Water Supply Paper 2254.

Helsel, D. and Hirsch, R., 2002, *Techniques of Water-Resources Investigation of the United States Geological Survey - Chapter A3 - Statistical Methods in Water Resources*, United States Geological Survey.

Hill, B. and Gilliom, R., 1993, *Streamflow, dissolved solids, suspended sediment, and trace elements, San Joaquin River, California, June 1985–September 1988*, United States Geological Survey Water-Resources Investigations Report 93–4085, Retrieved July 6, 2005 from at http://pubs.er.usgs.gov/pubs/wri/wri934085

Hodgson, S. and Quinn, T., 2002, T*he timing of adult sockeye salmon migration into fresh water: adaptations by populations to prevailing thermal regimes*, Canadian Journal of Zoology, Volume 80(3), pp. 542-55.

Hockey, J., Owens, I., and Tapper, N., 1982, *Empirical and theoretical models to isolate the effect of discharge on summer water temperature in the Hurunui River*, Journal of Hydrology, 21(1), pp. 1-12.

Johnson, F., 1971, *Stream temperatures in an alpine area*, Journal of Hydrology, 14, pp. 322-336.

Johnson, R. and Bagshaw, M., 1974, *The Effect of Serial Correlation on the Performance of CUSUM Tests*, Technometrics, 16, pp. 103-112.

Juran, J., 1997, *Early SQC: a historical supplement*, Journal of Quality Progress, Volume 30, pp 73-81.

Kentucky Watershed Watch, 2009, *Aluminum and Water Quality*, Retrieved February 30, 2009 from http://www.kywater.org/ww/ramp/rmal.htm

Khan, A., Abdel-Gawad, S, and Khan, H., 2008, *A Real Time Water Quality Monitoring Network and Water Quality Indices for River Nile,* World Water Congress 2008, Retrieved February 20, 2010 from http://www.worldwatercongress2008.org/resource/authors/abs894_article.pdf

Lagergaard Pedersen, N. and Sand-Jensen, K., 2007, *Temperature in lowland Danish streams: contemporary patterns, empirical models and future scenarios*, Journal of Hydrological Processes, 21, pp. 348-358.

Langan, S., Johnston, L., Donaghy, M., Younson, A., Hay, D. and Soulsby, C., 2001, *Variation in River Water Temperature in an Upland Stream over a 30 year period*, The Science of the Total Environment, 265, pp. 195-207.

Lu, C., and Reynolds, M., 2001, *CUSUM Charts for Monitoring an Autocorrelated Process*, Journal of Quality Technology, Vol. 33, pp. 316-334.

Lund, S., Caissie, D., Cunjak, R., Vijayan, M. and Tufts, B., 2002, *The effects of environmental heat stress on heatshock mRNA and protein expression in Miramichi Atlantic salmon,* Canadian Journal of Fisheries and Aquatic Sciences, Vol. 59, pp. 1553–1562.

Mac Nally, R., and Hart, B., 1997, *Use of CUSUM Methods for Water-Quality Monitoring in Storages*, Journal of Environmental Science and Technology, 31, pp. 2114-2119.

Manly, B., 1994, *CUSUM Methods for detecting changes in monitored environmental variables*, Statistics in Ecology and Environmental Monitoring, University of Otago Press, Dunedin, New Zealand, pp. 225-238.

Manly, B., 1997, *Randomization, Bootstrap and Monte Carlo Methods in Biology - 2nd Edition*, Chapman and Hall, London.

Manly, B., 2002, *CUSUM Analysis Tool 2.2*, Developed by Darryl MacKenzie, Department of Mathematics and Statistics, University of Otago, New Zealand, Retrieved April 22, 2009 from http://www.proteus.co.nz/home.html

Manly, B. and MacKenzie, D., 2000, *A cumulative sum type of method for environmental monitoring*, Environmetrics, 11, pp. 151-166.

Manly, B. and Mackenzie, D., 2003, *CUSUM environmental monitoring in time and space,* Journal of Environmental and Ecological Statistics, 10, pp. 231-247.

Mohn, E. and Volden, R., 1985, *Acid precipitation: effects on small lake chemistry. In Data Analysis and Real Life Environment: Ins and Outs of Solving Problems*, J.F. Marcotorchino, J.M. Proth, and J. Janssen (Editors), Elsevier, Amsterdam, pp. 191-196.

Mohseni, O. and Stefan, H., 1999, *Stream temperature/air temperature relationship: a physical interpretation*, Journal of Hydrology, Volume 218, pp. 128-141.

Mohseni, O. Erickson, T. and Stefan, H., 2002, *Upper Bounds for Stream Temperatures in the Contiguous United States*, Journal of Environmental Engineering, Volume 128(1) pp. 4-11.

Mohseni, O., Stefan, H., and Erickson, T., 1998, A *nonlinear regression model for weekly stream temperatures*, Journal of Water Resources Research, Volume 34(10), pp. 2685-2692.

Morin, G. and Couillard, D., 1990, *Predicting River Temperatures with a Hydrological Model*, Encyclopedia of Fluid Mechanics, 10th Edition, Gulf Publishing Company, Texas, pp. 171-209.

Morrill, J., Bales, R. and Conklin, M., 2005, *Estimating Stream Temperature from Air Temperature:Implications for Future Water Quality*, Journal of Environmental Engineering, Volume 131(1), pp. 139-146.

Mueller, D. and Helsel, D., 1996, *Nutrients in the Nation's Waters - too much of a good thing?* United States Geological Survey Survey Circular 1136.

Neumann, D., Rajagopalan, B., and Zagona, E., 2003, *Regression Model for Daily Maximum Stream Temperature*, Journal of Environmental Engineering, Volume 129(7), pp. 667-674.

Page, E.S., 1954, *Continuous inspection schemes*, Biometrika, 41, pp. 100–114.

Pilgrim, J. and Stefan, H., 1995, *Correlation of Minnesota stream water temperatures with air temperatures*, Project Report 382, St. Anthony Falls Laboratory, University of Minnesota.

Pilgrim, J., Fang, X., and Stefan, H., 1998, *Stream Temperature Correlations with Air Temperatures in Minnesota: Implications for Climate Warming*, Journal of the American Water Resources Association, 345(5), pp.1109-1121.

Rounds, S., 2002, *Development of a neural network model for dissolved oxygen in the Tualatin River, Oregon*, Paper presented at the Second Federal Interagency Hydrologic Modeling Conference, Las Vegas, Nevada.

Saffran, K. and Anderson, A., 1996, *An Empirical Analysis of Water Temperature and Dissolved Oxygen Conditions in the Red Deer River*, Water Sciences Branch, Water Management Division, Alberta Environmental Protection.

Smeti, E., Koronakis, D., and Golfinopolous, S. 2007. *Control charts for the toxicity of finished water - modeling the structure of toxicity*, Water Research, 41(12), 2679-2689.

Smith, K., 1979, *Man's Impact on the Hydrological Cycle in the United Kingdom,* Hollis GE.

Smith, K., 1981, *The Prediction of River Water Temperatures*, Hydrological Sciences Bulletin, 26, pp. 19-32.

Song, C., Pabst, A., and Bowers, C., 1973, *Stochastic Analysis of Air and Water Temperatures*, Journal of Environment Engineering, 99, pp. 785-800.

St. Hilaire, A., Morin, G., El-Jabi, N., and Caissie, D., 2000, *Water Temperature Modeling in a Small Forested Stream:Implication of Forest Canopy and Soil Temperature*, Canadian Journal of Civil Engineering, 27, pp. 1095-1108.

Stefan, H. and Preud'homme, E., 1993. *Stream Temperature Estimation from Air Temperature*, Water Resources Bulletin, 29(1), pp. 27-45.

Texas State University, 2009, *Lesson Five - Water conductivity and total dissolved solids water quality sampling*, Retrieved November 1, 2009 from http://txstreamteam.rivers.txstate.edu/Publications/Curriculum/contentParagraph/0/text_files/file4/Curriculum%20lesson%205.pdf

Teillet, P., Gauthier, R., Chichagov, A., and Fedosejevs, G., 2002, *Towards integrated Earth sensing: advanced technologies for in situ sensing in the context of Earth observation*, Canadian Journal of Remote Sensing, 28, pp. 713 - 718.

Turner, J. and Woodham, W., 1980, *Evaluation of remote hydrologic data-acquisition systems*, USGS Water Resources Investigations 79-102, West-Central Florida, USGS, Reston, VA.

United States Geological Survey, 2008, *What is Continuous Real Time Water Quality*, Retrieved January 15, 2009 from http://water.usgs.gov/waterwatch/wqwatch/index.php?qid=1&c=faq

USEPA, 1986, *Quality criteria for water*, United States Environmental Protection Agency Report 440/5-86-001.

Vander Wiel, S., 1996, *Monitoring Processes that Wander Using Integrated Moving Average Models*, Technometrics, 38, pp. 139-151.

Vasilopoulos, A. and Stamboulis, A., 1978, *Modification of Control Chart Limits in the Presence of Data Correlation*, Journal of Quality Technology, Volume 10(1), pp. 20-30.

Wardell, D., Moskowitz, H. and Plante, R., 1992, *Control charts in the Presence of Data Correlation*, Journal of Management Science, Volume 38, pp. 1084-1105.

Webb, B., 1987, *The Relationship Between Air and Water Temperatures for a Devon River*, Reports and Transactions of the Devonshire Association for the Advancement of Science, Exeter, United Kingdom, 119, pp. 197-222.

Webb, B. and Nobilis, F., 1997, *A long term perspective on the nature of the air-water temperature relationship: a case study*, Hydrological Processes 11, pp. 137-147.

Webb, B., Clack, P., and Walling, D., 2003, *Water Air Temperature Relationships in a Devon River System and the Role of Flow*, Hydrological Processes, 17, pp. 3069-3084.

Yourstone, S. and Montgomery, D., 1989, *A Time-series Approach to Discrete Real-Time Process Quality Control*, Journal of Quality and Reliability Engineering International, Volume 5, pp. 309-317

Appendix A

Formatting the ADRS Records and

Accounting for Sensor Drift

The Excel files of historical water quality downloaded from the ADRS contain the date and time of measurement, water temperature, pH, specific conductance, dissolved oxygen, and stage but there is a great deal of formatting that needs to be done to them before the information necessary for carrying out the research in this thesis can be obtained.

Take for example the partial screenshot of the historical record shown in Figure A.1 for real time station NF02YL0012 - Humber River at Humber Village Bridge. In its original form the Excel file contains thousands of rows of water quality data but will also contain blank rows (periods of time when the sensor was not in operation) and rows containing zeroes for all parameters (periods of time when the sensor was not operating effectively).

On a routine basis the sensors are taken out of the water for maintenance and calibration and during those periods the historical record will contain blank rows for the water quality parameters. It is possible to scroll down through the record and remove the blank rows manually - but the records are generally 30,000 rows long and this would take a considerable amount of time. A much more time efficient option is to use the column filter function in Excel to remove the blank rows - as Figure A.1 demonstrates.

	A	B	C	D	E	F	G	H
1	LocalDaTi	Stage	Temp-Water	pH	Conductance	Diss-Solids	Percent-Satur	Diss-Oxy
29	12/21/2003 5:50	1.853	4.1	7.17	0.0342	0.0219	88.8	11.37
30	12/21/2003 4:50	1.837	4.2	7.17	0.0342	0.0219	88.7	11.35
31	12/21/2003 3:50	1.842	4.1	7.12	0.0344	0.022	88.5	11.34
32	12/21/2003 2:50	1.834	4.1	7.17	0.0345	0.0221	88.7	11.36
33	12/21/2003 1:50	1.845	4.1	7.17	0.0346	0.0222	88.5	11.33
34	12/11/2003 11:50							
35	12/11/2003 10:50							
36	12/11/2003 9:50							
37	12/11/2003 14:50	1.906	4.9	7.19	0.0345	0.0221	88.6	11.11
38	12/11/2003 13:50	1.906	4.9	7.19	0.0344	0.022	88.2	11.07
39	12/11/2003 12:50	1.886	4.9	7.19	0.0344	0.022	87.7	11.02

1	LocalDaTi	Stage	Temp-Wat	pH	Conductance	Diss-Solids	Percent-Satur	Diss-Oxy
29	12/21/2003 5:50	1.853	4.1	7.17	0.0342	0.0219	88.8	11.37
30	12/21/2003 4:50	1.837	4.2	7.17	0.0342	0.0219	88.7	11.35
31	12/21/2003 3:50	1.842	4.1	7.12	0.0344	0.022	88.5	11.34
32	12/21/2003 2:50	1.834	4.1	7.17	0.0345	0.0221	88.7	11.36
33	12/21/2003 1:50	1.845	4.1	7.17	0.0346	0.0222	88.5	11.33
37	12/11/2003 14:50	1.906	4.9	7.19	0.0345	0.0221	88.6	11.11
38	12/11/2003 13:50	1.906	4.9	7.19	0.0344	0.022	88.2	11.07
39	12/11/2003 12:50	1.886	4.9	7.19	0.0344	0.022	87.7	11.02

Figure A.1 Using a Column Filter in Excel to Remove Blank Rows of Data

The identification and removal of the zeroes in the historical record takes much more effort. The identification of the location of the zeroes in the dataset can be achieved by reviewing maintenance and calibration records kept on the Department of Environment and Conservation webpage for each of the stations (http://www.env.gov.nl.ca/wrmd/RTWQ/CalibrationSchedule.asp). Each of these records contain the dates the Hydrolab sensor was removed from the monitoring site and when the sensor was reinstalled. Although a rather time consuming process, removing these zeroes is an essential correction step as these zeroes will throw off statistical properties like the mean. Take for example the historical record of the Humber River station where the mean value of the dissolved oxygen values from the original unformatted historical record for the month of January, 2004 is 4.75 mg/L. Once the zeroes have been picked out and removed the mean value for dissolved oxygen increases to 11.78 mg/L - a much safer level for aquatic health and a better indicator of what the true value was for that month.

Once the blank rows and zeroes have been removed from the historical record it is then possible to correct the data for drift over time. The WRMD has come to find that measurements taken by the sensor tend to be less accurate the longer the sensor is left in the water. At the start of the deployment period the sensor will have been recently calibrated and will take accurate measurements of water quality parameters. Over the next few weeks this calibration will be lost and the measurements taken by the sensor will slowly drift away from the true value. The WRMD can gain an estimate of the size of this drift by sending personnel out to the sampling station with a recently calibrated Hydrolab Minisonde to take a companion set of measurements. The accurate readings taken by the Minisonde can be compared to the final Datasonde readings to see how they differ (i.e. the Minisonde records dissolved oxygen as 11.55 mg/L and the last Datasonde readings show dissolved oxygen as 11.10 mg/L - therefore the difference between the two is 0.45 mg/L). The WRMD has setup an Excel spreadsheet that can correct the Datasonde measurements over the previous deployment period according to equation A.1:

$$\text{Corrected Value} = (\text{DS value}) + \{(\text{Field MS value}) - (\text{Field DS value})\} * \frac{rank}{n} \qquad \text{Equation A.1}$$

where DS stands for Datasonde and MS stands for Minisonde. The Datasonde measurements are ranked according to time (i.e. the first measurement in the deployment period has a ranking of 1, and the final measurement has a ranking of say 1200) so that the first measurements receive a minimal correction and the last measurements are brought closer to what the true value was likely to have been. All of this drift correction work is carried out by WRMD staff as it is necessary to use firsthand knowledge of the station to determine if the drift corrections being made are appropriate for that station.

The final product of making modifications for missing values and for sensor drift is referred to as a Drift *Corrected Historical Record* of sensor data for the station. It is the aim of the WRMD to keep these corrected records as recent as possible but at this point in time, corrected records are not available past 2008.

Appendix B

Minitab Macros for Obtaining

Mean, Maximum and Minimum Values

Minitab Macros were written to find mean, maximum and minimum values of water quality at the monthly, weekly and daily time scales. The complete code for each of these macros is attached on the following pages.

In order for the macros to work properly it is necessary that the RTWQ data is setup in Minitab as follows:

- Column 1 has the date and time of the sample

- Column 2 has the water temperature

- Column 3 has the pH

- Column 4 has the specific conductance

- Column 5 has the dissolved solids

- Column 6 has the percent saturation

- Column 7 has the dissolved oxygen

- Column 8 has the turbidity

Separate macros were written for stage and air temperature. Similar code to the ones for the real time sensor measurement were used but for the sake of space these will not be included in this appendix. It's possible to recreate the macros for stage and air temperature by copying out sections of the attached macros (i.e. have the date and time in column 1, the stage in column 2 and only use the parts of the macro that work with column 2 to find the mean, maximum and minimum).

In order for the macros to work properly it is necessary that the historical record of real time data has a value for every day of the year (if not the macro will stop unexpectedly). The historical records are known to have large gaps and for any days that are missing insert a * for the missing entry. For example if 12/6/2003 is missing put an entry in column one for 12/6/2003 0:00:00 and enter a * for water temperature (column 2), pH (column 3), and so on.

Macros are designed to account for leap years in 2003 and 2009 (29 days in aFebruary).

Macro for Daily Mean, Maximum and Minimum RTWQ Data

The macro will ask the user to enter the month and the year. May 2004 would be entered as 5 2004. The macro will then go and find all the daily mean, maximum, and minimum values for that particular month. Once the macro is finished, proceed to the next month of interest. The macro is rather lengthy - only way to handle the different number of days in each month.

```
(PAGE 1)
gmacro
Daily
ERASE k1-k1000
ERASE c10-c1000
Note Set Month and the Year
SET c50;
File "terminal";
NOBS 2.
Copy c50 k50 k51
File "terminal";
NOBS 2.
Copy c50 k50 k51
NUMERIC 'Date' c10;
Year;
FourDigit.
Name c10 'Year'
NUMERIC 'Date' c11;
Month.
Name c11 'Month'
NUMERIC 'Date' c12;
Day.
Name c12 'Day'

If k50=1 OR k50=2 OR k50=3 OR
K50=5 OR k50=7 OR k50=8 OR
k50=10 OR k50=12
do k100 = 1:31
Let c15(k100)=k51
Name c15 'Y'
Let c16(k100)=k50
Name c16 'M'
Let c17(k100)=k100
Name c17 'D'
Copy c2 c100;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Name c100 'Copied WT'
```

```
(PAGE 2)
Let k1 = mean(c100)
Let k2 = max(c100)
Let k3 = min(c100)
Let c18(k100) = k1
Let c19(k100) = k2
Let c20(k100) = k3
Name c18 'Mean WT'
Name c19 'Max WT'
Name c20 'Min WT'
Copy c3 c101;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Name c101 'Copied pH'
Let k4 = mean(c101)
Let k5 = max(c101)
Let k6 = min(c101)
Let c21(k100) = k4
Let c22(k100) = k5
Let c23(k100) = k6
Name c21 'Mean pH'
Name c22 'Max pH'
Name c23 'Min pH'
Copy c4 c102;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k7 = mean(c102)
Let k8 = max(c102)
Let k9 = min(c102)
Let c24(k100) = k7
Let c25(k100) = k8
Let c26(k100) = k9
Name c24 'Mean SC'
Name c25 'Max SC'
Name c26 'Min SC'
Copy c5 c103;
Include;
```

```
(PAGE 3)
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k10 = mean (c103)
Let k11 = max(c103)
Let k12 = min(c103)
Let c27(k100) = k10
Let c28(k100) = k11
Let c29(k100) = k12
Name c27 'Mean DS'
Name c28 'Max DS'
Name c29 'Min DS'
Copy c6 c104;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k13 = mean(c104)
Let k14 = max(c104)
Let k15 = min(c104)
Let c30(k100) = k13
Let c31(k100) = k14
Let c32(k100) = k15
Name c30 'Mean PS'
Name c31 'Max PS'
Name c32 'Min PS'
Copy c7 c105;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k16 = mean(c105)
Let k17 = max(c105)
Let k18 = min(c105)
Let c33(k100) = k16
Let c34(k100) = k17
Let c35(k100) = k18
Name c33 'Mean DO'
Name c34 'Max DO'
Name c35 'Min DO'
enddo
```

(Page 4)

```
ELSEIF k50=4 OR k50=6 OR
k50=9 OR k50=11
do k100 = 1:30
Let c15(k100)=k51
Name c15 'Y'
Let c16(k100)=k50
Name c16 'M'
Let c17(k100)=k100
Name c17 'D'
Copy c2 c100;
Include;
Where "'Month' = k50 And 'Year'
= k51 AND 'Day' = k100".
Name c100 'Copied WT'
Let k1 = mean(c100)
Let k2 = max(c100)
Let k3 = min(c100)
Let c18(k100) = k1
Let c19(k100) = k2
Let c20(k100) = k3
Name c18 'Mean WT'
Name c19 'Max WT'
Name c20 'Min WT'
Copy c3 c101;
Include;
Where "'Month' = k50 And 'Year'
= k51 AND 'Day' = k100".
Name c101 'Copied pH'
Let k4 = mean(c101)
Let k5 = max(c101)
Let k6 = min(c101)
Let c21(k100) = k4
Let c22(k100) = k5
Let c23(k100) = k6
Name c21 'Mean pH'
Name c22 'Max pH'
Name c23 'Min pH'
Copy c4 c102;
Include;
Where "'Month' = k50 And 'Year'
= k51 AND 'Day' = k100".
Let k7 = mean(c102)
```

(Page 5)

```
Let k8 = max(c102)
Let k9 = min(c102)
Let c24(k100) = k7
Let c25(k100) = k8
Let c26(k100) = k9
Name c24 'Mean SC'
Name c25 'Max SC'
Name c26 'Min SC'
Copy c5 c103;
Include;
Where "'Month' = k50 And 'Year'
= k51 AND 'Day' = k100".
Let k10 = mean (c103)
Let k11 = max(c103)
Let k12 = min(c103)
Let c27(k100) = k10
Let c28(k100) = k11
Let c29(k100) = k12
Name c27 'Mean DS'
Name c28 'Max DS'
Name c29 'Min DS'
Copy c6 c104;
Include;
Where "'Month' = k50 And 'Year'
= k51 AND 'Day' = k100".
Let k13 = mean(c104)
Let k14 = max(c104)
Let k15 = min(c104)
Let c30(k100) = k13
Let c31(k100) = k14
Let c32(k100) = k15
Name c30 'Mean PS'
Name c31 'Max PS'
Name c32 'Min PS'
Copy c7 c105;
Include;
Where "'Month' = k50 And 'Year'
= k51 AND 'Day' = k100".
Let k16 = mean(c105)
Let k17 = max(c105)
Let k18 = min(c105)
Let c33(k100) = k16
Let c34(k100) = k17
Let c35(k100) = k18
Name c33 'Mean DO'
Name c34 'Max DO'
Name c35 'Min DO'
enddo
```

(Page 6)

```
ELSEIF k50 = 2 AND K51=2003
do k100 = 1:29
Let c15(k100)=k51
Name c15 'Y'
Let c16(k100)=k50
Name c16 'M'
Let c17(k100)=k100
Name c17 'D'
Copy c2 c100;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Name c100 'Copied WT'
Let k1 = mean(c100)
Let k2 = max(c100)
Let k3 = min(c100)
Let c18(k100) = k1
Let c19(k100) = k2
Let c20(k100) = k3
Name c18 'Mean WT'
Name c19 'Max WT'
Name c20 'Min WT'
Copy c3 c101;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Name c101 'Copied pH'
Let k4 = mean(c101)
Let k5 = max(c101)
Let k6 = min(c101)
Let c21(k100) = k4
Let c22(k100) = k5
Let c23(k100) = k6
Name c21 'Mean pH'
Name c22 'Max pH'
Name c23 'Min pH'
Copy c4 c102;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k7 = mean(c102)
Let k8 = max(c102)
Let k9 = min(c102)
Let c24(k100) = k7
Let c25(k100) = k8
Let c26(k100) = k9
Name c24 'Mean SC'
Name c25 'Max SC'
Name c26 'Min SC'
```

```
(Page 7)
Copy c5 c103;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k10 = mean (c103)
Let k11 = max(c103)
Let k12 = min(c103)
Let c27(k100) = k10
Let c28(k100) = k11
Let c29(k100) = k12
Name c27 'Mean DS'
Name c28 'Max DS'
Name c29 'Min DS'
Copy c6 c104;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k13 = mean(c104)
Let k14 = max(c104)
Let k15 = min(c104)
Let c30(k100) = k13
Let c31(k100) = k14
Let c32(k100) = k15
Name c30 'Mean PS'
Name c31 'Max PS'
Name c32 'Min PS'
Copy c7 c105;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k16 = mean(c105)
Let k17 = max(c105)
Let k18 = min(c105)
Let c33(k100) = k16
Let c34(k100) = k17
Let c35(k100) = k18
Name c33 'Mean DO'
Name c34 'Max DO'
Name c35 'Min DO'
enddo
```

```
(Page 8)
ELSEIF k50 = 2 AND K51=2008
do k100 = 1:29
Let c15(k100)=k51
Name c15 'Y'
Let c16(k100)=k50
Name c16 'M'
Let c17(k100)=k100
Name c17 'D'
Copy c2 c100;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Name c100 'Copied WT'
Let k1 = mean(c100)
Let k2 = max(c100)
Let k3 = min(c100)
Let c18(k100) = k1
Let c19(k100) = k2
Let c20(k100) = k3
Name c18 'Mean WT'
Name c19 'Max WT'
Name c20 'Min WT'
Copy c3 c101;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Name c101 'Copied pH'
Let k4 = mean(c101)
Let k5 = max(c101)
Let k6 = min(c101)
Let c21(k100) = k4
Let c22(k100) = k5
Let c23(k100) = k6
Name c21 'Mean pH'
Name c22 'Max pH'
Name c23 'Min pH'
Copy c4 c102;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k7 = mean(c102)
Let k8 = max(c102)
Let k9 = min(c102)
Let c24(k100) = k7
Let c25(k100) = k8
Let c26(k100) = k9
Name c24 'Mean SC'
Name c25 'Max SC'
Name c26 'Min SC'
```

```
(Page 9)
Copy c5 c103;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k10 = mean (c103)
Let k11 = max(c103)
Let k12 = min(c103)
Let c27(k100) = k10
Let c28(k100) = k11
Let c29(k100) = k12
Name c27 'Mean DS'
Name c28 'Max DS'
Name c29 'Min DS'
Copy c6 c104;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k13 = mean(c104)
Let k14 = max(c104)
Let k15 = min(c104)
Let c30(k100) = k13
Let c31(k100) = k14
Let c32(k100) = k15
Name c30 'Mean PS'
Name c31 'Max PS'
Name c32 'Min PS'
Copy c7 c105;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k16 = mean(c105)
Let k17 = max(c105)
Let k18 = min(c105)
Let c33(k100) = k16
Let c34(k100) = k17
Let c35(k100) = k18
Name c33 'Mean DO'
Name c34 'Max DO'
Name c35 'Min DO'
enddo
```

(Page 10)	(Page 11)	(Page 12)

```
(Page 10)
ELSE
do k100 = 1:28
Let c15(k100)=k51
Name c15 'Y'
Let c16(k100)=k50
Name c16 'M'
Let c17(k100)=k100
Name c17 'D'
Copy c2 c100;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Name c100 'Copied WT'
Let k1 = mean(c100)
Let k2 = max(c100)
Let k3 = min(c100)
Let c18(k100) = k1
Let c19(k100) = k2
Let c20(k100) = k3
Name c18 'Mean WT'
Name c19 'Max WT'
Name c20 'Min WT'
Copy c3 c101;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Name c101 'Copied pH'
Let k4 = mean(c101)
Let k5 = max(c101)
Let k6 = min(c101)
Let c21(k100) = k4
Let c22(k100) = k5
Let c23(k100) = k6
Name c21 'Mean pH'
Name c22 'Max pH'
Name c23 'Min pH'
```

```
(Page 11)
Copy c4 c102;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k7 = mean(c102)
Let k8 = max(c102)
Let k9 = min(c102)
Let c24(k100) = k7
Let c25(k100) = k8
Let c26(k100) = k9
Name c24 'Mean SC'
Name c25 'Max SC'
Name c26 'Min SC'
Copy c5 c103;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k10 = mean (c103)
Let k11 = max(c103)
Let k12 = min(c103)
Let c27(k100) = k10
Let c28(k100) = k11
Let c29(k100) = k12
Name c27 'Mean DS'
Name c28 'Max DS'
Name c29 'Min DS'
Copy c6 c104;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k13 = mean(c104)
Let k14 = max(c104)
Let k15 = min(c104)
Let c30(k100) = k13
Let c31(k100) = k14
Let c32(k100) = k15
Name c30 'Mean PS'
Name c31 'Max PS'
Name c32 'Min PS'
```

```
(Page 12)
Copy c7 c105;
Include;
Where "'Month' = k50 And 'Year' =
k51 AND 'Day' = k100".
Let k16 = mean(c105)
Let k17 = max(c105)
Let k18 = min(c105)
Let c33(k100) = k16
Let c34(k100) = k17
Let c35(k100) = k18
Name c33 'Mean DO'
Name c34 'Max DO'
Name c35 'Min DO'

enddo
endif
endmacro.
```

Macro for Weekly Mean, Maximum and Minimum RTWQ Data

The macro will ask the user to enter desired year and the weeks of interest. If you wanted every week in 2004 you would enter 2004 1 53 in Minitab. Minitab uses 53 weeks in a year (where the last week is made up of left over days). Once the macro is finished, proceed to the next year of interest.

```
(Page 1)
gmacro
weekly
ERASE k1-k1000
ERASE c10-c1000
Note Set desired year and weeks (i.e.
2004 1 53)
Set c50;
File "terminal";
Nobs 3.
Copy c50 k50 k51 k52
do k100 = k51:k52

NUMERIC 'Date' c10;
Year;
FourDigit.
Name c10 'Y'
NUMERIC 'Date' c11;
Week.
Name c11 'W'
NUMERIC 'DATE' c12;
Month.
Name c12 'M'

Let c14(k100) = k50
Name c14 'Year'
Copy 'M' c99;
Include;
Where "Y=k50 and W=k100".
Let k200 = mean(c99)
Let c15(k100) = k200
Name c15 'Month'
Let c16(k100) = k100
Name c16 'Week'
```

```
(Page 2)
Copy c2 c100;
Include;
Where "Y=k50  And W=k100".
Let k1 = mean(c100)
Let k2 = max(c100)
Let k3 = min(c100)
Let c17(k100) = k1
Let c18(k100) = k2
Let c19(k100) = k3
Name c17 'Mean WT'
Name c18 'Max WT'
Name c19 'Min WT'

Copy c3 c101;
Include;
Where "Y=k50  And W=k100".
Let k4 = mean(c101)
Let k5 = max(c101)
Let k6 = min(c101)
Let c20(k100) = k4
Let c21(k100) = k5
Let c22(k100) = k6
Name c20 'Mean pH'
Name c21 'Max pH'
Name c22 'Min pH'

Copy c4 c102;
Include;
Where "Y=k50  And W=k100".
Let k7 = mean(c102)
Let k8 = max(c102)
Let k9 = min(c102)
Let c23(k100) = k7
Let c24(k100) = k8
Let c25(k100) = k9
Name c23 'Mean SC'
Name c24 'Max SC'
Name c25 'Min SC'
```

```
(Page 3)
Copy c5 c103;
Include;
Where "Y=k50 and W=k100".
Let k10 = mean (c103)
Let k11 = max(c103)
Let k12 = min(c103)
Let c26(k100) = k10
Let c27(k100) = k11
Let c28(k100) = k12
Name c26 'Mean DS'
Name c27 'Max DS'
Name c28 'Min DS'

Copy c6 c104;
Include;
Where "Y=k50 and W=k100".
Let k13 = mean(c104)
Let k14 = max(c104)
Let k15 = min(c104)
Let c29(k100) = k13
Let c30(k100) = k14
Let c31(k100) = k15
Name c29 'Mean PS'
Name c30 'Max PS'
Name c31 'Min PS'

Copy c7' c105;
Include;
Where "Y=k50  And W=k100".
Let k16 = mean(c105)
Let k17 = max(c105)
Let k18 = min(c105)
Let c32(k100) = k16
Let c33(k100) = k17
Let c34(k100) = k18
Name c32 'Mean DO'
Name c33 'Max DO'
Name c34 'Min DO'

enddo
endmacro.
```

Macro for Monthly Mean, Maximum and Minimum RTWQ Data

The macro will ask the user to enter desired year and the months of interest. If you wanted every month in 2004 you would enter 2004 1 12 in Minitab. Once the macro is finished, proceed to the next year of interest.

```
(Page 1)
gmacro
Monthly
ERASE k1-k1000
ERASE c9-c1000
Note Set desired year and months
(i.e. 2004 1 12)
Set c50;
File "terminal";
Nobs 3.
Copy c50 k50 k51 k52

do k100 = k51:k52
NUMERIC 'Date' c10;
Year;
FourDigit.
Name c10 'Y'
NUMERIC 'Date' c11;
Month.
Name c11 'M'
Let c15(k100) = k50
Name c15 'Year'
Let c16(k100) = k100
Name c16 'Month'

Copy c2 c100;
Include;
Where "Y=k50  And M=k100".
Let k1 = mean(c100)
Let k2 = max(c100)
Let k3 = min(c100)
Let c17(k100) = k1
Let c18(k100) = k2
Let c19(k100) = k3
Name c17 'Mean WT'
Name c18 'Max WT'
Name c19 'Min WT'
```

```
(Page 2)
Copy c3 c101;
Include;
Where "Y=k50  And M=k100".
Let k4 = mean(c101)
Let k5 = max(c101)
Let k6 = min(c101)
Let c20(k100) = k4
Let c21(k100) = k5
Let c22(k100) = k6
Name c20 'Mean pH'
Name c21 'Max pH'
Name c22 'Min pH'

Copy c4 c102;
Include;
Where "Y=k50  And M=k100".
Let k7 = mean(c102)
Let k8 = max(c102)
Let k9 = min(c102)
Let c23(k100) = k7
Let c24(k100) = k8
Let c25(k100) = k9
Name c23 'Mean SC'
Name c24 'Max SC'
Name c25 'Min SC'

Copy c5 c103;
Include;
Where "Y=k50 and M=k100".
Let k10 = mean (c103)
Let k11 = max(c103)
Let k12 = min(c103)
Let c26(k100) = k10
Let c27(k100) = k11
Let c28(k100) = k12
Name c26 'Mean DS'
Name c27 'Max DS'
Name c28 'Min DS'
```

```
(Page 3)
Copy c6 c104;
Include;
Where "Y=k50 and M=k100".
Let k13 = mean(c104)
Let k14 = max(c104)
Let k15 = min(c104)
Let c29(k100) = k13
Let c30(k100) = k14
Let c31(k100) = k15
Name c29 'Mean PS'
Name c30 'Max PS'
Name c31 'Min PS'

Copy c7 c105;
Include;
Where "Y=k50  And M=k100".
Let k16 = mean(c105)
Let k17 = max(c105)
Let k18 = min(c105)
Let c32(k100) = k16
Let c33(k100) = k17
Let c34(k100) = k18
Name c32 'Mean DO'
Name c33 'Max DO'
Name c34 'Min DO'

enddo
endmacro.
```

Appendix C

Statistical Overview of the Datasets Used for Water Temperature Regression

Tables and plots are attached in this appendix to provide a more complete statistical overview of the datasets developed for developing regression models for water temperature at the RTWQ stations.

Figure C.1 presents a side by side comparison of water temperature over time at the RTWQ stations. Note how mean daily water temperatures are highest during the summer months - with mean values going up over 20 degrees Celsius in July and August. Note that mean daily water temperature at the stations never really gets overly high as Newfoundland does have a rather colder climate.

Figure C.2 presents a side by side comparison of air temperature over time at the RTWQ stations. Note how During the summer months mean daily air temperature at each station can get up to around 20 degrees Celsius but these periods of time are usually not much longer than one month. There are no gaps in the data plotted in this figure because the historical record of air temperature used for the stations is continuous (unlike the real time sensors, the weather monitoring stations are rarely taken offline).

Figure C.3 presents a comparison of mean daily stage levels recorded at the RTWQ stations. Note how mean daily stage at Humber River is significantly higher than the other three - with an overall mean level for the entire dataset of 2.11 meters. The mean daily stage levels at Peter's River are the next highest (overall mean of the entire dataset of 1.14 meters). Leary's Brook and Waterford River are the smallest of the four (overall mean of 0.76 and 0.56, respectively).

Tables C.1 to C.6 present a detailed statistical overview of the mean, maximum and minimum water temperature, air temperature and stage levels at the stations.

Table C.7 presents the Pearson's correlation coefficients for the mean water temperature, air temperature and stage at the RTWQ stations.

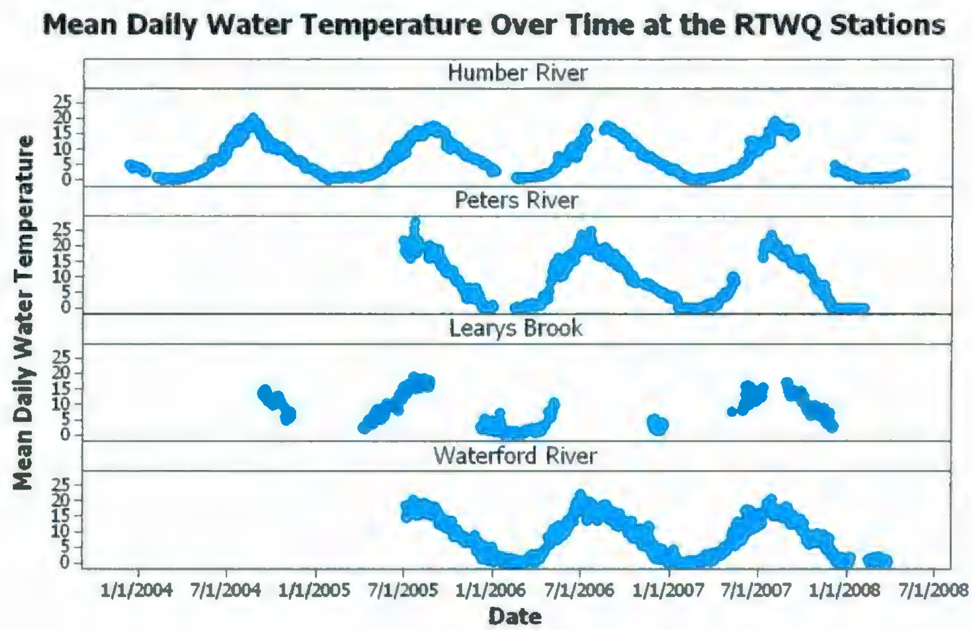


Figure C.1 - Comparison of mean daily water temperature at the RTWQ stations

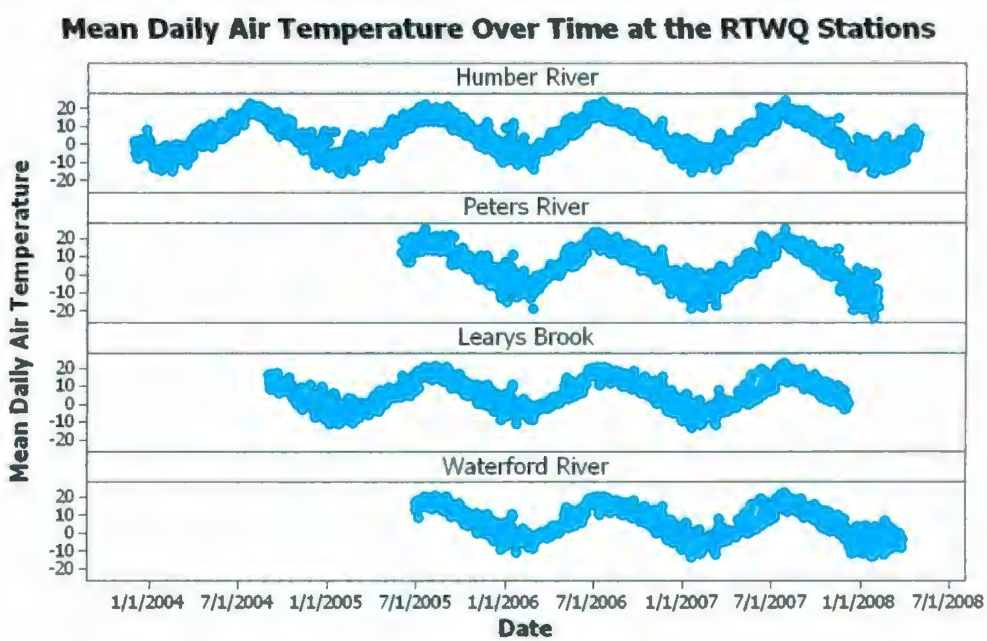


Figure C.2 - Comparison of mean daily air temperature at the RTWQ stations

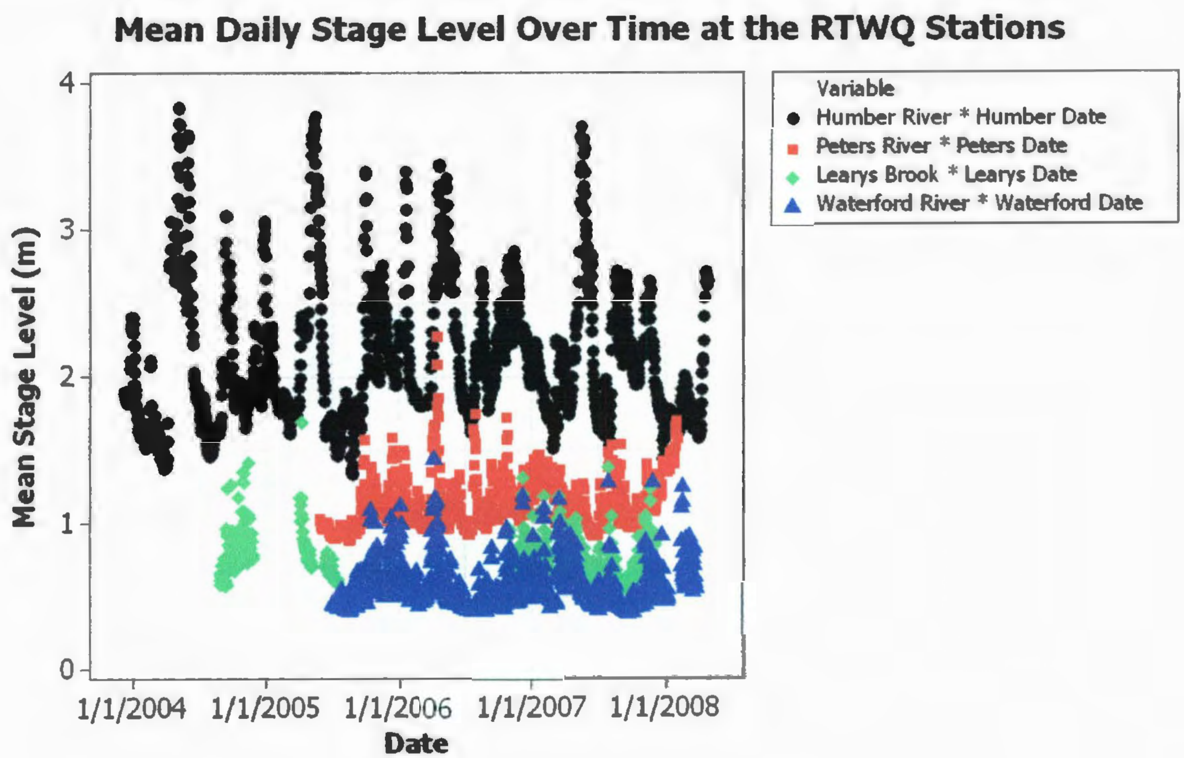


Figure C.3 Comparison of mean daily stage at the RTWQ stations

Table C.1 - General Overview of the Datasets Developed for NF02YL0012 - Humber River

		Water Temperature			Air Temperature			Stage		
Dataset	Obs.	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Monthly Models - For Regression (Dec 03 to Dec 06)										
Mean Monthly	37	7.068	0.384	17.845	4.89	-8.99	18.07	2.123	1.521	3.297
Max Monthly	37	9.06	1.11	20.67	17.75	2	29.1	2.617	1.768	3.867
Min Monthly	37	5.004	-0.1	16.1	-5.73	-21	7.9	1.784	1.331	2.557
Monthly Models - For Prediction (Jan 07 to April 08)										
Mean Monthly	16	5.05	0.336	17.14	2.16	-7.75	18.02	2.0556	1.7026	2.995
Max Monthly	16	6.89	1.18	20.27	15.91	1.7	31.5	2.417	1.82	3.723
Min Monthly	16	2.23	-0.08	15.34	-9.48	-20.3	7.3	1.7115	1.399	2.15
Weekly Models - For Regression (Dec 03 to Dec 06)										
Mean Weekly	149	7.106	0.253	18.529	5.141	-12.94	20.508	2.133	1.388	3.658
Max Weekly	149	7.97	0.57	20.67	14.101	-3	29.1	2.2831	1.435	3.867
Min Weekly	149	6.149	-0.1	17.34	-2.564	-21	14.3	1.9904	1.331	3.584
Weekly Models - For Prediction (Jan 07 to April 08)										
Mean Weekly	74	4.564	0.18	18.377	1.83	-12.29	20.77	2.0598	1.5612	3.4516
Max Weekly	74	5.405	0.57	20.27	11.21	-4.9	31.5	2.2009	1.665	3.723
Min Weekly	74	3.389	-0.08	16.58	-6.41	-20.3	13	1.9372	1.399	3.118
Daily Models - For Regression (Dec 03 to Dec 06)										
Mean Daily	986	7.253	0.0233	20.124	5.385	-16.56	23.033	2.1413	1.3367	3.8299
Max Daily	986	7.577	0.19	20.67	9.563	-15	29.1	2.1721	1.344	3.867
Min Daily	986	6.935	-0.1	19.77	1.263	-21	20	2.1104	1.331	3.807
Daily Models - For Prediction (Jan 07 to April 08)										
Mean Daily	484	4.663	-0.026	19.401	2.19	-16.32	24.729	2.0545	1.4335	3.6953
Max Daily	484	5.002	0.12	20.27	6.286	-14.3	31.5	2.0821	1.458	3.723
Min Daily	484	4.3	-0.08	19.25	-1.864	-20.3	19.7	2.0273	1.399	3.67

Note - the maximum value of the max monthly, max weekly and max daily datasets will match each other. The minimum value of the datasets will not (as the minimum value for max monthly will be determined from 12 values while the minimum value for max weekly will be determined from 53 values (Minitab counts 53 weeks in the year). The same logic applies to the minimum value of the min monthly, min weekly and min daily datasets being equal but the max value is not.

Table C.2 - Statistical Summary of the Humber River Warming and Cooling Season Datasets

Cooling season - August to January and Warming season - February to July

Dataset	Obs.	Mean WT	Min WT	Max WT	Mean AT	Min AT	Max AT	Mean Stage	Min Stage	Max Stage
Monthly (December 2003 to December 2006)										
Cooling Mean	19	9.06	1.21	17.85	4.44	-8.99	17.69	2.1123	1.5732	2.5696
Warming Mean	18	4.97	0.384	14.11	5.36	-6.33	18.07	2.134	1.521	3.297
Cooling Max	19	10.93	2.71	20.67	17.46	2	27.9	2.641	1.812	3.421
Warming Max	18	7.08	1.11	17.51	18.06	3.2	29.1	2.592	1.768	3.867
Cooling Min	19	6.89	-0.05	16.1	-5.35	-20.2	6.7	1.8076	1.331	2.213
Warming Min	18	3.014	-0.1	10.77	-6.12	-21	7.9	1.7594	1.347	2.557
Weekly (December 2003 to December 2006)										
Cooling Mean	76	9.454	0.302	18.529	5.253	-12.94	19.161	2.1066	1.4815	2.908
Warming Mean	73	4.662	0.253	15.353	5.024	-9.947	20.508	2.1618	1.3882	3.658
Cooling Max	76	10.181	0.67	20.67	13.747	-3	29	2.2629	1.533	3.412
Warming Max	73	5.667	0.57	17.42	14.47	-2.8	29.1	2.3042	1.435	3.867
Cooling Min	76	8.51	-0.05	17.34	-1.842	-20.2	12.7	1.9634	1.331	2.574
Warming Min	73	3.69	-0.1	14.03	-3.32	-21	14.3	2.018	1.347	3.584
Daily (December 2003 to December 2006)										
Cooling Mean	500	9.504	0.0233	20.124	5.233	-16.56	21.446	2.1108	1.3367	3.393
Warming Mean	486	4.938	0.0346	16.907	5.541	-15.55	23.033	2.1731	1.3626	3.83
Cooling Max	500	9.747	0.25	20.67	8.877	-15	27.9	2.1414	1.344	3.412
Warming Max	486	5.345	0.19	17.51	10.27	-12.4	29.1	2.2039	1.378	3.867
Cooling Min	500	9.246	-0.05	19.77	1.644	-20.2	19.9	2.0806	1.331	3.324
Warming Min	486	4.558	-0.1	16.6	0.871	-21	20	2.1414	1.347	3.807

Table C.3 - Statistical Properties of the Warming and Cooling Datasets Used for Prediction

Humber River - High and Low Season Datasets developed for predicting maximum water temperature

Dataset	Obs.	Mean WT	Min WT	Max WT	Mean AT	Min AT	Max AT	Mean Stage	Min Stage	Max Stage
Monthly Models for Prediction (Jan 07 to April 08)										
Cooling Season - August to January										
Mean Monthly	5	7.93	17.14	1.08	2.14	16.1	-6.1	2.0174	2.319	1.703
Max Monthly	5	9.628	20.27	1.98	14.94	25.5	7.6	2.3478	2.72	1.82
Min Monthly	5	3.316	15.34	0	-8.92	7.3	-20.3	1.6652	2.043	1.399
Warming Season - February to July										
Mean Monthly	9	3.44556	13.52	0.34	1.62222	18	-7.8	2.04022	2.995	1.73
Max Monthly	9	5.36889	18.92	1.18	15.7889	31.5	1.7	2.446	3.723	1.976
Min Monthly	9	1.63	8.36	-0.08	-10.6	5.7	-19.3	1.66733	2.15	1.495
Weekly Models for Prediction (Jan 07 to April 08)										
Cooling Season - August to January										
Mean Weekly	25	6.4056	18.38	0.35	0.152	18.7	-12.3	1.97016	2.642	1.561
Max Weekly	25	7.0812	20.27	0.73	8.064	25.5	-4.9	2.09192	2.72	1.665
Min Weekly	25	4.808	16.58	0	-7.72	11.1	-20.3	1.858	2.49	1.399
Warming Season - February to July										
Mean Weekly	38	3.35184	14.63	0.18	1.76579	20.8	-11.7	2.06489	3.452	1.635
Max Weekly	38	4.30158	17.57	0.57	11.9211	31.5	-3.9	2.22997	3.761	1.694
Min Weekly	38	2.45553	11.85	-0.08	-7.0895	13	-19.3	1.92929	3.118	1.495
Daily Models for Prediction (Jan 07 to April 08)										
Cooling Season - August to January										
Mean Daily	130	7.07646	19.4	0.26	1.30923	19.9	-16.3	1.96398	2.71	1.433
Max Daily	130	7.31623	20.27	0.42	4.74769	25.5	-13.3	1.99091	2.72	1.458
Min Daily	130	6.66046	19.25	0	-2.1808	15.9	-20.3	1.93615	2.692	1.399
Warming Season - February to July										
Mean Daily	268	3.49295	18.3	-0.03	1.71642	24.7	-14.9	2.03859	3.695	1.504
Max Daily	268	3.87948	18.92	0.12	6.24254	31.5	-12.9	2.06802	3.723	1.517
Min Daily	268	3.15478	17.71	-0.08	-2.7892	19.7	-19.3	2.00994	3.67	1.495

Table C.4 - General Overview of the Datasets Developed for NF02YO0121 - Peters River

		Water Temperature			Air Temperature			Stage		
Dataset	Obs.	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Monthly Models - For Regression (July 2005 to May 2007)										
Mean Monthly	23	8.07	-0.099	20.58	4.28	-8.95	17.86	1.1415	0.9352	1.4796
Max Monthly	23	13.11	-0.045	30.19	18.92	3.2	33.6	1.4357	0.995	2.454
Min Monthly	23	5.14	-0.603	14.53	-12.13	-31.6	0.7	0.9841	0.884	1.075
Monthly Models - For Prediction (July 2007 to February 2008)										
Mean Monthly	8	8.69	-0.208	20.59	3.4	-8.61	17.31	1.2045	1.0399	1.4287
Max Monthly	8	12.17	-0.149	27.88	17.93	4.6	34	1.4941	1.184	1.717
Min Monthly	8	5.84	-0.23	15.86	-13.51	-32.6	2.6	1.0375	0.917	1.332
Weekly Models - For Regression (Week 28, 2005 - July to Week 20, 2007 - May)										
Mean Weekly	91	8.341	-0.133	22.527	4.401	-13.76	20.82	1.1636	0.9286	1.8396
Max Weekly	91	11.197	-0.089	30.194	14.51	-3.6	33.6	1.2965	0.953	2.454
Min Weekly	91	6.558	-0.603	18.33	-7.19	-31.6	12.2	1.0548	0.884	1.4
Weekly Models - For Prediction (Week 28, 2007 - July to Week 6, 2008 - February)										
Mean Weekly	32	8.76	-0.206	21.06	3.75	-15.84	20.48	1.2016	0.9821	1.5588
Max Weekly	32	10.88	-0.197	27.88	13.64	-2.7	34	1.322	0.995	1.717
Min Weekly	32	7.35	-0.23	18.2	-8.4	-32.6	11.1	1.1128	0.961	1.431
Daily Models - For Regression (July 1, 2005 to May 15, 2007)										
Mean Daily	595	8.448	-0.287	27.876	4.609	-19.08	25.458	1.1445	0.884	2.2744
Max Daily	595	9.819	-0.162	30.194	9.713	-14.1	33.6	1.1722	0.884	2.454
Min Daily	595	7.434	-0.603	23.871	-1.223	-31.6	21	1.1185	0.884	2.014
Daily Models - For Prediction (July 10, 2007 to February 5, 2008)										
Mean Daily	196	8.764	-0.22	23.622	4.325	-23.87	25.183	1.1895	0.9639	1.6967
Max Daily	196	9.734	-0.219	27.88	9.421	-13.8	34	1.2128	0.967	1.717
Min Daily	196	8.023	-0.23	21.86	-1.467	-32.6	18.5	1.1665	0.961	1.684

Table C.5 - General Overview of the Datasets Developed for NF02ZM0178 Leary's Brook

		Water Temperature			Air Temperature			Stage		
Dataset	Obs.	Mean	Min	Max	Mean	Min	Max	MEan	Min	Max
Monthly Models - For Regression (September 2004 to December 2006)										
Mean Monthly	16	6.93	0.705	16.58	5.12	-3.65	16.98	0.789	0.6317	0.9535
Max Monthly	16	11.04	2.73	20.94	16.75	2.9	25.7	1.45	0.897	2.121
Min Monthly	16	3.94	-0.161	14.02	-4.77	-13.5	8.9	0.6528	0.564	0.73
Monthly Models - For Prediction (May 2007 to December 2007)										
Mean Monthly	8	10.52	2.75	17.02	8.73	-3.64	17.09	0.7145	0.5777	0.8542
Max Monthly	8	14.65	4.16	19.7	20.91	8.4	29.6	1.21	0.892	1.949
Min Monthly	8	7.23	1.67	14.9	-0.663	-14.5	8.7	0.5931	0.542	0.646
Weekly Models - For Regression (Week 37, 2004 - September to Week 51, 2006 - December)										
Mean Weekly	57	6.923	0.268	17.498	4.981	-6.403	18.744	0.8008	0.5935	1.1827
Max Weekly	57	9.491	1.248	20.936	13.107	-1.8	25.7	1.1166	0.669	2.121
Min Weekly	57	5.071	-0.161	14.62	-1.765	-13.5	13.2	0.696	0.564	1.012
Weekly Models - For Prediction (Week 19, 2007 - May to Week 49, 2007 - December)										
Mean Weekly	23	10.464	2.629	17.146	10.292	-0.212	19.453	0.7025	0.5505	0.9458
Max Weekly	23	13.705	3.763	19.699	19.58	3.9	29.6	1.015	0.558	1.949
Min Weekly	23	7.911	1.673	14.899	3.465	-5.4	13.8	0.6316	0.542	0.821
Daily Models - For Regression (September 11, 2004 to December 19, 2006)										
Mean Daily	347	6.789	0.0738	19.131	4.944	-10.87	21.221	0.8075	0.5741	1.3517
Max Daily	347	7.928	0.263	20.936	8.518	-8.9	25.7	0.8855	0.577	2.121
Min Daily	347	5.834	-0.161	17.997	1.574	-13.5	18.7	0.7582	0.565	1.211
Daily Models - For Prediction (May 11, 2007 to December 6, 2007)										
Mean Daily	136	10.643	2.332	17.318	10.523	-3.2	23.233	0.6969	0.5432	1.3797
Max Daily	136	12.174	2.832	19.699	14.558	-0.8	29.6	0.7623	0.544	1.949
Min Daily	136	9.34	1.673	16.498	6.667	-4.7	20	0.66	0.542	1.012

Table C.6 General Overview of the Datasets Developed for NF02ZM0009 - Waterford River

		Water Temperature			Air Temperature			Stage		
Dataset	Obs.	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Monthly Models - For Regression (July 2005 to March 2007)										
Mean Monthly	21	8.09	0.221	17.71	5.94	-5.74	17.11	0.5635	0.4156	0.9088
Max Monthly	21	12.83	1.26	24.05	17.1	2.9	27.4	1.102	0.607	1.977
Min Monthly	21	4.06	-0.26	13.73	-4.33	-16.6	8.9	0.4388	0.386	0.579
Monthly Models - For Prediction (April 2007 to March 2008)										
Mean Monthly	12	7.57	0.665	16.58	4.95	-3.62	17.09	0.5671	0.412	0.765
Max Monthly	12	12.43	2.2	24.3	17.39	7.9	29.6	1.048	0.513	1.956
Min Monthly	12	4.14	-0.13	11.99	-4.66	-16.2	8.7	0.4473	0.38	0.537
Weekly Models - For Regression (Week 28, 2005 - July to Week 13, 2007 - March)										
Mean Weekly	90	8.1	-0.168	18.733	5.917	-9.265	18.744	0.5617	0.4049	1.0525
Max Weekly	90	11.05	0.04	24.05	13.759	-2.1	27.4	0.7911	0.429	1.977
Min Weekly	90	5.709	-0.26	15.94	-0.733	-16.6	13.6	0.4789	0.386	0.93
Weekly Models - For Prediction (Week 14, 2007 - April to Week 13, 2008 - March)										
Mean Weekly	48	7.997	0.235	18.456	4.88	-8.24	19.45	0.5708	0.3924	0.9339
Max Weekly	48	11.28	0.93	24.3	13.8	0.8	29.6	0.7999	0.403	1.956
Min Weekly	48	5.597	-0.13	16.29	-1.95	-16.2	13.8	0.49	0.38	0.769
Daily Models - For Regression (July 6, 2005 to March 31, 2007)										
Mean Daily	587	8.121	-0.185	22.46	5.924	-13.46	21.221	0.5588	0.3938	1.4346
Max Daily	587	9.356	-0.17	24.05	9.275	-10.6	27.4	0.607	0.396	1.977
Min Daily	587	7.062	-0.26	20.51	2.764	-16.6	18.7	0.522	0.386	1.236
Daily Models - For Prediction (April 1, 2007 to March 27, 2008)										
Mean Daily	306	8.217	-0.061	20.913	5.076	-13.4	23.233	0.5866	0.3817	1.2778
Max Daily	306	9.74	-0.05	24.3	8.703	-11.8	29.6	0.616	0.383	1.956
Min Daily	306	7	-0.13	19.16	1.647	-16.2	20	0.5319	0.38	1.033

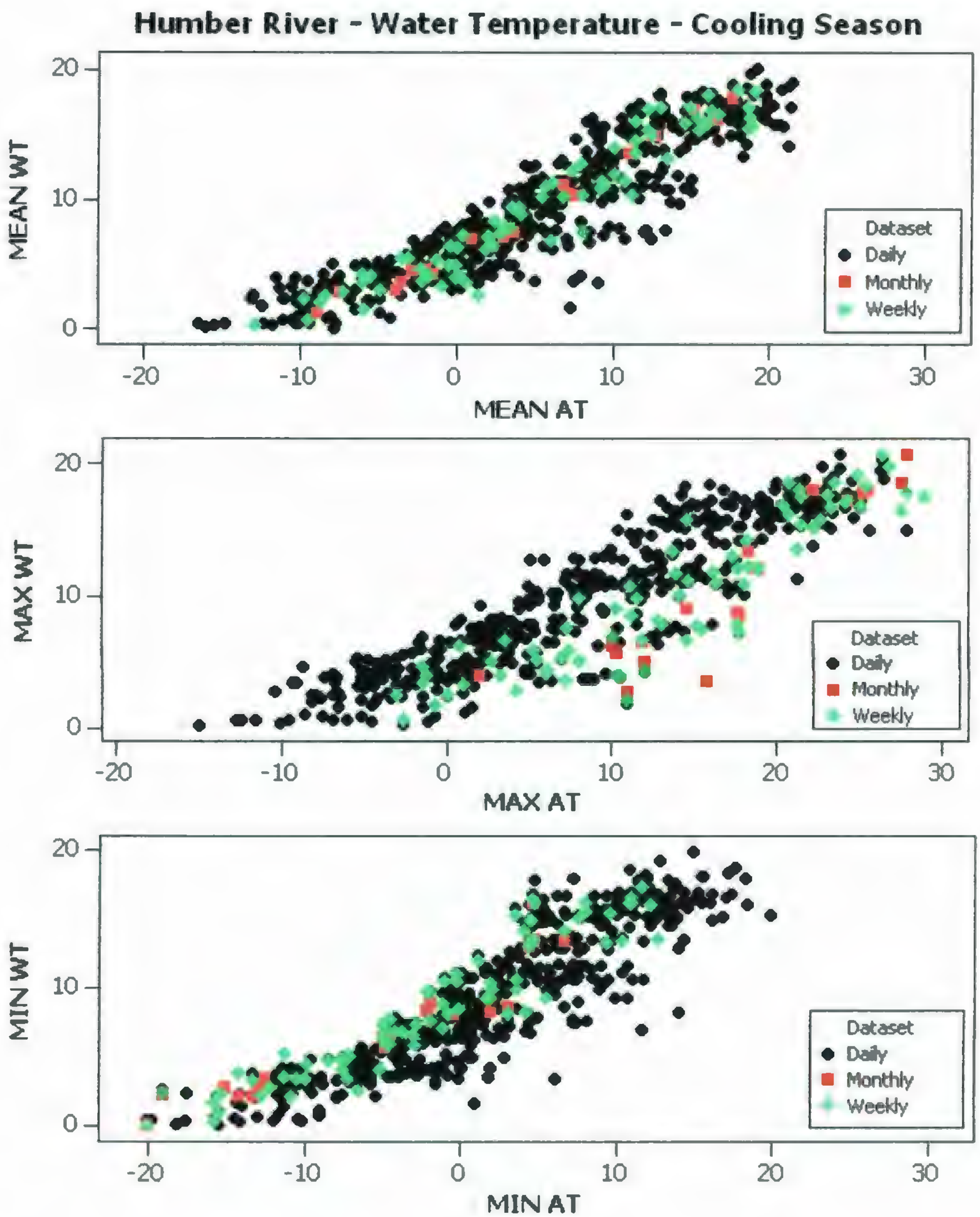
Table C.7 - Pearson Correlation and P-Value for RTWQ Stations - WT, AT and Stage

Humber River			Peter's River			Leary's Brook			Waterford River		
Monthly											
	Mean WT	Mean AT		Mean WT	Mean AT		Mean WT	Mean AT		Mean WT	Mean AT
Mean AT	0.874 0.000		Mean AT	0.973 0.000		Mean AT	0.976 0.000		Mean AT	0.994 0.000	
Mean Stage	-0.175 0.300	0.012 0.942	Mean Stage	-0.547 0.007	-0.538 0.008	Mean Stage	-0.629 0.070	-0.740 0.023	Mean Stage	-0.641 0.002	-0.584 0.005
Weekly											
	Mean WT	Mean AT		Mean WT	Mean AT		Mean WT	Mean AT		Mean WT	Mean AT
Mean AT	0.841 0.000		Mean AT	0.948 0.000		Mean AT	0.970 0.000		Mean AT	0.981 0.000	
Mean Stage	-0.190 0.020	-0.009 0.910	Mean Stage	-0.434 0.000	-0.398 0.000	Mean Stage	-0.595 0.001	-0.569 0.002	Mean Stage	-0.502 0.000	-0.424 0.000
Daily											
	Mean WT	Mean AT		Mean WT	Mean AT		Mean WT	Mean AT		Mean WT	Mean AT
Mean AT	0.801 0.000		Mean AT	0.909 0.000		Mean AT	0.945 0.000		Mean AT	0.956 0.000	
Mean Stage	-0.207 0.000	-0.028 0.378	Mean Stage	-0.447 0.000	-0.399 0.000	Mean Stage	-0.469 0.000	-0.391 0.000	Mean Stage	-0.413 0.000	-0.298 0.000

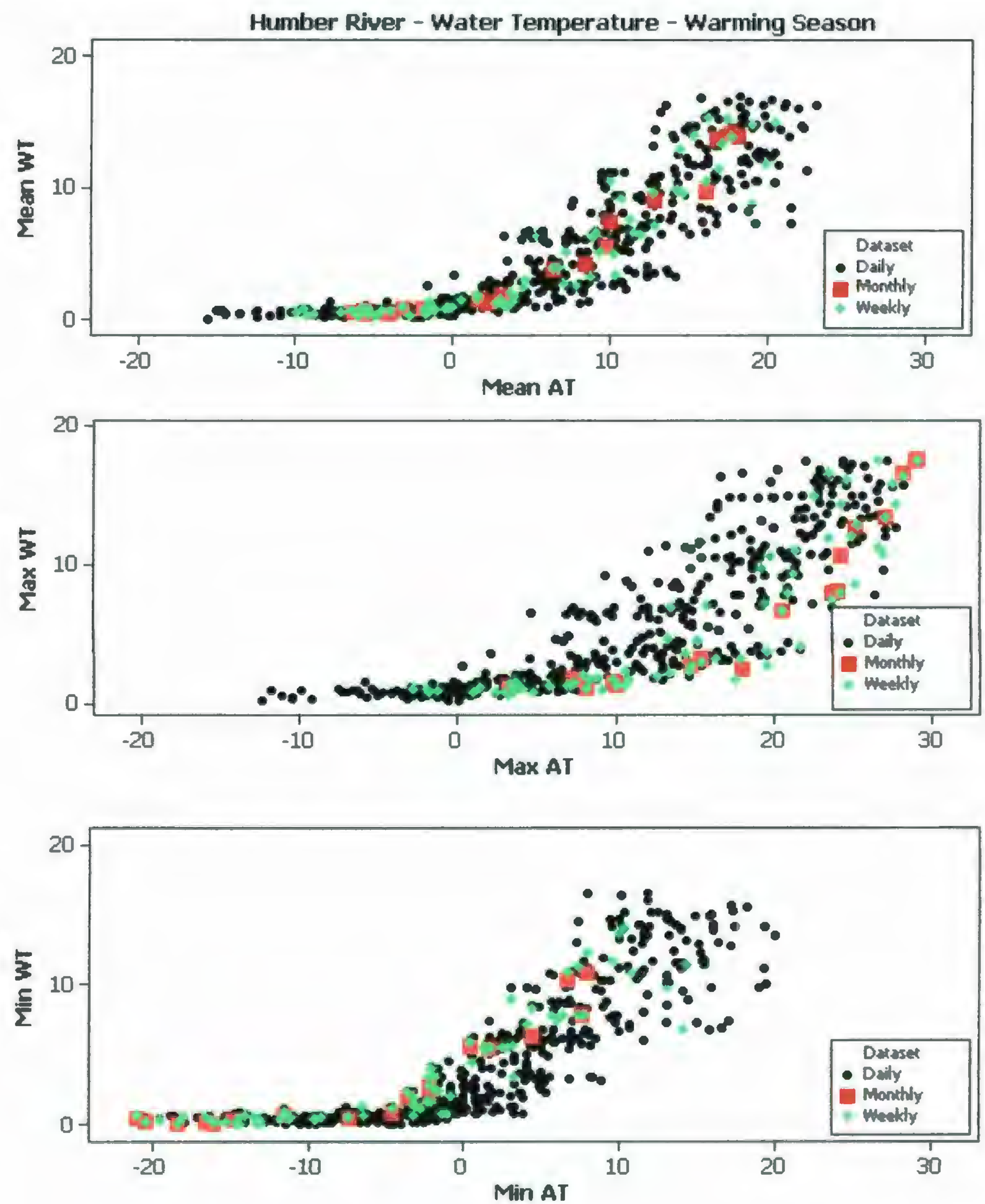
Appendix D

Graphical Plots of Water Temperature and

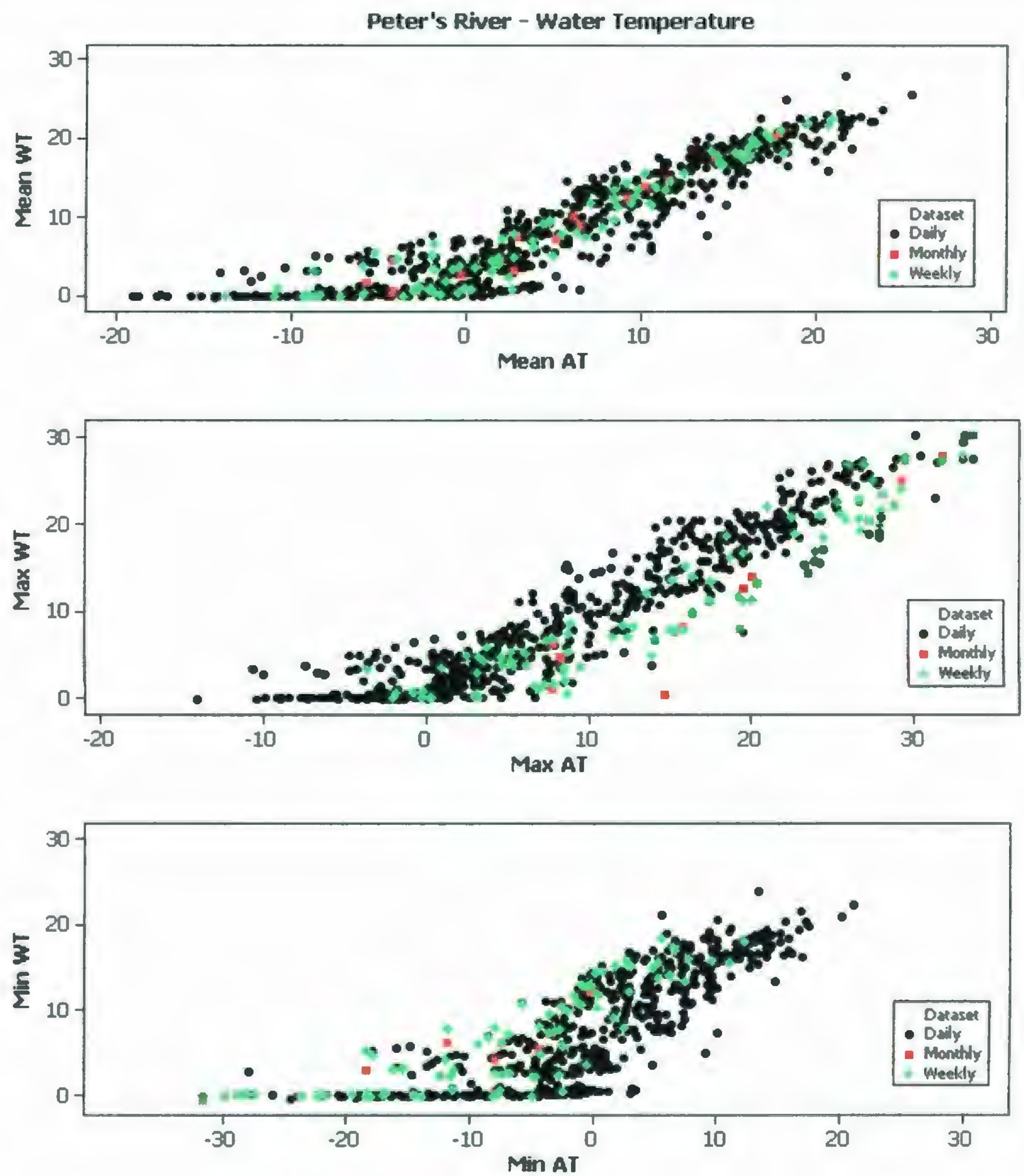
Air Temperature

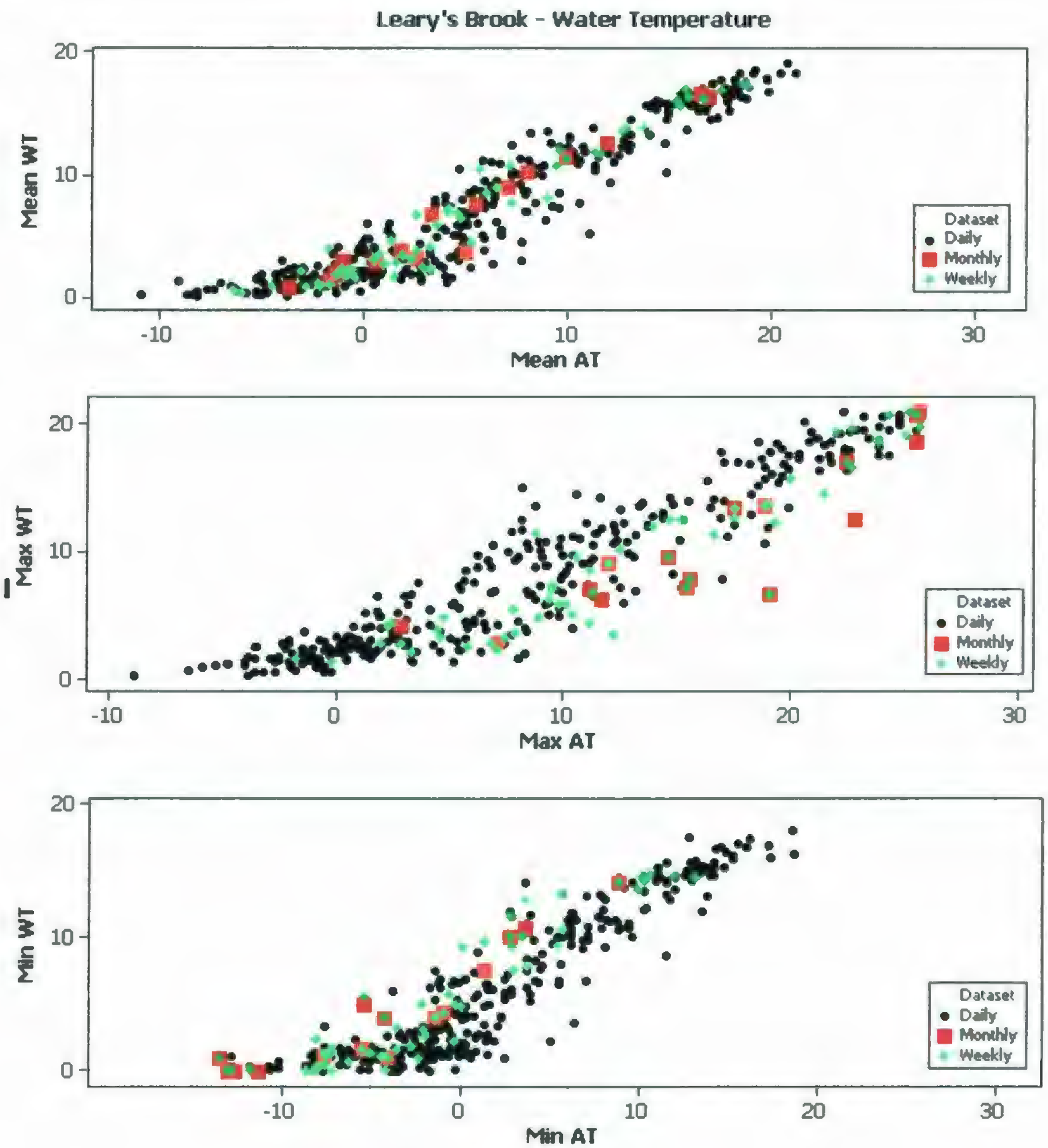


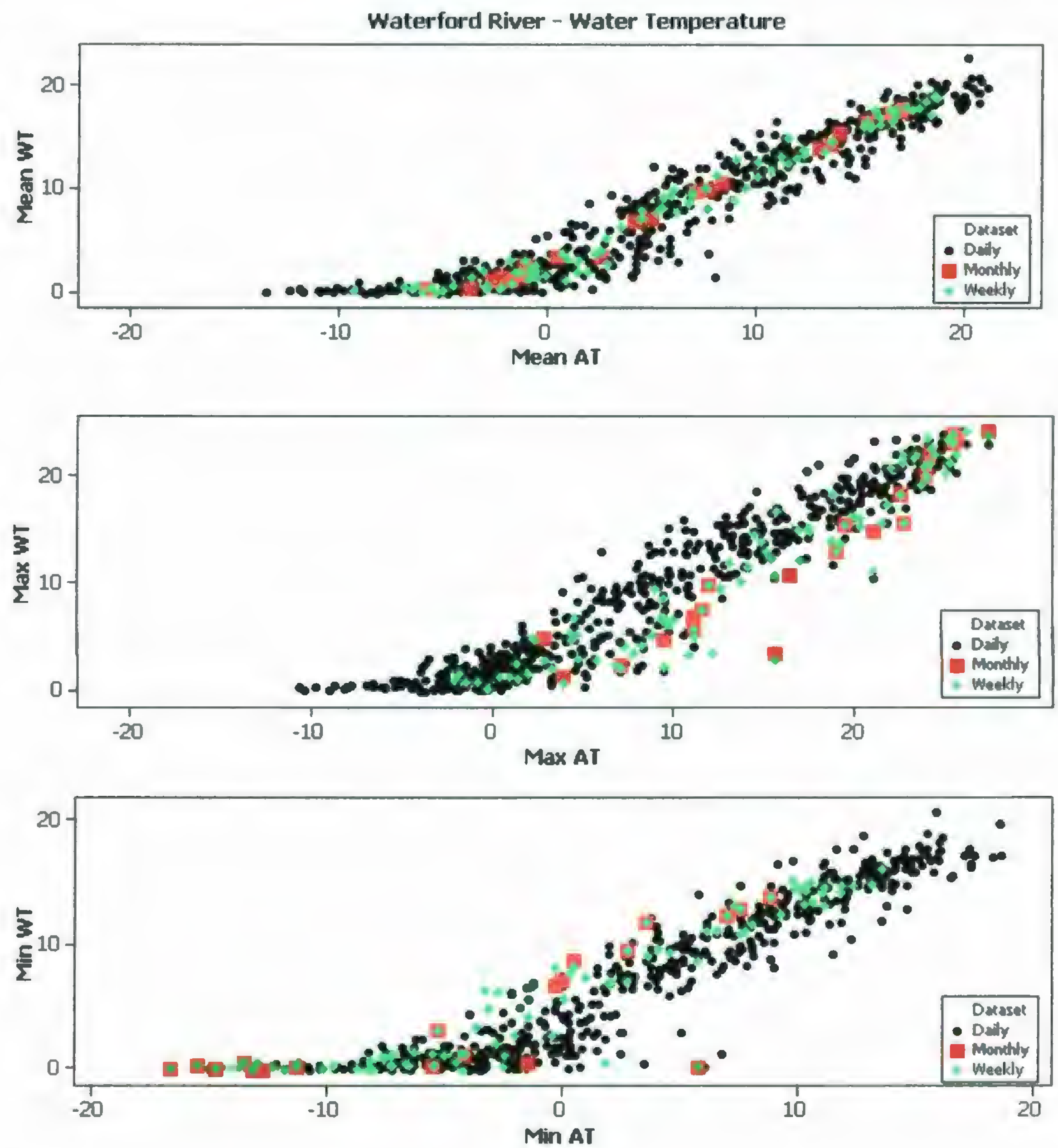
Humber River - Water Temperature - Cooling Season
MEAN WT
MEAN AT
MAX WT
MAX AT
MIN WT
MIN AT
Dataset
Daily
Monthly
Weekly



Humber River - Water Temperature - Warming Season
Mean WT
Mean AT
Max WT
Max AT
Min WT
Min AT
Dataset
Daily
Monthly
Weekly


Peter's River - Water Temperature
Mean WT
Mean AT
Max WT
Max AT
Min WT
Min AT
Dataset
Daily
Monthly
Weekly
30
20
10
0
-30
-20
-10


Leary's Brook - Water Temperature
Mean WT
Mean AT
Max WT
Max AT
Min WT
Min AT
Dataset
Daily
Monthly
Weekly


Waterford River - Water Temperature
Mean WT
Mean AT
Max WT
Max AT
Min WT
Min AT
Dataset
Daily
Monthly
Weekly

Appendix E

Points to Remember When Using Curve Fitting Software

When working with the curve fitting software (like Minitab or Datafit) for determining the goodness of fit of the model and the appropriates of the model for predicting water temperature, there are a number points that should be considered:

1. Look at the residual scatter plot output in Datafit and make sure the residuals are randomly scattered around zero and show no discernable pattern. If there are groupings of residuals with similar signs, or if the increase (or decrease) of the residuals is a factor of the size of the independent variable then it is likely that another functional approximation probably exists to better describe the data.
2. Check the normality of residual plots in Datafit to ensure the residuals are normally distributed This plot show a plot of the normalized residuals on the vertical axis and the normal quantiles on the horizontal axis). If the residuals are normally distributed around zero then the plot will be a straight line with a 45 degree slope that passes through the origin.
3. Check the plot of the regression model and the data points - the data points should be randomly scattered above and below the curve.
4. Check to see how well the developed regression model describes the data. This can be done be observing the following calculated parameters:
   a. R^2 which is a measure of variation in the data points Yi that the regression model explains. An R^2 value equal to 1 would mean that the developed curve passes through every data point while an R^2 value equal to 0 would mean that the regression model does not describe the data any better than a horizontal line that passes through the average of all data points.
   b. The Residual Sum of Squares (RSS) which is the sum of the squares of the differences between the data that was entered and the curve generated by fitting the regression model. A model that perfectly fits the data would have an RSS value equal to 0.
   c. The Standard Error of the Estimate is the standard deviation of the differences between the entered data and the curve generated from the fitted model. The standard error gives an idea about how scattered the residuals are around the average. As the standard error approaches zero then its possible to be reasonably certain that the model accurately describes the data. A perfectly fit model would have a standard error equal to zero.

Appendix F

Curve Fitting Results for Water Temperature

Regression Models

Table F.1 - Regression Models for Predicting Humber River Cooling Season Water Temperature

	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Humber River Cooling Season - High Water Temperature Months (August to January) *(19 Monthly Observations, 76 Weekly Observations and 500 Daily Observations)*									
*Linear Model: Tw = a*Ta + b*									
a	0.64	0.61	0.55	0.78	0.59	0.51	0.50	0.55	0.56
b	6.20	6.27	6.62	-2.67	2.09	5.23	9.54	9.53	8.33
RSS	10.09	181.77	2336.13	109.51	430.71	2811.90	46.32	224.35	2615.68
R^2 Adj	0.98	0.91	0.83	0.82	0.81	0.80	0.87	0.87	0.80
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	21.41	20.31	20.50	27.87	27.54	22.06	37.65	18.26	18.85
b	0.15	0.15	0.13	0.15	0.11	0.11	0.11	0.16	0.15
c	7.20	6.56	6.78	21.25	19.74	11.57	11.11	-0.51	2.17
RSS	5.13	147.35	2150.71	69.37	361.24	2693.80	32.99	187.45	2349.29
R^2 Adj	0.99	0.93	0.84	0.88	0.84	0.81	0.90	0.89	0.82
Logistic 2 Model: Tw = d + (a-d)/(1 + exp (b(c-Ta)))*									
a	21.45	19.34	19.88	18.44	21.05	22.28	258.07	17.43	17.90
b	0.15	0.17	0.14	1.01	0.20	0.11	0.07	0.18	0.18
c	7.20	6.59	6.73	18.60	17.37	11.61	47.35	-0.47	2.27
d	-0.03	0.99	0.55	5.24	3.09	-0.18	-1.90	0.85	1.05
RSS	5.13	145.99	2148.62	44.02	344.91	2693.69	32.46	186.09	2337.03
R^2 Adj	0.99	0.93	0.84	0.92	0.85	0.81	0.90	0.89	0.82
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	NS	NS	8.24	NS	NS	7.06	NS	NS	10.00
b			0.54			0.50			0.55
c			-0.76			-0.83			-0.80
RSS			2291.78			2755.01			2568.66
R^2 Adj			0.83			0.81			0.80

Table F.2 - Regression Models for Predicting Humber River Warming Season Water Temperature

	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Humber River Warming Season - Low Water Temperature Months (February to July) *(18 Monthly Observations, 73 Weekly Observations and 486 Daily Observations)*									
*Linear Model: Tw = a*Ta + b*									
a	0.55	0.51	0.48	0.64	0.50	0.44	0.32	0.37	0.45
b	2.02	2.12	2.30	-4.51	-1.54	0.81	4.98	4.93	4.17
RSS	49.60	313.56	2989.49	102.28	582.41	3707.99	53.65	296.85	3190.17
R^2 Adj	0.88	0.80	0.74	0.83	0.71	0.70	0.75	0.74	0.70
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	21.56	15.68	15.77	366.91	32.80	18.90	11.11	11.39	13.10
b	0.19	0.24	0.22	0.13	0.14	0.15	0.30	0.33	0.30
c	14.96	11.48	11.81	51.47	28.95	19.06	1.92	2.29	5.35
RSS	9.68	143.42	1911.98	12.75	384.49	2886.83	10.04	110.13	1756.17
R^2 Adj	0.97	0.91	0.83	0.98	0.81	0.77	0.95	0.90	0.83
Logistic 2 Model: Tw = d + (a-d)/(1 + exp (b(c-Ta)))*									
a	21.56	14.87	15.22	29.61	26.70	17.96	11.07	11.30	12.91
b	0.19	0.28	0.24	0.24	0.16	0.17	0.31	0.35	0.33
c	14.96	11.29	11.67	27.82	26.41	18.56	1.91	2.38	5.45
d	0.00	0.47	0.34	1.24	0.40	0.29	0.02	0.22	0.32
RSS	9.68	140.38	1904.27	10.63	383.78	2883.77	10.04	108.99	1745.67
R^2 Adj	0.97	0.91	0.84	0.98	0.80	0.77	0.95	0.90	0.83
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	7.88	6.38	5.98	0.75	3.75	4.84	9.13	8.86	7.42
b	0.57	0.53	0.50	0.72	0.54	0.46	0.33	0.39	0.47
c	-2.81	-2.02	-1.77	-2.57	-2.57	-1.96	-2.34	-1.92	-1.55
RSS	13.66	212.09	2340.63	44.84	374.05	2882.28	41.13	226.80	2709.59
R^2 Adj	0.96	0.87	0.80	0.92	0.81	0.77	0.80	0.80	0.74

Table F.3 - Regression Models for Predicting Peter's River Water Temperature

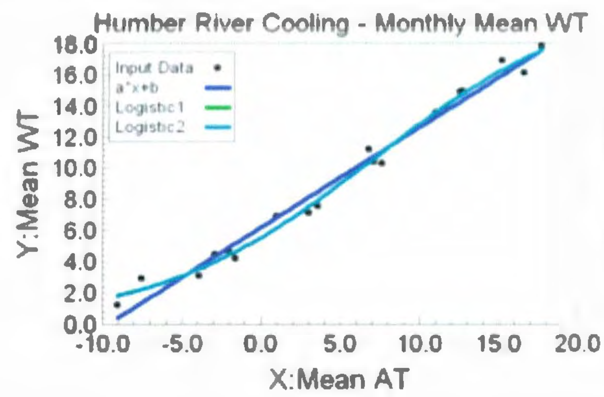
	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Peter's River (23 Monthly Observations, 91 Weekly Observations, 595 Daily Observations)									
*Linear Model: Tw = a*Ta + b*									
a	0.83	0.78	0.72	1.00	0.84	0.76	0.42	0.49	0.55
b	4.52	4.91	5.15	-5.78	-0.99	2.43	10.25	10.07	8.11
RSS	61.89	477.92	5976.11	197.43	819.56	6549.54	154.47	950.50	9067.82
R^2 Adj	0.94	0.90	0.81	0.90	0.89	0.84	0.75	0.71	0.65
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	21.13	22.89	23.16	43.07	34.59	28.19	68.64	18.19	19.52
b	0.24	0.21	0.20	0.14	0.14	0.15	0.16	0.22	0.21
c	7.63	8.73	9.15	27.12	22.04	15.46	9.08	-1.56	3.06
RSS	31.03	237.02	3962.91	97.50	599.18	5386.46	52.19	612.91	6495.23
R^2 Adj	0.97	0.95	0.88	0.95	0.92	0.87	0.91	0.81	0.75
Logistic 2 Model: Tw = d + (a-d)/(1 + exp (b(c-Ta)))*									
a	21.74	23.26	23.54	37.24	35.99	29.59	335.26	17.24	19.23
b	0.22	0.20	0.19	0.17	0.13	0.14	0.14	0.27	0.22
c	7.71	8.79	9.18	25.52	22.46	15.67	23.37	-1.66	3.05
d	-0.48	-0.27	-0.32	1.43	-0.51	-1.01	-0.38	0.72	0.27
RSS	30.66	236.49	3957.73	95.46	598.49	5361.10	51.69	605.70	6489.75
R^2 Adj	0.97	0.95	0.88	0.95	0.91	0.87	0.91	0.81	0.75
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	NS	NS	8.58	NS	NS	6.85	NS	21.23	12.78
b			0.70			0.74		0.45	0.54
c			-2.97			-3.63		-10.84	-4.24
RSS			5752.56			6167.02		849.88	8662.33
R^2 Adj			0.82			0.85		0.74	0.67

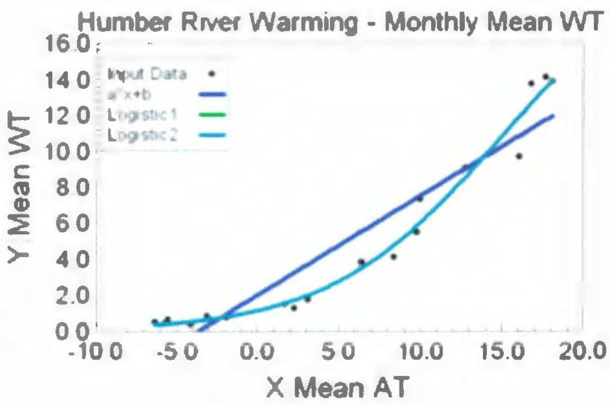
Table F.4 - Regression Models for Predicting Leary's Brook Water Temperature

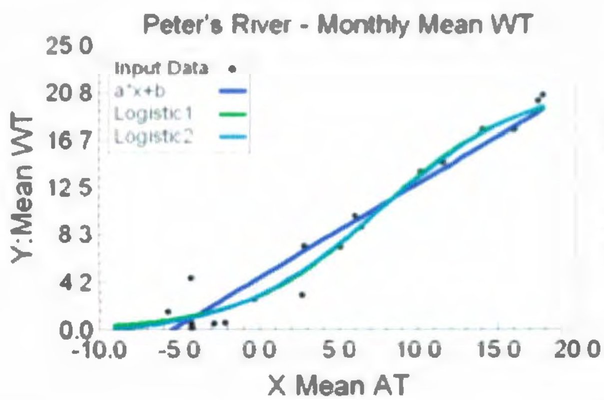
	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Leary's Brook (16 Monthly Observations, 57 Weekly Observations, 347 Daily Observations)									
*Linear Model: Tw = a*Ta + b*									
a	0.81	0.79	0.73	0.77	0.75	0.68	0.60	0.69	0.71
b	2.78	2.97	3.18	-1.79	-0.34	2.12	6.78	6.30	4.72
RSS	18.87	99.31	1153.85	105.07	286.74	1706.74	42.49	210.46	1510.14
R^2 Adj	0.95	0.94	0.89	0.78	0.86	0.86	0.84	0.85	0.84
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	18.28	18.41	19.06	1181.08	41.12	24.22	16.28	14.69	16.49
b	0.24	0.23	0.21	0.08	0.11	0.15	0.28	0.35	0.27
c	7.74	7.75	8.72	76.55	25.47	14.77	1.78	1.30	4.80
RSS	15.17	76.14	873.54	72.09	220.36	1523.39	17.57	108.29	1024.05
R^2 Adj	0.96	0.95	0.92	0.84	0.89	0.88	0.93	0.92	0.89
Logistic 2 Model: Tw = d + (a-d)/(1 + exp (b(c-Ta)))*									
a	18.75	19.43	19.57	921.62	74.62	28.39	16.30	14.58	16.56
b	0.22	0.20	0.20	0.10	0.08	0.11	0.28	0.36	0.26
c	7.75	7.76	8.80	66.23	36.30	16.31	1.79	1.33	4.80
d	-0.47	-1.06	-0.40	1.79	-2.10	-1.81	-0.01	0.17	-0.08
RSS	15.11	74.48	871.40	71.36	219.48	1510.48	17.57	108.05	1023.90
R^2 Adj	0.95	0.95	0.92	0.82	0.89	0.88	0.93	0.92	0.89
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	NS	NS	NS	NS	NS	NS	NS	NS	NS
b									
c									
RSS									
R^2 Adj									

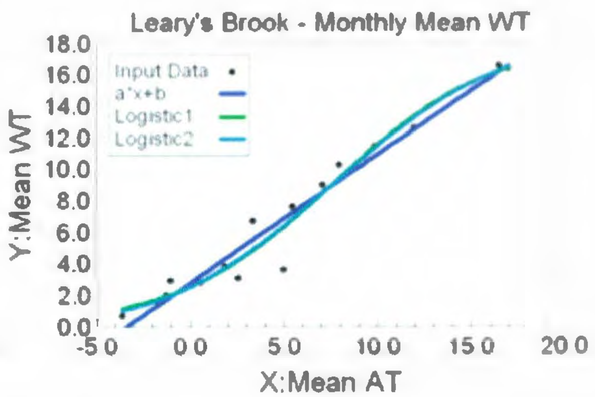
Table F.5 - Regression Models for Predicting Waterford River Water Temperature

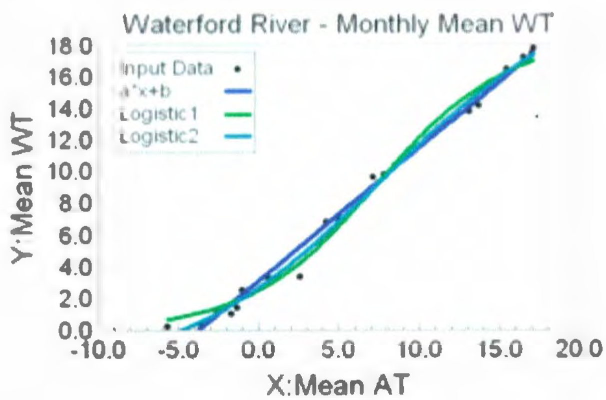
	Mean Datasets			Maximum Datasets			Minimum Datasets		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Waterford River *(21 Monthly Observations, 90 Weekly Observations, 587 Daily Observations)*									
*Linear Model: Tw = a*Ta + b*									
a	0.84	0.81	0.76	0.97	0.85	0.76	0.57	0.63	0.71
b	3.12	3.32	3.62	-3.69	-0.63	2.34	7.02	6.17	5.10
RSS	9.55	142.43	2225.48	146.26	575.65	3103.96	94.32	437.63	2765.96
R^2 Adj	0.99	0.96	0.91	0.88	0.89	0.90	0.81	0.85	0.88
Logistic 1 Model: Tw = a/(1+exp(b(c-Ta)))*									
a	18.47	18.95	19.56	70.24	32.34	22.82	11.97	16.34	17.44
b	0.25	0.25	0.23	0.11	0.14	0.18	1.64	0.25	0.26
c	7.41	7.77	8.36	32.13	19.94	12.24	-0.20	3.66	5.41
RSS	8.68	83.73	1596.62	77.11	433.75	2709.58	23.22	294.36	1705.31
R^2 Adj	0.99	0.98	0.94	0.93	0.92	0.91	0.95	0.90	0.92
Logistic 2 Model: Tw = d + (a-d)/(1 + exp (b(c-Ta)))*									
a	22.22	20.27	20.85	39.32	38.11	25.13	11.96	17.56	18.06
b	0.15	0.20	0.19	0.17	0.12	0.14	1.69	0.21	0.23
c	7.75	7.73	8.37	24.46	21.97	12.46	-0.18	3.83	5.35
d	-3.23	-1.41	-1.27	2.38	-1.35	-1.94	0.21	-0.92	-0.71
RSS	5.55	75.57	1546.69	75.30	431.50	2648.92	22.71	286.34	1678.33
R^2 Adj	0.99	0.98	0.94	0.93	0.92	0.91	0.95	0.90	0.92
*Multiple Regression Model: Tw = a + b*Ta + c*Stage*									
a	6.33	6.38	7.13	0.20	2.28	5.44	NS	10.46	9.26
b	0.79	0.77	0.73	0.92	0.82	0.73		0.60	0.67
c	-5.24	-5.06	-5.93	-2.80	-3.17	-4.76		-8.98	-7.75
RSS	5.04	108.07	1766.77	125.12	485.64	2498.05		391.20	2324.47
R^2 Adj	0.99	0.97	0.93	0.89	0.91	0.92		0.87	0.89

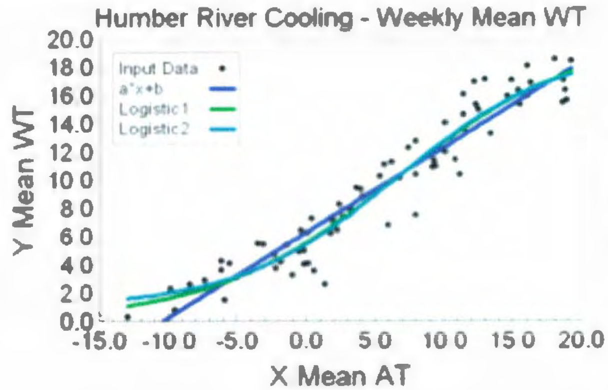

Humber River Cooling - Monthly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y:Mean WT
X:Mean AT
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0

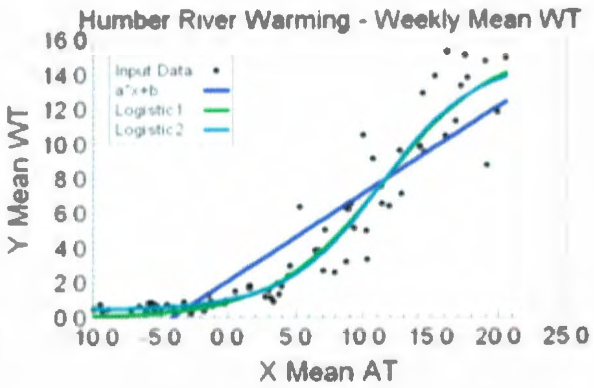

Humber River Warming - Monthly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y:Mean WT
X:Mean AT
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0

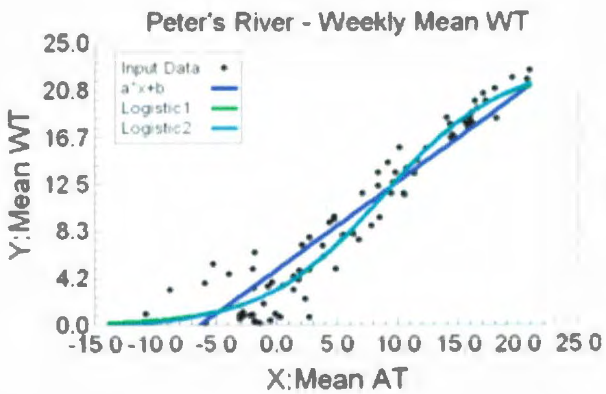

Peter's River - Monthly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y:Mean WT
X:Mean AT
25.0
20.8
16.7
12.5
8.3
4.2
0.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0

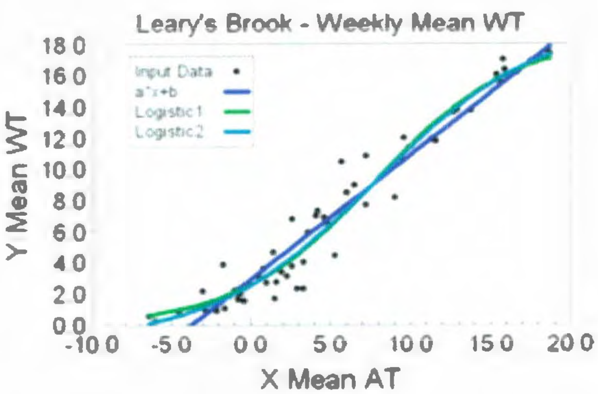

Leary's Brook - Monthly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y:Mean WT
X:Mean AT
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-5.0
0.0
5.0
10.0
15.0
20.0

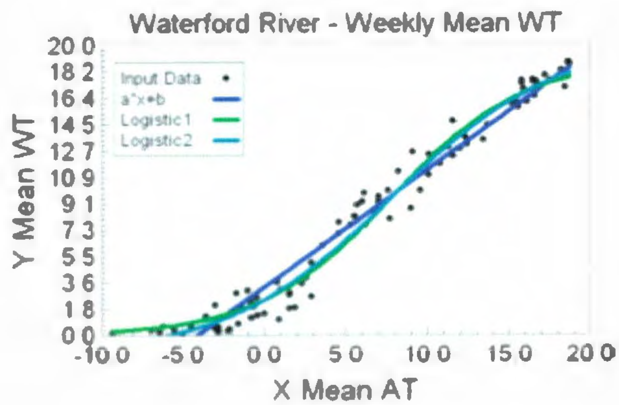

Waterford River - Monthly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y:Mean WT
X:Mean AT
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0

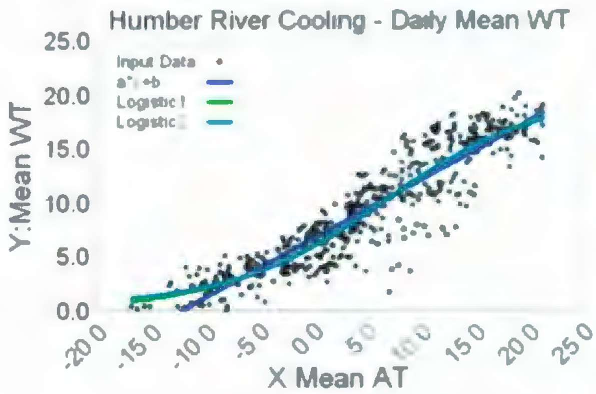

Humber River Cooling - Weekly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y Mean WT
X Mean AT
20.0
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-15.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0

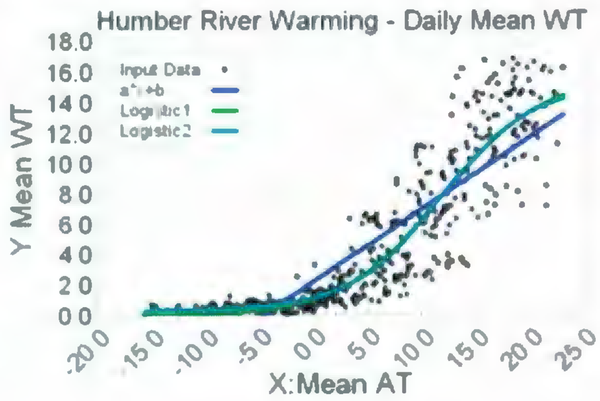

Humber River Warming - Weekly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y Mean WT
X Mean AT
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0
25.0

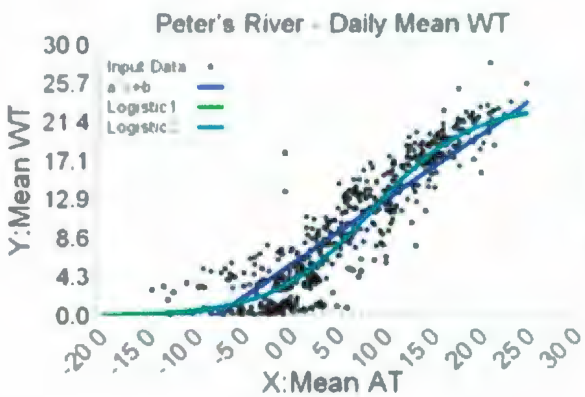

Peter's River - Weekly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y:Mean WT
X:Mean AT
25.0
20.8
16.7
12.5
8.3
4.2
0.0
-15.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0
25.0

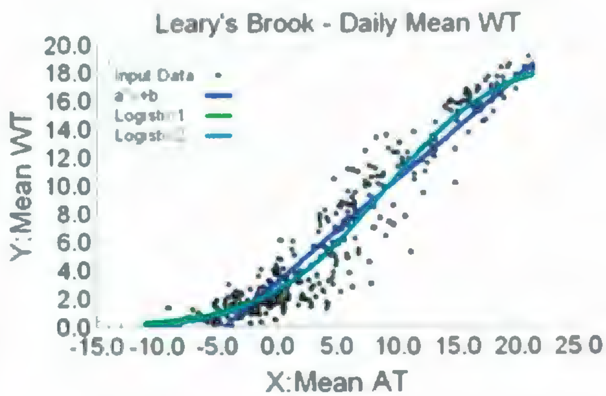

Leary's Brook - Weekly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y Mean WT
X Mean AT
18.0
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0

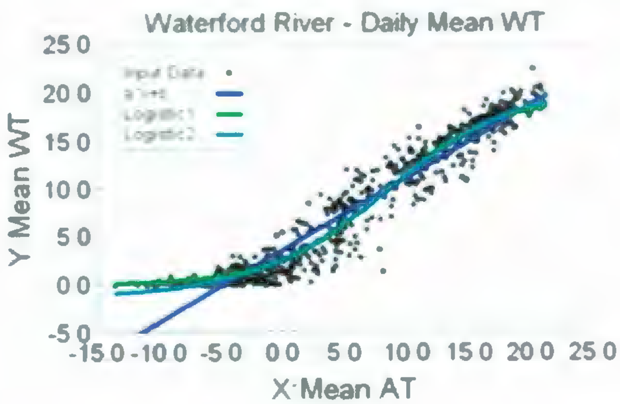

Waterford River - Weekly Mean WT
Input Data
a*x+b
Logistic1
Logistic2
Y Mean WT
X Mean AT
20.0
18.2
16.4
14.5
12.7
10.9
9.1
7.3
5.5
3.6
1.8
0.0
-10.0
-5.0
0.0
5.0
10.0
15.0
20.0


Humber River Cooling - Daily Mean WT
Input Data
Logistic1
Y:Mean WT
X Mean AT


Humber River Warming - Daily Mean WT
Input Data
Logistic1
Logistic2
Y Mean WT
X:Mean AT


Peter's River - Daily Mean WT
Input Data
Logistic1
Y:Mean WT
X:Mean AT


Leary's Brook - Daily Mean WT
Input Data
Y:Mean WT
X:Mean AT


Waterford River - Daily Mean WT
Y Mean WT
X Mean AT

Appendix G

An Alternative Approach to Handling Hysteresis

An alternative approach to developing separate regression models for the warming and cooling seasons is to add an additional explanatory variable to the regression models to account for the time of the year the sample of water quality was taken.

Table G.1 presents the curve fitting results for these modified models. Note that the additional variable X in each model is set equal to 1 for data collected during August to January and is set to 0 for data collected during February to July. The modified logistic models provide a better fit to the data than the modified linear - with only a slight difference between logistic 1 and 2.

Similar to the models developed for the separated seasonal datasets the modified models fit best to the monthly datasets but high scatter in the daily datasets drives up the residual sum of squares term and lowers the adjusted R^2 values.

The curve fitting results for these models are similar to those found for the seasonally divided datasets in that the logistic model best describes the relationship between air temperature and water temperature. The fit of the model is best at longer time scales (monthly) and poorer at shorter time scales (daily)

Figure G.1 presents a model plot for the modified first logistic model for the Humber River mean water temperature models.

Table G. 1 - Modified Regression Models for Handling Hysteresis in Humber River Datasets

Linear Models for Predicting Mean Water Temperature: WT = a*AT + bX + c
(where X =1 for August to january and X = 0 for February to July)

Dataset	a	b	c	St. Error	RSS	R^2	R^2 Adj
Monthly Mean WT	0.608	4.728	1.637	1.336	55.358	0.944	0.941
Weekly Mean WT	0.567	4.719	1.755	1.866	476.7	0.886	0.884
Daily Mean WT	0.53	4.417	2.233	2.514	5865.7	0.791	0.79
Monthly Max WT	0.714	4.382	-5.92	2.604	210.281	0.831	0.821
Weekly Max WT	0.559	5.022	-2.521	2.667	974.4	0.793	0.79
Daily Max WT	0.49	4.766	0.557	2.778	7163.7	0.754	0.753
Monthly Min WT	0.418	3.62	5.51	1.464	62.15	0.908	0.896
Weekly Min WT	0.47	4.169	5.207	1.999	547.6	0.843	0.848
Daily Min WT	0.51	0.133	6.37	3.245	9776.9	0.636	0.635

Logistic 1 Models for Predicting Mean Water Temperature: WT = a/(1+exp(b*(c-AT))) +dX
(where X =1 for August to january and X = 0 for February to July)

	a	b	c	d	St. Error	RSS	R^2	R^2 Adj
Monthly Mean WT	15.23	0.248	10.08	4.282	1.276	48.86	0.951	0.946
Weekly Mean WT	14.23	0.259	9.622	4.428	1.697	391.5	0.906	0.904
Daily Mean WT	14.79	0.223	9.839	4.156	2.38	5252	0.812	0.812
Monthly Max WT	21.79	0.2297	24.08	5.218	1.892	107.354	0.914	0.905
Weekly Max WT	16.81	0.214	19.8	4.867	2.445	812.8	0.828	0.824
Daily Max WT	15.69	0.181	14.96	4.445	2.73	6908.2	0.762	0.762
Monthly Min WT	11.71	0.2833	1.026	3.333	1.552	72.243	0.894	0.883
Weekly Min WT	11.69	0.31	1.533	4.039	1.68	383.9	0.89	0.888
Daily Min WT	16.35	0.184	3.896	0.177	3.065	8710	0.675	0.674

Table G.1 Continued

Logistic 2 Models for Predicting Mean Water Temperature: WT = d + (a-d)/(1+exp(b*(c-AT))) +eX Where

	a	b	c	d	e	St. Error	RSS	R^2	R^2 Adj
Monthly Mean WT	28.75	0.101	16.32	-4.325	4.807	1.108	35.594	0.96	0.959
Weekly Mean WT	15.76	0.186	9.701	-1.65	4.734	1.646	365.8	0.91	0.91
Daily Mean WT	16.64	0.16	10.05	-1.82	4.46	2.342	5078.8	0.82	0.818
Monthly Max WT	24.31	0.198	24.93	-0.721	5.352	1.911	105.91	0.92	0.903
Weekly Max WT	19.57	0.1589	20.85	-1.401	5.105	2.43	797.32	0.83	0.826
Daily Max WT	20.24	0.108	16.56	-2.96	4.79	2.686	6678.6	0.77	0.769
Monthly Min WT	14.15	0.193	2.204	-1.466	3.906	1.464	62.149	0.91	0.896
Weekly Min WT	19.57	0.1589	20.85	-1.401	5.105	2.43	797.32	0.83	0.826
Daily Min WT	15.45	0.232	4.017	1.148	0.051	3.054	8638.2	0.68	0.677

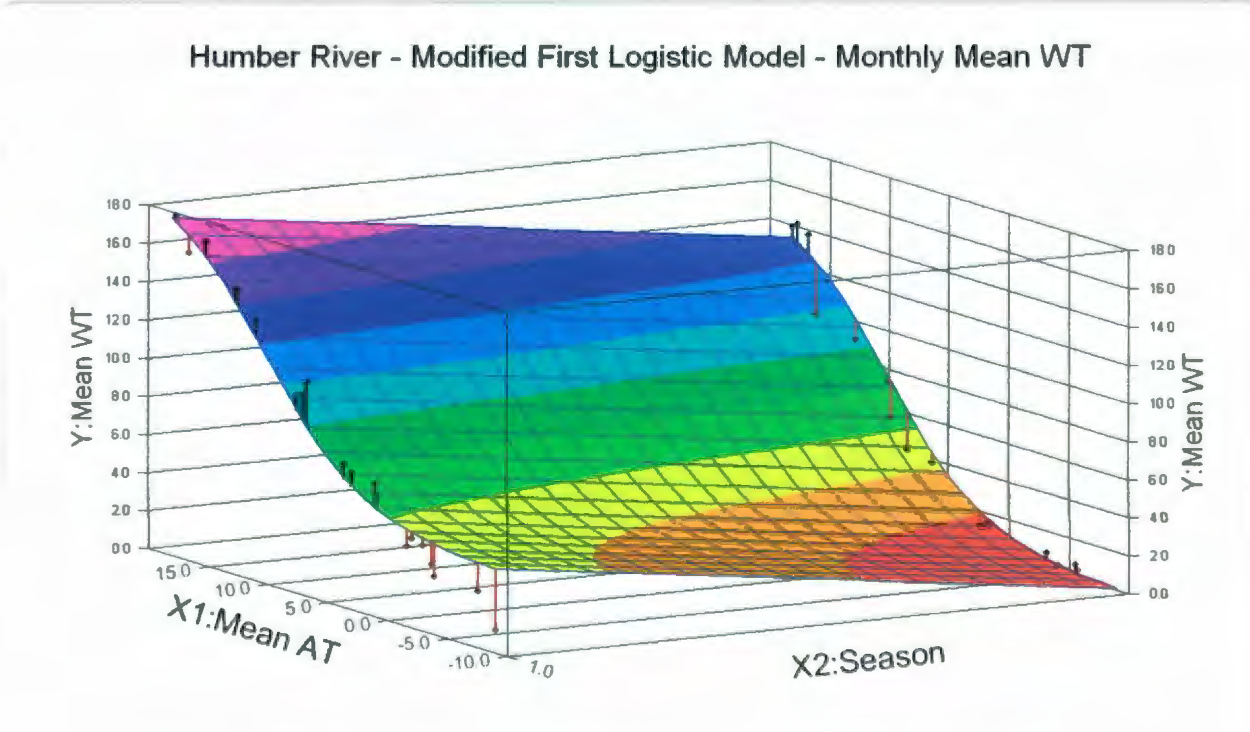


Figure G.1 Modified First Logistic Model for Humber River Monthly Mean WT

Appendix H

Model Testing Results for Water Temperature

The first logistic models were tested using the water temperature datasets reserved for model testing purposes. The following tables and plots are contained in this appendix:

- Table H.1 - Using the Logistic 1 Model for Predicting Humber River Mean Water Temperature. Model testing results are presented for the original (before hysteresis was accounted for), warming and cooling Humber River datasets
- Table H.2 - Using the Logistic 1 Model for Predicting Humber River Maximum Water Temperature
- Table H.3 - Using the Logistic 1 Model for Predicting Humber River Minimum Water Temperature
- Table H.4 - Using the Logistic 1 Model for Predicting Peter's River Water Temperature
- Table H.5 - Using the Logistic 1 Model for Predicting Leary's Brook Water Temperature
- Table H.6 - Using the Logistic 1 Model for Predicting Waterford River Water Temperature

Another way to compare the ability of the models to predict water temperature is to examine the absolute value of the differences between observed and model predicted values. Figure H.1 presents a boxplot comparison of the absolute value of the differences for each of the RTWQ stations. At the monthly time scale the spread of the difference in the Humber River warming season data (when water temperatures are lower) and Waterford River are quite small.

Table H.1 - Using the Logistic I Model for Predicting Humber River Mean Water Temperature								
		Absolute Value of Difference Abs[Pred - Obs]				**Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%**		
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**
Original Dataset Monthly Mean WT = 20.05/(1+exp(0.154*(10.275- Mean AT)))								
0.771	14	1.95	4.5	0.42	1.23	50.49	97.8	11.98
Cooling Season Monthly Mean WT = 21.4135/(1+exp(0.146*(7.19713- Mean AT)))								
0.989	5	0.85	1.61	0.29	0.67	19.9	59.8	1.86
Warming Season Monthly Mean WT = 21.56393/(1+exp(0.1915*(14.9611- Mean AT)))								
0.975	9	0.25	0.54	0.003	0.2	26.6	78.9	0.03
Original Weekly Mean WT = 19.208/(1+exp(0.159*(9.817- Mean AT)))								
0.734	63	1.85	7.09	0.04	1.35	57.6	325.3	7.2
Cooling Season Weekly Mean WT = 20.3125/(1+exp(0.148653*(6.56463- Mean AT)))								
0.93	25	1.14	3.25	0.44	0.67	33.46	111.4	2.6
Warming Season Weekly Mean WT = 15.67955/(1+exp(0.243939*(11.4797- Mean AT)))								
0.91	38	0.634	4.66	0.011	0.8	86.24	557.92	1.73
Original Daily Mean WT = 18.518/(1+exp(0.1497*(9.538- Mean AT)))								
0.67	398	2.11	9.58	0.007	1.72	55.84	410.34	0.051
Cooling Season Daily Mean WT = 20.50325/(1+exp(0.13209*(6.78741- Mean AT)))								
0.84	130	1.786	5.53	0.012	1.35	37.92	126.3	0.34
Warming Season Daily Mean WT = 15.77483(1+exp(0.21703*(11.8132- Mean AT)))								
0.83	268	1.04	7.77	3.82	1.43	79.39	837.4	0.02

Table H.2 - Using the Logistic 1 Model for Predicting Humber River Maximum Water Temperature								
		Absolute Value of Difference Abs[Pred - Obs]				Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%		
Adj R^2	Obs.	Mean	Max	Min	St Dev	Mean	Max	Min
Original Dataset Monthly Max WT = 28.547/(1+exp(0.134*(24.893- Max AT)))								
0.746	14	2.606	6.21	0.12	1.79	41.75	95.49	6.39
Cooling Season Monthly Max WT = 27.86925/(1+exp(0.147781*(21.2548- Max AT)))								
0.989	5	1.22	2.1	0.13	0.85	23.31	52.91	2.93
Warming Season Monthly Max WT = 366.9056/(1+exp(0.132958*(51.44699- Max AT)))								
0.977	9	1.241	5.175	0.1626	1.57	38.88	173.5	7.99
Original Dataset Weekly Max WT = 24.045/(1+exp(0.123*(21.414- Max AT)))								
0.653	63	2.42	8.69	0.03	2.06	53.97	280.2	0.209
Cooling Season Weekly Max WT = 27.54347/(1+exp(0.109373*(19.7377- Max AT)))								
0.84	25	1.55	5.45	0.05	1.26	32.52	106.8	2.2
Warming Season Weekly Max WT = 32.79584/(1+exp(0.140317*(28.9461- Max AT)))								
0.81	38	1.29	7.11	0.01	1.48	38.28	162.7	0.47
Original Dataset Daily Max WT = 19.909/(1+exp(0.118*(14.933- Max AT)))								
0.609	398	2.52	9.39	0.007	1.93	52.11	231.8	0.48
Cooling Season Daily Max WT = 22.0563/(1+exp(0.10974*(11.569- Max AT)))								
0.81	130	2.05	8.5	0.006	1.57	36.63	93.09	0.043
Warming Season Daily Max WT = 18.89737(1+exp(0.154291*(19.0636- Max AT)))								
0.77	268	1.31	8.03	0.005	1.69	46.82	399.16	0.77

Table H.3 - Using the Logistic 1 Model for Predicting Humber River Minimum Water Temperature								
		Absolute Value of Difference Abs[Pred - Obs]				**Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%**		
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**
Original Dataset Monthly Min WT = 15.365/(1+exp(0.1771*(1.0168- Min AT)))								
0.71	14	1.84	9.19	0.047	2.53	66.98	113.9	8.32
Cooling Season Monthly Min WT = 37.64699/(1+exp(0.105106*(11.1109- Min AT)))								
0.9	5	3.204	11.67	0.24	4.76	71.73	100	1.57
Warming Season Monthly Min WT = 11.10682/(1+exp(0.304584*(1.92339- Min AT)))								
0.95	9	0.455	1.6	0.043	0.62	229.9	688.5	0.91
Original Dataset Weekly Min WT = 15.5146/(1+exp(0.1894*(1.1589- Min AT)))								
0.71		1.7	9.4	0.003	1.64	72.58	552.61	0.02
Cooling Season Weekly Min WT =18.25805/(1+exp(0.156347*(-0.508- Min AT)))								
0.89		1.4	11.81	0.022	2.31	45.59	131.95	0.692
Warming Season Weekly Min WT = 11.39423/(1+exp(0.33499*(2.2864- Min AT)))								
0.9		0.58	2.31	0.01	0.63	489.4	4877.2	1.67
Original Dataset Daily Min WT = 16.269/(1+exp(0.195*(3.888- Min AT)))								
0.69		1.88	10.84	0.02	1.7	71.51	1123.1	0.6
Cooling Season Daily Min WT =18.85133/(1+exp(0.152521*(2.17453- Min AT)))								
0.82		1.98	12.93	0.077	1.75	46.87	244.59	0.481
Warming Season Daily Min Min WT = 13.0989/(1+exp(0.301678*(5.35169- Min AT)))								
0.83		0.969	6.32	0.005	1.27	272.31	5216.2	0.14

Table H.4 - Using the Logistic 1 Model for Predicting Peter's River Water Temperature

		Absolute Value of Difference Abs|Pred - Obs|				Absolute Value of % Error Abs|(Pred - Obs)/Pred|*100%		
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**ST Dev**	**Mean**	**Max**	**Min**
Monthly Mean Mean WT = 21.13/(1+exp(0.24*(7.63- Mean AT)))								
0.97	8	0.874	3.202	0.044	1.03	57.63	144	0.239
Weekly Mean WT = 22.89/(1+exp(0.21*(8.73- Mean AT)))								
0.95	32	1.213	6.61	0.021	1.42	63.42	252.5	0.27
Daily Mean WT = 23.16/(1+exp(0.20*(9.15- Mean AT)))								
0.88	196	1.923	10.8	0.024	1.82	78.41	662.94	0.12
Monthly Max WT = 43.07/(1+exp(0.14*(27.12- Max AT)))								
0.95	8	2.017	3.79	0.36	1.21	40.95	107.18	3.91
Weekly Max WT = 34.59/(1+exp(0.14*(22.04 - Max AT)))								
0.92	32	1.923	8.72	0.153	1.61	50.37	125.8	0.62
Daily Max WT = 28.19/(1+exp(0.15*(15.46 - Max AT)))								
0.87	196	2.186	7.99	0.06	1.79	58.31	197.71	0.32
Monthly Min WT = 68.64/(1+exp(0.16*(9.08 - Min AT)))								
0.91	8	0.836	1.91	0.21	0.63	121.45	366.6	1.34
Weekly Min WT = 18.19/(1+exp(0.22*(-1.56- Min AT)))								
0.81	32	1.74	5.49	0.15	1.5	200.86	1365.5	2.3
Daily Min WT = 19.52/(1+exp(0.21*(3.06- Min AT)))								
0.75	196	2.37	12.03	0.04	2.2	142.9	2136.7	0.22

Table H.5 - Using the Logistic I Model for Predicting Leary's Brook Water Temperature

		Absolute Value of Difference Abs[Pred - Obs]				Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%		
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**
Monthly Mean WT = 18.28/(1+exp(0.24*(7.74- Mean AT)))								
0.96	8	1.168	2.682	0.191	0.83	26.687	131.3	1.422
Weekly Mean WT = 18.41/(1+exp(0.23*(7.75- Mean AT)))								
0.95	23	0.843	2.34	0.014	0.73	11.15	37.5	0.1
Daily Mean WT = 19.06/(1+exp(0.21*(8.72- Mean AT)))								
0.92	136	1.639	5.01	0.01	1.19	25.54	140.3	0.14
Monthly Max WT = 1181.08/(1+exp(0.08*(76.55- Max AT)))								
0.84	8	3.099	8.38	0.532	2.63	21.29	31.23	2.63
Weekly Max WT = 41.12/(1+exp(0.11*(25.47- Max AT)))								
0.89	23	1.564	6.15	0.01	1.44	15.99	110.9	0.05
Daily Max WT = 24.22/(1+exp(0.15*(14.77- Max AT)))								
0.88	136	2.06	7.22	0.02	1.61	28.04	133.8	0.09
Monthly Min WT = 16.28/(1+exp(0.28*(1.78 - Min AT)))								
0.93	8	1.37	3.19	0.1	0.98	123.9	828.6	1.63
Weekly Min WT = 14.69/(1+exp(0.35*(1.30- Min AT)))								
0.92	23	1.27	2.92	0.27	0.73	21.15	72.4	3.96
Daily Min WT = 16.49/(1+exp(0.27*(4.80- Min AT)))								
0.89	136	1.55	5	0.01	1.13	29.1	155.3	0.08

Table H.6 - Using the Logistic 1 Model for Predicting Waterford River Water Temperature

		Absolute Value of Difference Abs|Pred - Obs|				**Absolute Value of % Error Abs|(Pred - Obs)/Pred|*100%**		
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**
Monthly Mean WT = 18.47/(1+exp(0.25*(7.4q- Mean AT)))								
0.99	12	0.497	1.244	0.026	0.36	14.45	39.62	0.191
Weekly Mean WT = 18.95/(1+exp(0.25*(7.77- Mean AT)))								
0.98	48	0.803	2.49	0.03	0.61	22.09	138	0.4
Daily Mean WT = 19.56/(1+exp(0.23*(8.36- Mean AT)))								
0.94	306	1.36	6.87	0.01	1.22	32.4	204.4	0.05
Monthly Max WT = 70.24/(1+exp(0.11*(32.13- Max AT)))								
0.93	12	2.091	5.78	0.21	1.64	20.95	56.18	1.33
Weekly Max WT = 32.34/(1+exp(0.14*(19.94- Max AT)))								
0.92	48	1.712	5.7	0.02	1.26	26.91	133.5	0.19
Daily Max WT = 22.82/(1+exp(0.18*(12.24- Max AT)))								
0.91	306	1.8	9.61	0	1.66	32.84	231.04	0
Monthly Min WT = 9.90/(1+exp(2.44*(-0.44 - Min AT)))								
0.72	12	1.172	4.3	0.01	1.57	--	--	--
Weekly Min WT = 16.34/(1+exp(0.14*(19.94- Min AT)))								
0.9	48	0.8	2.01	0.01	0.54	41.9	145.2	1
Daily Min WT = 17.44/(1+exp(0.26*(5.42- Min AT)))								
0.92	306	1.29	6.15	0	1.14	42.69	223	0.14

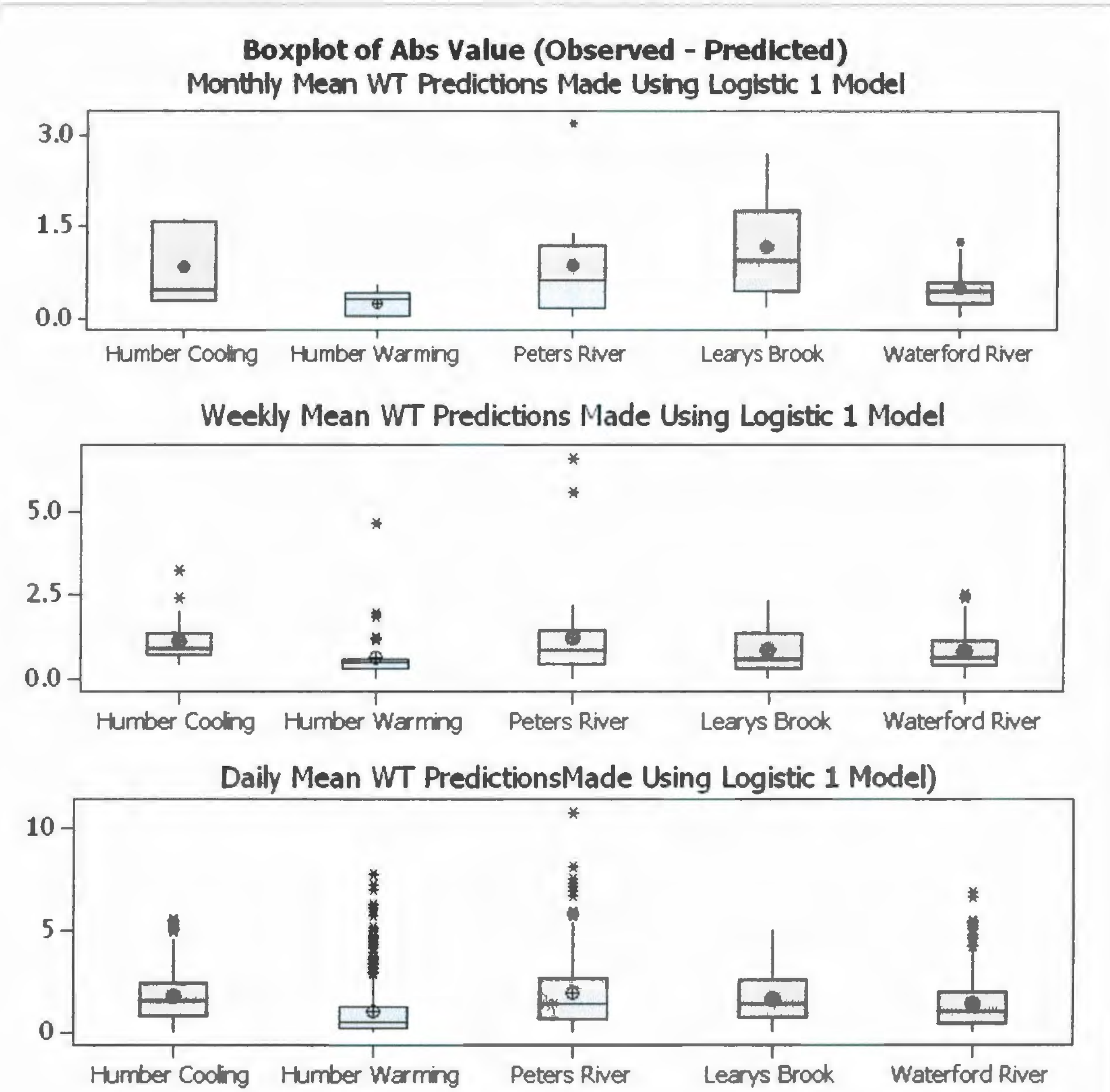


Figure H.1 Boxplot of Observed vs. Model Predicted Water Temperature

Appendix I

Statistical Overview of the Datasets Used for Dissolved Oxygen Regression

Tables and plots are attached in this appendix to provide a more complete statistical overview of the datasets developed for developing regression models for dissolved oxygen at the RTWQ stations.

Tables I.1 to I.4 present a detailed statistical overview of the mean, maximum and minimum dissolved oxygen, water temperature and stage levels at the stations.

Table I.5 presents the Pearson's correlation coefficients for the mean dissolved oxygen, water temperature and stage at the RTWQ stations.

Table I.1 - General Overview of the Dissolved Oxygen Datasets Developed for NF02YL0012 - Humber River

		Water Temperature			Stage			Dissolved Oxygen		
Dataset	Obs.	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Monthly Models - For Regression (Dec 03 to Dec 06)										
Mean Monthly	37	7.068	0.384	17.845	2.123	1.521	3.297	12.092	8.603	19.481
Max Monthly	37	9.06	1.11	20.67	2.617	1.768	3.867	12.897	8.81	20.01
Min Monthly	37	5.004	-0.1	16.1	1.784	1.331	2.557	11.359	8.43	18.56
Monthly Models - For Prediction (Jan 07 to April 08)										
Mean Monthly	14	5.05	0.336	17.14	2.0556	1.7026	2.995	12.662	10.205	14.558
Max Monthly	14	6.89	1.18	20.27	2.417	1.82	3.723	13.379	10.66	14.82
Min Monthly	14	2.23	-0.08	15.34	1.7115	1.399	2.15	11.912	9.41	14.39
Weekly Models - For Regression (Dec 03 to Dec 06)										
Mean Weekly	149	7.106	0.253	18.529	2.133	1.388	3.658	12.054	8.556	19.606
Max Weekly	149	7.97	0.57	20.67	2.2831	1.435	3.867	12.364	8.68	20.01
Min Weekly	149	6.149	-0.1	17.34	1.9904	1.331	3.584	11.725	8.43	19.19
Weekly Models - For Prediction (Jan 07 to April 08)										
Mean Weekly	63	4.564	0.18	18.377	2.0598	1.5612	3.4516	12.683	9.878	14.614
Max Weekly	63	5.405	0.57	20.27	2.2009	1.665	3.761	12.975	10.34	14.82
Min Weekly	63	3.389	-0.08	16.58	1.9372	1.399	3.118	12.379	9.41	14.47
Daily Models - For Regression (Dec 03 to Dec 06)										
Mean Daily	986	7.253	0.0233	20.124	2.1413	1.3367	3.8299	11.983	8.497	19.942
Max Daily	986	7.577	0.19	20.67	2.1721	1.344	3.867	12.077	8.57	20.01
Min Daily	986	6.935	-0.1	19.77	2.1104	1.331	3.807	11.883	8.43	19.85
Daily Models - For Prediction (Jan 07 to April 08)										
Mean Daily	398	4.663	-0.026	19.401	2.0545	1.4335	3.6953	12.732	9.583	14.713
Max Daily	398	5.002	0.12	20.27	2.0821	1.458	3.723	12.813	9.65	14.82
Min Daily	398	4.3	-0.08	19.25	2.0273	1.399	3.67	12.64	9.41	14.61

Note - the maximum value of the max monthly, max weekly and max daily datasets will match each other. The minimum value of the datasets will not (as the minimum value for max monthly will be determined from 12 values while the minimum value for max weekly will be determined from 53 values (Minitab counts 53 weeks in the year). The same logic applies to the minimum value of the min monthly, min weekly and min daily datasets being equal but the max value is not.

Table I.2 - General Overview of the Dissolved Oxygen Datasets Developed for NF02YO0121 - Peters River

		Water Temperature			Stage			Dissolved Oxygen		
Dataset	Obs.	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Monthly Models - For Regression (July 2005 to May 2007)										
Mean Monthly	23	8.07	-0.099	20.58	1.1415	0.9352	1.4796	11.001	7.658	13.762
Max Monthly	23	13.11	-0.045	30.19	1.4357	0.995	2.454	12.195	8.996	14.957
Min Monthly	23	5.14	-0.603	14.53	0.9841	0.884	1.075	9.368	1.282	12.9
Monthly Models - For Prediction (July 2007 to February 2008)										
Mean Monthly	8	8.69	-0.208	20.59	1.2045	1.0399	1.4287	10.82	8.719	12.995
Max Monthly	8	12.17	-0.149	27.88	1.4941	1.184	1.717	12.05	9.696	14.249
Min Monthly	8	5.84	-0.23	15.86	1.0375	0.917	1.332	9.706	7.267	11.898
Weekly Models - For Regression (Week 28, 2005 - July to Week 20, 2007 - May)										
Mean Weekly	91	8.341	-0.133	22.527	1.1636	0.9286	1.8396	10.953	6.988	14.258
Max Weekly	91	11.197	-0.089	30.194	1.2965	0.953	2.454	11.624	8.134	14.957
Min Weekly	91	6.558	-0.603	18.33	1.0548	0.908	1.4	10.155	1.282	13.812
Weekly Models - For Prediction (Week 28, 2007 - July to Week 6, 2008 - February)										
Mean Weekly	32	8.76	-0.206	21.06	1.2016	0.9821	1.5588	10.85	8.172	13.584
Max Weekly	32	10.88	-0.197	27.88	1.322	0.995	1.717	11.445	8.626	14.249
Min Weekly	32	7.35	-0.23	18.2	1.1128	0.961	1.431	10.202	7.267	12.516
Daily Models - For Regression (July 1, 2005 to May 15, 2007)										
Mean Daily	595	8.448	-0.287	27.876	1.1445	0.884	2.2744	10.936	6.699	14.441
Max Daily	595	9.819	-0.162	30.194	1.1722	0.884	2.454	11.218	7.133	14.957
Min Daily	595	7.434	-0.603	23.871	1.1185	0.884	2.014	10.624	1.282	14.248
Daily Models - For Prediction (July 10, 2007 to February 5, 2008)										
Mean Daily	196	8.764	-0.22	23.622	1.1895	0.9639	1.6967	10.87	7.751	14.037
Max Daily	196	9.734	-0.219	27.88	1.2128	0.967	1.717	11.135	8.272	14.249
Min Daily	196	8.023	-0.23	21.86	1.1665	0.961	1.684	10.605	7.267	13.954

Table 1.3- General Overview of the Dissolved Oxygen Datasets Developed for NF02ZM0178 Leary's Brook

		Water Temperature			Stage			Dissolved Oxygen		
Dataset	Obs.	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Monthly Models - For Regression (September 2004 to December 2006)										
Mean Monthly	16	6.93	0.705	16.58	0.789	0.6317	0.9535	11.642	6.448	13.866
Max Monthly	16	11.04	2.73	20.94	1.45	0.897	2.121	16.67	10.25	52.82
Min Monthly	16	3.94	-0.161	14.02	0.6528	0.564	0.73	8.094	0.682*	12.145
Monthly Models - For Prediction (May 2007 to December 2007)										
Mean Monthly	8	10.52	2.75	17.02	0.7145	0.5777	0.8542	10.868	9.147	12.965
Max Monthly	8	14.65	4.16	19.7	1.21	0.892	1.949	11.933	9.899	13.335
Min Monthly	8	7.23	1.67	14.9	0.5931	0.542	0.646	9.165	7.592	12.533
Weekly Models - For Regression (Week 37, 2004 - September to Week 51, 2006 - December)										
Mean Weekly	57	6.923	0.268	17.5	0.8008	0.5935	1.1827	11.51	5.235	14.929
Max Weekly	57	9.491	1.248	20.94	1.1166	0.669	2.121	13.897	8.287	52.821
Min Weekly	57	5.071	-0.161	14.62	0.696	0.564	1.012	9.388	2536	14.048
Weekly Models - For Prediction (Week 19, 2007 - May to Week 49, 2007 - December)										
Mean Weekly	23	10.464	2.629	17.15	0.7025	0.5505	0.9458	10.87	9.136	12.976
Max Weekly	23	13.705	3.763	19.7	1.015	0.558	1.949	11.731	9.899	13.335
Min Weekly	23	7.911	1.673	14.9	0.6316	0.542	0.821	9.59	7.592	12.533
Daily Models - For Regression (September 11, 2004 to December 19, 2006)										
Mean Daily	347	6.789	0.0738	19.13	0.8075	0.5741	1.3517	11.494	2.848	16.708
Max Daily	347	7.928	0.263	20.94	0.8855	0.577	1.853	12.461	3.3	52.821
Min Daily	347	5.834	-0.161	18	0.7582	0.564	1.211	10.619	2.536	15.77
Daily Models - For Prediction (May 11, 2007 to December 6, 2007)										
Mean Daily	136	10.643	2.332	17.32	0.6969	0.5432	1.3797	10.816	8.735	13.144
Max Daily	136	12.174	2.832	19.7	0.7623	0.544	1.949	11.299	9.294	13.335
Min Daily	136	9.34	1.673	16.5	0.66	0.542	1.012	10.17	7.592	12.992

Historical records for dissolved oxygen at Leary's Brook can be odd in that dissolved oxygen levels are recorded but no water temperature, pH, and specific oxygen data are recorded. The monthly minimum value of 0.682 is taken from a day where this occurred.

Table I.4 General Overview of the Dissolved Oxygen Datasets Developed for NF02ZM0009 - Waterford River

		Water Temperature			Stage			Dissolved Oxygen		
Dataset	Obs.	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Monthly Models - For Regression (July 2005 to March 2007)										
Mean Monthly	21	8.09	0.221	17.71	0.5635	0.4156	0.9088	11.09	6.593	14.217
Max Monthly	21	12.83	1.26	24.05	1.102	0.607	1.977	15.25	8.05	59.31
Min Monthly	21	4.06	-0.26	13.73	0.4388	0.386	0.579	8.351	0.36	12.87
Monthly Models - For Prediction (April 2007 to March 2008)										
Mean Monthly	12	7.57	0.665	16.58	0.5671	0.412	0.765	10.82	2.49	17.15
Max Monthly	12	12.43	2.2	24.3	1.048	0.513	1.956	15.84	11.35	24.64
Min Monthly	12	4.14	-0.13	11.99	0.4473	0.38	0.537	7.46	-0.53*	15.39
Weekly Models - For Regression (Week 28, 2005 - July to Week 13, 2007 - March)										
Mean Weekly	90	8.1	-0.168	18.733	0.5617	0.4049	1.0525	11.095	5.953	14.896
Max Weekly	90	11.05	0.04	24.05	0.7911	0.429	1.977	12.727	6.77	59.31
Min Weekly	90	5.709	-0.26	15.94	0.4789	0.386	0.93	9.766	0.36	13.54
Weekly Models - For Prediction (Week 14, 2007 - April to Week 13, 2008 - March)										
Mean Weekly	48	7.997	0.235	18.456	0.5708	0.3924	0.9339	11.024	1.053	17.967
Max Weekly	48	11.28	0.93	24.3	0.7999	0.403	1.956	13.166	3.14	24.64
Min Weekly	48	5.597	-0.13	16.29	0.49	0.38	0.769	9.159	-0.53*	17.23
Daily Models - For Regression (July 6, 2005 to March 31, 2007)										
Mean Daily	587	8.121	-0.185	22.46	0.5588	0.3938	1.4346	11.071	5.23	15.528
Max Daily	587	9.356	-0.17	24.05	0.607	0.396	1.977	11.626	5.63	59.31
Min Daily	587	7.062	-0.26	20.51	0.522	0.386	1.236	10.564	0.36	15.33
Daily Models - For Prediction (April 1, 2007 to March 27, 2008)										
Mean Daily	306	8.217	-0.061	20.913	0.5866	0.3817	1.2778	10.763	-0.34	19.14
Max Daily	306	9.74	-0.05	24.3	0.616	0.383	1.956	11.514	-0.12	24.6
Min Daily	306	7	-0.13	19.16	0.5319	0.38	1.033	10.017	-0.53*	18.34

In February and March of 2008 there were unusually low dissolved oxygen values. ANy values that dropped below 0 mg/L were removed from the Prediction Dataset.

Table I.5 - Pearson Correlation and P-Value for RTWQ Stations - WT, DO and Stage

Humber River			Peter's River			Leary's Brook			Waterford River		
						Monthly					
	Mean WT	Mean DO		Mean WT	Mean DO		Mean WT	Mean DO		Mean WT	Mean DO
Mean DO	-0.950 0.000		Mean DO	-0.972 0.000		Mean DO	-0.924 0.000		Mean DO	-0.943 0.000	
Mean Stage	-0.196 0.267	0.139 0.434	Mean Stage	-0.547 0.007	0.506 0.014	Mean Stage	-0.629 0.070	0.686 0.041	Mean Stage	-0.641 0.002	0.604 0.004
						Weekly					
	Mean WT	Mean DO		Mean WT	Mean DO		Mean WT	Mean DO		Mean WT	Mean DO
Mean DO	-0.939 0.000		Mean DO	-0.962 0.000		Mean DO	-0.903 0.000		Mean DO	-0.903 0.000	
Mean Stage	-0.230 0.006	0.187 0.027	Mean Stage	-0.434 0.000	0.399 0.000	Mean Stage	-0.595 0.001	0.643 0.000	Mean Stage	-0.502 0.000	0.469 0.000
						Daily					
	Mean WT	Mean DO		Mean WT	Mean DO		Mean WT	Mean DO		Mean WT	Mean DO
Mean DO	-0.933 0.000		Mean DO	-0.956 0.000		Mean DO	-0.845 0.000		Mean DO	-0.895 0.000	
Mean Stage	-0.243 0.000	0.194 0.000	Mean Stage	-0.447 0.000	0.383 0.000	Mean Stage	-0.469 0.000	0.544 0.000	Mean Stage	-0.413 0.000	0.410 0.000

Appendix J

Graphical Plots of Dissolved Oxygen and

Water Temperature

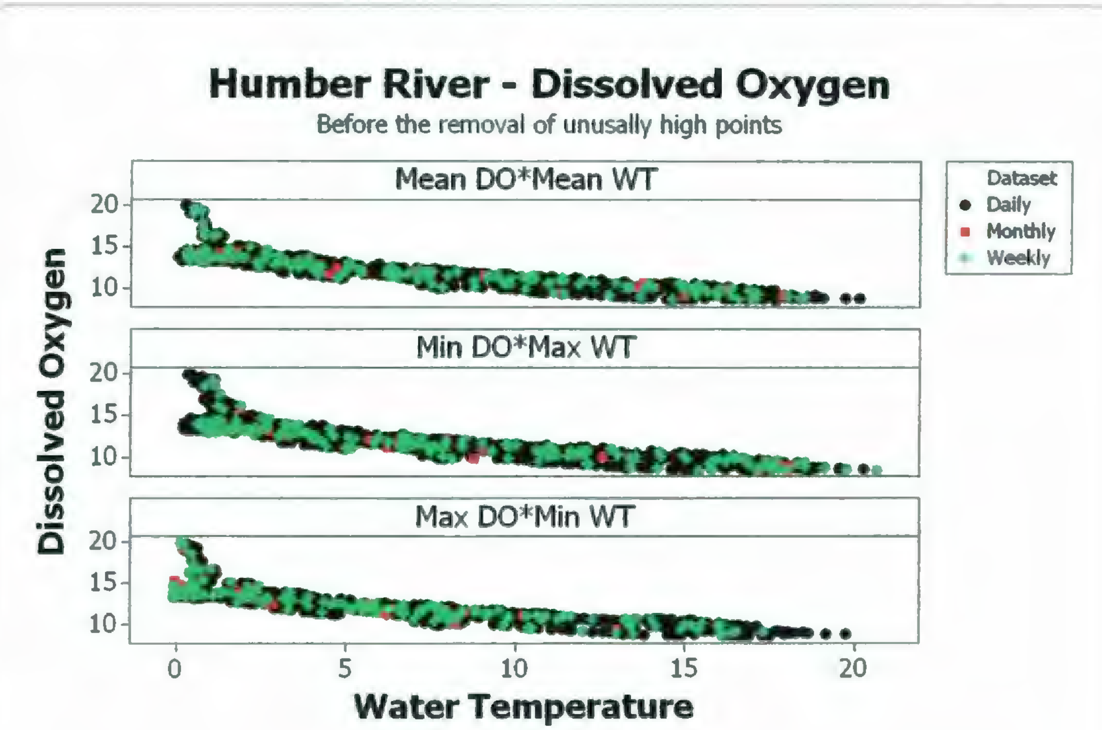


Figure J.1 Humber River Water Temperature and Dissolved Oxygen

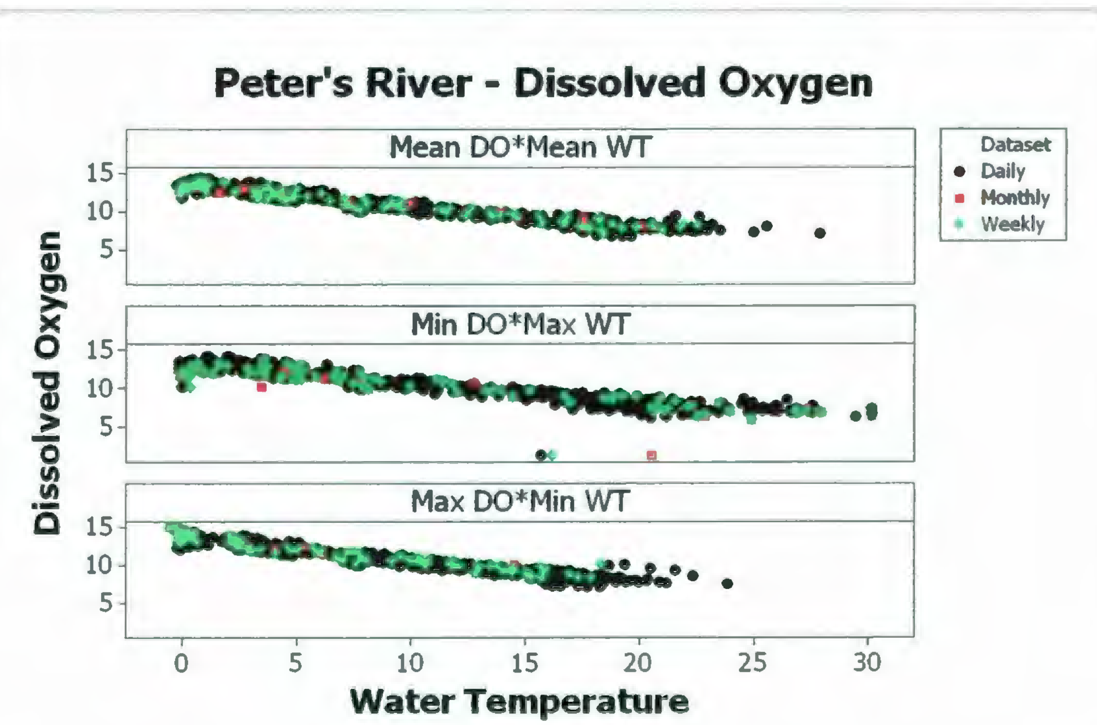


Figure J.2 Peter's River Water Temperature and Dissolved Oxygen

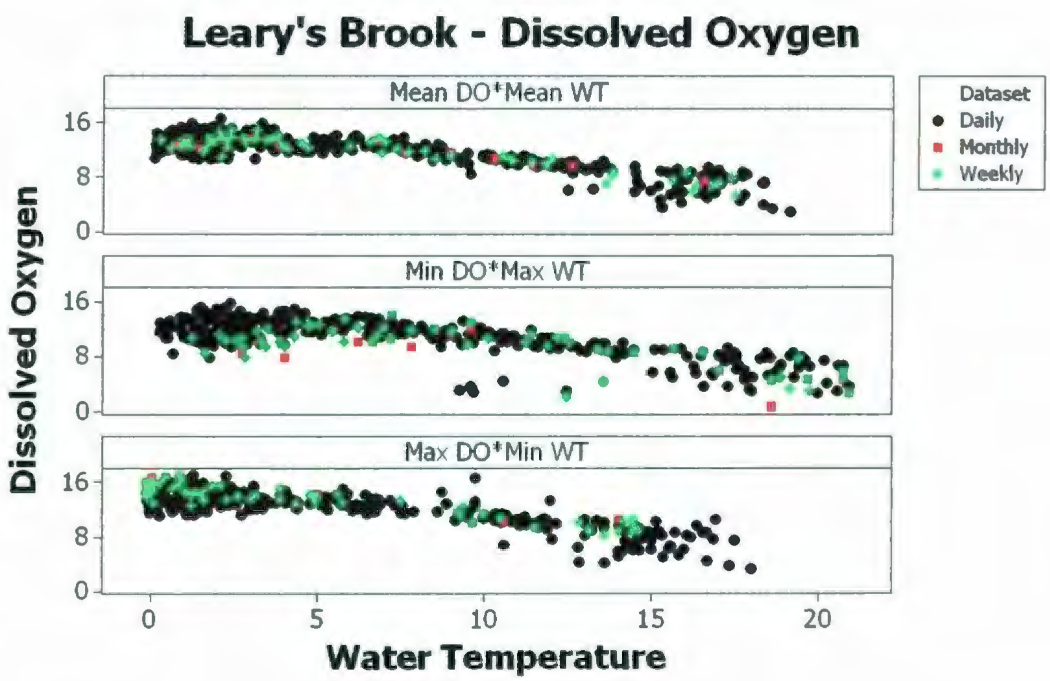


Figure J.3 Leary's Brook Water Temperature and Dissolved Oxygen

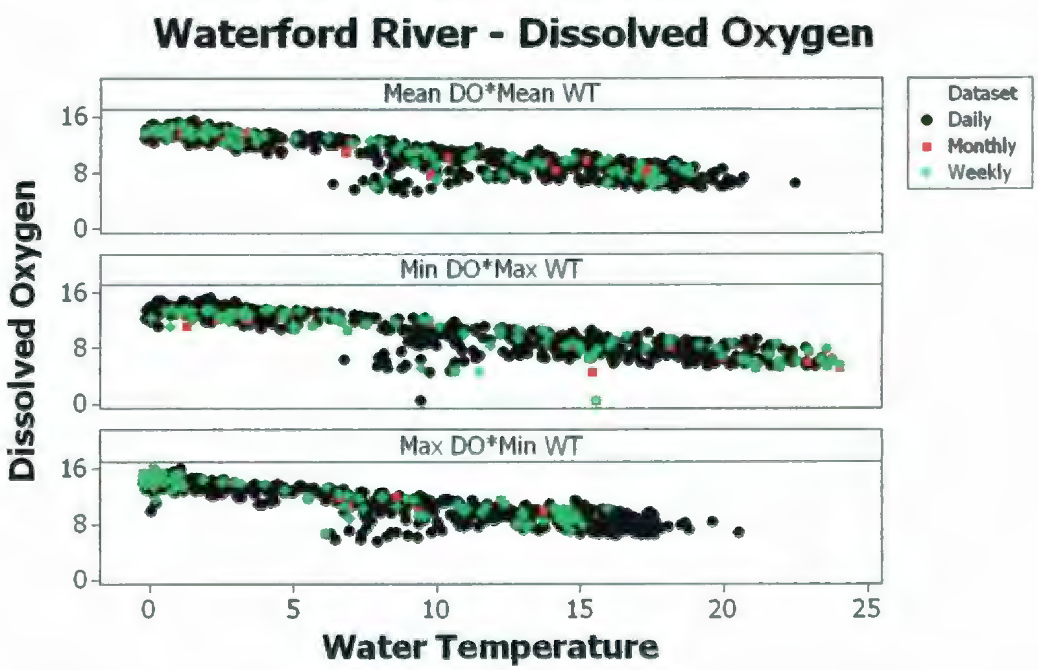


Figure J.4 Waterford River Water Temperature and Dissolved Oxygen

Appendix K

Curve Fitting Results for Dissolved Oxygen

Regression Models

Table K.1 - Regression Models for Predicting Humber River Dissolved Oxygen

	Mean Datasets Mean Dissolved Oxygen Mean Water Temperature Mean Stage			Minimum Datasets Minimum Dissolved Oxygen Maximum Water Temperature Minimum Stage			Maximum Datasets Maximum Dissolved Oxygen Minimum Water Temperature Maximum Stage		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Humber River									
(37 Monthly Observations, 149 Weekly Observations and 986 Daily Observations)									
*Linear Model: DO = a*Tw + b*									
a	-0.291	-0.291	-0.288	-0.258	-0.277	-0.277	-0.347	-0.305	-0.290
b	13.871	13.900	13.857	13.489	13.740	13.740	14.268	13.992	13.867
RSS	9.05	47.07	342.93	9.16	50.50	50.50	10.44	48.44	341.22
R^2 Adj	0.90	0.88	0.87	0.90	0.88	0.88	0.88	0.87	0.87
*Exponential Decay Model: DO = exp(a + b*Tw)*									
a	2.643	2.644	2.642	2.620	2.636	2.636	2.668	2.649	2.641
b	-0.026	-0.026	-0.026	-0.024	-0.025	-0.025	-0.029	-0.026	-0.026
RSS	8.26	43.61	317.68	8.29	46.45	46.45	9.42	45.38	316.88
R^2 Adj	0.91	0.89	0.88	0.91	0.89	0.89	0.90	0.88	0.88
*MLR with Stage: DO = a + b*Tw + c*ST*									
a	NS	NS	14.12	NS	NS	14.10	NS	NS	14.14
b			-0.29			-0.29			-0.29
c			-0.11			-0.11			-0.12
RSS			339.78			347.21			337.79
R^2 Adj			0.87			0.87			0.87

Table K.2 - Regression Models for Predicting Peter's River Dissolved Oxygen

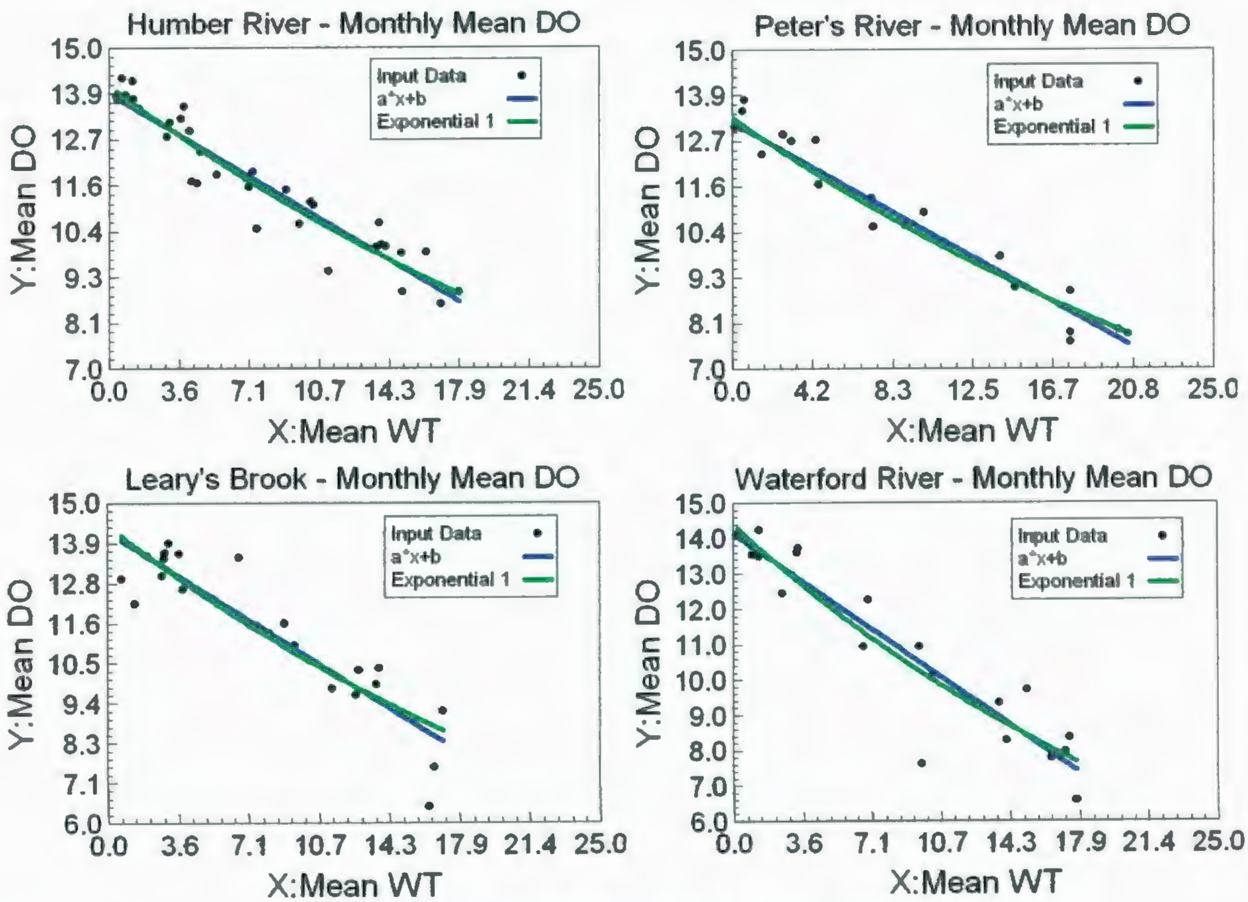
	Mean Datasets Mean Dissolved Oxygen Mean Water Temperature Mean Stage			Minimum Datasets Minimum Dissolved Oxygen Maximum Water Temperature Minimum Stage			Maximum Datasets Maximum Dissolved Oxygen Minimum Water Temperature Maximum Stage		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Peter's River *(23 Monthly Observations, 91 Weekly Observations, 595 Daily Observations)*									
*Linear Model: DO = a*Tw + b*									
a	-0.271	-0.270	-0.270	-0.216	-0.230	-0.245	-0.330	-0.303	-0.286
b	13.189	13.208	13.214	12.498	12.811	13.030	13.887	13.615	13.346
RSS	4.92	27.51	220.58	12.23	46.61	365.03	5.96	27.01	203.54
R^2 Adj	0.94	0.93	0.91	0.88	0.89	0.87	0.92	0.92	0.91
*Exponential Decay Model: DO = exp(a + b*Tw)*									
a	2.588	2.591	2.591	2.545	2.566	2.581	2.634	2.618	2.599
b	-0.025	-0.026	-0.026	-0.023	-0.023	-0.024	-0.028	-0.027	-0.026
RSS	5.42	28.56	224.11	13.34	52.68	374.09	5.76	25.25	198.11
R^2 Adj	0.94	0.92	0.91	0.87	0.88	0.87	0.92	0.92	0.92
*MLR with Stage: DO = a + b*Tw + c*ST*									
a	NS	NS	13.73	NS	NS	13.45	NS	NS	13.81
b			-0.27			-0.25			-0.29
c			-0.43			-0.36			-0.37
RSS			216.13			362.36			199.50
R^2 Adj			0.92			0.87			0.91

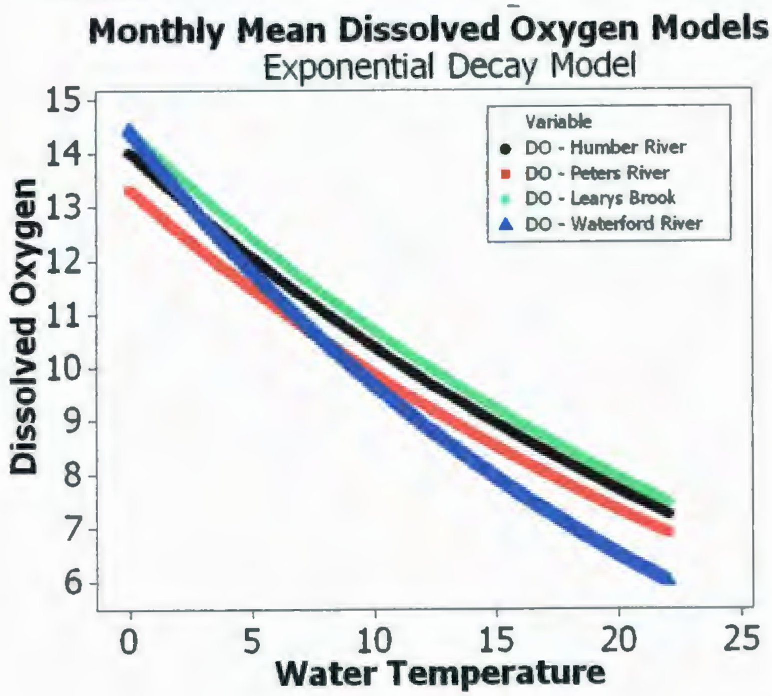
Table K.3 - Regression Models for Predicting Leary's Brook Dissolved Oxygen

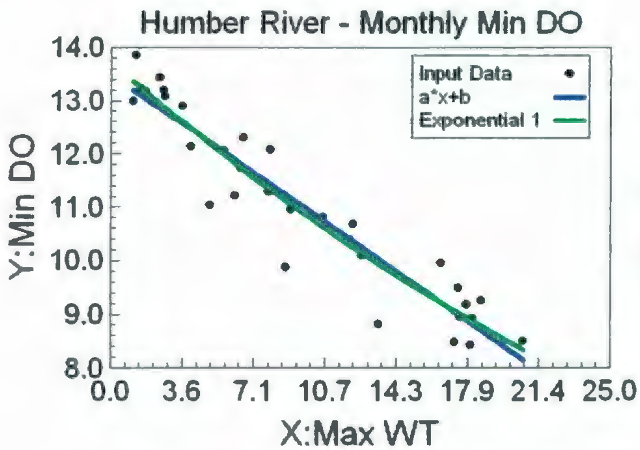
	Mean Datasets Mean Dissolved Oxygen Mean Water Temperature Mean Stage			Minimum Datasets Minimum Dissolved Oxygen Maximum Water Temperature Minimum Stage			Maximum Datasets Maximum Dissolved Oxygen Minimum Water Temperature Maximum Stage		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Leary's Brook *(16 Monthly Observations, 57 Weekly Observations, 347 Daily Observations)*									
*Linear Model: DO = a*Tw + b*									
a	-0.347	-0.407	-0.398	-0.341	-0.355	-0.351	-0.442	-0.418	-0.402
b	14.207	14.328	14.199	12.626	12.757	13.401	15.895	15.302	14.689
RSS	15.05	63.15	687.79	118.71	238.54	1093.68	11.26	48.52	2130.05
R^2 Adj	0.83	0.81	0.71	0.41	0.51	0.59	0.82	0.84	0.42
*Exponential Decay Model: DO = exp(a + b*Tw)*									
a	2.666	2.673	2.662	2.567	2.557	2.607	2.772	2.734	2.691
b	-0.030	-0.036	-0.035	-0.037	-0.035	-0.033	-0.034	-0.033	-0.033
RSS	17.28	76.56	770.17	126.67	265.69	1180.39	10.93	51.53	2209.88
R^2 Adj	0.80	0.77	0.68	0.37	0.46	0.55	0.82	0.83	0.40
*MLR with Stage: DO = a + b*Tw + c*ST*									
a	NS	NS	14.05	NS	NS	13.23	NS	NS	14.52
b			-0.41			-0.36			-0.41
c			0.66			0.81			0.91
RSS			662.87			1061.50			3782.27
R^2 Adj			0.72			0.60			0.29

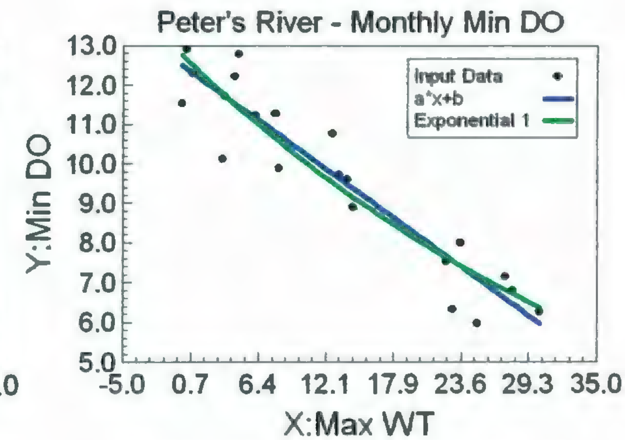
Table K.4 - Regression Models for Predicting Waterford River Dissolved Oxygen

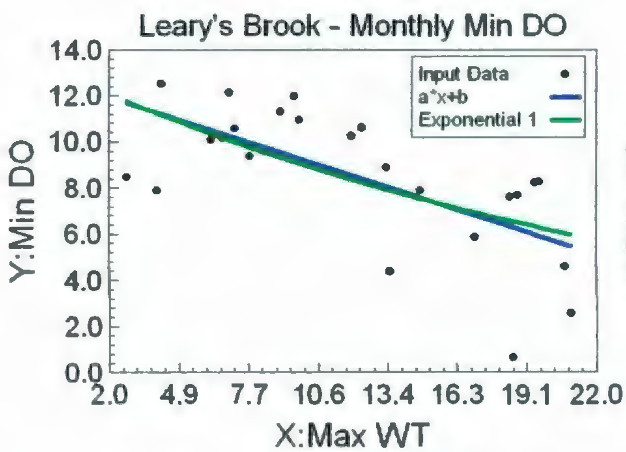
	Mean Datasets Mean Dissolved Oxygen Mean Water Temperature Mean Stage			Minimum Datasets Minimum Dissolved Oxygen Maximum Water Temperature Minimum Stage			Maximum Datasets Maximum Dissolved Oxygen Minimum Water Temperature Maximum Stage		
	Monthly	Weekly	Daily	Monthly	Weekly	Daily	Monthly	Weekly	Daily
Waterford River *(21 Monthly Observations, 90 Weekly Observations, 587 Daily Observations)*									
*Linear Model: DO = a*Tw + b*									
a	-0.378	-0.371	-0.373	-0.335	-0.330	-0.349	-0.409	-0.397	-0.383
b	14.148	14.097	14.098	12.649	13.412	13.828	14.791	14.496	14.249
RSS	15.13	120.07	886.03	102.32	229.78	1132.24	15.77	104.83	851.31
R^2 Adj	0.88	0.81	0.80	0.55	0.71	0.77	0.84	0.81	0.79
*Exponential Decay Model: DO = exp(a + b*Tw)*									
a	2.667	2.663	2.663	2.617	2.631	2.651	2.697	2.680	2.668
b	-0.035	-0.035	-0.035	-0.043	-0.035	-0.035	-0.033	-0.030	-0.035
RSS	14.73	116.89	861.44	96.34	226.17	1090.66	15.33	101.64	825.25
R^2 Adj	0.89	0.82	0.81	0.58	0.72	0.78	0.85	0.82	0.80
*MLR with Stage: DO = a + b*Tw + c*ST*									
a	NS	NS	0.81	NS	NS	12.69	NS	NS	13.33
b			-0.36			-0.33			-0.36
c			0.87			1.92			1.33
RSS			877.14			1107.64			3283.55
R^2 Adj			0.80			0.77			0.48

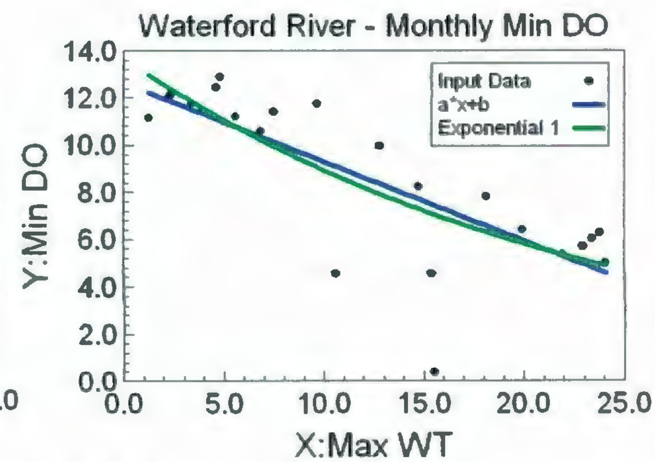

Humber River - Monthly Mean DO
Peter's River - Monthly Mean DO
Leary's Brook - Monthly Mean DO
Waterford River - Monthly Mean DO
Input Data
a*x+b
Exponential 1
Y:Mean DO
X:Mean WT

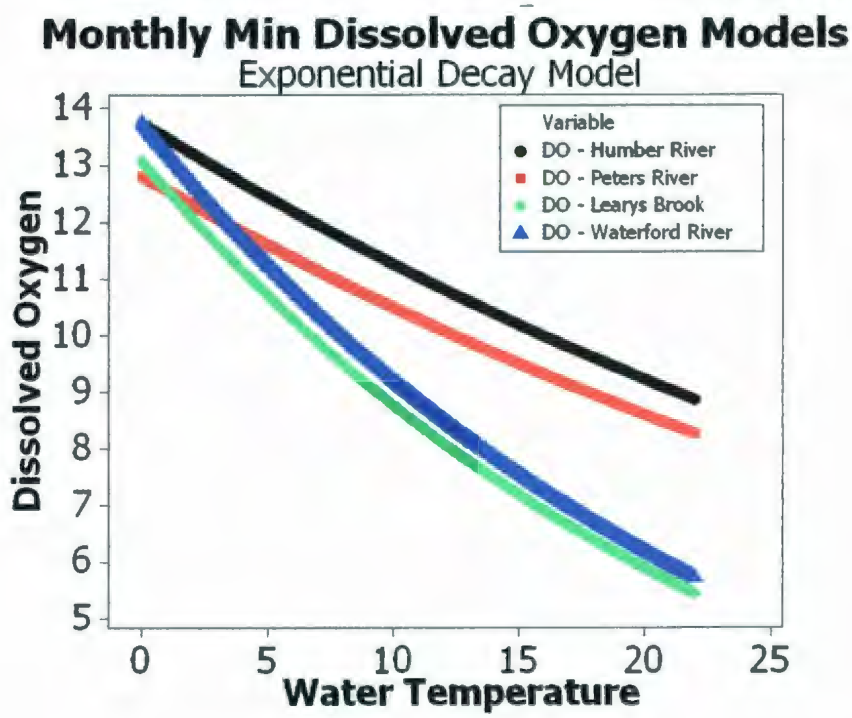

Monthly Mean Dissolved Oxygen Models
Exponential Decay Model
Variable
DO - Humber River
DO - Peters River
DO - Learys Brook
DO - Waterford River
Dissolved Oxygen
Water Temperature

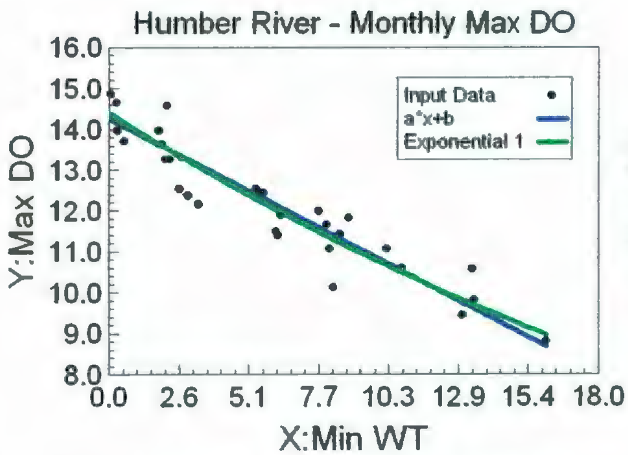

Humber River - Monthly Min DO
Input Data
a*x+b
Exponential 1
Y:Min DO
X:Max WT
14.0
13.0
12.0
11.0
10.0
9.0
8.0
0.0
3.6
7.1
10.7
14.3
17.9
21.4
25.0

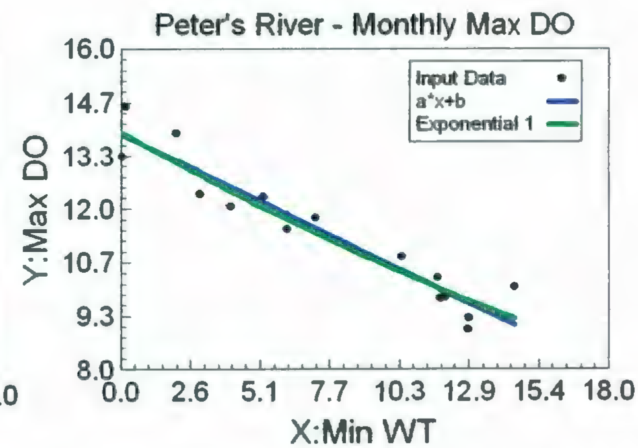

Peter's River - Monthly Min DO
Input Data
a*x+b
Exponential 1
Y:Min DO
X:Max WT
13.0
12.0
11.0
10.0
9.0
8.0
7.0
6.0
5.0
-5.0
0.7
6.4
12.1
17.9
23.6
29.3
35.0

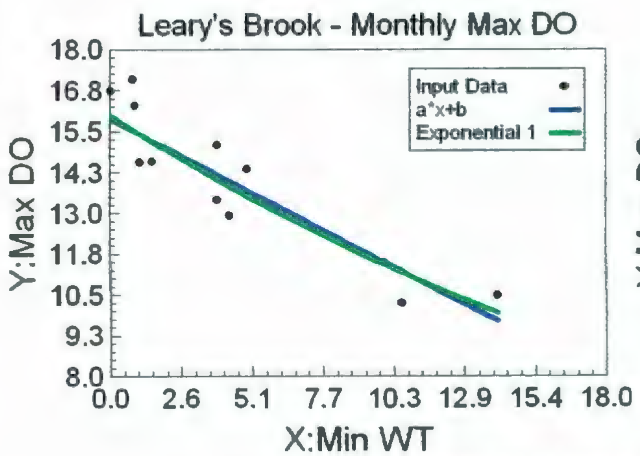

Leary's Brook - Monthly Min DO
Input Data
a*x+b
Exponential 1
Y:Min DO
X:Max WT
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
2.0
4.9
7.7
10.6
13.4
16.3
19.1
22.0

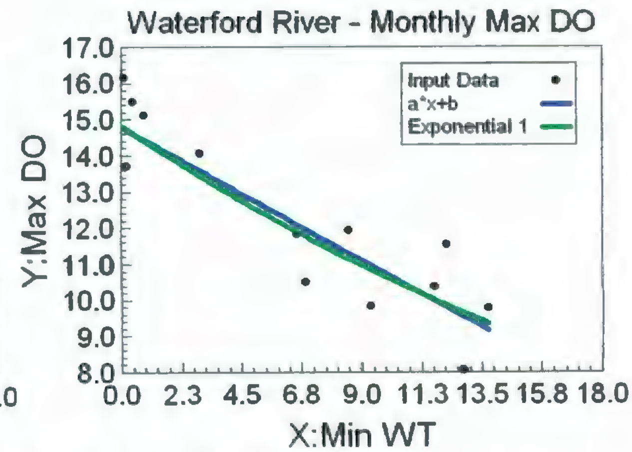

Waterford River - Monthly Min DO
Input Data
a*x+b
Exponential 1
Y:Min DO
X:Max WT
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0.0
5.0
10.0
15.0
20.0
25.0

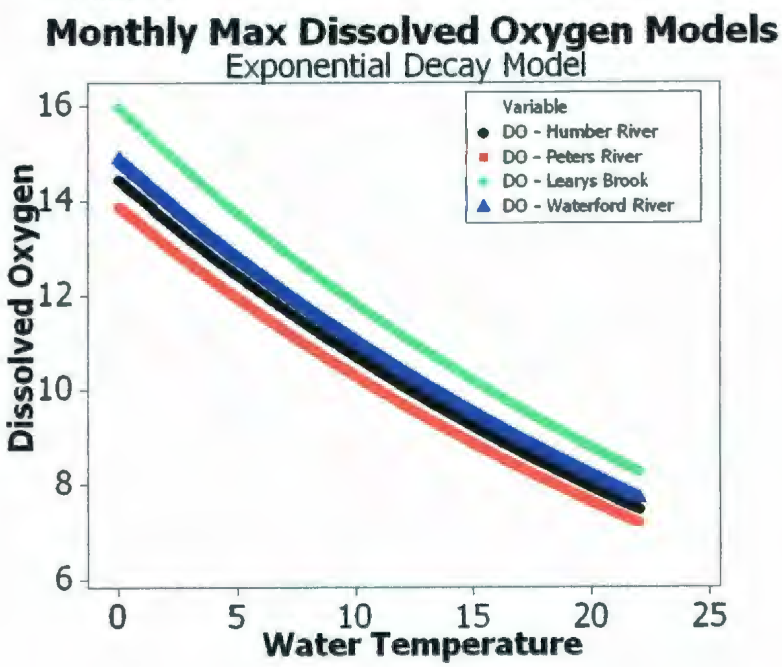

Monthly Min Dissolved Oxygen Models
Exponential Decay Model
Variable
DO - Humber River
DO - Peters River
DO - Learys Brook
DO - Waterford River
Dissolved Oxygen
Water Temperature
14
13
12
11
10
9
8
7
6
5
0
5
10
15
20
25

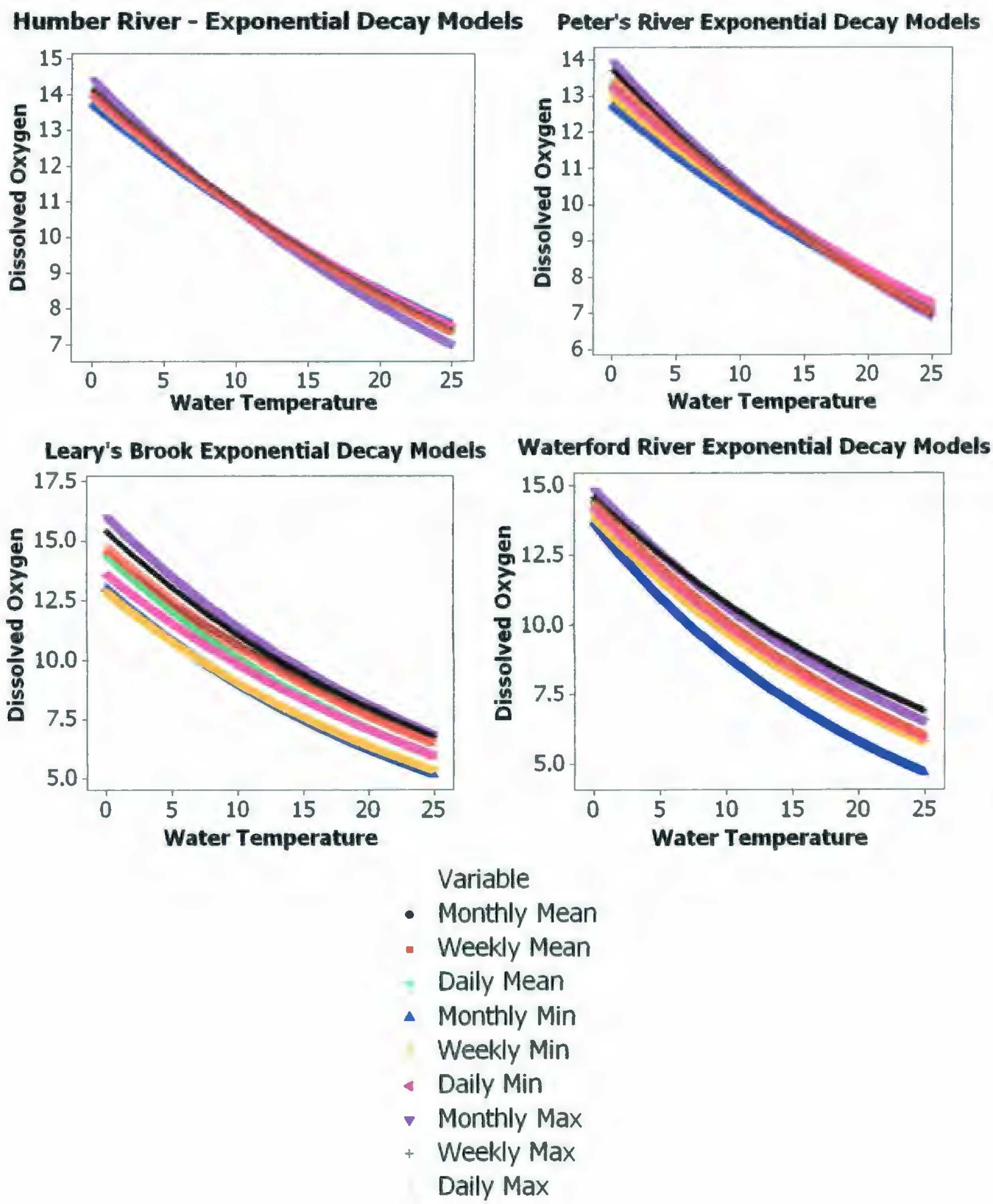

Humber River - Monthly Max DO
Y:Max DO
X:Min WT
Input Data
a*x+b
Exponential 1
16.0 15.0 14.0 13.0 12.0 11.0 10.0 9.0 8.0
0.0 2.6 5.1 7.7 10.3 12.9 15.4 18.0


Peter's River - Monthly Max DO
Y:Max DO
X:Min WT
Input Data
a*x+b
Exponential 1
16.0 14.7 13.3 12.0 10.7 9.3 8.0
0.0 2.6 5.1 7.7 10.3 12.9 15.4 18.0


Leary's Brook - Monthly Max DO
Y:Max DO
X:Min WT
Input Data
a*x+b
Exponential 1
18.0 16.8 15.5 14.3 13.0 11.8 10.5 9.3 8.0
0.0 2.6 5.1 7.7 10.3 12.9 15.4 18.0


Waterford River - Monthly Max DO
Y:Max DO
X:Min WT
Input Data
a*x+b
Exponential 1
17.0 16.0 15.0 14.0 13.0 12.0 11.0 10.0 9.0 8.0
0.0 2.3 4.5 6.8 9.0 11.3 13.5 15.8 18.0


Monthly Max Dissolved Oxygen Models
Exponential Decay Model
Dissolved Oxygen
Water Temperature
Variable
DO - Humber River
DO - Peters River
DO - Learys Brook
DO - Waterford River
16 14 12 10 8 6
0 5 10 15 20 25


Humber River - Exponential Decay Models
Peter's River Exponential Decay Models
Leary's Brook Exponential Decay Models
Waterford River Exponential Decay Models
Dissolved Oxygen
Water Temperature
Variable
Monthly Mean
Weekly Mean
Daily Mean
Monthly Min
Weekly Min
Daily Min
Monthly Max
Weekly Max
Daily Max

Appendix L

Model Testing Results for Dissolved Oxygen

The exponential decay models were tested using the dissolved oxygen datasets reserved for model testing purposes. The following tables and plots are contained in this appendix:

- Table L.1 - Using the Exponential Model for Predicting Humber River Dissolved Oxygen.
- Table L.2 - Using the Exponential Model for Predicting Peter's River Dissolved Oxygen
- Table L.3 - Using the Exponential Model for Predicting Leary's Brook Dissolved Oxygen
- Table L.4 - Using the Exponential Model for Predicting Waterford River Dissolved Oxygen
- Figure L.1 Presents a boxplot of the absolute value of the observed mean dissolved oxygen minus the predicted mean dissolved oxygen found using the exponential model.

Table L.1 - Using the Logistic 1 Model for Predicting Dissolved Oxygen - Humber River

		Absolute Value of Difference Abs[Pred - Obs]				Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%			
Adj R^2	Obs.	Mean	Max	Min	St Dev	Mean	Max	Min	St Dev
Humber River									
Mean Dissolved Oxygen Models									
Monthly Mean DO = exp(2.643 - 0.0258 Mean WT)									
0.91	14	0.76	1.77	0.04	0.53	6.51	16.11	0.3	5.02
*Weekly Mean DO = exp(2.644 - 0.0258 * Mean WT)*									
0.89	63	0.74	2.9	0.01	0.59	6.18	21.48	0.04	5.05
*Daily Mean DO = exp(2.642 - 0.0256 * Mean WT)*									
0.88	398	0.76	2.97	0	0.59	6.36	22.21	0	5
Minimum Dissolved Oxygen Models									
Monthly Min DO = exp(2.620 - 0.024 Max WT)									
0.91	14	0.859	2.245	0.007	0.629	7.57	17.84	0.058	5.308
Weekly Min DO = exp(2.6236 - 0.025 Max WT)									
0.89	63	0.792	2.83	0.013	0.602	6.728	21.47	0.102	5.08
Daily Min DO = exp(2.636 - 0.025 Max WT)									
0.89	398	0.761	2.983	0.00002	0.602	6.425	22.24	0.0001	5.14
Maximum Dissolved Oxygen Models									
Monthly Max DO = exp(2.668 - 0.029 Min WT)									
0.9	14	0.804	2.664	89	0.726	6.182	18.82	0.631	5.729
Weekly Max DO = exp(2.649 - 0.026 Min WT)									
0.88	63	0.782	2.914	0.008	0.606	6.223	20.898	0.058	4.817
Daily Max DO = exp(2.641 - 0.026 Min WT)									
0.88	398	0.767	3	0.001	0.592	6.329	22.34	0.011	4.947

Table L.2 - Using the Logistic 1 Model for Predicting Dissolved Oxygen - Peter's River

		Absolute Value of Difference Abs[Pred - Obs]				Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%			
Adj R^2	Obs.	Mean	Max	Min	St Dev	Mean	Max	Min	St Dev
				Peter's River					
				Mean Dissolved Oxygen Models					
				Monthly Mean DO = exp(2.588 - 0.0255 Mean WT)					
0.940	8	0.720	1.420	0.020	0.500	6.690	10.770	0.140	4.140
				Weekly Mean DO = exp(2.591 - 0.0256 Mean WT)					
0.92	32	0.68	1.94	0.02	0.58	6.38	21.87	0.17	5.75
				Daily Mean DO = exp(2.591 - 0.0256 Mean WT)					
0.91	196	0.72	2.32	0	0.59	6.78	24.7	0.02	5.82
				Minimum Dissolved Oxygen Models					
				Monthly Min DO = exp(2.545 - 0.023 Max WT)					
0.87	8	0.699	1.77	0.121	0.538	6.86	13.82	1.15	4.29
				Weekly Min DO = exp(2.566 - 0.023 Max WT)					
0.88	31	0.712	2.04	0.0004	0.615	7.05	26.68	0.004	6.687
				Daily Min DO = exp(2.581 - 0.024 Max WT)					
0.87	196	0.728	2.252	0.004	0.565	6.947	27.71	0.042	5.66
				Maximum Dissolved Oxygen Models					
				Monthly Max DO = exp(2.634 - 0.028 Min WT)					
0.92	8	0.791	1.902	0.206	0.638	6.825	16.21	1.47	5.53
				Weekly Max DO = exp(2.618 - 0.027 Min WT)					
0.92	32	0.759	2.1	0.026	0.583	6.806	20.28	0.28	5.64
				Daily Max DO = exp(2.599 - 0.026 Min WT)					
0.92	196	0.756	2.32	0.013	0.605	7.08	27.74	0.103	6.07

Table L.3 - Using the Logistic 1 Model for Predicting Dissolved Oxygen - Leary's Brook

		Absolute Value of Difference Abs[Pred - Obs]				**Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%**			
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**	**St Dev**
Leary's Brook									
Mean Dissolved Oxygen Models									
Monthly Mean DO = exp(2.666 - 0.0302 Mean WT)									
0.8	8	0.37	0.86	0.1	0.26	3.78	9.03	0.82	2.91
*Weekly Mean DO = exp(2.673 - 0.0357 * Mean WT)*									
0.77	23	0.84	1.71	0.21	0.39	8.85	19.44	1.58	4.92
*Daily Mean DO = exp(2.662 - 0.0351 * Mean WT)*									
0.68	132	0.88	1.83	0.05	0.41	9.51	22.09	0.37	5.2
Minimum Dissolved Oxygen Models									
Monthly Min DO = exp(2.567 - 0.037 Max WT)									
0.37	8	1.44	1.95	0.33	0.58	19.27	31.07	4.44	8.8
Weekly Min DO = exp(2.557 - 0.035 Max WT)									
0.46	23	1.56	2.52	0.16	0.56	19.98	39.08	2.14	8.48
Daily Min DO = exp(2.607 - 0.033 Max WT)									
0.55	132	1.05	2.12	0.04	0.46	12.04	28.67	0.49	6.48
Maximum Dissolved Oxygen Models									
Monthly Max DO = exp(2.772 - 0.034 Min WT)									
0.82	8	0.83	1.87	0.23	0.56	6.19	12.37	2.19	3.52
Weekly Max DO = exp(2.734 - 0.033 Min WT)									
0.83	23	0.42	1.32	0.027	0.3	3.41	9.03	0.25	2.24
Daily Max DO = exp(2.691 - 0.033 Min WT)									
0.4	132	0.87	2.02	0.02	0.5	8.89	22.88	0.143	5.67

Table L.4 - Using the Logistic 1 Model for Predicting Dissolved Oxygen - Waterford River

		Absolute Value of Difference Abs[Pred - Obs]				Absolute Value of % Error Abs[(Pred - Obs)/Pred]*100%			
Adj R^2	**Obs.**	**Mean**	**Max**	**Min**	**St Dev**	**Mean**	**Max**	**Min**	**St Dev**
				Waterford River					
				Mean Dissolved Oxygen Models					
				Monthly Mean DO = exp(2.667 - 0.0353 Mean WT)					
0.89	10	1.49	3.09	0.05	1.17	13.88	36.8	0.5	11
				Weekly Mean DO = exp(2.663 - 0.0347 Mean WT)					
0.82	42	1.71	4.09	0.07	1.15	16.3	41.81	0.65	11.22
				Daily Mean DO = exp(2.663 - 0.0351 Mean WT)					
0.81	272	1.7	8.14	0	1.25	16.28	74	0.01	12.35
				Minimum Dissolved Oxygen Models					
				Monthly Min DO = exp(2.617 - 0.043 Max WT)					
0.58	10	1.24	2.93	0.02	1	15.57	35.78	0.36	11.61
				Weekly Min DO = exp(2.631 - 0.035 Max WT)					
0.72	42	1.57	3.95	0.03	1.22	17.3	48.64	0.51	14.18
				Daily Min DO = exp(2.651 - 0.035 Max WT)					
0.78	272	1.69	10.72	0.004	1.28	17.25	81.46	0.04	13.11
				Maximum Dissolved Oxygen Models					
				Monthly Max DO = exp(2.697 - 0.033 Min WT)					
0.85	10	2.66	11.35	0.06	3.3	20.49	85.36	0.47	24.67
				Weekly Max DO = exp(2.680 - 0.030 Min WT)					
0.82	42	2.21	12.08	0.03	1.98	18.75	96.22	0.23	16.24
				Daily Max DO = exp(2.668 - 0.035 Min WT)					
0.8	272	1.8	13.5	0.002	1.41	16.35	121.18	0.02	13.05

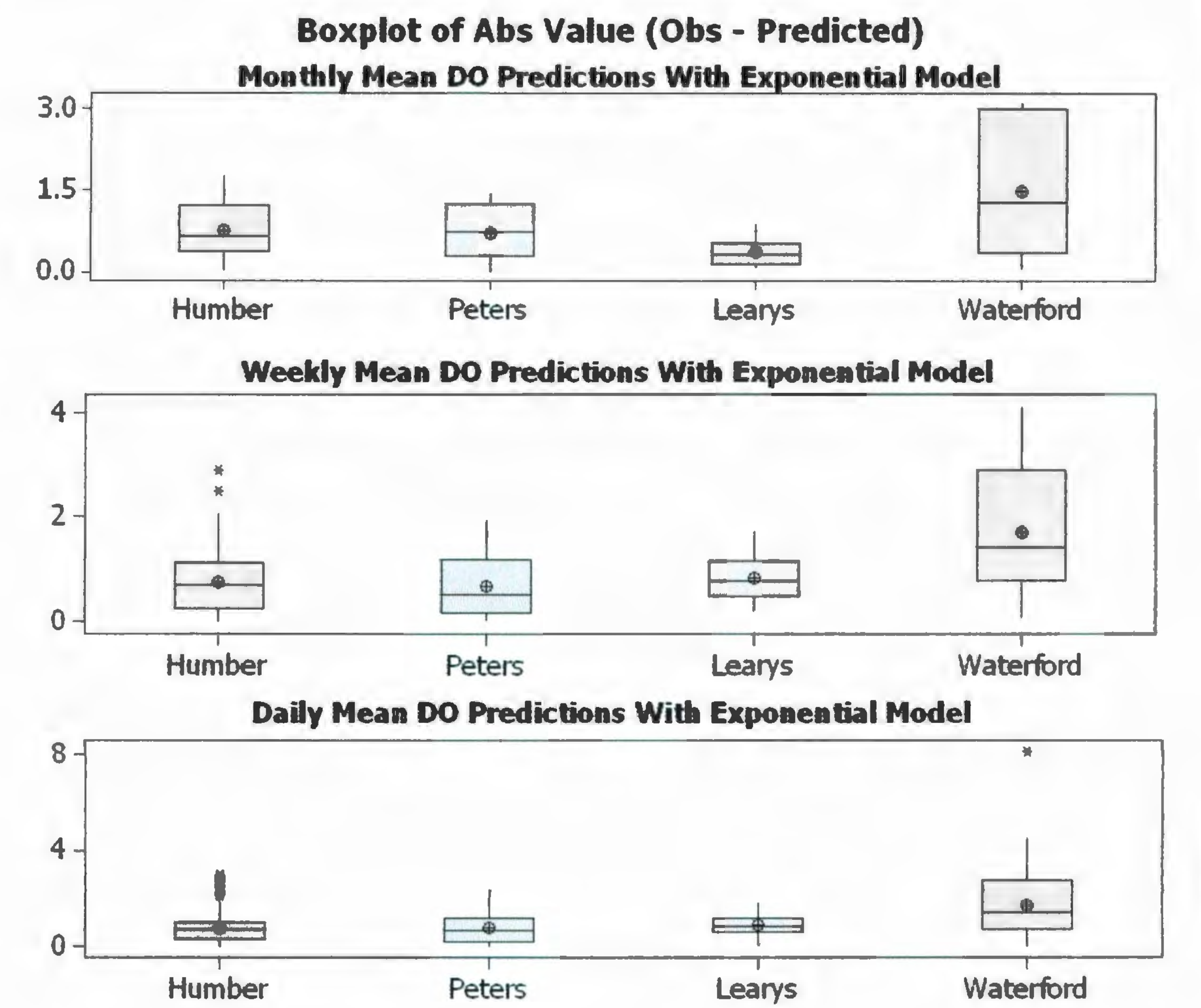


Figure L.1 Boxplot of the Abs Value (Observed - Predicted) for Mean DO Found Using the Exponential Model

Appendix M

Minitab Macro for Plotting Air Temperature, Water Temperature and Dissolved Oxygen

```
(Page 1)
gmacro
ATWTDO
Note Author - Richard Harvey

Erase c1-c100
Erase k1-k100

Note macro uses logistic model for WT
and AT
Note i.e. Tw = 20.92/
(1+exp(0.26*(6.97-Ta)))
Note or   Tw = a/1+exp(b*(c-Ta)))
Note enter a, b, and c below (i.e. 20.92
0.26 6.97)
Set c50;
File "terminal";
Nobs 3.
Copy c50 k50 k51 k52.

Note macro uses exponential model for
WT and DO
Note i.e. DO = exp(a + b*Tw)
Note or   DO = exp(d + e*Tw)
Note enter d and e below (i.e. 2.64 and
-0.026)
set c51;
File "terminal";
Nobs 2.
Copy c51 k53 k54.

Note  enter a value for air temperature
(i.e. 10)
Set c52;
File "terminal";
Nobs 1.
Copy c52 k55.

Name c1 'Air Temperature'
Set c1
-5:25/0.1
End

Name c2 'Water Temperature'
Let c2= k50/(1+exp(k51*(k52-c1)))
Name c3 'Dissolved Oxygen'
Let c3 = exp(k53 + k54*c2)
Let c100 = k50/(1+exp(k51*(k52-k55)))
Copy c100 k1.
Let c101 = exp(k53 + k54*k1)
Copy c101 k2.
Name c5 'Plot Line Left WT1'
Set c5
0:k1/0.01
End
```

```
(Page 2))
Let k60 = count(c5)
Name c6 'Plot Line Left AT1'
Set c6
k60(k55)
ENd

Name c7 'Plot Line Left AT2'
Set c7
k55:25/0.01
End
Let k61 = count(c7)
Name c8 'Plot Line Left WT2'
Set c8
k61(k1)
End


Name c9 'Plot Line Right WT1'
Set c9
0:k1/0.01
End
Let k62 = count(c9)
Name c10 'Plot Line Right DO1'
Set c10
k62(k2)
End

Name c11 'Plot Line Right DO2'
Set c11
0:k2/0.01
End
Let k63 = count(c11)
Name c12 'Plot Line Right WT2'
Set c12
k63(k1)
End

Layout;
Title "Using AT to find WT & WT to
find DO".
Plot 'Water Temperature' * &
'Air Temperature' 'Plot Line Left
WT1' * &
'Plot Line Left AT1' &
'Plot Line Left WT2' * &
'Plot Line Left AT2';
Scale 1;
Tick -5 0 5 10 15 20 25;
Min -5;
Max 25;
Scale 2;
Tick 0 2.5 5 7.5 10 12.5 15 17.5
20;
Min 0;
```

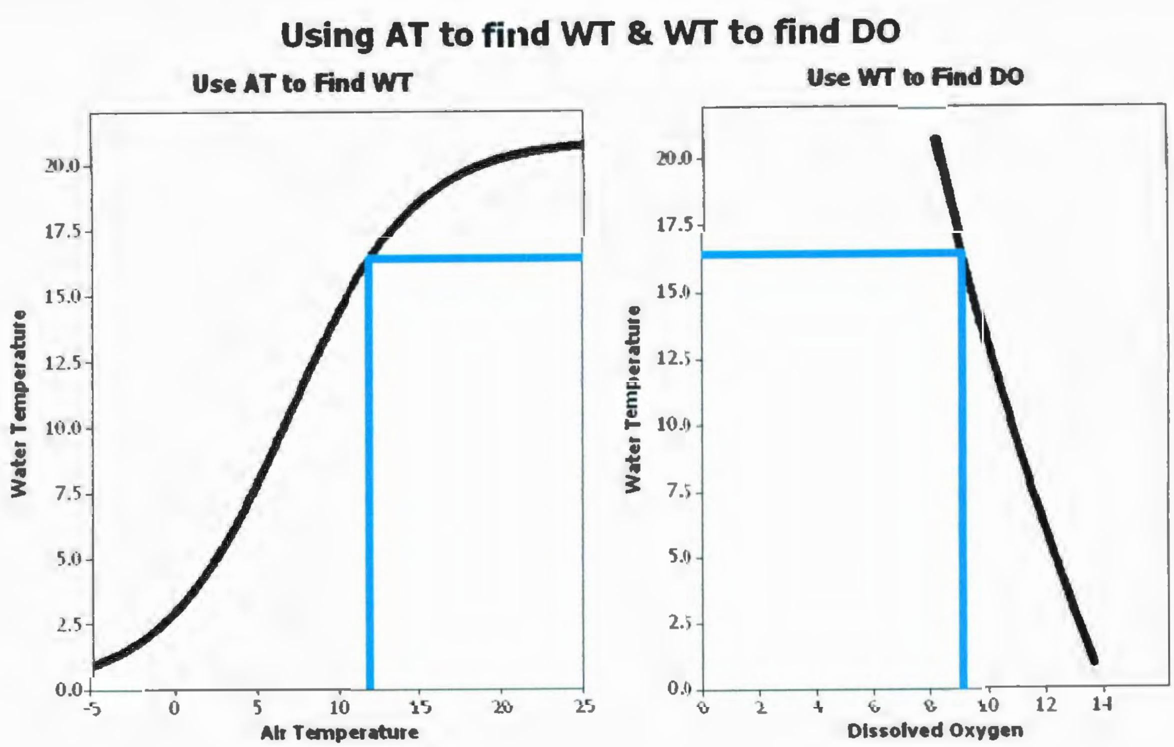
```
(Page 3)
AxLabel 1 "Air Temperature";
ADisplay 1;
AxLabel 2 "Water Temperature";
ADisplay 1;
NoLegend;
Overlay;
NoJitter;
Symbol;
Type 6 6 6 20 23 26 29 2 3 4;
Color 16 66 66 74 42 86 84;
Size 1 1 1;
Grid 1;
MGrid 1;
Grid 2;
MGrid 2;
Footnote;
FPanel;
Title "Use AT to Find WT";
NoDTitle;
Figure 0 0.53 0.02 0.98.

Plot 'Water Temperature' * &
'Dissolved Oxygen' 'Plot Line Right
WT1' &
* 'Plot Line Right DO1' &
'Plot Line Right WT2' * &
'Plot Line Right DO2';
Scale 1;
Min 0;
Tick 0 2 4 6 8 10 12 14;
Scale 2;
Tick 0 2.5 5 7.5 10 12.5 15 17.5 20;
Min 0;
AxLabel 1 "Dissolved Oxygen";
ADisplay 1;
AxLabel 2 "Water Temperature";
ADisplay 1;
NoLegend;
Overlay;
NoJitter;
Symbol;
Type 6 6 6 20 23 26 29 2 3 4;
Color 16 66 66 74 42 86 84;

Size 1 1 1;
Grid 1;
MGrid 1;
Grid 2;
MGrid 2;
Footnote;
FPanel;
Title "     Use WT to Find DO";
NoDTitle;
Figure 0.5 1 0.02 0.98.
Endlayout.

endmacro
```

```
MTB > %ATWTDO
Executing from file: C:\Program Files\MINITAB 14\MACROS\ATWTDO.MAC
Author - Richard Harvey - July 2009
macro is used to develop a three way plot
Step 1 - use air temperature to find water temperature
Step 2 - use water temperature to find dissolved oxygen
In this macro enter a value for AT then the macro will find DO for you
macro uses logistic model for water temperature and air temperature
i.e. Tw = 20.92/(1+exp(0.26*(6.97-Ta)))
or   Tw = a/1+exp(b*(c-Ta)))
please enter values for a, b, and c below (i.e. 20.92 0.26 6.97)
DATA> 20.92 0.26 6.97
macro uses the exponential model for water temperature and dissolved oxygen
i.e. DO = exp(a + b*Tw)
or   DO = exp(d + e*Tw)
plese enter values for d and e below (i.e. 2.64 and -0.026)
DATA> 2.64 -0.026
please enter a value for air temperature (i.e. 10)
DATA> 12
```



Appendix N

Statistical Overview of the Grab Sample Datasets

Table N-1 - Statistical Properties of the Humber River Grab Sample Dataset						
	# of Samples	Minimum	Maximum	Mean	Median	Zeroes
Real Time Measurements						
Water Temperature (ºC)	31	0.627	18.799	7.610	6.806	*
pH (pH units)	31	6.669	7.557	7.039	6.950	*
Specific conductance (μS/cm)	31	24.010	42.898	35.720	37.499	*
Dissolved Oxygen (mg/L)	31	8.585	19.299	12.231	11.440	*
Stage (m)	31	1.482	3.499	2.206	2.236	*
Grab Sample Measurements						
Alkalinity (mg/L CaCO3)	31	10.0	20.0	13.226	13.000	0.000
Color (TCU)	31	22.0	112.0	37.710	38.000	0.000
Conductivity (uS/cm)	31	39.0	56.0	43.903	43.000	0.000
Hardness (mg/L CaCO3)	31	7.0	17.0	10.839	10.000	0.000
pH (pH units)	31	6.6	7.6	6.880	6.840	0.000
TDS (mg/L)	31	25.0	36.0	28.548	28.000	0.000
TSS (mg/L)	*	*	*	*	*	*
Turbidity (NTU)	31	0.400	4.200	0.777	0.600	0.000
Boron (mg/L)	31	0.000	0.030	0.002	0.000	29.000
Bromide (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Calcium (mg/L)	31	3.000	5.000	4.097	4.000	0.000
Chloride (mg/L)	31	3.000	5.000	3.903	4.000	0.000
Flouride (mg/L)	31	0.000	0.110	0.004	0.000	30.000
Potassium (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Sodium (mg/L)	31	0.000	3.000	1.387	2.000	11.000
Sulphate (mg/L)	31	3.000	4.000	3.226	3.000	0.000
Ammonia (mg/L)	31	0.000	0.240	0.025	0.000	19.000
DOC	31	0.800	14.000	5.194	4.900	0.000
Nitrate(ite) (mg/L)	31	0.000	0.130	0.042	0.000	15.000

Table N-1 - continued

	# of Samples	Minimum	Maximum	Mean	Median	Zeroes
Kjeldahl Nitrogen (mg/L)	31	0.000	0.370	0.152	0.130	1.000
Total Phosphorus (mg/L)	31	0.000	0.090	0.019	0.010	11.000
Aluminum (mg/L)	31	0.050	0.170	0.073	0.070	0.000
Antimony (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Arsenic (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Barium (mg/L)	31	0.000	0.010	0.000	0.000	30.000
Cadmium (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Chromium (mg/L)	31	0.000	0.001	0.000	0.000	30.000
Copper (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Iron (mg/L)	31	0.040	0.130	0.073	0.070	0.000
Lead (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Magnesium (mg/L)	31	0.000	1.000	0.161	0.000	27.000
Manganese (mg/L)	31	0.000	0.030	0.001	0.000	30.000
Mercury (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Nickel (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Selenium (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Uranium (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Zinc (mg/L)	31	0.000	0.000	0.000	0.000	31.000
Source Water Temperature (ºC)	16	0.770	16.300	6.599	4.250	*

Table N.2 - Statistical Properties of the Peter's River Grab Sample Dataset (Only reinstallation)

	# of Samples	Minimum	Maximum	Mean	Median	Zeroes
Real Time Measurements						
Water Temperature (ºC)	15	-0.2	28.7	9.5	5.1	*
pH (pH units)	15	4.9	8.1	6.7	6.8	*
Specific conductance (µS/cm)	15	34.0	84.9	49.0	43.0	*
Dissolved Oxygen (mg/L)	15	7.5	14.3	11.0	11.9	*
Stage (m)	15	0.9	1.5	1.1	1.1	*
Grab Sample Measurements						
Alkalinity (mg/L CaCO3)	15	8.000	34.000	20.700	20.000	0
Color (TCU)	15	15.000	74.000	41.900	39.000	0
Conductivity (uS/cm)	15	41.000	89.000	58.100	56.000	0
Hardness (mg/L CaCO3)	15	16.000	33.000	23.200	23.000	0
pH (pH units)	15	6.500	7.600	7.200	7.100	0
TDS (mg/L)	15	25.000	58.000	36.500	33.000	0
TSS (mg/L)	**	*	*	*	*	
Turbidity (NTU)	15	0.400	0.700	0.600	0.600	0
Boron (mg/L)	15	0.000	0.00000000	0-0.02	0.000	7
Bromide (mg/L)	15	0.000	0.000	0.000	0.000	15
Calcium (mg/L)	15	4.800	10.000	7.100	7.000	0
Chloride (mg/L)	15	2.000	6.000	3.700	4.000	0
Flouride (mg/L)	15	0.000	0.000	0.000	0.000	14
Potassium (mg/L)	15	0.000	0.300	0.100	0.000	8
Sodium (mg/L)	15	0.000	3.500	2.000	2.600	4
Sulphate (mg/L)	15	0.000	4.000	1.600	2.000	7
Ammonia (mg/L)	15	0.000	0.100	0.000	0.000	11
DOC	15	3.800	11.000	7.000	6.100	0
Nitrate(ite) (mg/L)	15	0.000	1.400	0.200	0.200	1

Table N.2 Continued

	# of Samples	Minimum	Maximum	Mean	Median	Zeroes
Kjeldahl Nitrogen (mg/L)	15	0.100	0.400	0.200	0.200	3
Total Phosphorus (mg/L)	15	0.000	0.100	0.000	0.000	10
Aluminum (mg/L)	15	0.000	0.100	0.100	0.100	0
Antimony (mg/L)	15	0.000	0.000	0.000	0.000	15
Arsenic (mg/L)	15	0.000	0.000	0.000	0.000	15
Barium (mg/L)	15	0.000	0.000	0.000	0.000	3
Cadmium (mg/L)	15	0.000	0.000	0.000	0.000	15
Chromium (mg/L)	15	0.000	0.000	0.000	0.000	12
Copper (mg/L)	15	0.000	0.000	0.000	0.000	14
Iron (mg/L)	15	0.100	0.300	0.200	0.200	0
Lead (mg/L)	15	0.000	0.000	0.000	0.000	15
Magnesium (mg/L)	15	1.000	2.000	1.300	1.300	0
Manganese (mg/L)	15	0.000	0.000	0.000	0.000	2
Mercury (mg/L)	15	0.000	0.000	0.000	0.000	15
Nickel (mg/L)	15	0.000	0.000	0.000	0.000	15
Selenium (mg/L)	15	0.000	0.000	0.000	0.000	15
Uranium (mg/L)	15	0.000	0.000	0.000	0.000	15
Zinc (mg/L)	15	0.000	0.000	0.000	0.000	11
Source Water Temperature (°C)	6	-0.200	19.300	6.000	0.500 *	

Table N.3 - Statistical Properties of the Leary's Brook Grab Sample Dataset

	# of Samples	Minimum	Maximum	Mean	Median	Zeroes
Real Time Measurements						
Water Temperature (ºC)	20	0.6	17.3	5.7	3.8	*
pH (pH units)	20	5.3	14.0	6.8	6.4	*
Specific conductance (µS/cm)	19	167.0	1329.0	589.7	416.0	*
Dissolved Oxygen (mg/L)	20	9.3	16.3	12.5	12.9	*
Stage (m)	10	0.6	0.9	0.7	0.8	*
Grab Sample Measurements						
Alkalinity (mg/L CaCO3)	20	0.0	13.0	5.2	6.0	8.000
Color (TCU)	20	0.0	24.0	12.8	12.5	0.000
Conductivity (uS/cm)	20	210.0	2100.0	656.0	465.0	0.000
Hardness (mg/L CaCO3)	20	10.0	53.0	27.7	28.0	0.000
pH (pH units)	20	6.1	7.1	6.6	6.6	0.000
TDS (mg/L)	20	107.0	959.0	375.1	236.5	0.000
TSS (mg/L)	*	*	*	*	*	*
Turbidity (NTU)	20	0.3	19.2	3.0	1.7	0.000
Boron (mg/L)	20	0.0	-0.100	0.0	0.0	10.000
Bromide (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Calcium (mg/L)	20	4.0	18.0	9.1	9.3	0.000
Chloride (mg/L)	20	50.0	510.0	174.9	109.5	0.000
Flouride (mg/L)	20	0.0	0.1	0.0	0.0	15.000
Potassium (mg/L)	20	0.0	5.0	1.3	1.0	7.000
Sodium (mg/L)	20	32.0	390.0	115.5	80.5	0.000
Sulphate (mg/L)	20	7.0	27.0	12.4	9.5	0.000
Ammonia (mg/L)	20	0.0	0.3	0.1	0.0	5.000
DOC	20	1.4	5.6	3.1	3.0	0.000
Nitrate(ite) (mg/L)	20	0.2	0.6	0.4	0.4	0.000

Table N.3 Continued

	# of Samples	Minimum	Maximum	Mean	Median	Zeroes
Kjeldahl Nitrogen (mg/L)	19	0.1	0.4	0.2	0.2	1.000
Total Phosphorus (mg/L)	20	0.0	0.1	0.0	0.0	8.000
Aluminum (mg/L)	20	0.0	0.5	0.2	0.1	0.000
Antimony (mg/L)	20	0.0	0.000000	0.00000	0.0	20.000
Arsenic (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Barium (mg/L)	20	0.0	0.1	0.0	0.0	2.000
Cadmium (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Chromium (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Copper (mg/L)	20	0.0	0.000	0.0000	0.0	1.000
Iron (mg/L)	20	0.1	1.3	0.4	0.3	0.000
Lead (mg/L)	20	0.0000	0.0093	0.0000	0.0000	11.000
Magnesium (mg/L)	20	0.0	0-0.00	1.1	1.0	4.000
Manganese (mg/L)	20	0.0	0.3	0.1	0.1	0.000
Mercury (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Nickel (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Selenium (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Uranium (mg/L)	20	0.0	0.0	0.0	0.0	19.000
Zinc (mg/L)	20	0.0	0.1	0.0	0.0	0.000
Source Water Temperature (°C)	Source water temperature not recorded in the grab samples					

Table N.4 - Statistical Properties of the Waterford River Grab Sample Dataset

	# of Samples	Minimum	Maximum	Mean	Median	Zeroes
Real Time Measurements						
Water Temperature (ºC)	20	0.5	20.4	9.2	10.0	*
pH (pH units)	20	5.8	11.2	7.1	6.9	*
Specific conductance (µS/cm)	20	235.0	1060.0	504.3	429.5	*
Dissolved Oxygen (mg/L)	20	7.1	24.7	12.2	10.9	*
Stage (m)	20	0.4	1.2	0.6	0.5	*
Grab Sample Measurements						
Alkalinity (mg/L CaCO3)	20	6.0	21.0	13.7	14.5	0.000
Color (TCU)	20	8.0	26.0	15.9	15.0	0.000
Conductivity (uS/cm)	20	219.0	1200.0	517.0	447.0	0.000
Hardness (mg/L CaCO3)	20	17.0	52.0	34.2	33.0	0.000
pH (pH units)	20	6.6	7.4	7.0	7.0	0.000
TDS (mg/L)	20	142.0	625.0	295.0	255.5	0.000
TSS (mg/L)	*	*	*	*	*	*
Turbidity (NTU)	20	0.5	3.8	1.7	1.6	0.000
Boron (mg/L)	20	0.0	0.0	0.0	0.0	7.000
Bromide (mg/L)	20	0.0	1.1	0.1	0.0	17.000
Calcium (mg/L)	20	5.0	17.0	10.5	10.0	0.000
Chloride (mg/L)	20	51.0	360.0	132.3	110.0	0.000
Flouride (mg/L)	20	0.0	0.5	0.0	0.0	17.000
Potassium (mg/L)	20	1.0	2.6	1.5	1.4	0.000
Sodium (mg/L)	20	33.0	210.0	85.8	66.5	0.000
Sulphate (mg/L)	20	7.0	18.0	11.5	11.0	0.000
Ammonia (mg/L)	20	0.0	0.2	0.1	0.1	0.000
DOC	20	2.2	7.8	3.6	3.3	0.000
Nitrate(ite) (mg/L)	20	0.5	1.2	0.8	0.7	0.000

Table N.4 Continued

	# of Samples	Minimum	Maximum	Mean	Median	Zeroes
Kjeldahl Nitrogen (mg/L)	18	0.0	0.6	0.3	0.3	2.000
Total Phosphorus (mg/L)	20	0.0	0.3	0.0	0.0	9.000
Aluminum (mg/L)	20	0.0	0.2	0.1	0.1	0.000
Antimony (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Arsenic (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Barium (mg/L)	20	0.0	0.0	0.0	0.0	4.000
Cadmium (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Chromium (mg/L)	20	0.0	0.0	0.0	0.0	17.000
Copper (mg/L)	20	0.0	0.0	0.0	0.0	4.000
Iron (mg/L)	20	0.1	0.4	0.2	0.2	0.000
Lead (mg/L)	20	0.00	0.00	0.00	0.00	13.000
Magnesium (mg/L)	20	1.0	2.5	1.9	2.0	0.000
Manganese (mg/L)	20	0.0	0.2	0.1	0.1	0.000
Mercury (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Nickel (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Selenium (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Uranium (mg/L)	20	0.0	0.0	0.0	0.0	20.000
Zinc (mg/L)	20	0.00	0.03	0.0127	0.01	8.000
Source Water Temperature (ºC)	0*	*	*	*		

Appendix O

Using the ARIMA Approach for Developing Control Charts for Observations with Seasonality

Overview

The ARIMA model fitting approach for control chart work will also work well for seasonal data that does not show large amounts of autocorrelation - i.e. monthly mean dissolved oxygen levels recorded at the RTWQ stations. The following figure presents a scatterplot of the monthly mean dissolved oxygen levels at the Humber River station - note the twelve month time period in the data.

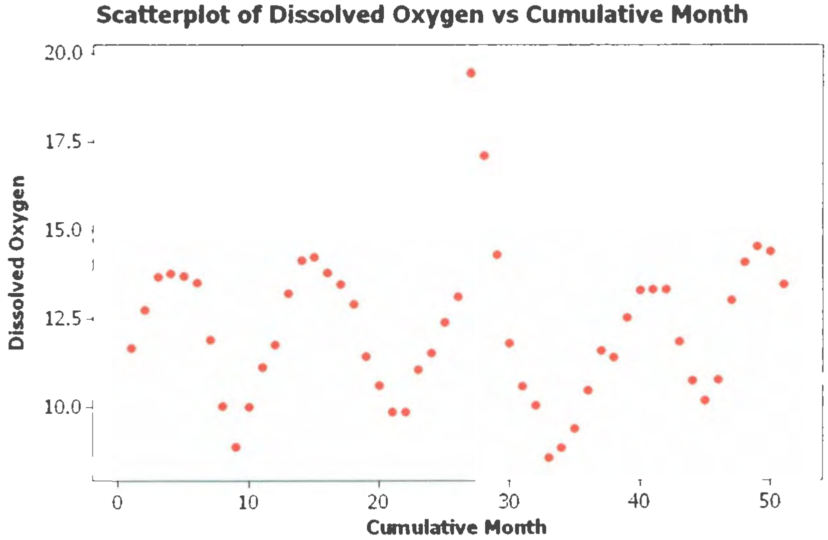


Figure O.1 - Scatterplot of the Monthly Mean DO

Minitab was used to check the monthly mean observations for linear trends (no major trends) and normality (significantly normal). The ACF plot for the observations showed lag 1 and lag 6 to be important. The PACF plot for the observations showed lag 1 had a significant positive partial autocorrelation coefficient while lag 2 had a significant negative partial autocorrelation coefficient.

Minitab was used to check the various available models - the AR(1) model with a seasonal component was found to fit well to the data.

Type	Coef	SE Coef	T	P
AR 1	0.4791	0.1553	3.08	0.004
SAR 12	1.0068	0.1802	5.59	0.000
Constant	-0.0377264	-0.2243473	0.17	0.867
Mean	10.68	63.53		

Once the seasonal AR(1) model was fit to the data it was then possible to use a control chart to examine the data. The Shewhart control chart (subgroup size of 2) shows that subsample 14 (February and March 2006) had an out of control dissolved oxygen level. Referring.

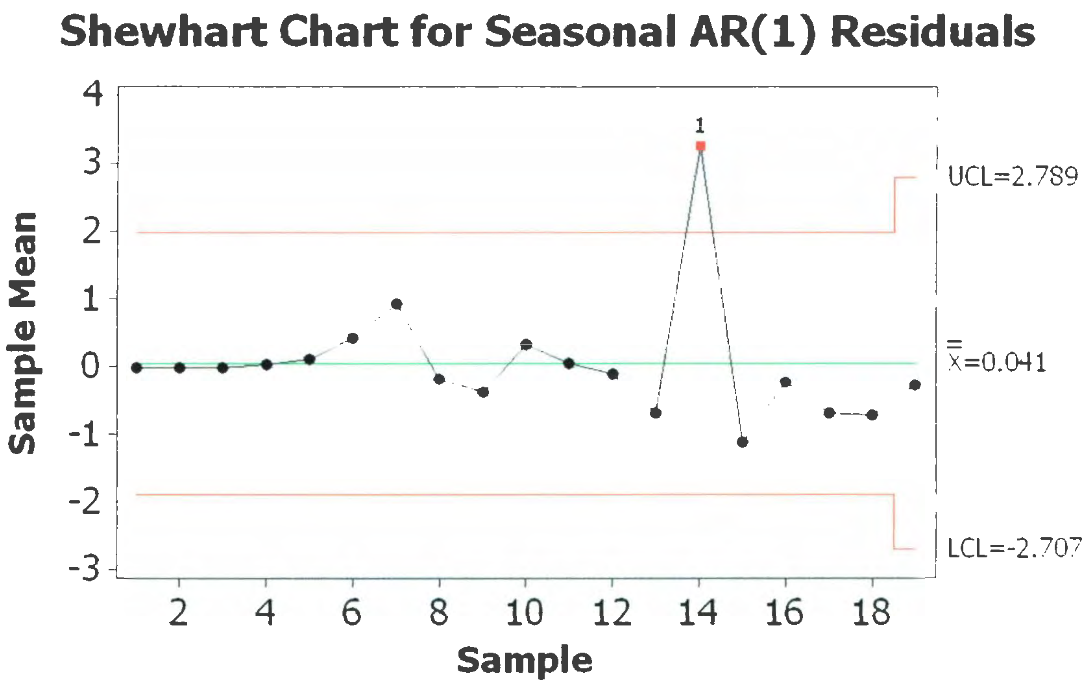


Figure O.3 - Shewhart Chart for Seasonal AR(1) Residuals

Appendix P

Using Spectral Analysis to Find the Time Period of a Dataset

Overview

Although everyday experience tells us that monthly mean water temperature should follow a twelve month cycle, we can use a technique known as spectral analysis to identify the time period of the Humber River monthly mean water temperature observations.

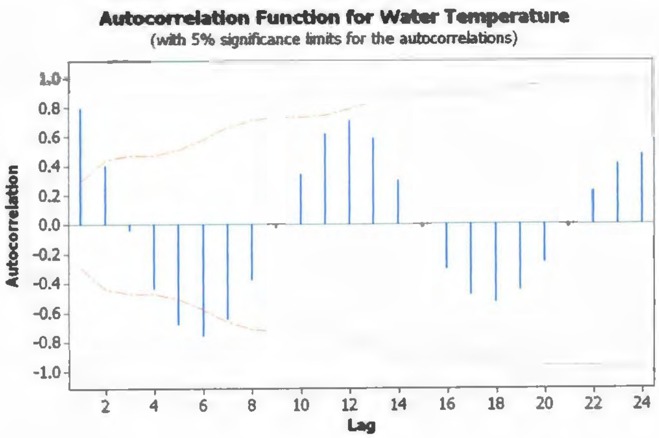
A minitab macro (**spectrum.mac** - author Dr. Leonard Lye of Memorial University) can be used to carry out this spectral analysis. In order for the macro to work properly the water temperature needs to be in column one while the time of the observation needs to be in column two of the Minitab worksheet.

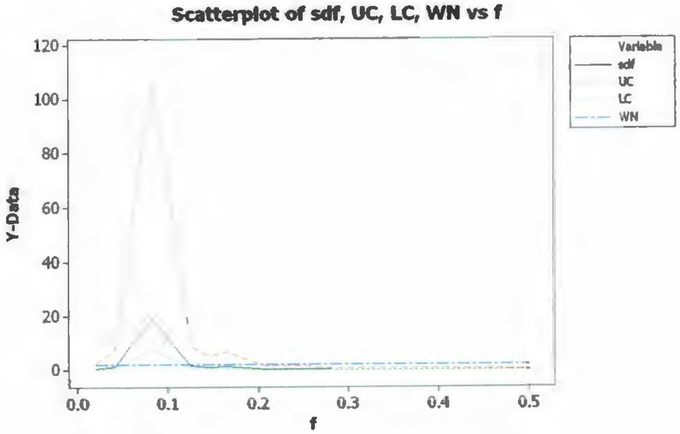
The macro will ask for the number of time lags - if we use 24 it will be easy to see the 12 month cycles in the data. The macro will also develop a plot of the spectral density function - it is pretty broad as we are only using 24 lags. Note that a frequency of 1/12 or 0.083 stands out from the rest,

The results from the spectral analysis of the mean monthly water temperature can be compared with 1000 randomly generated values using a normal distribution. Using the spectral analysis with 100 lags we find that the scatterplot shows that the random data is all over the place - which is different than what was found with the water temperature data where one particular frequency stuck out from the others.

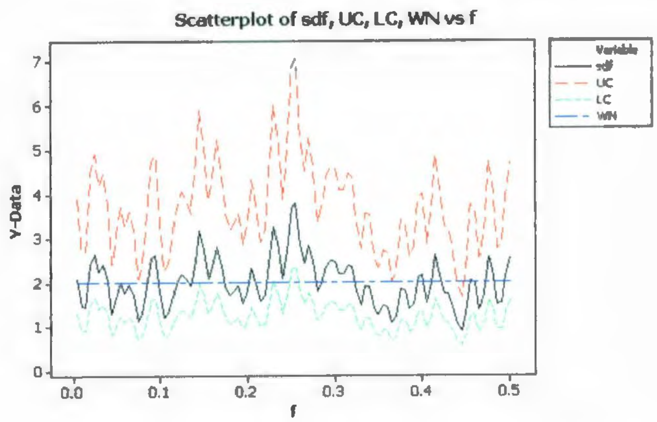
Spectrum.mac Minitab macro

```
GMACRO
spectrum
Note Macro performs spectral analysis of the data
Note Author - Dr. Leonard Lye, Memorial University
erase c3-c10
Note
Note Data in C1 and Index or Year in C2.
Note
name c3 'r(k)' c4 'k' c5 'w(k)' c6 'f' c7 'r x w'
name c8 'rwcos' c9 'sdf' c10 'period'
Note: Enter number of lags to use.  E.g. 24
set c50;
file "terminal";
nobs 1.
let k2=c50(1)
acf k2 c1 c3
set c4
1:k2
end
let c5=0.5*(1+cos(3.14159*c4/k2))
let c6=c4/(2*k2)
let c7=c5*c3
let c10=1/c6
do k1=1:k2
       let c8=c7*cos(2*3.14159*c6(k1)*c4)
       let c9(k1)=2*(1+2*sum(c8))
       enddo
let c11=loge(c9)
name c11 'logsdf'
let k5=8*n(c1)/(3*k2)
invcdf 0.025 k6;
chisquare k5.
invcdf 0.975 k7;
chisquare k5.
let c14=c9*k5/k6
let c15=c9*k5/k7
name c14 'UC' c15 'LC'
Let c12=c11+loge(k5/k6)
let c13=c11+loge(k5/k7)
name c12 'UCL' c13 'LCL'
set c16
k2(2)
end
name c16 'WN'
let c17= loge(c16)
Plot 'sdf'*'period' 'UC'*'period' 'LC'*'period' 'WN'*'period';
  Connect;
  Overlay.
Plot 'sdf'*c6 'UC'*c6 'LC'*c6 'WN'*c6;
  Connect;
  Overlay.
name c17 'logWN'
Plot 'logsdf'*'f' 'UCL'*'f' 'LCL'*'f' 'logWN'*'f';
  Connect;
  Overlay.
ENDMACRO
```





Spectral Density Function for Monthly Mean Water Temperature (Frequency = 1/12)



Spectral Density Function for 1000 Randomly Generated Values (nothing stands out)

Appendix Q

Macros for Modified Control Charts

Overview of the Macros

Macro for Modified Dissolved Oxygen Control Chart - Limit on Lower Side

Macro is used to plot up a modified control chart with limits on the lower side. c1 contains the date and time, c2 contains the water temperature, c3 contains the pH, c4 contains specific conductance, c5 contains dissolved solids, c6 contains percent saturation, c7 contains dissolved oxygen and c8 contains turbidity. The green line plotted on the chart is the mean of the dataset. The desired limits are defined by the user.

Macro for Modified Water Temperature Control Chart - Limit on Higher Side

Macro is similar to the one written for dissolved oxygen but instead flags points outside of a user defined threshold.

Macro for Modified pH Control Chart - Two-Sided Limits

Macro is used to to plot a control chart with limits on two sides. Points outside of these lines are flagged in red.

Macro for Modified Dissolved Oxygen Control Chart - Limit on Lower Side

```
(Page 1)
gmacro
ControlDO
Note Author - R.Harvey
Note Set The desired lower limit (i.e.
5 and 5.5)

Erase c9-c100
Erase k1-k100
Set c50;
File "terminal";
Nobs 2.
Copy c50 k50 k51
Let k1 = min(c7)
Let k2 = max(c7)
Let k3 = mean(c7)
If (k1 > k51)
  Plot c7*c1;
  AxLabel 1 "Time";
  ADisplay 1;
  AxLabel 2 "Dissolved Oxygen";
  ADisplay 1;
  NoLegend;
  NoJitter;
  Symbol;
  Type 6;
  Color 16;
  Size 1;
  Grid 1;
  Grid 2;
  Reference 2 k50;
  Type 1;
  COlor 17;
  Size 5;
  MODEL 1;
Reference 2 k51;
  Type 1;
  Color 17;
  Size 5;
  MODEL 1;
  Reference 2 k3;

(Page 2)
  Type 1;
  Color 53;
  Size 5;
  MODEL 1;
  Footnote;
  FPanel;
  Title "Modified Control Chart for
Dissolved Oxygen";
  NoDTitle.
Elseif k1 > k50 AND k1 <= k51
  Copy c1 c7 c10 c11;
  Include;
  Where "c7 <= k51";
  Varnames.
  Plot c7*c1 c11*c10;
  AxLabel 1 "Time";
  ADisplay 1;
  AxLabel 2 "Dissolved Oxygen";
  ADisplay 1;
  NoLegend;
  Overlay;
  NoJitter;
Symbol;
  Type 6 16 16 20 23 26 29 2 3 4;
  Color 16 25 52 74 42 86 84;
  Size 1 1;
  Grid 1;
  Grid 2; Type 1;
  Color 17;
  Size 5;
  MODEL 1;
  Reference 2 k51;
  Type 1;
  Color 17;
  Size 5;
MODEL 1;
  Reference 2 k3;
  Type 1;
  Color 53;
  Size 5;
 MODEL 1;
 Title "Modified Chart for
DO";Footnote;
  FPanel;

(Page 3)
  NoDTitle.
Else
Copy c1 c7 c10 c11;
  Include;
  Where "c7 <= k50";
  Varnames.
  Copy c1 c7 c13 c14;
  Include;
  Where "c7 <= k51";
  Varnames.
  Plot c7*c1 c11*c10 c14*c13;
  AxLabel 1 "Time";
  ADisplay 1;
  AxLabel 2 "Dissolved Oxygen";
  ADisplay 1;
  NoLegend;
  Overlay;
  NoJitter;
  Symbol;
Type 6 16 16 20 23 26 29 2 3 4;
  Color 16 25 25 74 42 86 84;
  Size 1 1 1;
  Grid 1;
  Grid 2;
  Reference 2 k50;
  Type 1;
  COlor 17;
  Size 5;
  MODEL 1;
  Reference 2 k51;
  Type 1;
  Color 17;
  Size 5;
  MODEL 1;
  Reference 2 k3;
  Type 1;
  Color 53;
  Size 5;
  MODEL 1;
  Footnote;
  FPanel;
  Title "Modified Chart for DO";
  NoDTitle.
ENDIF
endmacro
```

Macro for Modified Water Temperature Control Chart - Limit on Higher Side

```
gmacro
ControlWT
Note - Author - R. Harvey
Note Set The desired upper
limit (i.e. 18 and 20)

Erase c9-c100
Erase k1-k100

Set c50;
File "terminal";
Nobs 2.
Copy c50 k50 k51

Let k1 = min(c2)
Let k2 = max(c2)
Let k3 = mean(c2)

If k1 >= k51 OR k1 >=k50
   Plot c2*c1;
   AxLabel 1 "Time";
   ADisplay 1;
   AxLabel 2 "Water
Temperature";
   ADisplay 1;
   NoLegend;
   NoJitter;
   Symbol;
   Type 16;
   Color 25;
   Size 1;
   Grid 1;
   Grid 2;
   Reference 2 k50;
   Type 1;
   COlor 17;
   Size 5;
   MODEL 1;
   Reference 2 k51;
Type 1;
   Color 17;
   Size 5;
   MODEL 1;
   Reference 2 k3;
   Type 1;
Color 53;
   Size 5;
   MODEL 1;
   Footnote;
   FPanel;
 Title "Modified Control Chart for
Water Temperature";
   NoDTitle.

ELSEIF k2>=k51 AND k1 < k50

   Copy c1 c2 c10 c11;
   Include;
   Where "c2 >= k50";
   Varnames.
   Plot c2*c1 c11*c10;
   AxLabel 1 "Time";
   ADisplay 1;
   AxLabel 2 "Water
Temperature";
   ADisplay 1;
   NoLegend;
   Overlay;
   NoJitter;
 Symbol;
   Type 6 16 16 20 23 26 29 2 3
4;
   Color 16 25 52 74 42 86 84;
   Size 1 1;
   Grid 1;
   Grid 2;
   Reference 2 k50;
   Type 1;
   Color 17;
   Size 5;
   MODEL 1;
   Reference 2 k51;
   Type 1;
   Color 17;
   Size 5;
   MODEL 1;
   Reference 2 k3;
   Type 1;
   Color 53;
   Size 5;
MODEL 1;
   Title "Modified Control Chart for
Water Temperature";
   Footnote;
   FPanel;
   NoDTitle.

ELSEIF k2 >=k50 AND k1 <=
k50
   Copy c1 c2 c10 c11;
   Include;
   Where "c2 >= k50";
   Varnames.
   Plot c2*c1 c11*c10;
   AxLabel 1 "Time";
   ADisplay 1;
   AxLabel 2 "Water Temperature";
   ADisplay 1;
   NoLegend;
   Overlay;
   NoJitter;
   Symbol;
Type 6 16 16 20 23 26 29 2 3 4;
   Color 16 25 52 74 42 86 84;
   Size 1 1;  Grid 1; Grid 2; Reference 2 k50;
   Type 1; Color 17; Size 5; MODEL 1;
   Reference 2 k51;
   Type 1;
   Color 17;
   Size 5;
   MODEL 1;
   Reference 2 k3;
   Type 1;
   Color 53;Size 5;
   MODEL 1;
   Title "Modified Control Chart for Water
Temperature";
   Footnote;
   FPanel;
   NoDTitle.
ELSEIF k2<k50
   Plot c2*c1;
   AxLabel 1 "Time";
   ADisplay 1;
   AxLabel 2 "Water Temperature";
   ADisplay 1;
   NoLegend;
   NoJitter;
   Symbol;
   Type 6; Color 16; Size 1; Grid 1;Grid 2;
   Reference 2 k50; Type 1; Color 17; Size 5;
   MODEL 1;
   Reference 2 k51; Type 1; Color 17;Size 5;
   MODEL 1;
   Reference 2 k3;
   Type 1;
   Color 53;
   Size 5;
   MODEL 1;
   Footnote;
   FPanel;
   Title "Modified Control Chart for WT";
   NoDTitle.
ENDIF
endmacro
```

Macro for Modified pH Control Chart - Two-Sided Limits

```
gmacro
ControlpH
Note Author - R. Harvey
Note Set The desired lower and
upper limit (i.e. 6 and 8)

Erase c9-c100
Erase k1-k100

Set c50;
File "terminal";
Nobs 2.
Copy c50 k50 k51

Let k1 = min(c3)
Let k2 = max(c3)
Let k3 = mean(c3)

If k1 > k50 AND k2 < k51
  Plot c3*c1;
  AxLabel 1 "Time";
  ADisplay 1;
  AxLabel 2 "pH Level";
  ADisplay 1;
  NoLegend;
  NoJitter;
  Symbol;
  Type 6;Color 16;
  Size 1; Grid 1;
  Grid 2;Reference 2 k50; Type 1;
COlor 17; Size 5; MODEL 1;
Reference 2 k51;   Type 1;
Color 17;
Size 5; MODEL 1;
Reference 2 k3;Type 1;Color 53; Size
5;
MODEL 1;
  Footnote;
  FPanel;
  Title "Modified Control Chart for
pH";
  NoDTitle.

Elseif k1 >= k51 OR k2 <= k50

  Plot c3*c1;
  AxLabel 1 "Time";
  ADisplay 1;
  AxLabel 2 "pH Level";
  ADisplay 1;
  NoJitter;
  Symbol;
  Type 6; Color 25;Size 1;Grid 1;
Grid 2;Reference 2 k50; Type 1;
Color 17; Size 5; MODEL 1;
Reference 2 k51; Type 1;Color 17;
  Size 5;
  MODEL 1;
  Reference 2 k3; Type 1;Color 53;
  Size 5;
  MODEL 1;
  Footnote;
  FPanel;
  Title "Modified Chart for pH";
  NoDTitle.

Elseif k2 > k50 AND k2 < k51 AND
k1 <= k50

  Copy c1 c3 c10 c11;
  Include;
  Where "c3 <= k50";
  Varnames.
  Plot c3*c1 c11*c10;
  AxLabel 1 "Time"; ADisplay 1;
  AxLabel 2 "pH Level";
  ADisplay 1; NoLegend;
  Overlay; NoJitter;
  Symbol;
  Type 6 16 16 20 23 26 29 2 3 4;
  Color 16 25 52 74 42 86 84;
  Size 1 1; Grid 1; Grid 2;
  Reference 2 k50; Type 1;
  Color 17; Size 5;
  MODEL 1;
  Reference 2 k51; Type 1; Color 17;
  Size 5;
     MODEL 1;
  Reference 2 k3;
  Type 1;
  Color 53;
  Size 5;
  MODEL 1;
  Footnote;
  FPanel;
  Title "Modified Control Chart for
pH";
  NoDTitle.

Elseif k2 >= k51 AND k1 <= k51
AND k1 > k50

  Copy c1 c3 c10 c11;
  Include;
  Where "c3 >=k51";
  Varnames.
  Plot c3*c1 c11*c10;
  AxLabel 1 "Time"; ADisplay 1;
  AxLabel 2 "pH Level"; ADisplay 1;
  NoLegend; Overlay; NoJitter;
  Symbol;
  Type 6 16 16 20 23 26 29 2 3 4;
  Color 16 25 52 74 42 86 84;
  Size 1 1; Grid 1; Grid 2;
  Reference 2 k50; Type 1; Color 17;
  Size 5; MODEL 1;
  Reference 2 k51; Type 1; Color 17;
  Size 5; MODEL 1;
  Reference 2 k3; Type 1; Color 53;
  Size 5; MODEL 1;
  Footnote; FPanel;
  Title "Modified Chart for pH";
  NoDTitle.

Elseif k2 >=k51 AND k1 <=k50
  Copy c1 c3 c10 c11; Include;
  Where "c3 >=k51"; Varnames.
  Copy c1 c3 c13 c14;
  Include; Where "c3 <= k50";
  Varnames.
  Plot c3*c1 c11*c10 c14*c13;
  AxLabel 1 "Time"; ADisplay 1;
  AxLabel 2 "pH Level"; ADisplay 1;
  NoLegend; Overlay; NoJitter;
  Symbol;
  Type 6 16 16 20 23 26 29 2 3 4;
  Color 16 25 25 74 42 86 84;
  Size 1 1; Grid 1; Grid 2;
  Reference 2 k50; Type 1; Color 17;
     Size 5; MODEL 1;
  Reference 2 k51;Type 1; Color 17;
  Size 5; MODEL 1;
  Reference 2 k3;Type 1; Color 53;
  Size 5;MODEL 1;
  Footnote;
  FPanel;
  Title "Modified Control Chart for
pH";
  NoDTitle.

ENDIF
endmacro
```