The Accuracy of Water Quality Monitoring Data: A Comparison Between Citizen Scientists and Professionals

by
Ashley Shelton

A Thesis Submitted to Saint Mary's University, Halifax, Nova Scotia,
in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Applied Science

June 14, 2013, Halifax, Nova Scotia

Approved: Dr. Cathy Conrad
Supervisor
Department of Geography

Approved: Dr. Shannon Sterling
Supervisor
Earth Sciences, Dalhousie University

Approved: Dr. Kevin Garroway
External Examiner
Water Monitoring & Reporting Specialist
Nova Scotia Environment

Approved: Dr. Peter Bush
Supervisory Committee Member
Nova Scotia Environment

Approved: Dr. Hai Wang
Graduate Studies Representative

Date: June 14, 2013

Abstract

The Accuracy of Water Quality Monitoring Data: A Comparison
Between Citizen Scientists and Professionals

By Ashley M. Shelton

This study compared water quality data of trained citizen scientists and a water professional. Side-by-side field measurements in Nova Scotia's freshwater streams were conducted to determine how professional measurements compared to citizen scientists and to identify what factors improve the ability of citizen scientists to collect accurate water quality data. It was expected that no significant difference would be found between citizen scientists and the professional scientist for all freshwater parameters, within mechanical error and government data correction criteria. Results identified similarities for volunteer and professional measurements including water temperature, pH, conductivity and discharge, while there were significant differences revealed for dissolved oxygen. Changes to address the differences found include further training in calibration and field procedures, to offer a better chance of integration of volunteer data with government run programs. The study aimed to demonstrate the value of volunteer data and whether it can be used to increase the overall knowledge of water resources.

Keywords. Citizen science, Community-based monitoring, Data accuracy, Data collection, Freshwater stream, Monitoring-program design, Surface water, Water quality

June 14, 2013

Acknowledgements

This project could not have been completed without many people and I wish to acknowledge their assistance and support during the course of this study. I am indebted to my supervisors, Dr. Shannon Sterling and Dr. Cathy Conrad, for all their direction, advice, patience and most of all enthusiasm for this study. I am very grateful to my committee member Dr. Peter Bush for providing me his guidance, time and academic experience. The contributions from Dr. Ron Russell and his vital expertise were instrumental in the statistical analysis of this study.

I wish to thank the volunteer participants who were involved in the water quality data collection, your time and effort was greatly appreciated. I must acknowledge Melissa Healey who was invaluable during every step of this project, she is a super star.

The informal support and encouragement of my family and friends has been indispensable to my sanity, as well as providing their time as proofreaders. Lastly, to Jason, my rock, I thank him.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

CABIN	Canadian Aquatic Biomonitoring Network
CAMP	Community Aquatic Monitoring Program
CBEMN	Community-Based Environmental Monitoring Network
CBM	Community-Based Monitoring
CCME	Canadian Council of Ministers of the Environment
CEW	Citizens' Environment Watch
DO	Dissolved Oxygen
EMAN	Ecological Monitoring and Assessment Network
GNL	Government of Newfoundland & Labrador
NGO	Non-Governmental Organizations
NSE	Nova Scotia Environment
QA/QC	Quality Assurance Quality Control
RCCA	Reef Check California Association
REB	Research Ethics Board
RWN	River Watch Network
USEPA	United States Environmental Protection Agency
USGS	United States Geological Survey
WQMN	Water Quality Monitoring Network
YSI ProPlus	YSI Professional Plus

Chapter One

Introduction and Literature Review

1.1 Motivation

Within the province of Nova Scotia, and throughout the world, the need for water quality data has become apparent, as both human and ecosystem health are intrinsically tied to this resource. Comparatively, as the public's environmental consciousness continues to rise (Conrad & Daoust, 2008; Savan *et al.*, 2003), so do water quality concerns such as the lack of information on water quality in rural areas (Fore, Paulsen & O'Laughlin, 2001; Roa Garcia & Brown, 2009); environmental pollution, particularly in water sources (Silva & Sacomani, 2000); and current and/or future water shortages (Asano, 2009; UNU-INWEH, 2012, p.19). To address these water quality and quantity issues, accurate water testing and monitoring are necessary to track and act on these concerns; however, the use of volunteer-based monitoring programs as a source of data collection have been historically considered unreliable (Breed, Stichter & Crone, 2012; Fore, Paulsen, O'Laughlin, 2001; Gillett *et al.*, 2011; Loperfido, Beyer, Just & Schnoor, 2010; Schmeller *et al.*, 2009). As the quality of this data has not yet been accepted among academic and governmental communities, this study sought to examine the accuracy of volunteer-based water quality data collection when compared to a professional water scientist to identify if volunteer data could be integrated in government run programs.

Water quality data can be collected by many sources, including government agencies, educational institutions and private consulting firms. The individuals collecting the data from these sources are termed "professionals" due to their formal training and educational background. Often within government agencies, those collecting the data have some form of professional certification, such as Certified Engineering Technologist (CET) accreditation (TechNova, 2011), or is a scientist with a formal degree from a university. Here the term “professional” was used to describe a person receiving payment for the collection of data by a government agency, an educational institution, or a private company.

Alternatively, water quality data can also be collected by volunteer citizen scientists. As science as a paid profession only became recognized late in the 19th century; previously, interested individuals, now referred to as citizen scientists, led scientific research (Silvertown, 2009). This concept has continued with volunteer-based initiatives, utilizing the public in scientific research (Hochachka *et al.*, 2012). The term “volunteer” is used in this study to describe an individual who is not receiving payment for data collection. Anecdotally, citizen scientists from environmental groups can have similar background training and education as professional scientists; nevertheless, they are working as volunteers. Often volunteer environmental monitoring involves community-based monitoring (CBM) initiatives, which is the involvement and collaboration between concerned citizens, government agencies, industry, academia, community groups and local institutions to monitor, track and respond to common community concerns (Conrad & Daoust, 2008; Whitelaw, Vaughan, Craig, & Atkinson, 2003).

There have been government-funding cuts for the environmental sector (Au *et al.*, 2000; De Souza, 2013), while pollution resulting from continued urbanization and other activities continue which can include sedimentation, petroleum spills, fossil fuel combustion, vehicle exhaust, urban runoff, discharge into rivers and estuaries, and atmospheric deposition (Cavalcante, Sousa, Nascimento, Silveira, & Freire, 2009; Xu and Wu, 2006). Government cutbacks affect the capacity to track and provide appropriate responses to environmental changes, leading to a decline in government investigations and prosecutions with a particular emphasis on water quality (Savan, Morgan, & Gore, 2003). With government agencies and scientists requiring information but lacking the resources necessary to gather it, for example establishing monitoring programs that are large enough to appropriately monitor marine ecosystems, low cost volunteer-based initiatives can provide an alternative for data collection (Pattengill-Semmens, Semmens, & Reef Environmental Education Foundation, 2003). The United Nations Environment Programme have displayed their support for and stressed the necessity of public participation in environmental management to achieve sustainability (Conrad & Sharpe, 2006; UN, 1992). With the proper support for volunteer monitoring programs, citizen scientists can collect valuable data that could be used to effectively monitor and track environmental changes. This form of participatory monitoring can also strengthen decision-making (Leopold, Cakacaka, Meo, Sikolia, & Lecchini, 2009), as citizen science promotes active engagement in policy making (Jordan, Gray, Howe, Brooks, & Ehrenfeld, 2011).

Research examining the comparability between “professionals” and “volunteers” is necessary as challenges and roadblocks are limiting the amount of valuable data collected by citizen scientist being used in government decision-making (Sharpe & Conrad, 2006). Concerns lie in whether the level of the quality of the data is adequate to be integrated with the efforts of professional scientists (Breed *et al.*, 2012; Gillett *et al.*, 2011). Do volunteer citizen scientists collect water quality data that is different from the data collected by professionals? Despite literature indicating that when properly trained, citizen volunteers can collect reliable data and make stream assessments that are comparable to professionals (Fore *et al.*, 2001), an evaluation of previous studies comparing citizen scientist water quality monitoring data with that collected professionals displayed a notable gap in the research focus. Biological indicators such as invertebrate communities have been previously assessed to evaluate accuracy of volunteer data when compared to professional scientists and results have varied; some results demonstrated no significant difference between field samples collected by professionals and volunteers (Fore *et al.*, 2001), while Gillet *et al.* (2010) noted similarities in the description of benthic invertebrate communities although difference were found in relative abundance measurements between volunteer and professionals. Studies evaluating data collected by citizen scientists have also been conducted for water quality parameters including phosphorus, turbidity, electrical conductivity and pH; however, comparative study analysis included historical datasets, with instances of many days between data collection leading to discrepancies in the water quality data and there was also a lack of equipment standardization used between the professionals and the volunteers, leading to further sources of error in data analysis (Nicholson *et al*., 2002).

This study focuses on the variability of water quality data between a treatment group comprising of eighteen volunteers of community groups within Nova Scotia (Table 1.0: Appendix A) and a control group that consisted of one professional water scientist. By using a side-by-side in-situ water quality sampling method, water quality professional field measurements were compared with community volunteer measurements to obtain a more thorough comparative analysis of data accuracy. For the purposes of this study, the measurement collected by a professional is assumed to be the “true value” and the volunteer measurement will be compared to various government data accuracy and rejection criteria and to the instrument cable and sensor accuracy specifications. The objectives of the proposed study were: (1) to identify if the volunteer measurements were within the instrument accuracy specifications of the professional measurements, (2) to identify sources of error in monitoring programs and variability in water quality parameters to determine favorable conditions for volunteer monitoring programs, and (3) to evaluate the level of training provided in this study for improved calibration and field procedure.

The study examined two research questions:

1) Do volunteer citizen scientists collect data that is significantly different from the data collected by professionals?

    a. I hypothesized no significant difference would be found between the water quality measurements collected by the citizen scientists and the professional scientist for all freshwater parameters, within the mechanical error of the equipment used and based on government data correction criteria.

2) What factors can improve the ability of citizen scientists to collect accurate water quality data?

    a. I hypothesized the main source of data error would be resulting from the calibration procedure and previous experience and training in water quality monitoring programs or relevant educational experience outside of volunteer efforts would lead to more robust data accuracy from the participants in the citizen science treatment group.

To address the first research question relating to the differences between the data collected by citizen scientists and a professional scientist when comparing to the in-situ measurement variability of the equipment were examined. The differences of the data collected by citizen scientists and a professional scientist were also compared to provincial, national and international data rejection criteria to interpret how reliable is the volunteer data on various government scales. The second research question encompassed the task of identifying what factors can improve the ability of citizen scientists to collect accurate water quality data. Identifying conditions that improved the data accuracy of volunteer monitoring programs were examined by detecting sources of error in monitoring programs and variability in water quality parameters and evaluating the level of training required for improved field and calibration procedure accuracy for volunteers.

With a better understanding of the accuracy of citizen scientist data collection there is a greater chance of data integration with government run programs, providing a larger reservoir of environmental data. With increased integration and use of citizen science data, there is a potential for increased understanding of various issues relating to the state of the environment, both locally and internationally, whether it relates to migration patterns of butterflies (Breed *et al*., 2012), or localized aquatic ecosystem health. This chapter will explore citizen science and water quality literature to identify sources of water quality variability, sources of error in comparative studies and knowledge gaps.

1.2 Perceptions and Challenges of Citizen Science

As a result of an increase in the public's environmental consciousness beginning in the early 1960s (Conrad & Daoust, 2007; Savan *et al.*, 2003), there has been a significant growth in the number of public participatory initiatives (Lasker and Weiss, 2003; Leopold *at al.*, 2009). One such example of participatory initiative growth can be seen through volunteer monitoring programs used by the United States Environmental Protection Agency (USEPA) (2013) for monitoring of rivers, lakes estuaries, beaches, wetlands and ground water. The number volunteer monitoring programs have seen an increase as Loperfido, Beyer, Just and Schnoor, (2010) noted up to 900 organizations active in the United States. The increase of environmental monitoring of waterways, in particular, by communities and volunteers are a result of the decline in environmental funding from governments (Au *et al.*, 2000). There has been a noticeable decrease in the government's ability to monitor the environment over recent decades, resulting from the increased complexity of environmental issues and decreases in environmental program funding (Au *et al*., 2000; Conrad & Daoust, 2007). With environmental monitoring playing an important role in sustainable development, through monitoring activities, community groups have been attempting to fill the gaps caused by these government budget cuts (Sharpe & Conrad, 2006).

There are many advantages of citizen science as a source for gathering data, including an increase in environmental democracy, scientific literacy, and public participation with local issues; while also advancing scientific knowledge (Bonney *et al.*, 2009; Conrad & Hilchey, 2010; Gillett *et al.*, 2011). As van Horen (2001) noted, since most residents have knowledge of their local area; they are well suited to conduct environmental monitoring, provided that they are properly trained and equipped. Citizen science programs have been documented and observed for community groups, non-governmental organizations (NGOs), and government agencies. As volunteer-based groups contribute to ecosystem monitoring by collecting data at a reduced cost, these programs also address the spatial and temporal gaps from academic and government run programs (Conrad & Daoust, 2008; Gillett *et al.*, 2011; Kremen, Ullman, & Thorp, 2011; Schmeller *et al.*, 2009).

Participation in citizen science involves a variety of topics ranging from climate change to water quality monitoring (Silvertown, 2009). Studies have shown citizen science as a form of data gathering have included fish populations (Leopold *et al*., 2009), benthic invertebrates (Gillett *et al*., 2010; Hart *et al*., 2001), water quality and water usage (Roa Garcia & Brown, 2009), tracking the effects of climate change (Beaubien & Hamann, 2011; Hurlbert, & Zhongfei, 2012), biodiversity monitoring (Schmeller *et al.*, 2009) and community-based forest management (Tole, 2010). The use of citizen monitoring for the purposes of information was noted by Bonney *et al*. (2009) and Hochachka *et al*. (2012) to span across many locations, habitats, and time. Citizen science projects also provide the participants with an increased knowledge of the scientific investigation process and the subject in which they are studying. This form of monitoring can include large-scale projects spanning continents and global data-gathering networks (Pollock & Whitelaw, 2005; Bonney *et al*., 2009). In particular, with respect to natural history observations, “laypersons” could possibly document distribution and species abundance where otherwise monitoring data does not exist (Breed *et al.*, 2012). The documentation of this data can be essential for tracking ecological trends world-wide, which would otherwise be difficult due to incomplete data sets or monitoring activities by professional scientists and government agencies (Conrad & Hilchey, 2010). Much of the data being collected is directed towards the position of environmental “watchdog”, problem identification, education and background information for decision-making. Specific tasks have also been directed at habitat remediation; changes to government protection; and regulation and policy (Savan *et al.*, 2003). Due to their personal interest in their local environment, volunteers are often ideal candidates for stream monitoring and identifying ecological changes (Fore *et al.*, 2001).

There has also been governmental support for such programs, an example being the USEPA. They provide conferences, manuals and other resources directed at water quality programs (Savan *et al.*, 2003), and many U.S. states include volunteer-collected water chemistry data in their biennial reports for the USEPA (Fore *et al.*, 2001). At the Canadian level, the Community Aquatic Monitoring Program (CAMP) demonstrates a government-developed monitoring program that utilizes an outreach program to interact with community groups and provides standardized sampling methodology and related protocols (Weldon, Courtenay, & Garbary, 2007).

Unfortunately, there are also limiting factors to CBM, such as limited funding and being dependent on only a few motivated individuals (Savan *et al.*, 2003). This impacts the equipment choices for groups, and in turn can affect the accuracy of the data (Nicholson, Ryan, & Hodgkins, 2002). Au *et al.* (2000) discussed the concerns of the validity of the water quality data, as official sampling protocols are very specific and the possibility of uncertainties can arise when these protocols are not followed precisely. Efforts to address these concerns include collaboration with academia and government agencies, as well the provision of available resources and protocols to ensure high standards of data collection. Some non-profit organizations, such as the River Watch

Network (RWN), provide dedicated training programs in water quality collection techniques aimed at citizen scientists (Savan *et al.*, 2003).

There are various resources available which can ensure high data quality to accompany a developing policy framework for CBM data. Environment Canada's Ecological Monitoring and Assessment Network (EMAN) was the first to introduce standardized benthic invertebrate monitoring protocols in Atlantic Canada (Sharpe & Conrad, 2006). EMAN and Nature Watch Programs were government initiatives led to coordinate standardized methods for data collection; management; and distribution of information (Whitelaw *et al.*, 2003). Although EMAN is no longer active, their protocols are still available electronically and provide a foundation for future resources and networks. The Canadian Aquatic Biomonitoring Network (CABIN) is another example of a resource made available through Environment Canada, which introduced high scientific standards and vigorous protocols. CABIN was made accessible to community watershed groups to monitor the abundance and diversity of benthic invertebrate communities and provided the training necessary to meet the level of accuracy needed. The data collected was held in a central database, thereby allowing the comparison of the results from different watersheds (Sharpe & Conrad, 2006). Available data reservoirs and availability of resources are very important for the development of citizen based initiatives and can also benefit from resource sharing with academia.

The number of partnerships between universities and communities monitoring the environment has seen an increase within recent years. The Citizens' Environment Watch (CEW) group, re-branded as EcoSpark in 2010, was based out of the University of Toronto and provided an excellent example of a group that collected surface water quality and collaborated with CBM programs spanning from 1997 (Savan *et al.*, 2003). The collaboration with universities provided support through the development of monitoring protocols; lab and data analysis; quality assurance quality control (QA/QC) procedures; training sessions; available office and lab space; student support; leadership; funding opportunities; and presents credibility and infrastructure to aid in establishing partnerships with governments, non-governmental organizations and scientific bodies (Savan *et al.*, 2003). By addressing concerns of available resources, scientific protocols and collaboration with government and academia for CBM programs, the factors influencing accuracy of fresh surface water quality-monitoring programs must also be examined.

It has been acknowledged that with modest training, useful observational data can be collected by citizen scientists (Kremen *et al.*, 2011), although defining this level of training has not been clearly identified. Literature has shown that volunteers with training can produce data that is comparable to professionals for a variety of parameters and habitats including: beach microbiology; subtropical reef fauna; birds; and freshwater macroinvertebrates (Gillett *et al.*, 2011; Nicholson *et al*., 2002). By observing previous studies, the benefits and challenges of such programs can be evaluated to determine where the data collection credibility issues originate, and what has yet to be researched in the field of the accuracy of the water quality data.

1.3 Water Quality Monitoring

The monitoring of water quality provides a range of information from ecosystem health to valuable information on the effectiveness of environmental restoration projects (Palmer, Allan, Meyer, & Bernhardt, 2007). There is a range of chemical, biological and physical parameters, which help in understanding ecosystem health, habitat potential for species and tracking environmental trends. These parameters are also tied to a range of variability and uncertainties depending on the equipment used. In this section, a review of the water quality parameters, their significance, and their potential for use in citizen science based monitoring programs will be explored.

1.3.1 Water Quality Parameters

There are a variety of water quality parameters collected in water monitoring activities to assess non-drinking freshwater. These include: biological (e.g. fecal coliform bacteria and benthic invertebrates), physical (e.g. temperature or total suspended solids), and chemical (e.g. dissolved oxygen and pH) (Sharpe & Conrad, 2006). Each provides critical information and is chosen based on the goals and objectives of a monitoring project. As noted by Savan *et al.* (2003), chemical parameters can provide a “snapshot” of the water quality of a sample site, whereas biological indicators are useful in providing cumulative assessment of the environmental quality. Basic parameters that were recommended by Nicholson *et al.* (2002) and Wenner, Sanger, Arendt, Holland, and Chen (2004) include: pH, temperature, dissolved oxygen, salinity, turbidity and water level. These parameters are easy to measure cheaply and can be measured quickly by non-scientists (Nicholson *et al.*, 2002). Comparatively, parameters such as dissolved oxygen (DO), temperature, and pH are critical indicators of freshwater ecosystem health and often influence the habitat quality for fish species (PASCO, 2007).

In determining the selection of parameters for monitoring programs, the most appropriate river health indicators for volunteers are cost effective, quick and easy to use

(Nicholson *et al*., 2002). As it is not feasible to measure all environmental variables, utilizing a few basic water quality parameters is the most cost effective method for gathering data over a continuous period. These parameters provide indicators of environmental stress and the quality of the habitat (Wenner *et al.*, 2004). Parameters used for measurements such as water chemistry and ecology are often chosen in correspondence with government regulations or agreements with local polluters. Chemical parameters can be compared directly to Canadian standards for surface water quality; unfortunately, there are no governmental guidelines or regulating standards for biological indicators (Savan *et al.*, 2003).

Volunteer based groups that monitor lakes and streams, such as the previously discussed Citizens' Environmental Watch, used water chemical parameters such as pH; temperature; turbidity; and ammonia and phosphate levels/concentrations, as they are relatively inexpensive, simple, and could be compared with historical data collected by government programs. Chemical parameters can provide details on the sources and transformation of pollutants, while biological parameters demonstrate the nature and value of biodiversity and ecosystem health. Biological values can also describe the impact of non-point source pollution and are simple and reliable tools for volunteers to assess river and lake health (Savan *et al*., 2003).

In order to increase data accuracy, some modifications to citizen science programs are necessary, as some parameters may require complex field sampling methods that may not be appropriate for volunteers. For example, Bonney *et al.* (2009) noted that citizen science data was more suited for determining the relative species abundance in such programs rather than estimations of absolute abundance. According to Fore *et al.* (2001), with proper methodology and training, river assessments conducted by volunteers can be comparable to professionals.

1.3.2 Sources of Error and Variability of Water Quality Data

To evaluate the degree of difference in the accuracy of water quality data collection, consideration must be made for the variability of the parameters and the equipment. The temporal and spatial variability of a sampling program and representative sampling of the ambient environment will have a great impact on the water quality measurements (i.e. when and where to sample). While comparing water samplings collected by two groups, the frequency and timing at which one samples will influence the accuracy of the comparison. Water chemistry is greatly impacted by daily and seasonal weather patterns (Wenner *et al*., 2004), as air temperature and rainfall are considered "external driving factors" influencing water quality (Buzzelli *et al.*, 2009). The influence of flowing freshwater can also affect fluctuations in salinity, temperature, turbidity, sediment, DO and nutrients (Xu and Wu, 2006). Cross-sectional variability for surface water quality parameters have also been evaluated in a study by Marron & Blanchard (1995), assessing urbanized streams in the Illinois River Basin and identified temperature, specific conductance and pH as parameters with low variability, with coefficients of variations not exceeding 6%. While DO displayed higher cross-section variability, with coefficients of variations as high as 19%.

Different parameters will also respond to and be influenced by varying sources. Water quality parameters such as conductivity, dissolved solids, coliform, nitrates and phosphates can demonstrate a variability as a result of land use changes; comparatively, temperature, pH, dissolved oxygen, turbidity, calcium and total hardness often reflect the impacts of topography and natural geology (Roa Garcia & Brown, 2009). Variability in salinity measurements reported to be a result of anthropogenic factors on a spatial scale, demonstrating watershed level variability due to flow regime changes. While salinity readings have been shown to fluctuate in response to runoff events, daily temperature fluctuations of 10 °C were noted in one case study (Wenner *et al*., 2004). There is also a direct relationship regarding increase of salinity reducing the saturation potential of DO; while DO can also be influenced by oxygen demanding wastewater discharges, reduced water flow, water temperature and excessive plant growth (Wilding, Brown, & Collier, 2012) which includes the production or consumption of DO by aquatic plants, chemical reactions, biological processes (Marron & Blanchard, 1995). Further water quality variability can be introduced by ground-water seepage into streams and point sources of effluent (Marron & Blanchard, 1995).

Choosing the most applicable parameters are important for volunteer-based monitoring programs, as restrictions such as cost of equipment, time, laboratory analysis and complexity of the data collection methods are important factors to consider. Laboratory-based parameters, for example, can be effective for volunteer-based monitoring programs. These can include water nutrient measurements that require laboratory analysis such as total phosphorus, a common scientific monitoring parameter measured by professionals and a useful parameter to monitor as its presence in high levels can indicate a point source of pollution (PASCO, 2007). Laboratory-based parameters however may not be a feasible option for all community groups due to laboratory costs and a higher complexity for volunteers compared to basic field measurements. A study evaluated by Nicholson *et al.* (2002), examined total phosphorus as a suitable parameter for volunteer-based program, and results indicated overestimations of total phosphorus concentrations by volunteers and potential chemical contamination.

In developing countries, physical and chemical methods are often the most favoured approaches to assessing water quality due to the low cost associated with field measurements. Hart *et al*. (2001) noted the value of assessing river health utilizing water-quality meters for underdeveloped areas. As previously discussed, by measuring chemical parameters a glimpse of the water quality in an area can be taken to determine the freshwater ecosystem health (Savan *et al.*, 2003), and some of the basic chemical and physical parameters such as pH, temperature and dissolved oxygen are critical indicators of the habitat quality for fish species (PASCO, 2007). While many portable in-situ meters can measure these parameters, the cost for equipment and training is small in contrast to the costs of setting up chemical analysis laboratory. The portable in-situ meters are often used for community-based monitoring (CBM) which provide real-time measurements. This type of equipment was used in the youth citizens water quality monitoring testing study discussed by Roa Garcia and Brown (2009) where turbidity, pH, temperature, conductivity, total dissolved solids and dissolved oxygen were measured.

Other forms of biological indicators such as coliform, can also vary, although they are still useful and accurate in establishing patterns of contamination. Au *et al.* (2000) demonstrated that even with simplified methodology, CBM data collection of total coliforms levels used to establish patterns of contamination and determine sources of pollution were comparable to modern monitoring methods used by professionals. Citizen science monitoring programs often produce large datasets that are well suited for easy interpretation of patterns in data and criteria can be created to aid in identifying systematic errors. While errors and biases can occur in CBM programs, including misidentification and misinterpretations of protocols, data identified with these errors can be omitted from analysis while still maintaining the goals of a research project (Bonney *et al.*, 2009).

Data confidence protocols, equipment, and data analysis can increase the level of accuracy in data collection (Nicholson *et al*., 2002). Water quality parameters including turbidity, conductivity, and pH were evaluated in a comparison study of the volunteer-based Waterwatch program and professionally collected water quality data and were chosen as they are the most commonly measured parameters used in the Waterwatch program. This study identified variables that need to be addressed in order to increase the confidence limits of community-collected water quality field-based measurement. This study had results varying temporally and spatially, with conductivity and pH demonstrating very similar values to professionally collected data. The same data collection protocols were followed for both groups when similar equipment was used, although there was some difference in equipment such as turbidity meters and turbidity tubes. Turbidity was determined as the parameter that displayed the greatest inaccuracies and the equipment used to monitor turbidity and total phosphorus appeared to have ranges of accuracy from limited to moderate (Nicholson *et al*., 2002).

Additional data inaccuracies can result from a lack of standardized methods, QA/QC procedures and participant objectivity (Pollock & Whitelaw, 2005). As noted by Zabiegala, Kot-Wasik, Urbanowicz, and Namiesnik (2010), the act of sampling alone introduces uncertainties into the measurement and sources of error can appear prior to field sampling, resulting from poor sampling design. By identifying all of the variables for data inaccuracies, an experimental monitoring program can attempt to address these concerns and decrease the level of uncertainty while increasing the accuracy.

1.4 Comparative Studies: Volunteer vs. Professional

Despite studies that have demonstrated decades of successful community monitoring programs, such as those described by Bonney *et al.* (2009), there are still questions relating to how volunteer data might vary from professionally collected measurements. The comparison of volunteer and professionally collected monitoring data has been researched over a variety of fields of study, from California's rocky reef kelp forests to the Waterwatch Victoria water quality program (Gillett *et al*., 2011; Nicholson *et al*., 2002). Unfortunately, there have been a limited number of field studies in water quality comparison research with different level of complexity in the volunteer sampling methodology used. Biological indicators such as benthic invertebrate sampling have also been examined (Gillett *et al*., 2010; Hart *et al*., 2001) and are able to provide some indication of the degree of difference in a comparison study, where the differences can appear, and how to model a project for the purposes of surface freshwater sampling.

As methodology was highlighted as one source of potential error in volunteer sampling, standardizing simplified field methods for a comparison study could minimize this error (Au *et al.*, 2000). Fore *et al.* (2001) evaluated the performances of volunteers through the collection benthic macroinvertebrates field samples while exercising professional protocols. This study demonstrated no significant difference between field samples collected by volunteers and professionals, while utilizing identical field methodology, standardized sampling protocols and equipment. This study indicates that trained volunteers can collect reliable biological data aiding in stream assessments, although overall this study required the sampling design and data interpretation of the data to be done by professional biologists (Fore *et al.*, 2001). The lesson of training and guidance had been highlighted by Arvanitidis *et al*. (2011), noting that guidance to citizen scientists through the all stages of the monitoring program was instrumental for the success of the project.

The use of simplified methodology for citizen scientists has been mirrored in other comparison studies. In a study by Au *et al.* (2000), Canadian high school students with brief environmental monitoring training acted as volunteers, to determine if the methods used by "environmental authorities" and the simplified methodology by the volunteers would result in similar information. This program involved the collection of water samples to identify the concentration of total coliform, and toxicity measuring dissolved oxygen, phosphate, ammonium ions, pH, dissolved oxygen and hardness in a lab setting. This study demonstrated that simplified methodologies provide comparable patterns to the modern accepted monitoring methods and concluded that with this form of monitoring, CBM can reliably alert environmental authorities to sources of contamination. In another comparison study, the Reef Check California Association (RCCA) program monitored biological and physical parameters, such as fish, and benthic invertebrates. In order to simplify the sampling process of the RCCA to make the program accessible to volunteers and to increase the precision of the data collected, several modifications to the sampling locations and recorded taxa were done to the RCCA program compared to the professional scientist program (Southern California Bight program). The results displayed varying degrees of difference between the two programs, with procedural differences believed to be the cause for the varying results of physical habitat data. Sampling design was highlighted as a source for procedural differences, with extrapolation procedures and spatial scale varying between the two programs (Gillett *et al.*, 2011). By evaluating these results, one could suggest that through a standardization of procedural methodology, such as sampling locations, the degrees of difference between the data collected could be lowered, although the complexity of the bias is dependent on the parameters being used. This form of simplified and standardized procedural methodology has yet to be performed on a water quality comparison study evaluating basic freshwater parameters.

Literature has shown that the level of complexity chosen for volunteer monitoring programs, such as field methodology and equipment choice, can be another source of error; however, this has not been explored fully for water quality programs. Gillett *et al*. (2011) noted observer error as a source of the difference between the volunteers' and professionals' data, including that trained volunteers with less experience were less accurate with taxonomic identification compared to trained professionals. Research has identified that volunteer programs that are the most effective have the guidance of experts, and laboratory analysis by professional taxonomists is preferred over volunteers. Nevertheless, it was recognized that the number of professional taxonomists are diminishing and therefore the role of citizen scientists in the successful data collection process will be vital (Arvanitidis *et al*., 2011). It is however more suitable for certain procedures to be left to professionals, and according to Fore *et al.* (2001), the most appropriate use of volunteer participation in biological monitoring is in field collection and laboratory analysis.

In a review by Savan *et al.* (2003) of Citizens' Environment Watch (CEW) data sets collected by volunteers in 1997, it was determined that the data quality varied. With the introduction and involvement of scientific advisors, extensive quality assurance and quality control measures were developed which included rigorous protocols, reagent preparation, distribution and the use of blank and standard sample testing for volunteer monitoring (Savan *et al*., 2003). Overall a key element to the maintenance of high data quality standards is based on appropriate management procedures including quality assurance, quality control procedures (Hochachka *et al*., 2012). Roa Garcia and Brown (2009) noted a similar approach in a youth participatory water quality program, with field monitoring, sample collection and laboratory analysis being conducted by volunteers under the supervision of an environmental chemist, although this program did not compare the accuracy of the youth volunteer samples.

A gap in comparison research of volunteer water quality accuracy was emphasized as previous research conducted utilized historical datasets and did not incorporate standardization of monitoring equipment which could have resulted in discrepancies in data. Introducing a lack of standardization in equipment can also introduce another variable into a study. A source of difference in the Reef Check California Association case study was concluded to be a result of the overall study design, as this study used a post-hoc method, with data not being collected simultaneously from the same reef, leading to small-scale spatial and temporal differences (Gillett *et al.*, 2011). Nicholson *et al.* (2002) also discussed comparative research where the same data collection protocols were used for both groups only when similar equipment was used. The turbidity measurements collected with different equipment also displayed the greatest inaccuracies. Although this study discussed by Nicholson *et al.* (2002) also utilized post-hoc methodology and therefore the concern of equipment could not be addressed, using synoptically collected data for future research can address this potential source of error by using identical equipment and this knowledge gap.

1.5 Knowledge Gaps

Literature indicates that, when properly trained, citizen volunteers can collect reliable biological invertebrate data and make stream assessments that are comparable to professionals (Fore *et al.*, 2001). Through an evaluation of the literature available for water quality monitoring and comparative studies on citizen scientists and professionals, there is a notable gap in the research examination. While biological indicators have been assessed, as well as some chemical water quality parameters, a few important variables still need to be addressed: (1) comparative study analysis included historical datasets, often with a minimum of 30 days between collection of the professional and volunteer samples; (2) need for water quality monitoring equipment standardization; and (3) need of simplified methodology and basic fresh-water chemical parameters.

In comparative studies, data was used that included historical records, which leads to a question of water quality variability. For example, parameters used in comparative studies indicated that turbidity and total phosphorus had the greatest variability after rain events when there is increased run-off (Nicholson *et al.*, 2002). When comparing surface water quality field measurements, the temporal scale is very important component of sampling program design as factors such as: rain events, changes in water level, water flow and time of day will greatly influence your results. By conducting side-by-side field measurements with water quality professionals and community volunteers, eliminating temporal and spatial scale errors, a more thorough comparative analysis can be done to determine the accuracy of volunteer data.

As Cohn (2008) observed, balance is needed between collecting data of high reliability and the goal of public education. Pairing citizen scientists with trained staff can help to compare data and determine accuracy for volunteer-based monitoring programs (Cohn, 2008). The scientific question must take into account the level of experience that researchers will have. Projects that require high skill level must include participant training and supporting material (Bonney *et al.*, 2009). This study incorporated this recommendation by providing baseline training and detailed supporting material in the project design. Further details on training will be described in Chapter 2.

To determine what program characteristics are necessary for volunteer based projects to be most effective, an evaluation of the accuracy of basic freshwater parameters are necessary, including pH, temperature, dissolved oxygen and conductivity (Nicholson *et al.*, 2002; Wenner *et al.*, 2004). This will identify a higher level of credibility at a competency level in which we expect a volunteer program to collect. As Gillett *et al.* (2011) suggested trained volunteers have the ability to receive the appropriate skills to produce similar data from professionals, keeping in mind sufficient guidance, supervision and rigorous sampling plan. By properly selecting the appropriate levels of volunteer involvement in environmental monitoring programs, sources for error can be reduced.

Another key component of volunteer data integration is acceptance. Hart *et al*. (2001) noted that acceptance is needed from the scientific community regarding technical methodology and from government officials of method and outputs. Branching from the literature, the larger context of this current research would be the goal of citizen science data integration to supplement Canadian government programs with a future model at a national scale including standardized methodology. Monitoring programs such as the water quality monitoring network (WQMN) program highlighted the need for methods, procedures, and equipment standardization with all involved partners to produce reliable and comparable data (Sigua & Tweedale, 2004). By standardizing the equipment, datasets can be used as a whole, which critical for an appropriate interpretation of results. To address the first level of acceptance addressed by Hart *et al*. (2001) relating to technical methodology, the use of standardized field methodology and the incorporation of national and regional standards of data collection will be adopted in this study.

Lastly, an assumption exists that professionally collected data is more accurate, although more research is required to prove the validity of this statement. Preliminary research describes discrepancies in professional analysis (Nicholson *et al.*, 2002). This study attempted to address the how comparable is citizen science data to that gathered by professionals, although future research may also need to address the accuracy of all water quality data collection.

Chapter 2

Methods

2.1 Study Design

This study was designed to address sources of error that could arise in comparative field water quality measurements to determine the margin of error between a volunteer citizen scientist and professional scientist. The level of accuracy that was used to compare the difference between measurements of the control and treatment groups included the sensor specifications of the equipment and calibration standard error, which represented the mechanical variability. Various levels of accuracy standards used by government agencies were also compared.

In this study the control group was the professional scientist, with an assumption that the professional data was the “true value”. The professional participant from the control group was selected to meet this study’s definition of a professional, “as an individual receiving payment for their services from a government agency, educational institution or private consulting firm” and also was an accredited CET certified professional. The participants in the treatment group were volunteer citizen scientists.

The study design addressed sources of variability by controlling the following variables: (1) spatial differences in the sampling location with respect to micro-scale variances in water chemistry, (2) date and time of sampling to address temporal variability of data collection, (3) age and type of equipment used, and (4) baseline level of training of each participant in the treatment group. The sources of variability expected in this study that may account for differences in measurements were: calibration procedures and calibration error, stabilization time of sensors and the sensor deployment of the treatment group.

The YSI Professional Plus (YSI ProPlus) was used for field measurements in this study. Data accuracy guidelines were employed to compare the differences between field measurements taken by volunteers and those taken by professionals, based on the sensor drift and calibration error (Table 2.0). When determining the potential equipment error in a comparison of water quality measurements, the accuracy value for each sensor and buffer calibration standard were needed to establish the equipment and calibration errors prior to evaluating the human errors made by the operators. For example, the pH sensor accuracy of the YSI ProPlus was +/- 0.2 pH units, thus in the comparison of two measurements from two sensors, the accuracy dropped to +/- 0.4 units. As the pH calibration standards are +/- 0.01 pH units, the pH calibration error for each sensor is +/- 0.02 units for a two-point pH calibration for one sensor. The overall equipment and calibration error became +/- 0.44 pH units for two sensors calibrating a two-point pH calibration (D. Parent, personal communication, November 16, 2012).

An examination of guidelines used for data accuracy in government fresh water monitoring programs identified where volunteer monitoring data could potentially be integrated within various levels of government (Table 2.1). The Science division, Water Resources Management Unit of Nova Scotia Environment and Environment Canada maintain strict QA/QC protocols and data rejection/correction criteria for the collection of fresh surface water. Provincial and federal water quality data verification for continuous water sensors maintains United States Geological Survey (USGS) protocols, including methodology for determining the accuracy of water sensors. Nova Scotia Environment's

(NSE) data rejection criteria has been adapted from USGS standards (Nova Scotia Environment, 2010a), however Environment Canada's allowable accuracy range for a continuous water quality sensors are taken directly from USGS standards (Wagner, Boulger, Oblinger, & Smith, 2006). The report from the Government of Newfoundland & Labrador (GNL) Department of Environment and Conservation (2012) was also examined, as these standards were adjusted specifically to reflect the conditions found in Atlantic Canada.

Table 2.0. Professional Plus System Cable and Sensor Specifications (Adapted from YSI, 2011)

Parameter	Calibration Error	Cable Accuracy
Dissolved Oxygen (%) (T: -5 to 45°C)	N/A	0 to 200% (± 2% air saturation*)
Dissolved Oxygen (mg L^{-1}) (T: -5 to 45°C)	N/A	0 to 20 mg L^{-1} (± 2% of reading or 0.2 mg L^{-1}*)
Temperature (Field rugged cables)	N/A	± 0.2 °C
Conductivity	N/A	± 0.5% of reading or 0.001 mS cm^{-1}*(1-, 4-m cable)
pH	+/- 0.01	± 0.2

*Whichever is greater

Consideration among standards for field measurement should also address acceptable accuracy limits between two operators who are handling the equipment; however, no specific guidelines were found for comparing two operators. The understanding of the uncertainty in measured discharge or streamflow and water quality data has not been well established (Harmel, Cooper, Slade, Haney, & Arnold, 2006). This uncertainty is directly linked to the understanding of the variability resulting from the operator of the measurements rather than equipment error.

Table 2.1. Data correction criteria and maximum allowable limits for water quality monitoring sensors values

Parameter	Nova Scotia Environment (NSE) & United States Geological Survey (USGS)	Environment Canada	GNL Department of Environment and Conservation
Temperature (°C)	± 2	± 0.2	± 0.2
Specific Conductivity (µS cm^{-1})	± 50 (or 30%)	± 5 (or 3%)	± 3
pH	± 2	± 0.2	± 0.2
Dissolved Oxygen (mg L^{-1})	± 2 (or 20%)	± 0.3	± 0.3 (or 0.3%)

Note: Data Rejection Criteria: Adapted from Government of Newfoundland & Labrador Department of Environment and Conservation, 2012; Nova Scotia Environment, 2010a; Wagner *et al.*, 2006.

The sample size for this study was determined using preliminary experimental data of one sample set. The Shapiro-Wilk paired t-test was used to determine the statistical significance with the study design using 18 participants in the volunteer treatment group and one professional scientist in the control group. Each volunteer collected ten sets of water quality measurements side-by-side with the professional from the control group and both groups collected one set of channel measurements for each sample site.

The treatment group and the control group, in total, collected 360 water measurements for each parameter. Each of the 18 treatment group participants collected 10 sets of water quality data (water temperature (°C), dissolved oxygen (%), dissolved oxygen (mg/L), pH, conductivity (uS/cm), specific conductivity (uS/cm) and total dissolved solids (mg/L)) for their sample site location as the control group participant collected side-by-side data. Channel measurements were also collected for each sample location, which included a total of 36 discharge (conditions at time of sampling) and bankfull discharge measurements (high flow conditions). A supplementary survey was also distributed to the treatment group for a qualitative assessment on each participant's levels of previous training, education, and years of monitoring experience to address research question #2. A total of 15 out of 18 participants responded.

The results of this study were expected to provide an overall qualitative and quantitative assessment of the difference between a volunteer citizen scientist and a professional, by removing temporal and spatial sources of variability. This study sought to determine under what conditions a citizen scientist's data collection accuracy can be improved.

2.2 Study Area

The study area of this project extended to twelve watersheds in Nova Scotia, Canada (Table 2.2, Figure 2.0); with field sampling taking place from May 31, 2012 to August 28, 2012. Nova Scotia is a province of diverse fresh water systems and with over five percent of the province's land covered by fresh water. This environment provides habitat for many species of fish, insects and vegetation (NS Museum of Natural History, 1996). There are 46 primary watersheds throughout the province (Nova Scotia

Environment, 2010); however, Nova Scotia Environment developed the provincial watershed boundaries used in Figure 2.0 with secondary watershed boundaries for a more accurate representation of the provincial watershed system (Guan, Sterling, Garroway, & Kennedy, 2013).

The site selection for this study was restricted to small streams to facilitate sampling. The study sites had a range of environmental stream characteristics including: (1) bed topography (ripple, pool and straight runs), (2) channel reaches (meandering and straight) and (3) stream profile (ungraded, graded and aggrading) (Charlton, 2008, pp. 131-132). Streams in Nova Scotia are also characteristically ungraded, along their length, with hydrologic features such as rapids and waterfalls common among these streams (Spooner, Fenton, & Myers, 1998).

The site characteristics of each sampling location were broken down into primary watersheds, provincial districts, upstream dominant land use, site location and surrounding river or stream name (Table 2.2). As described by the Ontario Ministry of Environment (2010), land-use activities which can impact water quality may include: application of nutrients to agricultural land; salting of roads; waste from local industries and sewage treatment plants; mining (e.g. metal); and urban development. For the purposes of characterizing these potential impacts as a result of land-use dominant land use practices for the sample sites were categorized by the following: agricultural practices (small and large scale), mining activities, urban development, industrial development (e.g. logging and pulp mill) and residential areas or recreational practices (rural or urban).

Table 2.2. Sample site characteristics

Site Code	Location	GPS coordinates	Primary Watershed	Provincial District	Dominant Land Use
1	Bennery Brook/ Shubenacadie R.	20T 0457732 4975159	Shubenacadie	Colchester-Musquodoboit Valley	Urban Development
2	Mossman/ Northfield	20T 0375664 4924153	Lahave	Lunenburg	Industrial (lumber mill downstream)
3	Mushamush R.	20T 0390283 4923660	Gold	Lunenburg	Rural Residential
4	Medway R.	20T 0362756 4895447	Herringcove	Queens	Rural Recreational/Residential, Industrial (lumber mill)
5	Medway R.	20T 0362744 4895448	Herringcove	Queens	Rural Recreational/Residential, Industrial (lumber mill)
6	Medway R.	20T 0362749 4895449	Herringcove	Queens	Rural Recreational/Residential, Industrial (lumber mill)
7	St. Mary's R.	20T 0574913 5025858	St Mary's	Guysborough-Sheet Harbour	Rural Recreational/Residential, Agriculture (small scale)
8	West Lochaber	20T 0575660 5029606	South	Antigonish	Rural Recreational/Residential
9	West Lochaber	20T 0575660 5029606	South	Antigonish	Rural Recreational/Residential
10	West Brook/ St. Croix R.	20T 0411498 4985772	St Croix	Hants West	Rural Residential, Agriculture (small scale)
11	Annapolis R.	20T 0342136 4979729	Annapolis	Annapolis	Rural Residential
12	Little Sackville R.	20T 0445538 4958093	Sackville	Sackville-Cobequid	Urban Residential/ Recreational (baseball field)

Table 2.2 (continued)

Site Code	Location	GPS coordinates	Primary Watershed	Provincial District	Dominant Land Use
13	West River Pictou	20T 0509399 5044897	Pictou	Pictou West	Rural Residential
14	West River Pictou	20T 0509622 5043368	Pictou	Pictou West	Rural Recreational (provincial park)
15	West River Pictou	20T 0509632 5043345	Pictou	Pictou West	Rural Recreational (provincial park)
16	Yarmouth/ Tusket R.	20T 0272259 4888563	Tusket	Argyle	Rural Residential, Agriculture (small scale)
17	Yarmouth/ Tusket R.	20T 0272259 4888563	Tusket	Argyle	Rural Residential, Agriculture (small scale)
18	Musquodoboit/ Fish R. - L. Charlotte	20T 0509033 4961781	Tangier	Eastern Shore	Rural Recreational/ Residential

Note:
(1) GPS coordinates and Dominant Land Use data were recorded at each sample site

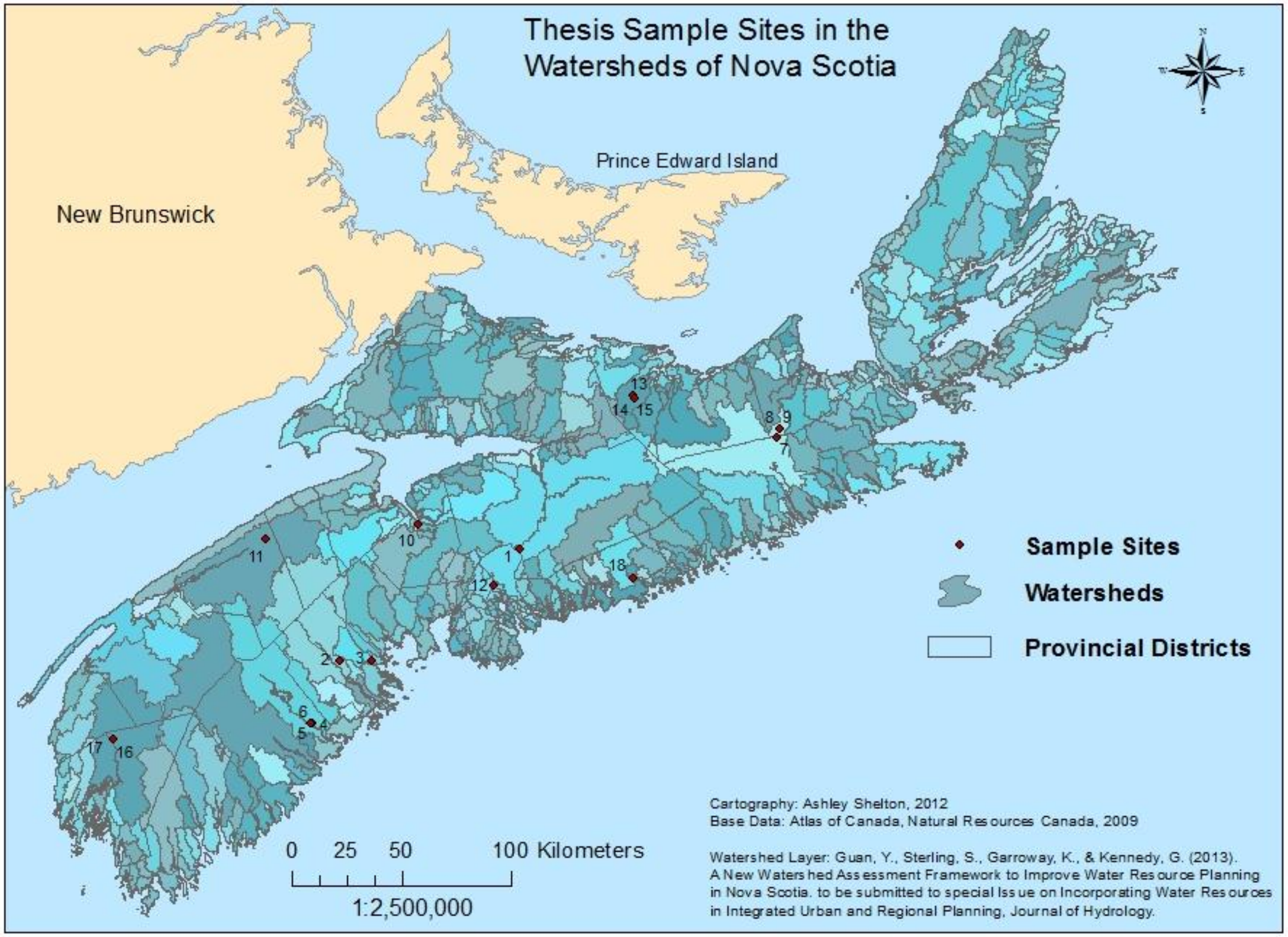


Figure 2.0. Distribution of sample sites within watershed boundaries and provincial districts. Adapted from Guan *et al.*, 2013

2.1.1 Water Quality of Study Area

There are many important aspects of water quality that impact stream ecosystem health including: acidity of water, electrical conductance, water temperature and dissolved oxygen content (Table 2.3: Appendix A). These surface water quality parameters can be influenced by many natural factors including: bedrock composition, watershed size, precipitation, land topography, vegetation and proximity to the ocean. The subsurface geology and soil characteristics of a local watershed also play an important role in determining the natural conditions of a region's water quality (NS Museum of Natural History, 1996).

For a baseline of water chemistry data, an inventory for the province's lakes is available through the Nova Scotia Lake Survey Program (Nova Scotia Environment, 2013a), while automated water quality data from December 2005 to 2008 can be found for five of the province's rivers including Shelburne River, North East Margaree River, Kelley River, St. Marys River and Lahave River (Nova Scotia Environment, 2013b), with sample site code #7 located in the same locations as the NSE St. Marys River station. According to 2012 NSE data (2013), expected water quality parameter ranges for sample site code #7 include pH seasonal range varying from 6.0 to 6.9; while conductivity (umho/cm) ranged 25.2 to 35.1. In Nova Scotia there are varying expected ranges for these water quality parameters. According to the NS Museum of Natural History (1996) the mean conductivity for Nova Scotia was reported to be 69.5 micromhos/cm (mmho/cm = level of dissolved solids), however in Lunenburg County, the location for sample sites 2 and 3, mean conductivity was 26.4 mmho/cm. The range of natural pH values also varies notably across the province, depending on the various factors including: surrounding geology, acidic precipitation and wastewater runoff and drainage from surround coniferous forests that lead to lower pH in streams (CABIN, 2010).

Nova Scotia surface waters are particularly sensitive to acid deposition (Underwood, Ogden, Kerekes, & Vanghan, 1985) and according to Figure T8.2.6 from the NS Museum of Natural History (1996), mean annual pH values across Southwest to Southeast Nova Scotia range from less than 4.7 to areas greater than 5.4, with study sample site codes #16 and #17 located in areas of particular risk of more acidic surface waters. Literature also indicated strong mineral acid concentrations in southwestern Nova Scotia's water systems, leading to further acidification where there were previously existing natural acidic conditions (Howell & El-Shaarawi, 1990).

2.1.2 Geology of Study Area

Geology and surrounding soil type has impacts on various aspect of water quality, including water chemistry and nutrient level, as the acidity of a freshwater environment may be a result of the buffering capacity of the soil or the local bedrock. For example, areas consisting of limestone, which contains calcium carbonate, are more capable of moderating the acidity of precipitation due to the interaction with magnesium and calcium carbonate (NS Museum of Natural History, 1996).

In large areas of Nova Scotia, granite and shale bedrock contain little buffering capacity. An estimated 78 percent of lakes and streams in Nova Scotia are located in areas with underlying granite and metamorphic bedrock, resulting in low conductivity values (Dennis, Scruton, Gilliss, & Clair, 2007; Underwood *et al.*, 1985). Therefore, implying a low concentration of dissolved solids, and leaving it subject to low pH values and vulnerable to low acid deposition conditions (Dennis *et al*., 2007; NS Museum of Natural

History, 1996; Underwood *et al.*, 1985). One of these regions include the Wolfville Formation bedrock, where Spooner *et al*. (1998) noted the geologic formation coincides with high conductivity and strong buffering capacity of base flow and the Annapolis River sample site lies above this formation. Base flow refers to the portion of the stream flow that originates from groundwater discharge (Tallaksen, 1994). In other regions of the province, higher conductivity values can be found in areas underlain with sedimentary rocks consisting of limestone and gypsum (NS Museum of Natural History, 1996). Figure 2.1 displays the acidification index for the province in relation to the location of thesis sample sites (Guan *et al*., 2013). The acidification index is based on the acid neutralization capacity (ANC) of the surface waters in the province, which was determined by Dennis *et al.* (2007) through the gran alkalinity titration method. This map displays sample sites codes 4,5,6,16 and 17 lie within the highest risk zones for ANC_G values ranging from -1.362030 to 0.816787 with the smallest acid neutralization capacity.

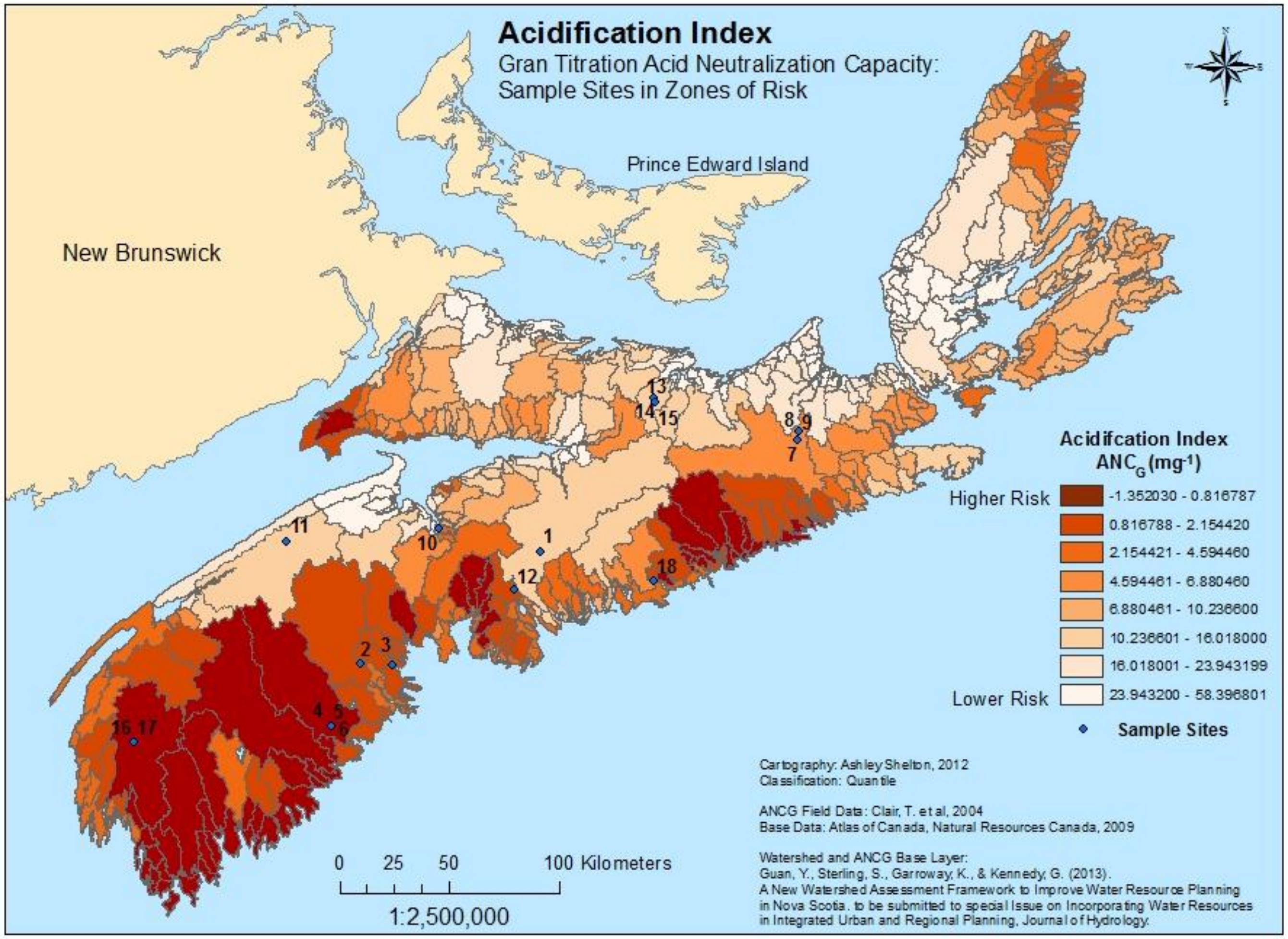


Figure 2.1. The distribution of thesis sample sites within ANC_G zones of risk. Adapted from Guan *et al.* 2013

2.3 Participant Recruitment and Training

Participants in the treatment group were recruited through a notice on the Community-Based Environmental Monitoring Network (CBEMN) website (http://www.envnetwork.smu.ca/) and an email notification sent from the CBEMN and Nova Scotia Adopt-A-Stream to various stewardship groups on their email contact list on behalf of the researcher. Ten community groups with various levels of involvement in water quality monitoring programs responded and confirmed their involvement in the study (Table 1.0: Appendix A). One to three volunteers were selected from each group for a total of 18 volunteers. No previous experience was needed to take part in the study, and volunteers were selected with evenly distributed range of monitoring experience, interest levels, and educational background.

The level of training among a volunteer group as identified by Fore *et al.* (2001), can provide a source for improved accuracy of monitoring data. Modeling this study off of Fore *et al.* (2001) recommendations, baseline training was provided to all volunteers with training divided into theoretical and practical instruction. The completion of an online training and certification course was required from all participants prior to field sampling to address theoretical education. The theoretical training criteria was previously identified through research conducted by the CURA H_2O project for the purpose of educating citizen scientists on skills needed for water quality monitoring programs. CURA H2O is a research project team composed of academia, community stewardship organizations, non-governmental environmental organizations (NGOs), government agencies, First Nations communities, public schools, the agricultural community, and the private sector. This project seeks to standardize data collection at the community level, through a water monitoring training course and an accompanying WetPro ™ toolkit (CURA H_20, 2013). The web-based course created as part of the WetPro certification and field kit project and the components of the training course include:

- Basics of freshwater systems;
- Water quality monitoring parameters;
- Guidelines for protection of aquatic life;
- Monitoring program design;
- Sampling methods and techniques; and
- Quality assurance and quality control.

(Wet-Pro Certification, 2013)

As calibration of water quality monitoring equipment was an expected source of error in the study design, a calibration manual created by the research was provided to the treatment group to detail the steps of the calibration process. Each of the volunteer treatment group participants calibrated the treatment group's YSI ProPlus to sampling. The YSI ProPlus has three parameters that required calibration: dissolved oxygen, pH and conductivity. On the day of sampling the researcher went through the manual with the volunteer and any questions relating to the calibration process were answered prior to calibration. The researcher observed the calibration process as the volunteer was instructed to attempt to calibrate and troubleshoot without requesting assistance. To help determine the effectiveness of the calibration, each volunteer logged a calibration record (Figure 2.2: Appendix A).

Each volunteer was provided a field manual prior to sampling, were asked to review the document and field procedures and were trained on the use of the water quality instrument and velocity meter. The researcher, prior to sampling, answered any questions from the treatment group participants relating to field procedures or equipment use.

2.4 Field Methods

The field procedures of this study were designed based on water quality data collection of the USGS and Canadian Council of Ministers of the Environment (CCME) water sampling protocols and the Canadian Aquatic Biomonitoring Network (CABIN) channel measurement procedures. The equipment set-up, field equipment and calibration training were controlled and standardized for each volunteer in the treatment group; however, the volunteer calibration was an uncontrolled variable.

There was an alternating sampling pattern concerning which participant would begin the collection of channel measurement data. The purpose of this alternating pattern design was to remove some of the data collection bias from the channel measurements including citizen scientists observing and following the same procedures as the professional rather than relying on training provided. The pattern began with the instructions for the treatment group participant, at sample site #1, to begin recording channel measurements and the professional proceeded once the volunteer had completed data collection. Once this sampling session ended, the next volunteer to sample at site code #2 was instructed to record channel measurements after the professional had completed recording channel measurements. Only one treatment group participant sampled at each sampling session with the control group participant.

While the channel measurements were being collected, the other participant was instructed to collect general site description notes including: their name, site code, GPS coordinates, date, weather conditions and local land use. The researcher recorded detailed site information including site drawings, photographs and notes. The researcher also recorded comprehensive observations on data collection. All field notes were recorded in the field sheets provided to the participants (Figure 2.3: Appendix A).

Once the treatment and control groups had completed channel measurements, the volunteer and the professional began collecting water measurements at transect #1 in the center of the stream, moving upstream until reaching transect #10. The data collection pattern could not be staggered for the water quality component as samples had to be recorded at the same time due to the water chemistry variability.

The following methodology illustrated in this chapter are the standard operating procedures for this study; however, as the method of recording data and sensor deployment for the citizen scientist was not a controlled variable, there may be variability in field procedures. This variability lead to some expected sources of error and observation data recorded by the researcher was used as a qualitative assessment to examine the variability and also the functionality of training provided to the citizen scientists.

2.4.1 Water Quality Field Measurements

An in-situ water quality probe and a flow meter were used to provide real-time water measurement readings of five water quality parameters: pH, temperature, dissolved oxygen, conductivity, and velocity. These parameters were chosen based on previous studies utilizing similar parameters (Nicholson *et al.*, 2002; Roa Garcia & Brown, 2009; Wenner *et al.*, 2004), identifying reduced parameter variability and were identified by Sharpe and Conrad (2006) as some of the parameters actively used among environmental community monitoring programs in Nova Scotia as indicators of environmental stress and the quality of the habitat.

The equipment was standardized for the treatment and control groups; there was one flow meter used by both treatment groups and two YSI Professional Plus multi-probes (YSI ProPlus), one for each group. One YSI ProPlus was handled and calibrated by the professional treatment group and the second unit was used and calibrated by the volunteer treatment group. The researcher conducted the maintenance required for the volunteer YSI ProPlus unit, such as DO membrane replacement.

When collecting a field measurement, participants were instructed to avoid disturbing the sediment or substrate at the bottom of the stream and to measure facing upstream into the current and away from the streambed. A two to three minute wait was necessary before beginning measurements to ensure that the disturbance from wading into the stream did not contaminate the sample and sediment has settled.

Measurements of water temperature, dissolved oxygen, pH and conductivity were taken at just below the surface of the water using the following CCME (2011) protocols for sampling depth:

- **Site with water depth <2 m:** In situ measurements taken just below surface of water (0.1m depth)
- **Site with water depth ≤2 m - ≥4 m:** In situ measurements taken at mid-depth

The instruments were placed at the appropriate water depth in an area where water was flowing, generally close to the center of the stream or in a main flow area, allowing equipment stabilization. Stabilization can range up to 10-15 minutes (YSI, 2011), which was based on the operator's discretion, and then the operator recorded the reading in the field logbook.

2.4.2 Channel and Velocity Measurements

Channel and velocity measurements were collected in order to calculate the volume of the water that passes through the channel cross-section at the sample site in a period of time, also referred as the discharge (Charlton, 2008, p. 3). The discharge measurements were compared between the treatment group and the control group to supplement the water quality data, as discharge is a very important aspect of many water quality programs (Harmel *et al*., 2006). The depth of the water was taken using a meter stick at each sampling location (CABIN, 2010; CCME, 2011); however, this was substituted with the velocity meter, which had depth markings along the side of the unit. The method used in this study to calculate discharge from channel measurements was the instantaneous measurement velocity-area method. This method involved taking velocity measurements at equidistant intervals across the width of a stream. Using a direct velocity measurement device, the Global Flow probe, an average velocity measurement was obtained at 0.6 of the total depth if the depth was <1 m (CCME, 2011; Charlton, 2008, p. 24). Refer to section 2.6 for further details on the discharge calculations.

The error margin for velocity as a result of variability of flow in the water column can be decreased by taking two velocity measurements at two depths, 0.2 and at 0.8 of depth, and the mean velocity can be calculated by averaging the two velocity measurements if depth is >1 m (CCME, 2011; Harmel *et al*., 2006). However, none of the sites selected for sampling had a depth greater than 1 m and one measurement was deemed sufficient for the purposes of this study.

Following CABIN (2010) recommendations for number of cross sections of a stream required for discharge measurements, this study identified that the reach at each site was relatively simple and uniform, and one cross section of a river or stream was sufficient to calculate average velocity and depth, with the velocity measurement collected in the center of the stream. The velocity measurement recorded the average speed at which the water was moving (CABIN, 2010). Proper handling of equipment was followed as per Global Water (2004) instructions including resetting the meter prior to collecting a measurement, orienting the propeller directly into the flow with the indicator arrow aiming downstream, and held in place for 40 seconds. The instrument was moved as needed if obstructions across the stream existed, such as boulders in the center of the stream (CABIN, 2010).

The widths and depths of the channel were recorded with five measurements: distance from shore (m), water depth (cm), bankfull width (m), wetted stream width (m) and bankfull-wetted depth (cm). The bankfull width measurement recognized high flow conditions of the two-three year peak flow and was identified from vegetation changes on the stream banks and where algae or marl have been scoured from the movement of boulders. Wetted width recorded the measurement of current flow conditions at the time of sampling and bankfull-wetted depth was recorded as the height between bankfull width and wetted width (CABIN, 2010; Charlton, 2008, p. 70).

2.4.3 Quality Assurance/Quality Control

According to Culp *et al*. (1999) (as cited in CABIN, 2010), participant training is the primary step for ensuring high quality data. This training was addressed through practical and theoretical training on freshwater basics, monitoring procedures, calibration procedures and equipment use. The main source of variability among measurements of the professional and the volunteer citizen scientists was suspected to be a result of equipment handling and calibration. Therefore, to avoid potential sampling errors during this study, the participants were provided with the training and calibration solutions necessary to calibrate equipment prior to sampling. The calibration procedure used in this study was modeled from the calibration procedures from YSI Inc. (2001), USGS standards and CCME guidelines.

For the purposes of this study a Quality Assurance/Quality Control (QA/QC) replicate sample was collected during sampling at each transect of real time temperature, pH, dissolved oxygen and conductivity readings. Replicate samples were taken at the same time and location as the other samples collected and can be taken with in situ monitoring equipment. In situ refers to measurements taken at the time of sampling and these measurements can be collected side-by-side (Chapman, 1996). The replicate sample was collected with an YSI 600QS multi-probe which was calibrated prior to each sampling session by the researcher.

Each participant was instructed to review of the water quality data on-site during data collection to prevent recording of false measurements. If a measurement appeared out of range compared to the previous readings at the site, before leaving the site a re-measurement was required. If there were concerns regarding instrument accuracy during the field sampling an end of the day verification was made to determine if the meter had drifted using a standard calibration solution and those readings were recorded in the field logbook. The researcher and professional control group performed this procedure. The treatment group participants were instructed to use their discretion on when this needed to be performed for their equipment. All operators were also provided training to ensure that data field sheets were filled out correctly prior to leaving the site (CABIN, 2010).

If there was concern that the velocity meter propeller was not turning freely, the operator was instructed to blow into the propeller for 5 to 10 seconds to verify the unit was functioning properly (Global Water, 2004).

2.5 Treatment Group Survey Design and Application

A participant survey, in the form of a questionnaire, was designed as a qualitative assessment of the previous experience of the treatment group participants. The purpose of this questionnaire was to provide qualitative data to determine the correlation between previous experience and improved accuracy values. Following approval by the Research Ethics Board (REB), the surveys were delivered in an email format to all the participants after field sampling. The REB file number for this project was 12-260 (Figure 4.4-Figure 4.6: Appendix A).

Researchers must understand and prevent or minimize bias in the design of a questionnaire. Sources of bias in a questionnaire can be found in the design of questions, such as ambiguous or complex questions, the use of technical jargon or providing scales for questions that force a choice, such as “yes or no” (Choi & Pak, 2004). The surveys were designed to avoid these biases by avoiding “yes or no” responses where possible and avoiding technical terminology. Nonresponse bias can also be a source of total survey error; therefore planning of the survey distribution was necessary (Groves, 2006). The distribution of the surveys was chosen through email correspondence to facilitate the survey process. According to Fanning (2005), the order by which you chose to ask your questions is important to maximize a respondent’s motivation with your survey.

Therefore, the format of questionnaire was designed to be simple, with non-complex wording, and evaluate three broad categories including:

- Previous experience with water quality monitoring programs;
- Relevant education related to environmental monitoring; and
- Relevant training prior to the online Wet-Pro training.

This data was used to further assess under what conditions a citizen scientist can collect the most accurate water quality data.

2.6 Analysis Methods

2.6.1 Calculating Discharge

The discharge measurements were calculated using the following equations:

$$Q = AV \quad (1)$$

The flow rate or discharge (*Q*) is the volume (cubic meters per second) of water that passes a flow section in a unit of time. This is calculated by multiplying the measured velocity (*V*) with the calculated area (*A*) in cubic meters per second (United States Department of Interior Bureau of Reclamation, 2001).

$$A = WD \quad (2)$$

The cross-section area (*A*) can be calculated by depth (*D*) multiplied by width (*W*), stretching tape across the channel at the cross section and measuring depth at location of velocity measurement (Charlton, 2008; EPA, 1995). The same equation was used to determine bankfull discharge (*BQ*) using bankfull width (*BW*) with bankfull wetted-depth (*BD*).

2.6.2 Normality and Distribution

Water quality analysis began by calculating the difference between the treatment group and the control group measurements (d_X) for each parameter using the following equation:

$$d_i = T_i - C_i \quad (3)$$

where i water quality parameter (i.e. temperature, dissolved oxygen, conductivity and pH), T is the measurement collect from the treatment group and C is the measurement collected from the control group.

The Anderson-Darling normality test was used to determine the normality of the data for the difference of each parameter. Each parameter (water temperature, pH, dissolved oxygen, dissolved oxygen percent saturation, conductivity, specific conductivity, total dissolved solids and discharge) were all non-normally distributed, therefore the data were transformed to determine if they could be normalized.

The derivation for calculating the difference between the treatment and control group measurements of temperature, dissolved oxygen, conductivity and discharge is given by:

$$d_i = \ln[|T_i - C_i| + 1] \quad (4)$$

The difference between the treatment and control group pH measurements were calculated according to:

$$d_{pH} = \ln(e^{[\,|T_{pH} - C_{pH}| \,+\, 1]}) \quad (5)$$

where i = water quality parameter (i.e. temperature, pH, dissolved oxygen, conductivity and discharge), T is the measurement collect from the treatment group and C is the measurement collected from the control group.

The Anderson-Darling normality test was used again on the transformed data to determine if the data was normalized. The results showed a *p*-value less than the significance level for water temperature, pH, dissolved oxygen, conductivity and discharge, therefore it was determined that the data could not be normalized and a non-parametric univariate test was required to compare the differences between the two groups (Figure 3.26 - Figure 3.43: Appendix A). The normality test results for the transformed bankfull discharge data presented a *p*-value greater than the significance value, therefore the data could be normalized; however, to maintain consistency with statistical analysis, the non-parametric test was also used for this parameter, as this dataset still met the assumptions of the Wilcoxon sign-rank test.

2.6.3 Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test was determined to be the most appropriate non-parametric test for comparing the median difference between the matched pairs for all water quality field samples collected, as this statistical method does not require a normal distribution (Moore, 2008; Steinijans & Diletti, 1983). An acceptable replacement for t-test (Moore, 2008), the Wilcoxon signed-rank involved gathering the differences of measurements between two groups and ranking them based on their absolute value, the sum of the positive differences and the sum of negative difference are then calculated, and then the sum is used as a test statistic (Conover, 1973; Crichton, 2000). With the assigned accuracies of the equipment error values and government standards, the two-tailed test was chosen to determine if the water quality measurements of the treatment and control groups were significantly different from each other. The Wilcoxon signed-rank confidence interval of 95% was calculated for each parameter, which determines the median value for each set of data and also indicates the highest and lower limits that would be accepted within the bounds of the median (Halperin, Hamdy, & Thall, 1989).

The Wilcoxon Signed-Rank test was used to compare the two continuous distributions of the treatment group and the control group, regardless of shape of the distribution by testing the hypothesis (Moore, 2008). There were, however, assumptions for this test including that the distribution of differences between the two groups are symmetric (Crichton, 2000). In this study the null hypothesis (H_0) and alternative hypothesis (H_a) are the following:

- H_0: There is no difference between the treatment and control measurements.
- H_a: There is a difference between treatment and control measurements.

The Wilcoxon test has a number of outputs including the Wilcoxon rank sum statistic (W), which is the sum of the ranks of one of the samples. The *p*-values for the Wilcoxon test are based on the sampling distribution of the rank sum statistic W when the null hypothesis (no difference in distributions) is true (Moore, 2008). When the *p*-value is greater than the significance value the null hypothesis is true, and there is no significant difference found between the two groups. Generally the significance level is set at $\alpha=0.05$; however, as noted by Selvin and Stuart (1966), in experimental situations it may be appropriate for testing procedures to include multiple hypotheses testing. With multiple testing it has been suggested that the probability of identifying at least one significant result due to chance will increase when more hypotheses are tested (i.e. in one dataset testing for pH, conductivity, water temperature and dissolved oxygen), leading to type I error. Therefore to offset this issue, the Bonferroni correction for multiple testing was used. Statistically speaking this makes the test more conservative, increasing the potential of type II error; however, due to the large statistical power of this dataset, it was deemed appropriate to use this correction (Napierala, 2012). The discharge data did not include multiple hypotheses testing, therefore a significance level of $\alpha=0.05$ was used.

For the purposes of this study the Bonferroni method was used for the water quality testing (water temperature, pH, conductivity and dissolved oxygen) using the following equation:

$$\alpha=0.05/n \qquad (6)$$

where n = number of water quality tests at a single event (i.e. 7 water quality measurements recorded at a single event).

Chapter Three

Results

3.1 Water Quality Data Normality and Distribution

As stated by Harnel et al. (2006), the understanding of the uncertainty in water quality measurement data has not been well established; however, this study aimed to increase the understanding of the variability resulting from the operator of water quality equipment by comparing data gathered by a treatment group (i.e. volunteer citizen scientists) and a control group (i.e. professional scientist).

3.1.1 Temperature

The temperature measurements by the treatment group were not significantly different from the control group at a hypothesis value of 0.00. The Wilcoxon signed-rank test was used to compare the median difference between the temperature measurements, taking into consideration the assigned accuracy values for the equipment and the different government standards. The two-tailed test examined the significance of three median test differences (0.00, 0.20 and 2.00) and the hypothesis test median value was equal to zero resulted in a *p*-value > 0.0071 (i.e. *p*-value=0.019) (Table 3.2). Therefore, the null hypothesis was not rejected, thus concluding that the temperature measurements collected by the treatment group were not significantly different from the control group.

The overall distribution of the difference in water temperature data was skewed slightly to the left (Figure 3.1) and outliers were found for sample site codes: 2, 8, 15, and 16 (Figure 3.13: Appendix B). Using sensitivity analysis to remove the outliers from the data and thn re-testing the d_T dataset resulted in no change to the confidence interval (95%), nor the decision to reject the null hypothesis. Therefore, the outliers did not impact the overall data results. Observational data did not indicate weather or site concerns that would explain variability in the data. However, the values of the outliers did not exceed 1.1°C, therefore daily fluctuations as discussed in Chapter 1, mechanical error or human error could potentially account for these events.

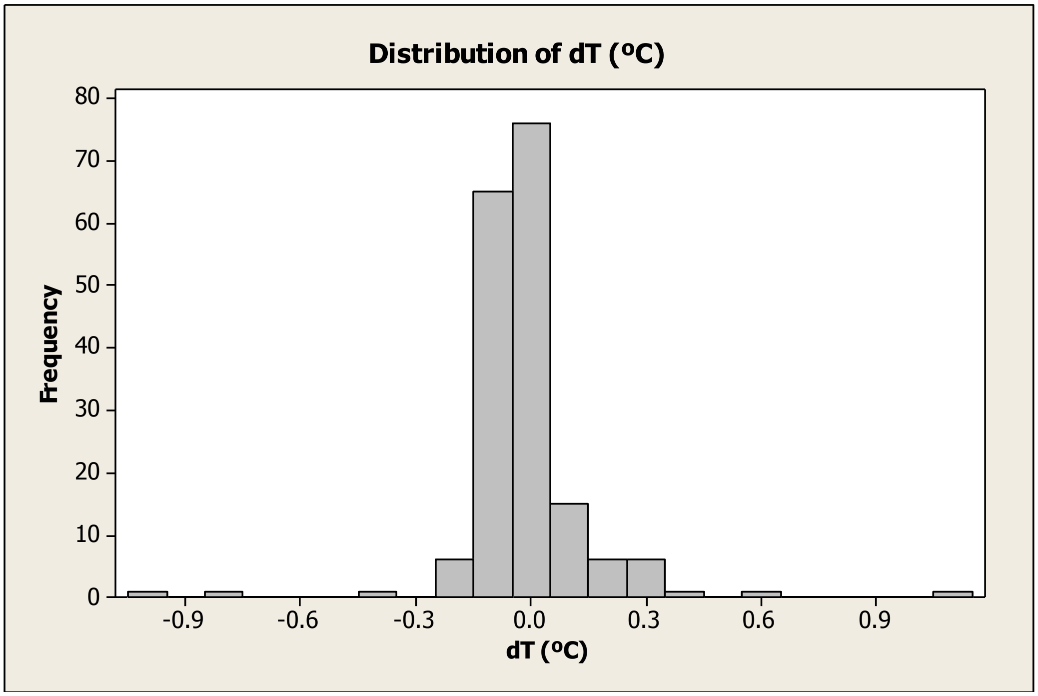


Figure 3.1. Distribution of the difference in the water temperature (°C) (d_T) between the measurement collected from the treatment group and the measurement collected from the control group.

Further analysis on the distribution of the water temperature data indicated that the treatment group temperature data parallels closely with the control group data (Figure 3.2). Using the Environment Canada and GNL Department of Environment and Conservation acceptable range of 0.2 units, 93.3% of the water temperature measurements from treatment group were within the 0.2 units from the control measurements. Additionally, 100% of water temperature treatment group measurements were within Nova Scotia Environment (NSE) & United States Geological Survey (USGS) acceptable range of 2.0 units from the control group measurements (Table 2.1).

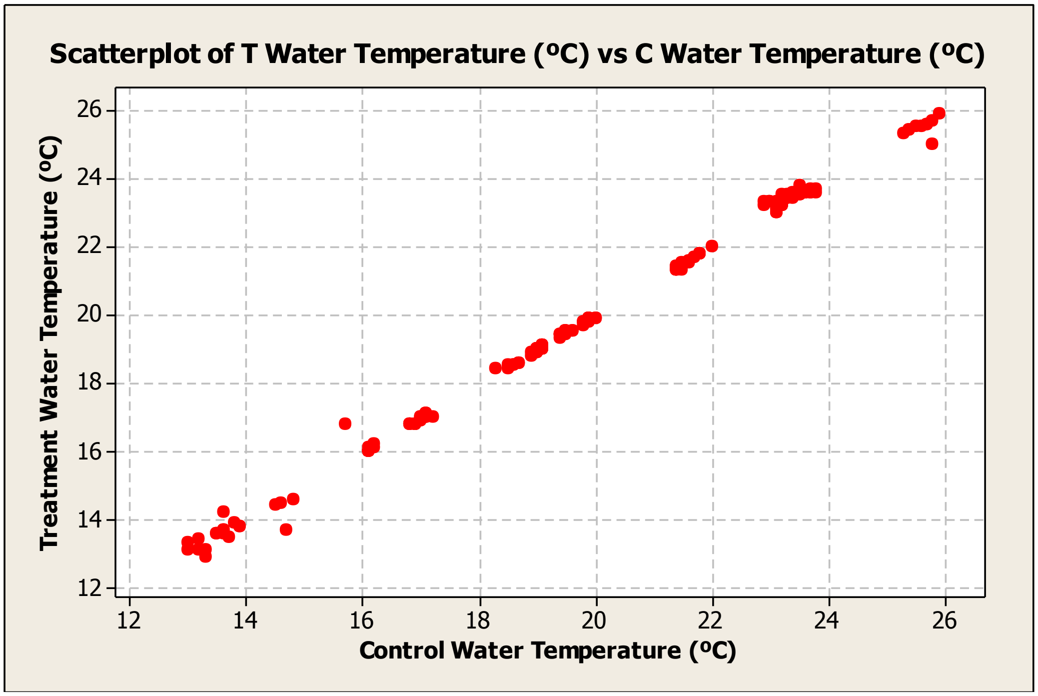


Figure 3.2. Scatterplot of control group water temperature measurements (C_T) versus treatment group water temperature measurements (T_T).

3.1.2 pH

The degree of difference between the treatment and control pH measurements for all median test values were within the equipment error, which included the individual instrument cable accuracy specifications (+/- 0.2 pH units) and the given source of error in the calibration buffer standard (+/- 0.01 pH unit), as well as within NSE, EC, and GNL standards (Table 3.2). Therefore, there was no significant difference between the treatment and control group measurements, with a p-value > 0.0071 (i.e. p-value =0.383) at a test median of 0.020, which lead to a conclusion of failing to reject the null hypothesis.

One extreme outlier data point, at a difference of 2.6 pH units between the treatment and control groups was identified (Figure 3.4, Figure 3.14: Appendix B). This may have been a result of a reduced stabilization time by the treatment group for this event as all other pH measurements made by that participant were within the pH-difference range of 0.32 - 0.85 with the control group. Although some of these values were outside of the equipment error range of +/- 0.44 pH units, they were still within the NSE criteria for pH of +/-2.0 (Table 2.0 and Table 2.1). A correlation to a lack of experience or training was not found to be the case, as the participant at sample site #17 had previous experience with water monitoring, previous knowledge and experience with the use of the equipment, as well as relevant work and education experience in the field of environmental monitoring. It is expected that as this event was the final measurement of the sampling day, human error was the cause of the outlier, as the measurement appears to have been rushed, with the time interval of 2 minutes between the final two measurements recorded.

The d_{pH} at sample site #17 had a consistently higher range of *d*-values (0.93 - 1.13 pH units) and is suspected to be the result of a calibration error. There were no reporting errors observed in the calibration logbook; however, observational data indicated that the treatment group participant displayed reduced confidence in the calibration process, attempted to ask a researcher questions during the process, and had to attempt the calibration of the pH twice. There is also a correlation to a lack of training and experience with water quality monitoring, however this participant did have a background scientific education (Table 3.5: Appendix B).

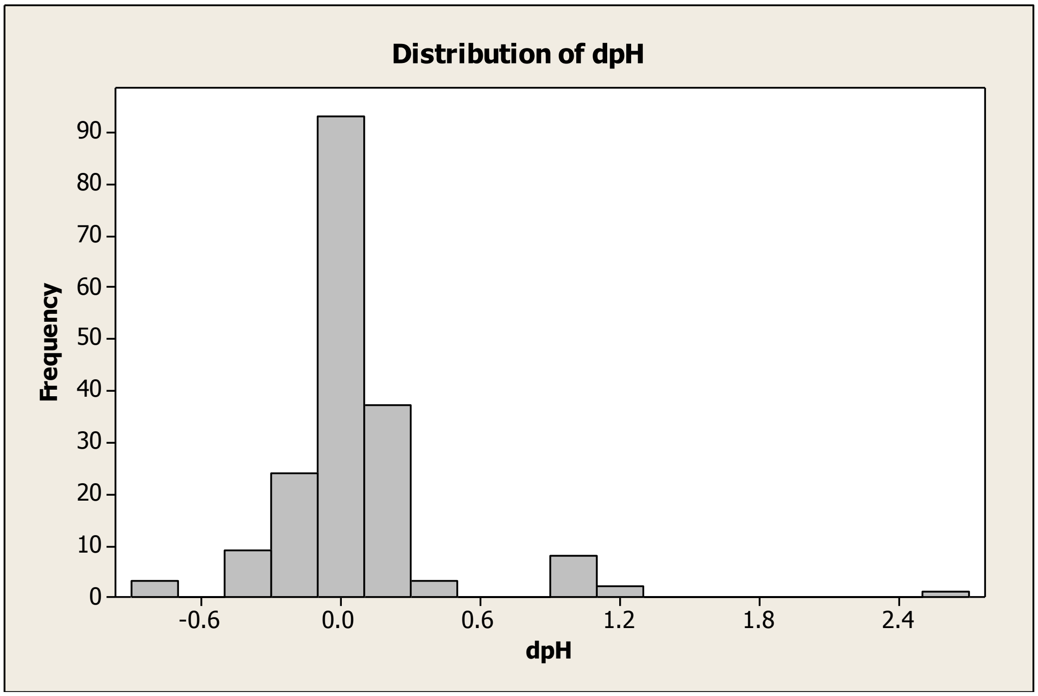


Figure 3.3. Difference in the pH data (d_{pH}) between the measurement collected from the treatment group and the measurement collected from the control group.

The pH measurements by the treatment group were not significantly different from the control group measurements at a test median of 0.020, as the *p*-value was greater than the significance value p-value < 0.0071 (p-value=0.383) (Table 3.2). This thereby prevents the rejection of the null hypothesis. The test median value was chosen as it was within the confidence interval and was at a higher level of accuracy than the pH values (Table 2.0 and Table 2.1).

Further analysis was conducted by breaking the pH measurement data into three subsections to determine if a correlation existed between the value of the pH measurement and the accuracy of the treatment group measurement compared to the control group. The samples were broken-up based on natural breaks in the data collected based on control group pH measurement values (<6, 6-7.5 and >7.5) (Figure 3.15- Figure 3.20: Appendix B).

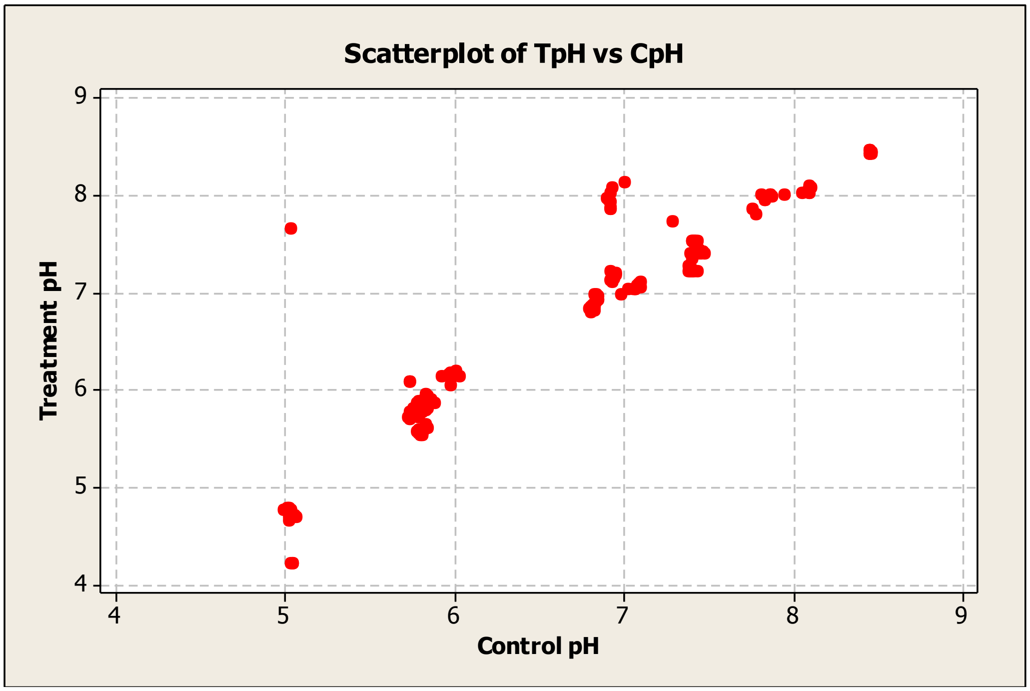


Figure 3.4. Scatterplot of control group pH measurements (C_{pH}) versus treatment group pH measurements (T_{pH}).

The results of the analysis show that only the pH measurement with a higher values (>7.5 pH) failed to reject the null hypothesis with a median value of 0.00 and 0.02 (*p*-value <0.0071), therefore the treatment group measurements collecting pH values greater than 7.5 was not significantly different from the control group (Table 3.3). Both the d_{pH} samples with control values at <6 and 6-7.5 rejected the null hypothesis, leading to conclude that there was a significant difference between the treatment group pH measurements and the control group for measurements collected at lower pH values than 7.5. Overall the distribution of the pH values are skewed to the left (Figure 3.3) and in particular for the lower pH values (Figure 3.15 and Figure 3.17: Appendix B).

3.1.3 Dissolved Oxygen

The results from the Wilcoxon Signed-Rank test demonstrated that dissolved oxygen measurements collected from the treatment group were significantly different from the measurements collected by the control group for both d_{DO} (mg/L) and d_{DO} (%). The two DO parameter measurements were collected and analyzed separately as the distribution of differences varied, although both distributions for d_{DO} (mg/L) and d_{DO} (%) were skewed to the left (Figure 3.5 and Figure 3.6).

Two extreme outliers were found in site codes #2 and #7 (Figure 3.21 and Figure 3.22: Appendix B). Both treatment group participants for these sampling events indicated little relevant training in water monitoring or relevant education in the survey responses prior to this study (Table 3.6: Appendix B). At site code #2 there was a reporting error on the calibration sheet; however, both sampling events followed proper calibration procedures with no observations of calibration procedure error. The sensitivity analysis resulted in no change to the significance result when removing these outliers, indicating that the outliers were not impacting the overall result of a significant difference between the treatment and control DO measurements.

Through the comparison testing of the DO measurements (mg/L and %) all median tests displayed significant difference between the measurements of the treatment group and the control group. The median difference of the DO (mg/L) measurements, with the assigned accuracy values of the equipment error values and government standards; the two-tailed test examined the significance of three median test differences (2.000, 0.300 and 0.200). The resulting *p*-values less than the significance value $p<0.0071$ (i.e. $p<0.001$), leading to reject the null hypothesis and concluding that DO (mg/L) measurement of the treatment group was significantly different from the control group for the test medians (Table 3.2). Using the selected median difference of the DO (%) measurements (20.00, 2.000 and 0.00), displayed results where the *p*-values were all less than the significance value $p<0.0071$ for all median hypothesis test, leading to reject the null hypothesis and concluding that DO (%) measurement of the treatment group is significantly different from the control group (Table 3.2).

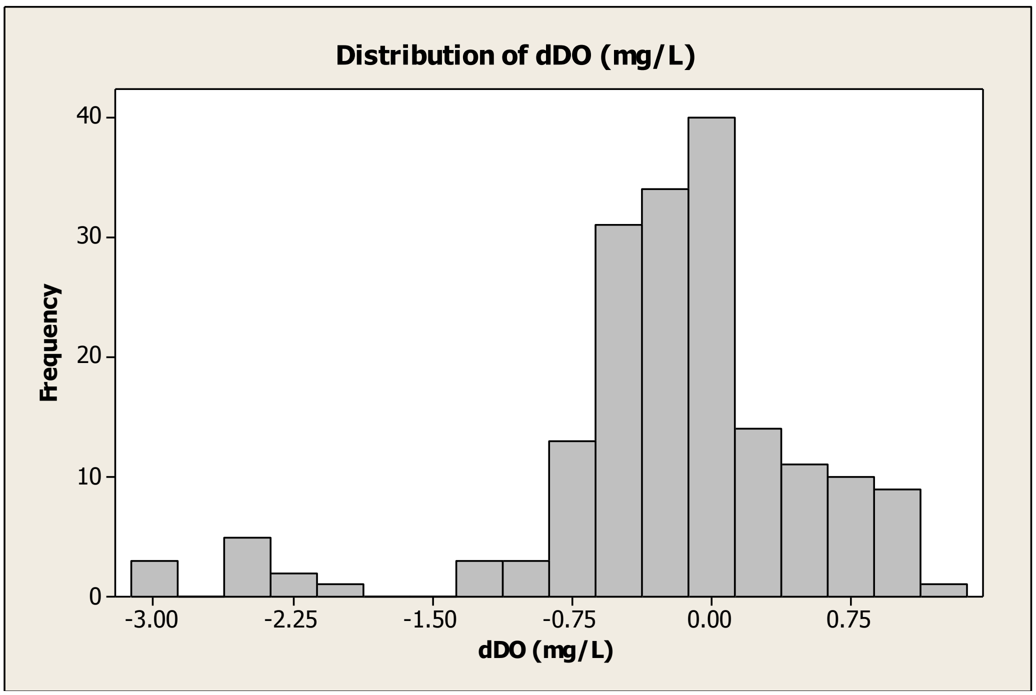


Figure 3.5. Difference in the DO (mg/L) data (d_{DO}) between the measurement collected from the treatment group and the measurement collected from the control group.

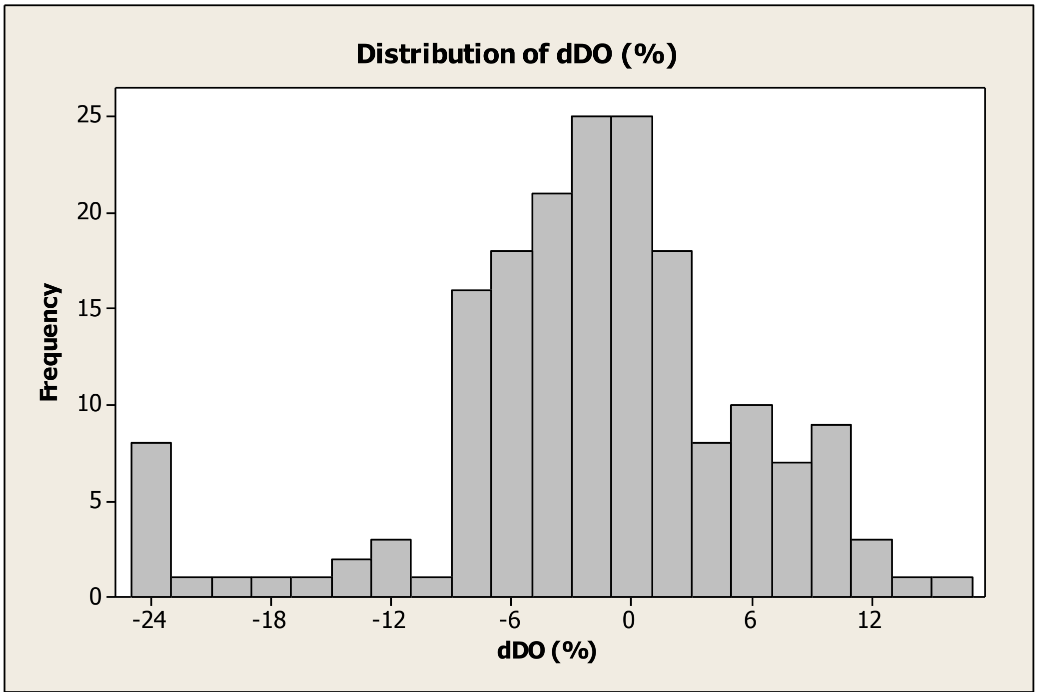


Figure 3.6. Difference of the DO (%) data (d_{DO}) between the measurement collected from the treatment group and the measurement collected from the control group.

3.1.4 Conductivity

Three different parameter measurements were collected for conductivity readings, conductivity (uS/cm) (d_C), specific conductivity (uS/cm) (d_{SPC}) and total dissolved solids (mg/L) (d_{TDS}). Each parameter was analyzed separately as the distribution of differences varied and each resulted in not rejecting the null hypothesis therefore the treatment group conductivity measurements were not significantly different than the control measurements.

Based on the graphical data, the distribution of the conductivity data skewed to the left, and an outlier is visible where one measurement had a large deviation from the treatment and control readings (Figures 3.7-Figure 3.9). The boxplot distribution indicated an extreme outlier at site code #10 (Figure 3.23- Figure 3.25: Appendix B). The

d_{SPC} value of 257 uS/cm is visible at site code #10, which is expected as a result of environmental conditions (Figure 3.24: Appendix B). Qualitative observation data collected on the day of sampling described fluctuation of SPC values and the QA/QC sample supports this assumption that the outlier at site code #10 was a result of the environmental conditions, rather than reporting or human error. When removing the outliers from the dataset, no change was observed for the significance value, and therefore these outlier events did not impact the overall dataset.

The Wilcoxon signed-rank test was used to compare the median difference of the conductivity measurements, with the assigned accuracy values of the equipment error values and government standards; the two-tailed test examined the significance of two median test differences (1.000 and 0.00). The test statistic values (Table 3.2) displays *p*-value is greater than the significance level of α=0.0071 (p=0.204), we fail to reject the null hypothesis, concluding that the conductivity measurements of the treatment group were not significantly different from the control group at a test median of 0.00. The result of the test median=1.000 to reject the null hypothesis was a result of the confidence interval being much smaller than the sensor drift error. The difference between the conductivity measurements of the treatment and control groups was smaller than the potential equipment error.

Through the comparison testing of the specific conductivity measurements, the median test differences (1.00, 0.050 and 0.00), displayed significant difference between the measurements of the treatment group and the control group. Through the comparison testing of the specific conductivity measurements final conclusion indicate that the SPC measurements of the treatment group were not significantly difference from those measurement collected by the control group at test medians of 0.050 and 0.100 (Table 3.2), which are within the accuracy standards discussed in Chapter 2 (Table 2.0 and Table 2.1). The test median 1.000 resulted in a *p*-value <0.001 leading to the rejection of the null hypothesis and conclusion that there was a significant difference at this test median. This result was potentially caused by the confidence interval of the ranks in medians and the spread of data being smaller than the sensor drift error highlighted by government standards. The difference between the specific conductivity measurements of the treatment and control groups was smaller than the potential equipment error.

The test statistic (Table 3.2) resulted in a *p*-value ($p=0.102$) is greater than the significance level of $\alpha=0.0071$, and leading to determine that the TDS measurement of the treatment group were not significantly different from the control group.

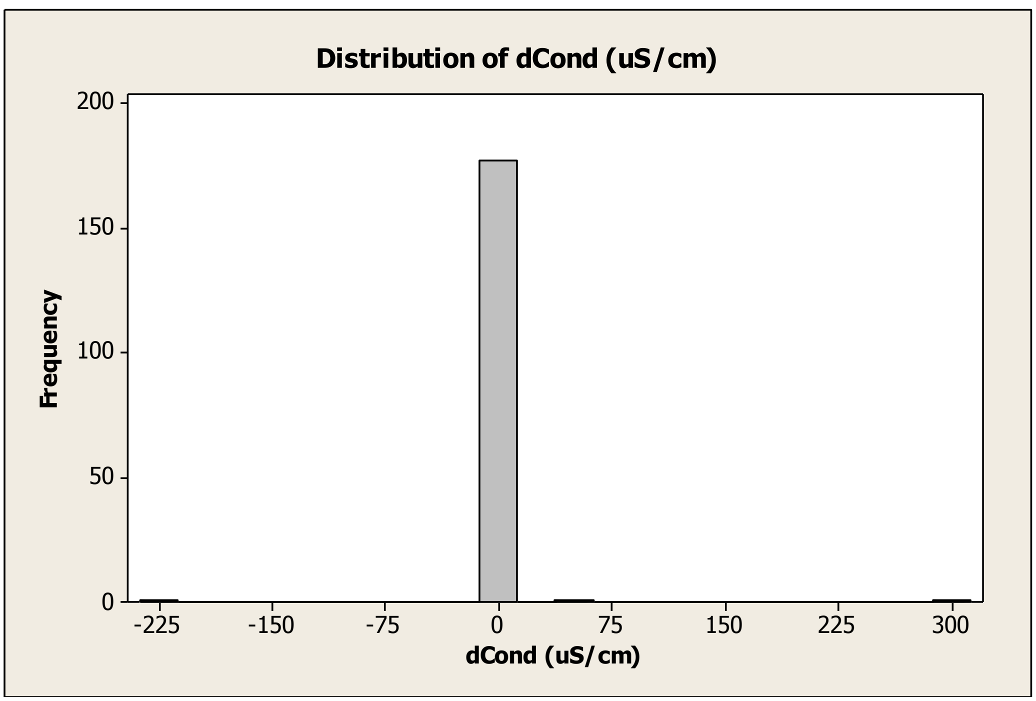


Figure 3.7. Difference of the conductivity (uS/cm) data (d_C) between the measurement collected from the treatment group and the measurement collected from the control group.

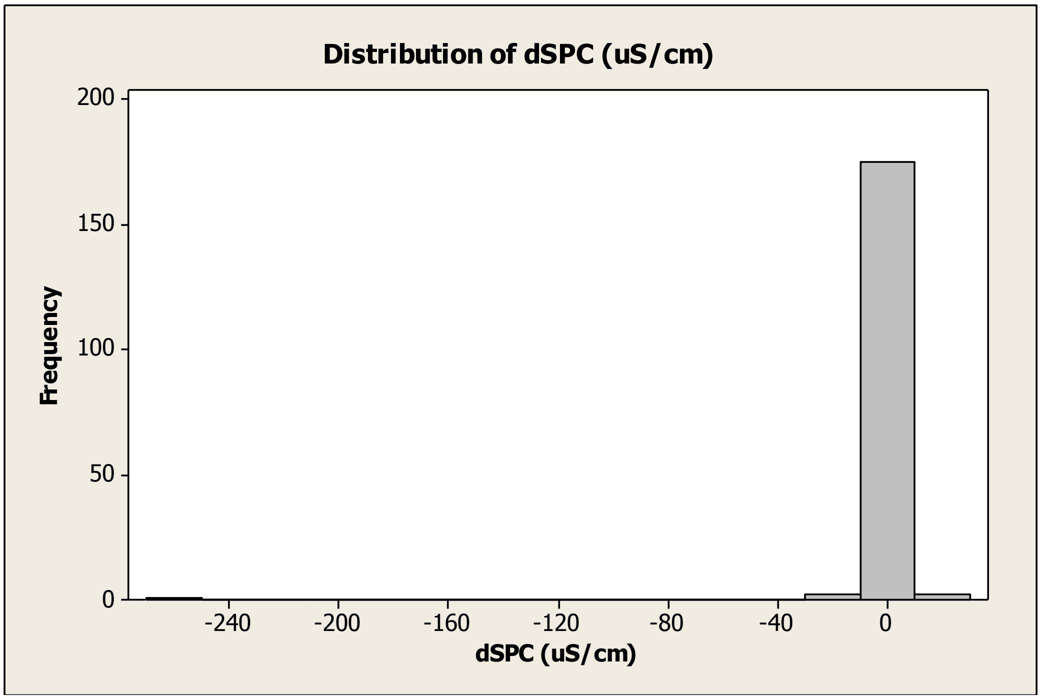


Figure 3.8. Difference of the SPC (uS/cm) data (d_{SPC}) between the measurement collected from the treatment group and the measurement collected from the control group.

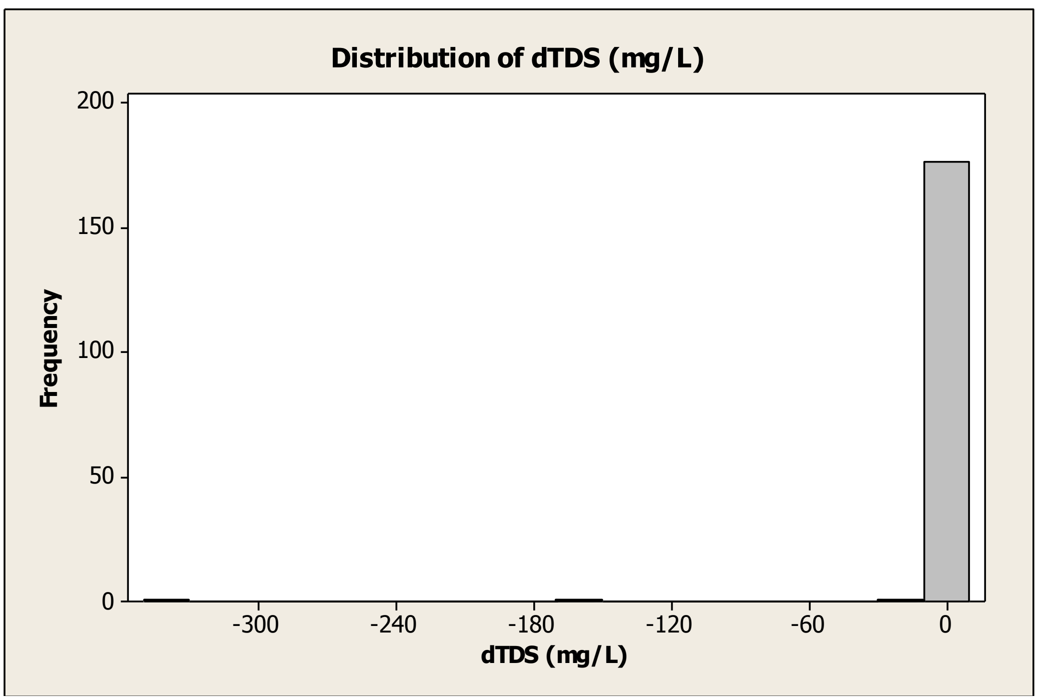


Figure 3.9. Difference of the TDS (mg/L) data (d_{TDS}) between the measurement collected from the treatment group and the measurement collected from the control group.

3.1.5 Discharge

Through the field data collected, a calculated discharge measurement for the cross section of each sample site was analyzed (Table 3.0 and Table 3.1). Using the calculated discharge and bankfull discharge measurements, results indicate no significant difference in the treatment discharge values compared to the control group discharge values (Table 3.4). The Wilcoxon Signed-Rank test was conducted on the discharge data with the hypothesized median value of 0.00, and resulted in *p*-values>0.05.

Extreme outliers were visible for d_D for sample sites #17 and #18 and d_{BD} for sample site code #16 (Figure 3.10- Figure 3.12). The outliers for sample site #17 appear to be a result sampling/recording error of the velocity measurement. The outlier for site code #18 with a d_D value of 21.499 was a result of human sampling/recording error of water depth. Using sensitivity analysis, removing the outliers to the dataset and re-testing resulted in no change to the significance values for discharge or bankfull discharge; therefore, the outliers did not impact the overall dataset.

Table 3.0. Treatment group channel and velocity measurements

Site Code	Treatment Group: Volunteer							
	Channel Data			**Velocity and Depth Data**			**Calculated Discharge**	
	A	**B**	**C**	**I**	**II**	**V**	**Discharge (m^3/s)**	**Bankfull Discharge (m^3/s)**
1	10.1	8.78	0.328	4.39	0.148	0.246	0.319	1.28
2	2.20	1.80	0.186	0.900	0.105	0.000	0.000	0.000
3	10.7	9.74	0.626	3.75	0.120	0.688	0.805	5.47
4	15.0	5.90	0.483	2.50	0.120	0.724	0.513	6.55
5	14.8	5.50	0.400	2.75	0.006	0.165	0.006	0.994
6	14.0	5.40	0.700	2.00	0.090	0.617	0.300	6.82
7	8.40	8.20	0.420	4.20	0.360	0.657	1.94	4.31
8	8.50	2.90	0.760	0.800	0.180	0.577	0.301	4.61
9	5.00	2.25	0.300	0.840	0.060	0.845	0.114	1.52
10	2.00	0.800	0.200	0.400	0.200	0.282	0.045	0.225
11	14.2	12.8	0.980	6.40	0.580	0.434	3.22	9.61
12	6.15	6.35	0.415	3.23	0.195	0.156	0.194	0.587
13	12.0	5.00	1.50	1.50	0.160	0.452	0.361	8.99
14	28.0	11.2	0.400	5.60	0.108	0.277	0.335	3.94
15	24.0	10.8	0.860	5.38	0.350	0.197	0.740	5.71
16	25.5	23.9	0.720	11.5	0.780	0.076	1.42	2.91
17	27.7	24.2	0.400	12.1	0.780	0.849	16.0	27.8
18	24.9	22.3	0.600	5.04	0.020	2.23	0.993	34.4

Notes:

(1) A=Bankfull Width (m), B= Wetted Stream Width (m), C= Bankfull-Wetted Depth (m), I= Distance from Shore (m), II= Depth (m), V= Average Velocity (m/s)

Table 3.1. Control group channel and velocity measurements

Site Code	Control Group: Professional							
	Channel Data			**Velocity and Depth Data**		**Calculated Discharge**		
	A	**B**	**C**	**I**	**II**	**V**	**Discharge (m^3/s)**	**Bankfull Discharge (m^3/s)**
1	10.3	8.40	N/A	4.20	0.122	0.845	0.866	N/A
2	2.50	2.10	0.365	1.25	0.200	0.000	0.000	0.000
3	9.70	9.50	0.170	3.75	0.100	0.590	0.561	1.55
4	15.0	5.90	0.690	3.00	0.150	0.738	0.653	9.29
5	15.9	5.50	0.690	2.75	0.220	0.702	0.849	10.2
6	15.9	5.60	0.760	2.80	0.160	0.586	0.525	8.57
7	8.70	8.00	0.420	4.00	0.360	0.715	2.06	4.85
8	7.90	2.50	1.04	1.25	0.115	0.662	0.190	6.04
9	6.50	2.20	0.730	1.10	0.095	0.921	0.192	4.94
10	2.70	1.00	0.820	0.500	0.090	0.148	0.013	0.362
11	13.8	12.8	0.980	6.40	0.570	0.528	3.85	11.2
12	6.20	6.40	0.320	6.20	0.190	0.188	0.228	0.594
13	9.80	4.50	1.90	2.25	0.250	0.787	0.885	16.6
14	21.4	11.3	1.10	4.90	0.180	0.429	0.869	11.8
15	22.9	10.8	1.35	5.40	0.290	0.174	0.546	6.55
16	24.7	26.6	0.135	13.3	0.270	0.054	0.385	0.537
17	24.7	26.8	0.138	13.4	0.780	0.063	1.31	1.42
18	23.4	21.8	0.360	5.04	0.490	2.11	22.5	41.9

Notes:

(1) A=Bankfull Width (m), B= Wetted Stream Width (m), C= Bankfull-Wetted Depth (m), I= Distance from Shore (m), II= Depth (m), V= Average Velocity (m/s)

(2) Site Code #1 there is a missing Bankfull-Wetted Depth professional measurement. Site Code #1 was excluded from the statistical analysis for comparison Bankfull Discharge measurements

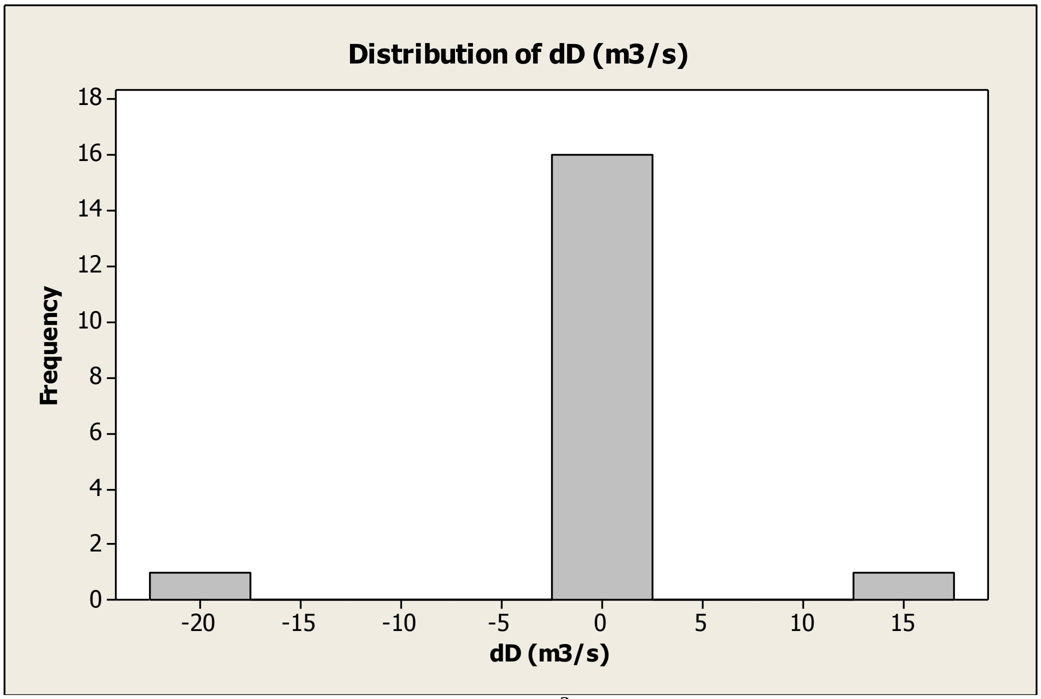


Figure 3.10. Difference of the discharge (m^3/s) data (d_D) between the measurement collected from the treatment group and the measurement collected from the control group.

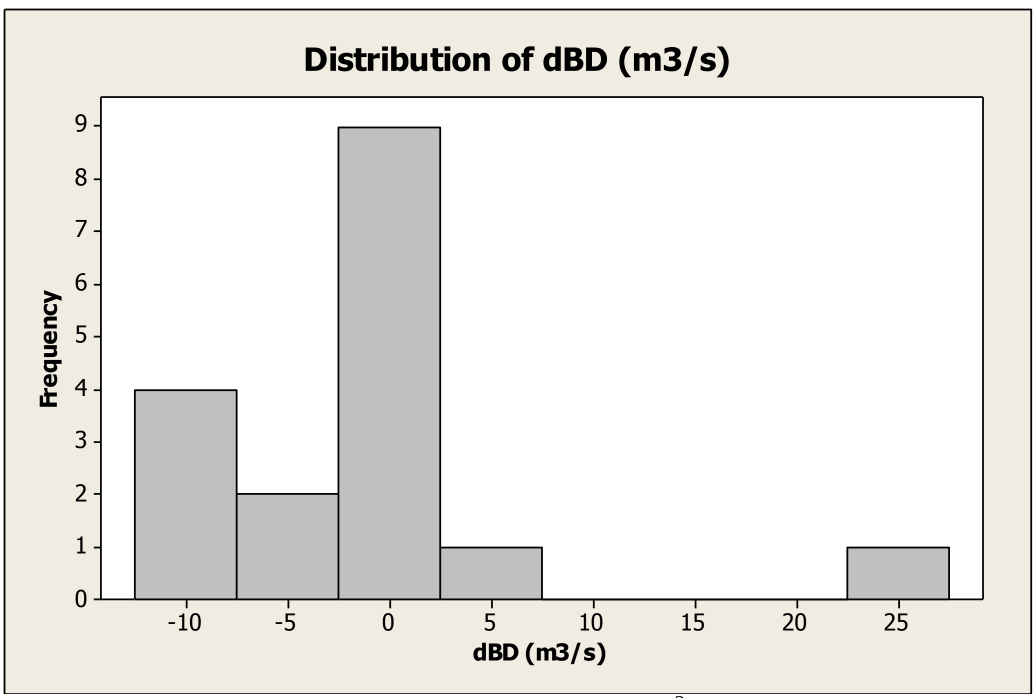


Figure 3.11. Difference of the bankfull discharge (m^3/s) data (d_{BD}) between the measurement collected from the treatment group and the measurement collected from the control group.

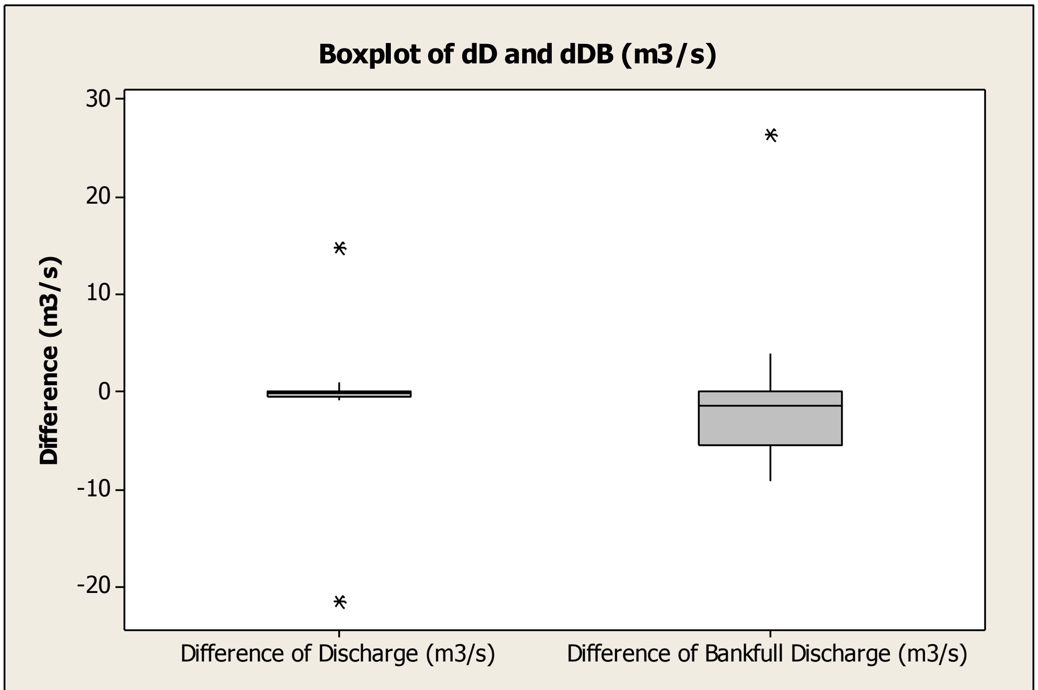


Figure 3.12. Difference in the discharge (m^3/s) data (d_D) and bankfull discharge (m3/s) data (d_{BD}) between the measurement collected from the treatment group and the measurement collected from the control group with whiskers marking the minimum and maximum data values for each sample site.

3.2 Summary of Findings

An evaluation of the water quality comparison results have showed the parameters including water temperature, pH, conductivity and discharge had no significant difference between the treatment group compared to the control group (Table 3.6: Appendix B). Dissolved oxygen displayed statistically significant difference for all hypothesis values. Upon further inspection, water temperature did reveal *p*-values less than 0.05, the generally accepted significance level for most testing; however, utilizing the Bonfferoni correction for multiple testing displayed a result of no significant difference (α=0.0071). When examining the actual measurements only 6.7% of the measurements were outside of the restriction for mechanical error of 0.2 units, which also coincides with Environment Canada and GNL Department of Environment and Conservation. Similarly, none of the values would have been rejected when compared to Nova Scotia Environment (NSE) & United States Geological Survey (USGS) acceptable range of variability of 2.0 units (Table 2.1).

Through an examination of the statistical test and its assumptions, a source of potential error was noted as a result of potentially failing to meet one of the assumptions of the Wilcoxon signed-rank test, the assumption of symmetry of the distributions between the paired samples (Crichton, 2000). If a skewed distribution was present in the population from which the paired differences were sampled, then there could have been a loss of significance and causing an incorrect rejection of the null hypothesis (type I error) (Rahbar, Chen, Jeon, Gardiner, & Ning, 2012).

To determine if the population of the paired differences was skewed, informal graphical assessment using histogram graphs were used for each water quality parameter (Figure 3.1, Figure 3.3, and Figure 3.5- Figure 3.12). The patterns varied; however, temperature, pH and dissolved oxygen displayed skewed distribution to the left and conductivity with overall symmetric distribution. As the paired sign test does not rely on symmetry (Prophet StatGuide, 1997), it was deemed an appropriate alternative test for the parameters with skewed distribution. Through further testing using the paired sign nonparametric test, results did not present new conclusions on the accuracy of the parameters noted to be significantly different. Therefore, concluding that the symmetry of the distribution for each parameter was not impacting the overall results of significance, and the assumptions of the Wilcoxon test were met.

Table 3.2. Wilcoxon signed-rank test results for water quality d_x data

Water Quality Parameter (d_x)	Confidence Interval (Lower and Upper Critical Values) and Estimated Median	Wilcoxon signed-rank: two-tail test results			
		Median (null hypothesis) and N	Wilcoxon Rank Sum Statistic (W)	*p*-value	Significance *
d_T (°C)	Lower = -0.0500 Upper = 0.000 Estimated Median = -0.0500	• 0.000 (N for test:104) • 0.2 (N for test:174) • 2.0 (N for test:180)	• 2,004.0 • 465.0 • 0.0	• 0.019 • <0.001 • <0.001	• No • Yes • Yes
d_{pH}	Lower = -0.0200 Upper = 0.0350 Estimated Median = 0.0100	• 0.020 (N for test:179)	• 7,449.0	• 0.383	• No
d_{DO} (mg L^{-1})	Lower = -0.2500 Upper = -0.100 Estimated Median = -0.1500	• 0.200 (N for test:170) • 0.3 (N for test:176) • 2.0 (N for test:180)	• 2,764.5 • 2,092.5 • 0.0	• <0.001 • <0.001 • <0.001	• Yes • Yes • Yes
d_{DO} (%)	Lower = -3.00 Upper = -1.00 Estimated Median = -2.00	• 0.00 (N for test:171) • 20.00 (N for test:180) • 2.0 (N for test:170)	• 4,809.0 • 0.0 • 2,820.0	• <0.001 • <0.001 • <0.001	• Yes • Yes • Yes
d_C (uS cm^{-1})	Lower = -0.050 Upper = 0.200 Estimated Median = 0.100	• 0.000 (N for test:165) • 1.000 (N for test:177)	• 7,629.0 • 2,888.5	• 0.204 • <0.001	• No • Yes
d_{SPC} (uS cm^{-1})	Lower = -0.050 Upper = 0.250 Estimated Median = 0.150	• 0.050 (N for test:180) • 0.100 (N for test:180) • 1.000 (N for test:178)	• 9,288.5 • 7,561.0 • 3,401.0	• 0.103 • 0.464 • <0.001	• No • No • Yes
d_{TDS} (mg L^{-1})	Lower = 0.000 Upper = 0.175 Estimated Median = 0.000	• 0.000 (N for test:179)	• 2,507.0	• 0.102	• No

Note:
(1) Values are omitted if equal to hypothesis median test value therefore "N for test" refers to sample size used for test
(2) * Using Bonforoni Correction for multiple testing, significance level α=0.0071

Table 3.3. Wilcoxon signed-rank test results for water quality d_{pH} data based on natural breaks in control data

Water Quality Parameter (d_x)	Confidence Interval (Lower and Upper Critical Values) and Estimated Median	Wilcoxon signed-rank: two-tail test results			
		Median (null hypothesis) and N	Wilcoxon Rank Sum Statistic (W)	*p*-value	Significance *
d_{pH} (>6pH)	Lower = -0.155 Upper= -0.040 Estimated Median= -0.110 N=66	• 0.00 (N for test: 65) • 0.20 (N for test: 65) • 0.02 (N for test: 66)	• 595.0 • 84.0 • 499.5	• 0.002 • <0.001 • <0.001	• Yes • Yes • Yes
d_{pH} (6-7.5pH)	Lower = 0.0350 Upper = 0.1050 Estimated Median= 0.070 N=85	• 0.00 (N for test: 83) • 0.20 (N for test: 84) • 0.02 (N for test: 84)	• 2589.0 • 883.0 • 2418.0	• <0.001 • <0.001 • 0.005	• Yes • Yes • Yes
d_{pH} (<7.5pH)	Lower = -0.0300 Upper = 0.0350 Estimated Median= -0.0050 N=29	• 0.00 (N for test: 28) • 0.20 (N for test: 29) • 0.02 (N for test: 29)	• 198.0 • 0.0 • 173.0	• 0.918 • <0.001 • 0.341	• No • Yes • No

Note:
(1) Values are omitted if equal to hypothesis median test value therefore “N for test” refers to sample size used for test
(2) * Using Bonforoni Correction for multiple testing, significance level α=0.0071

Table 3.4. Wilcoxon signed-rank test results for water quality discharge data

Water Quality Parameter (d_x)	Confidence Interval (Lower and Upper Critical Values) and Estimated Median	Wilcoxon signed-rank: two-tail test results			
		Median (null hypothesis) and N	Wilcoxon Rank Sum Statistic (W)	*p*-value	Significance *
D_D (m^3/s)	Lower = -0.406 Upper = 0.111 Estimated Median = -0.145 N=18	• 0.000 (N for test:17)	• 52.0	• 0.256	• No
D_{DB} (m^3/s)	Lower = -4.32 Upper = 0.37 Estimated Median = -1.64 N=17	• 0.000 (N for test:16)	• 35.0	• 0.093	• No

Note:
(1) Values are omitted if equal to hypothesis median test value therefore "N for test" refers to sample size used for test
(2) Site Code #1 there is a missing Bankfull-Wetted Depth professional measurement. Site Code #1 was excluded from the statistical analysis for comparison Bankfull Discharge measurements
(3) *Significance level α=0.05

Table 3.5. Summary of water quality comparison analysis between treatment and control

Water quality Parameter	Significant Difference (Yes/No)	Conditions for Robust or Non Robust Data Accuracy	Parameter Suitability	Recommendations for Improvements to Citizen Science Monitoring
Temperature (°C)	No	• No calibration required • Reduced spatial variability through the length of the river	High	• Maintaining detailed monitoring program design specifying exact time and locations of sampling
pH	No	• Procedural limitations • Calibration limitation	High	• Further training on procedural field methods • Trained scientist performing calibration
Conductivity, Specific Conductivity (μS cm^{-1}) & Total Dissolved Solids (mg L^{-1})	No	• Reduced spatial variability through the length of the river • Procedural Limitations	High	• Further training on procedural field methods
Discharge & Bankfull Discharge (m^3 s^{-1})	No	• Subjective observational data collection • Mechanical limitations	Moderate	• Training on procedural field methods and equipment troubleshooting

Table 3.5. (continued)

Water quality Parameter	Significant Difference (Yes/No)	Conditions for Robust or Non Robust Data Accuracy	Parameter Suitability	Recommendations for Improvements to Citizen Science Monitoring
Dissolved Oxygen ($mg\ L^{-1}$)(%)	Yes	• High cross-sectional variability • Mechanical limitations • Procedural limitations	Low	• Increased training on procedural field methods and equipment troubleshooting • Trained scientist performing calibration • Multiple Measurements • Detailed monitoring program design specifying exact time and locations of sampling locations

Chapter Four

Discussion and Conclusions

4.1 Discussion

4.1.1 Summary of Results

The expected results of the study were that the citizen science group measurements for all water quality parameters would be within the accuracy of the mechanical error of the YSI ProPlus and government correction/rejection criteria when compared to the professional measurements. However, through a comparative analysis of water quality and discharge data, results revealed some parameters with a higher robustness in data accuracy than others (Table 3.2-Table 3.4 and Table 3.5). There was no significant difference detected in the water quality values from the citizen science and professional field samples for water temperature, pH, conductivity, and discharge measurements; while differences were found for dissolved oxygen measurements.

By utilizing the parameters displaying robust data accuracy with no significant difference found between the citizen scientists and water professional field measurements, a sampling program representative to this study (i.e. restricting spatial and temporal variability by sampling in the same location at the same time, and maintaining similar field methods, calibration procedures and volunteer training), could employ community-based monitoring data collection. CBM could be used as a tool to provide meaningful data for various environmental problems with a high degree of confidence in the accuracy of the data. The use of water temperature data in citizen science monitoring programs could be instrumental for determining habitat potential for various aquatic species including macroinvertebrates, fish, and amphibians by providing a cost-effective and reliable source of data; for example, water temperature was one of the primary criteria used by governments to define habitat requirements for fish species, guiding habitat protection measures (Plumb & Blanchfield, 2009).

In contrast to the parameters displaying robust data accuracy, the DO measurements collected by the treatment group, both percent saturation and milligrams per liter, appear to be the most inaccurate of the water quality parameters observed in this study, showing a significant difference from the treatment and control field samples. Overall, the dissolved oxygen displayed significant variability in field measurements, which is expected resulting largely as this parameter is influenced by water temperature, plant growth, field procedures and environmental characteristics of the river or stream such as water flow. This parameter may not be as an appropriate parameter to be used for CBM as compared the other parameters examined in this study, due to the natural high variability of this parameter and more complex field procedures required.

4.1.2 General Patterns

By observing the results of the comparison tests, general patterns have emerged including data distribution and sources of potential errors and bias. As noted by Thomas and Juanes (1996), given a large enough sample size, any statistical hypothesis test is likely to result in statistical significance. The statistical analysis of the Wilcoxon signed-rank test was influenced by the sample size. Due to the large sample size of the study, the statistical power was very large, which in turn made it easier to detect a smaller difference in the measurements and potentially creating type I error and falsely rejecting the null hypothesis.

Further sources of error and bias were identified as a result of calibration process. Observational data highlighted a few challenges in confirming the completed calibration process, as the act of observing in itself could have influenced the participants' behaviours and actions, despite using unobtrusive direct observation methods. In one instance, the participant required additional physical space to complete the calibration process, therefore reducing the ability to fully observe the procedure or confirm calibration. Also the act of observing created some examples of anxiety in some participants and though the researcher indicated that they would not be able to communicate with the participant while calibrating, some participants demonstrated reduced independent actions and attempted to ask for help during the procedure. Although it is difficult to determine all errors in calibration, observational data did strongly suggest that participants with previous experience preforming equipment calibration prior to the study had more confidence in preforming the task without aid and had less instances of reporting errors (Table 3.6, Table 3.7 and Table 3.8: Appendix B).

4.1.3 Interpretation

By observing the essential characteristics, which determined whether a parameter was accurate or not accurate, limitations and areas of improvement were highlighted for further research (Table 3.5). The most variable water quality parameter in this study was DO, which may have been a result of the natural high spatial variability of this parameter in a cross-section of a river (Marron & Blanchard, 1995). The study identified where sources of variability and potential limitations for this parameter could exist including: high cross-sectional variability, mechanical and procedural limitations. DO is highly variable as it is influenced by many factors including salinity, wastewater discharges, reduced water flow, water temperature and excessive plant growth (Wilding, Brown, & Collier, 2012); therefore, the spatial and temporal scale of sampling was very important. For example if one sensor was placed in a division of the cross-section of the stream with a different flow velocity from the other sensor, then DO readings may have been significantly different despite reducing the temporal variability in this study. Procedural variability can result from a number of sources, in particular with respect to the placement of the YSI ProPlus sensor in a stream. This sensor required fresh water to flow across the membrane while sampling for an accurate reading; however, if the movement of water is too slow then the probe would require an up-and-down movement through the water column (YSI, 2011). Some participants in the treatment group did not perform this field sampling troubleshooting procedure where applicable, however it was noted that some sample sites made this procedure difficult, with shallow water without the required depth for vertical movement of the sensor (Table 3.7: Appendix B). These limitations underlie the need for simple equipment and procedures in community-based monitoring. The recording of dissolve oxygen data using the YSI ProPlus is simple, lower complexity of the procedural field methods and calibration, as well as more extensive field training may be needed. To address the cross-sectional variability further recommendations could also include taking multiple samples or collecting representative water samples (Marron & Blanchard, 1995), while a detailed monitoring program design could stress the importance of where to sample, with thorough notes on the sensor placement at each sampling site.

In further analysis of the water temperature comparison data (Figure 3.2), graphical evidence displayed the distribution of the temperature measurements of the treatment group plotted versus the control group, which followed closely to a 1:1 ratio. With 100% of the difference of the treatment and control groups temperature measurements falling within the Nova Scotia Environment's acceptable range of 2.0 units and 93.3% within 0.2 units of one unit's equipment sensor drift. The essential characteristics, which resulted in this parameter showing robust data accuracy, included the reduced temporal variability of sampling methods and high accuracy of the equipment. By reducing the time between samples collected by the two groups and considering that no calibration of the temperature probe was required, the potential sources of error resulting from environmental conditions and human error was reduced.

One objective of this study was to determine the influence of water quality data accuracy related to proper calibration and previous experience in field sampling. The sources where accuracy of this parameter could be limited included the environmental conditions of sampling, and procedural and calibration limitation. By analyzing the difference of pH measurement into three subsections based on natural breaks in the data (<6, 6-7.5 and >7.5), further interpretations can be made on the significance of environmental conditions to the accuracy of this parameter. Through the examination of the confidence interval, the lower pH magnitude subsections were within +/- 0.2 pH units and therefore the majority of the data spread fell within the acceptable mechanical drift. The results also indicated some data outliers in the pH measurements identified as a result of procedural error with reduced stabilization time by the treatment group (Figure 3.14). Observational data suggested that potential calibration error might have occurred for the pH measurements at sample site #17 as a result of reduced confidence in the calibration process and a lack of training and experience with water quality monitoring, although this participant did have a scientific education background. Overall, only one out of eighteen calibration events had a reported calibration error, and although no field experience or prior knowledge of water science existed, the participant did have background in a non-related scientific field (Table 3.6 and Table 3.7: Appendix B).

The conductivity dataset displayed little variability when comparing the volunteer and professional datasets. This is a result of quality assurance quality control measures, with calibration prior to each sampling site, and little variability of the conductivity measurements through the transects in each sampling site. One source of potential data error for conductivity was noted from observational data as a result of procedural and mechanical limitation. As the conductivity port sensor in the YSI ProPlus is located higher on the sensor sonde, it is possible for an operator to not submerge the sensor completely in the water column, leading to a false measurement in the conductivity measurement (Table 3.7: Appendix B). Through developing field experience and further training, a volunteer should be able to utilize reasoning skills in field sampling to identify when this issue occurs.

The discharge and bankfull discharge measurements displayed robust data accuracy, as no significant difference was found between the treatment group measurements compared to the “true value” of the control group. However, sources of data error were identified, as the outliers of the sampling data appeared to be caused from sampling and recording error (Table 3.7: Appendix B). As the data collection of bankfull discharge measurements does include more subjective observational data collection, more events of human error could be expected. The procedural differences as discussed by Gillett *et al.* (2011) may occur during sampling, leading to discrepancies in measurements. The mechanical limitation of the velocity meter may also account for some data error, as the Global Flow probe used in this study required the volunteer to reset the meter prior to every sample; however, observation data noted a sampling event where the treatment group participant had troubleshooting concerns with the equipment as a result of not resetting the unit (Table 3.7: Appendix B). As this equipment is quite simple, further field practical training could address this limitation.

4.2 Conclusions

As previously discussed, utilizing citizen scientists for the collection of data can answer a multitude of questions relating to various fields of study. Whether it is climate-driven changes such as the butterfly trends as a result of amateur naturalists' observations (Breed *et al*., 2012), or tree budding assessments based on citizen science in-field measurements (Schultz, 2013), to benthic macroinvertebrates species abundance (Fore *et al.*, 2001), this form of data gathering offers a wider range of knowledge of the current state of the world. However, this data has not been used in official data reporting by government agencies until recently as the issue of data reliability was still being considered (Breed *et al.*, 2012).

This project sought to examine if volunteer citizen scientists collected water quality data that is significantly different from the data collected by professionals by examining if the difference of measurements were within the expected mechanical error and compared with the government data rejection criteria. The theoretical and comparative literature examining citizen science reliability was inconclusive on several questions in the field of water science. The study sought to answer two of these questions:

1. Do volunteer citizen scientists collect data that is significantly different from the data collected by professionals?
    i. How does the difference of the data collected by citizen scientists and a professional scientist compare to the in situ measurement variability of the instrument?
    ii. How does the difference of the data collected by citizen scientists and a professional scientist compare to provincial, national and international data rejection criteria?
2. What factors can improve the ability of citizen scientists to collect accurate water quality data?

The implications of these results can be related to environmental policy as the question of whether integration with citizen science data and government run programs is possible. If water science is to progress to a level where a full understanding of watershed health is to exist, utilizing available resources should be explored. Citizen science provides one available route to utilize volunteer hours to collect valuable scientific data. However a certain level of reliability would be required for government to accept this data as accurate. It may be a possibility to utilize citizen science for very specific tasks of measurements or observational data collection based on chosen parameters that display the least variability or subjective measurements.

As discussed in previous literature, data collection that has inherent variability based on the methodology used in a particular study may not be appropriate for utilizing volunteer hours. However, in this study results have highlighted specific water quality parameters that display low variability and high accuracy when compared to a control group. This provided information on how to proceed with future research and collaborative approaches to earth science data collection.

Prior to this study, a complete understanding on how accurate citizen science data in the field of water quality was not cemented, as comparative studies previously conducted did not include side-by-side measurements therefore temporal and spatial variability in water quality data existed. This research aimed at reducing the sources of potential error in data sampling that could be resulting in citizen science data to be considered unreliable.

4.2.1 Main Findings

In this study, results have emphasized specific water quality parameters that display low variability and high accuracy of citizen science water quality data. By examining the results, the water quality parameters that were within accuracy requirements of the mechanical and government criteria noted in Chapter Two included water temperature, pH, conductivity, and discharge. Therefore, leading to the conclusion that citizen scientists can collection water temperature, pH, conductivity and discharge measurements that not statistically different from a professional scientist when utilizing the correction criteria of government agencies. These parameters would be acceptable to use in a citizen science based monitoring program, utilizing trained volunteers, for the purposes of integration with professionally collected data.

The parameter that has shown a significant difference from the treatment group and control group measurements is dissolved oxygen and therefore would require further training on equipment use and field sampling procedures. The relationship between a “reliable” parameter for data collection by citizen scientists may need to be further broadened as government rejection criteria in itself does not take into account operator error, but rather taking into consideration the mechanical and potential spatial variability of this parameter.

The second research question of this study involved determining the factors can improve the ability of citizen scientists to collect accurate water quality data. This question examined observational data the researcher collected during the course of this study by examining training, calibration process and field sampling design. An assumption of this study was that the level of training provided in this study would be sufficient to claim that the volunteer was “trained”. When comparing literature, the term “trained” is not clearly defined. This made it difficult to determine what level of training is necessary to increase the accuracy of citizen science data to a level to be considered reliable. Through observational data, the online and field training may need to incorporate different types of learning styles with an emphasis to hands-on-training, in particular with use of equipment in the field and calibration. Observational data indicated the confidence in the treatment group participant who had prior hands on knowledge of the equipment prior to the study was greater than the first time user of the equipment. In particular, the calibration process, although a written manual and verbal explanation was provided, hands on training seemed to be preferred by participants (Table 3.7: Appendix B). As the velocity flow meter was the piece of equipment that had more instances of reporting errors, as participants had to reset the meter prior to each sample, more field practical training with this equipment may be necessary (Table 3.7: Appendix B).

4.2.2 Limitations and Recommendations

This study experienced a number of limitations that often affect research projects. The ideal scenario for this comparison study would be to increase the number of professionals comparing with the volunteer treatment group in order to increase statistical power of the overall results. However, a comparison between the professionals would also be needed demonstrate the comparability between professionals. This presents an additional research question, "Are professional scientists comparable within a set range of accuracy standards?" Although this is a valid research question, it broadens the scope of this study and with funding and resources limited, this question will need to be addressed in future studies. As time was also limited, both with researchers and volunteer participants in the treatment group, optimizing the volunteers' time was key, and therefore training, and field sampling was limited to half a day for each participant. Although longer field practical training sessions would have been ideal, training was limited to an online course prior to sampling and field training ranging from 1-2 hours prior to sampling.

Through the limitations and results presented in this study, specific recommendations for future research projects and collaboration with government agencies have become evident. A focus for future studies should be directed to address data reliability of professional scientists by analyzing the human operator error in field surface water sampling. The recommendations for future citizen science monitoring programs include examining the level of training available to volunteers. As noted in the observational data, there were deviations from the study field procedures, which could lead to increased uncertainty in the measurements, and the importance of training was noted when evaluating accuracy of participants (Table 3.7 and Table 3.8: Appendix B).

A proposed solution to address the concerns with various aspects of a monitoring program design such as: site selection of monitoring sites; regular calibration of equipment; and proper calibration records to ensure QA/QC standards for auditing purposes, could be addressed through a government managed program, modeled off of similar projects such as the CAMP program run by the Department of Fisheries and Oceans Canada (Fisheries and Oceans Canada, 2012; Weldon, Courtenay, & Garbary, 2007). With government scientists conducting data analysis and coordinating data procurement, citizen scientists can be used as a resource for data gathering while achieving high data accuracy.

4.2.3 Final Conclusions

Despite the reported and theoretical benefits of citizen science as a source of data collection, on-going debate continues to exist as to whether or not the level of the quality of the data is adequate enough to be integrated with the efforts of professional scientists (Gillett *et al.*, 2011). By analyzing side-by-side in-situ water quality data, we conclude that volunteer citizen scientists can collect water quality data that is not significantly different from that gathered by professionals. The selection of ideal parameters and comprehensive training is necessary. Further research will be necessary to clearly define the standard level of training required for basic field sampling. With current government environmental monitoring funding cuts (Au *et al.*, 2000), citizen science is both beneficial and essential to continued scientific research. The small degree of difference found in this study between citizen scientists' data compared with the professional's further fuels the need to apply this means of cost effective data gathering. With the lack of funding available for environmental monitoring, it is imperative that government agencies make use of the skills offered through volunteer-based initiatives.

On a broader scope, citizen science as a source for data gathering has potential for various field of study around the world; however, in all events, the variability of the parameters being examined and the complexity of the field methods required of the volunteers should be acknowledged and evaluated to determine where the most applicable use of a volunteer’s time should be used. The results of this study also suggest that monitoring program design and highly technical or subjective measurements may not be suitable to citizen science researchers. Although not all water quality parameters would be suitable for citizen scientist based programs, selected parameters with reduced variability and simple methods of data collection that provide indicators of ecosystem health or habitat suitability can be useful data for government agencies and research programs. The same approach can be applied to other topics to address various environmental issues and increase the overall scientific knowledge base of the global research community.

References

Arvanitidis, C., Faulwetter, S., Chatzigeorgiou, G., Dailianis, T., Pafilis, E., Chatzinikolaou, E., Fanini, L., Vasileiadou, C., Pavloudi, C., Koulouri, P., Dounas, C., Penev, L., Banki, O., Kouratoras, M., & Vavilis, P. (2011). Engaging the broader community in biodiversity research: The concept of the COMBER pilot project for divers in vibrant. *Zookeys, 150,* 211-229.

Asano, T. (2009). Chapter One: Water Issues: Current Status and the Role of Water Reclamation and Reuse. *Water reuse: Issues, technologies, and applications*. (pp.1-36). New York, NY: McGraw-Hill.

Au, J., Bagchi, P., Chen, B., Martinez, R., Dudley, S. A., & Sorger, G. J. (2000). Methodology for public monitoring of total coliforms, Escherichia coli, and toxicity in waterways by Canadian high school students. *Environmental Management, 58*, 213-230.

Beaubien, E. G., & Hamann, A. (2011). Plant phenology networks of citizen scientists: recommendations from two decades of experience in Canada. *International Journal of Biometeorology, 55,* 6, 833-841.

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen science: A developing tool for expanding science knowledge and scientific literacy. *Bioscience Washington, 59*(11), 977-984.

Breed, G. A., Stichter, S., & Crone, E. (2012). Climate-driven changes in northeastern US butterfly communities. *Nature Climate Change, 3,* 142-145.

Buzzelli, C., Akman, O., Buck, T., Koepfler, E., Morris, J., & Lewitus, A. (2009). Relationships among Water-Quality Parameters from the North Inlet Winyah Bay National Estuarine Research Reserve, South Carolina. *Journal of Coastal Research, 20*(45), 59-74.

CABIN (Canadian Aquatic Biomonitoring Network) (2010). *Field manual: wadeable streams.* Environment Canada.

Cavalcante, R. M., Sousa, F. W., Nascimento, R. F., Silveira, E. R., & Freire, G. S. (2009). The impact of urbanization on tropical mangroves (Fortaleza, Brazil): evidence from PAH distribution in sediments. *Journal of Environmental Management, 91*(2), 328-335.

CCME (Canadian Council of Ministers of the Environment) (2007). *Canadian water quality guidelines for the protection of aquatic life.* Canadian Environmental

Quality Guidelines. Winnipeg: Canadian Council of Ministers of the Environment.

CCME (Canadian Council of Ministers of the Environment) (2011). *Protocols manual for water quality sampling in Canada.*

Chapman, D. V., Unesco., World Health Organization., & United Nations Environment Programme. (1996). *Water quality assessments: A guide to the use of biota, sediments, and water in environmental monitoring*. London: E & FN Spon.

Charlton, R. (2008). *Fundamentals of fluvial geomorphology*. (pp. 1-234). New York, NY: Routledge.

Choi, B.C.K., a& Pak, A.W.P. (2004). A catalog of biases in questionnaires. *Preventing Chronic Disease*, *2*(1), A13.

Cohn, J. P. (2008). Citizen Science: can volunteers do real research? *Bioscience*, 58(3), 192-197.

Conover, W.J. (1973). On methods of handling ties in the Wilcoxon Signed-Rank test, *Journal of the American Statistical Association*, *68*(344), 985-988.

Conrad, C. & Daoust, T. (2008). Community-based monitoring frameworks: increasing the effectiveness of environmental stewardship. *Environmental Management*, *41*, 358-366.

Conrad, C. & Hilchey, K. (2011). A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring Assessment, 176*, 273-291.

Crichton, N. (2000). Information point: Wilcoxon signed-rank test. Blackwell Science Ltd, *Journal of Clinical Nursing, 9,* 584.

Culp, J. M., Cash, K. J., Halliwell, D. B., & Fraser River Action Plan (Canada). (1997). *Volunteer-based monitoring program for the Salmon River basin: Using benthic indicators to assess stream ecosystem health*. Saskatoon, Sask: National Hydrology Research Institute, Environment Canada.

Dennis, I., Scruton, D., Gilliss, M., & Clair, T. (2007). Freshwater acidification research in Atlantic Canada: a review of results and predictions for the future. *Environmental Reviews, 15,* 153-167.

De Souza, M. (2013, 03 13). Federal budget cuts undermine environment canada's mandate to enforce clean air regulations: emails. *National Post*. Retrieved from http://news.nationalpost.com/2013/03/17/federal-budget-cuts-undermine-environment-canadas-mandate-to-enforce-clean-air-regulations-emails/

EPA (Environmental Protection Agency) (1995). *Volunteer stream monitoring: a methods manua*l (EPA 841 D95-001). Office of Wetlands, Oceans and Watersheds, 4503F Washington DC 20460.

EPA (Environmental Protection Agency) (2006). Chapter 7: In Field Monitoring. In 2nd Ed, *Volunteer estuary monitoring: a methods manual.* US Environmental Protection Agency.

Fanning, E. (2005). Formatting a paper-based survey questionnaire: best practices. *Practical Assessment, Research and Evaluation, 10*(12), 1-14.

Fisheries and Oceans Canada. (2012). Community Aquatic Monitoring Program (CAMP). Retrieved March 13, 2013 from http://www.glf.dfo-mpo.gc.ca/e0006182

Fore, L. S., Paulsen, K., & O'Laughlin, K. (2001). Assessing the performance of volunteers in monitoring streams. *Freshwater Biology, 46*(1), 109-123.

Gillett, D., Pondella, D., Freiwald, J., Schiff , K., Caselle, J., Shuman,C., Weisberg, S. (2011). Comparing volunteer and professionally collected monitoring data from the rocky subtidal reefs of Southern California, USA. *Environmental Monitoring and Assessment, 183*,1-19. doi: 10.1007/s10661-011-2185-5

Global Water (2004). *Fp101-Fp201 Global Flow Probe User Manual.* Global Water, Gold River CA.

Government of Newfoundland & Labrador Department of Environment and Conservation (2012). *Real Time Water Quality Report Humber River at Humber Village*. pp. 2.

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675.

Guan, Y., Sterling, S., Garroway, K., & Kennedy, G. (2013). A New Watershed Assessment Framework to Improve Water Resource Planning in Nova Scotia. to be submitted to special Issue on Incorporating Water Resources in Integrated Urban and Regional Planning, Journal of Hydrology.

Halperin, M., Hamdy, M. I., & Thall, P. F. (1989). Distribution-free confidence intervals for a parameter of Wilcoxon-Mann-Whitney type for ordered categories and progressive censoring. *Biometrics, 45*(2), 509-21.

Harmel, R. D., Cooper, R. J., Slade, R. M., Haney, R. L., & Arnold J. G. (2006). Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *American Society of Agricultural and Biological Engineers*, *49*(3), 689-701.

Hart, B.T., Davies, P.E., Humphrey, C.L., Norris, R.N., Sudaryanti, S., & Trihadiningrum, Y. (2001). Application of the Australian river bioassessment system(AUSRIVAS) in the Brantas River, East Java, Indonesia. *Journal of Environmental Management, 62*, 93-100.

Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K., & Kelling, S. (2012). Data-intensive science applied to broad-scale citizen science. *Trends in Ecology & Evolution, 27,* 2, 130-7.

Howell, G. & El-Shaarawi, A.H. (1991). An overview of the acidification of lakes in Atlantic Canada. *Environmental Monitoring & Assessment*, *17*(2-3), 323-338.

Hurlbert, A. H., & Zhongfei, L. (2012). Spatiotemporal Variation in Avian Migration Phenology: Citizen Science Reveals Effects of Climate Change. *Plos One*, 7(2), 1-11.

Jordan, R. C., Gray, S. A., Howe, D. V., Brooks, W. R., & Ehrenfeld, J. G. (2011). Knowledge Gain and Behavioral Change in Citizen-Science Programs. *Conservation Biology, 25,* 6, 1148-1154.

Kremen, C., Ullman, K. S., & Thorp, R. W. (2011). Evaluating the Quality of Citizen Scientist Data on Pollinator Communities. *Conservation Biology,* 25(3), 607-617.

Lasker, R. D., & Weiss, E. S. (2003). Broadening participation in community problem solving: a multidisciplinary model to support collaborative practice and research. *Journal of Urban Health: Bulletin of the New York Academy of Medicine, 80*(1), 14-47.

Leopold, M., Cakacaka, A., Meo, S., Sikolia, J., & Lecchini, D. (2009). Evaluation of the effectiveness of three underwater reef fish monitoring methods in Fiji. *Biodiversity and Conservation, 18*(13), 3367-3382.

Loperfido, J. V., Beyer, P., Just, C. L., & Schnoor, J. L. (2010). Uses and biases of volunteer water quality data. *Environmental Science & Technology*, *44*(19), 7193-7199. doi:10.1021/es100164c

Marron, D. C., & Blanchard, S. F., 1995, Surface-water-quality assessment of the upper Illinois River Basin in Illinois, Indiana, and Wisconsin: Cross-sectional and depth variation of water-quality constituents and properties in the upper Illinois River Basin, 1987-88, U.S. Geological Survey Water-Resources Investigations Report 95-4021. 19 pp.

McDonald, J. (2009). Wilcoxon signed rank test. *Handbook of biological Statistics*. Retrieved April 23, 2013 from http://udel.edu/~mcdonald/statpaired.html.

Moore, D. (2008). Chapter 25: Nonparametric Tests. *The basic practice of statistics*. (5th ed.). (pp.1-38). New York, NY: W.H. Freeman and Company.

Napierala, M. A. (2012). What is the Bonferroni correction? *AAOS Now, 7(4). Retrieved April 23, 2013 from http://www.aaos.org/news/aaosnow/apr12/research7.asp.*

Nicholson, E., Ryan, J., & Hodgkins, D. (2002). Community data - Where does the value lie? Assessing confidence limits of community collected water quality data. *Water Science and Technology*, *45*(11), 193-200.

Nova Scotia Environment (2010a). *Nova Scotia Environment's Automated Surface Water Quality Monitoring Network: Data Analysis and Interpretative Report*, pp. 147.

Nova Scotia Environment (2010b). *Water for life: Nova Scotia's Water Resource Management Strategy*. 1-34.

Nova Scotia Environment (2013a). *Nova Scotia Lake Survey Program*. Retrieved January 15, 2013 from http://www.gov.ns.ca/nse/surface.water/lakesurveyprogram.asp

Nova Scotia Environment (2013b). *Surface Water Quality Monitoring Network Data*. Retrieved March 12, 2013 from http://www.gov.ns.ca/nse/surface.water/automatedqualitymonitoringdata.asp

Nova Scotia Museum of Natural History (1996). Topic 8.2 *Freshwater environments.* Topics and Habitats. pp. 157-169.

Ontario Ministry of the Environment (2010). *Provincial water quality monitoring network.* Retrieved on December 2, 2012 from http://www.ene.gov.on.ca/environment/en/monitoring_and_reporting/provincial_water_quality_monitoring_network/index.htm

Palmer, M., Allan, J. D., Meyer, J., & Bernhardt, E. S. (2007). River Restoration in the Twenty-First Century: Data and Experiential Knowledge to Inform Future Efforts. *Restoration Ecology, 15(*3), 472-481.

PASCO (2007). *Water Quality Field Guide.* Pasco Scientific, Roseville CA.

Pattengill-Semmens, C. V., Semmens, B. X., & Reef Environmental Education Foundation. (2003). Conservation and management applications of the REEF volunteer fish monitoring program. *Environmental Monitoring and Assessment, 81*.

Plumb, J., & Blanchfield, P. (2009). Performance of temperature and dissolved oxygen criteria to predict habitat use by lake trout (Salvelinus namaycush). *Canadian Journal of Fisheries and Aquatic Sciences, 66*(11), 2011-2023.

Pollock, R., & Whitelaw, G. (2005). Community-based monitoring in support of local sustainability. *Local Environment*, *10*(2), 211-228.

Pratt, J.W. (1959) Remarks on zeros and ties in the Wilcoxon signed rank procedures. *Journal of the American Statistical Association*, *54*(287), 655-667.

Prophet StatGuide (1997). Prophet StatGuide: Do you data violate Wilcoxon paired signed rank test assumptions? Retrieved April 15, 2013 from http://www.basic.northwestern.edu/statguidefiles/srank_paired_ass_viol.html#Skewness

Rahbar, M. H., Chen, Z., Jeon, S., Gardiner, J. C., & Ning, J. (2012). A nonparametric test for equality of survival medians. *Statistics in Medicine*, *31*(9), 844-54.

Roa Garcia, C.E., & Brown, S. (2009). Assessing water use and quality through youth participatory research in a rural Andean watershed. *Journal of Environmental Management*, *90*, 3040-3047.

Savan B., Morgan A., & Gore C. (2003). Volunteer environmental monitoring and the role of the universities: The case of Citizens' Environment Watch. *Environmental Management*, *31*(5), 561-568.

Schmeller, D. S., Henry, P., Julliard, R., Gruber, B., Clobert, J., Dziock, F., & ... Henle, K. (2009). Advantages of volunteer-based biodiversity monitoring in Europe. *Conservation Biology*, *23*(2), 307-316. doi:10.1111/j.1523-1739.2008.01125.x

Schultz, C. *(*2013*),* Citizen science finds spring could come a month early to the United States, Eos Trans. *AGU, 94(*15*),* 148.

Selvin, H. C., & Stuart, A. (1966). Data-dredging procedures in survey analysis. *The American Statistician*, *20*(3), 20-23.

Sharpe, A. & Conrad, C. (2006). Community based ecological monitoring in Nova Scotia: challenges and opportunities. *Environmental Monitoring and Assessment*, *113*, 395–409.

Sigua, G.C, & Tweedale, W. A. (2004). Assessing redesigned effectiveness of the water quality monitoring program in the Indian River Lagoon, Florida. *Aquatic Conservation: Marine and Freshwater Ecosystems, 14*(1), 49-64.

Silva, A. & Sacomani, L. (2000). Using chemical and physical parameters to define the quality of Pardo river water (Botucatu-Sp-Crazil). *The Journal of Water Resources, 35*(6),1609-1616. doi: 10.1016/S0043-1354(00)00415-2

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution, 24*(9), 467-471.

Spooner, I., Fenton, H., & Myers, M. (1998). Influence of regional geology and hydrogeology on river habitat quality: examples from Mill Brook and Elderkin Brook, Kings County, Nova Scotia. *Atlantic Geology, 34*, 185-19

Steinijans, V. W., & Diletti, E. (1983). Statistical analysis of bioavailability studies: parametric and nonparametric confidence intervals. *European Journal of Clinical Pharmacology, 24*(1), 127-36.

Tallaksen, L.M. (1994). A review of baseflow recession analysis. *Journal of Hydrology, 165*, 349-370.

TechNova. *Certified Technology Professionals.* Retrieved January 19, 2012 from http://www.technova.ca/certification.html

Thomas, L., & Juanes F. (1996). The importance of statistical power analysis: an example from Animal Behaviour. *Animal Behaviour, 52*, 856-859.

Tole, L. (2010). Reforms from the ground up: A review of community-based forest management in tropical developing countries. *Environmental Management*, *45*, 1312-1331.

Underwood, J.K., Ogden, J.G., Kerekes, J.J., and Vanghan, H.H. (1985). Acidification of Nova Scotia Lakes: III Atmospheric deposition of SO_4 and NO_3 and effects on urban and rural lakes. *Water, Air and Soil Pollution, 32*, 77-88.

UN (United Nations). (1992). Earth Summit Agenda 21: the United Nations Programme of Action from Rio. New York, New York, USA.

United States Environmental Protection Agency (USEPA). (2013). National directory of volunteer monitoring programs. Retrieved April 1, 2013 from http://yosemite.epa.gov/water/volmon.nsf/Home?openform

United Nations University Institute for Water, Environment and Health (UNU-INWEH). (2012). *The Global Water Crisis: Addressing an Urgent Security Issue.*

United States Department of Interior Bureau of Reclamation. (2001). Discharge-area-velocity relationship. In R. Dodge (Ed.), *Water Measurement Manual: A Water*

Resources Techinical Publication. Retrieved from http://www.usbr.gov/pmts/hydraulics_lab/pubs/wmm/

van Horen, B. (2001). Developing community-based watershed management in Greater São Paulo: the case of Santo André. *Environment and Urbanization, 13*(1), 209-222.

Wagner, R., Boulger, R., Oblinger, C., & Smith, B. (2006). *Guidelines and standard procedures for continuous water-quality monitors: station operation, record computation, and data reporting.* Retrieved November 1, 2012, from http://pubs.usgs.gov/tm/2006/tm1D3/

Weldon, J. W., Courtenay, S. C., & Garbary, D. J. (2007). The Community Aquatic Monitoring Program (CAMP) for measuring marine environmental health in coastal waters of the southern Gulf of St. Lawrence: 2007. Ottawa: Fisheries and Oceans Canada.

Wenner, E., Sanger, D., Arendt, M., Holland, A. F., & Chen, Y. (2004). Variability in dissolved oxygen and other water-quality variables within the national estuarine Research Reserve System. *Journal of Coastal Research*, *20*, 17-38.

Wet-Pro Certification. (2013). *A community perspective*. Retrieved from http://wetpro.smu.ca/

Whitelaw, G., Vaughan, H., Craig, B., & Atkinson, D. (2003). Establishing the Canadian community monitoring network. *Environmental Monitoring and Assessment*, *88*, 409–418.

Wilding, T. K., Brown, E., and Collier, K. J. (2012). Identifying dissolved oxygen variability and stress in tidal freshwater streams of northern New Zealand. *Environmental Monitoring and Assessment*, *184*(10), 6045-60.

Xu, Y. J., and Wu, K. (2006). Seasonality and interannual variability of freshwater inflow to a large oligohaline estuary in the Northern Gulf of Mexico. *Estuarine Coastal & Shelf Science*, *68*, 619-626.

YSI Inc. (2011). *Professional Plus: Handheld Multiparameter Instrument.* Retrieved December 12, 2011, from http://www.ysi.com/productsdetail.php?Professional-Plus-18 .

Zabiegala, B., Kot-Wasik, A., Urbanowicz, M., & Namiesnik, J. (2010). Passive sampling as a tool for obtaining reliable analytical information in environmental quality monitoring. *Analytical and Bioanalytical Chemistry, 396*(1), 273-296.

Appendix A

Table 1.0. List of community groups involved in study

Name of Community Group	**Year of Establishment**
Avon Peninsula Watershed Preservation Society	2006
Bluenose Coastal Action Foundation	1993
Clean Annapolis River Project	1990
Eastern Shore Forest Watch	1998
Lochaber Watershed Association	2012
Medway River Salmon Association	2007
Pictou County Rivers Association	1990
Shubenacadie Watershed Environmental Protection Society	1993
Sackville Rivers Association	1988
Tusket River Environmental Protection Association	1986

Table 2.3. Water quality monitoring parameters used in study

Parameter	Definition	Importance	CCME Guidelines
Dissolved Oxygen (DO)	The measure oxygen gas dissolved in water (mg/L). Varies with water temperature and air pressure.	DO concentration is critical for determining water quality and greatly influences aquatic ecosystems health.	Warm Water: 5.5 mg/L - 6 mg/L Cold Water: 6.5 mg/L - 9.5 mg/L
pH	The measure of hydrogen ion concentration in water. Determines how acidic or basic the solution is. Range from 0-14 (0-most acidic, 14- most basic)	pH is important for overall water quality and affects solubility of metals and nutrients. Aquatic life cannot handle extremely low or high pH values.	7.0-8.7 pH units
Conductivity	The measure of ability of water to conduct an electrical charge (µs/cm). It is dependent on the concentration of dissolved ions in the water.	Can be used to quickly estimate the amount of total dissolved solids (TDS) in the water by multiplying conductivity measurement by 0.5 for natural waters at 25°C.	Freshwater aquatic life desirable levels 150-500 µs/cm
Water Temperature	The measure of average energy resulting from movement of microscopic particles in a substance (°C).	Impacts abilities of life functions in aquatic organisms and human use of the water.	- Human activities should not change temperature by +/- 1 °C - Max. human induced water temperature change should not surpass 0.5 °C per hour.

Note: Author: Ashley Shelton, Modified from Sources: (CCME, 2007; PASCO, 2007)

Figure 2.2. Calibration sheet

YSI Calibration Sheet

<table>
<tr><th colspan="4">Calibration Sheet</th></tr>
<tr><td>Date/Time</td><td colspan="3"></td></tr>
<tr><td>Name of Operator</td><td colspan="3"></td></tr>
<tr><td>Sonde Serial Number</td><td colspan="3"></td></tr>
<tr><td>Parameter</td><td>Buffer Standard Used</td><td>Pre-Calibration</td><td>Post-Calibration</td></tr>
<tr><td>Specific Conductivity (µS/cm)</td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">pH 4</td><td rowspan="2"></td><td></td><td>N/A</td></tr>
<tr><td colspan="2">mV:</td></tr>
<tr><td rowspan="2">pH7</td><td rowspan="2"></td><td></td><td></td></tr>
<tr><td colspan="2">mV:</td></tr>
<tr><td rowspan="2">Dissolved Oxygen (% Sat)</td><td rowspan="2">N/A</td><td></td><td></td></tr>
<tr><td colspan="2">Temp:</td></tr>
<tr><td>Observations/Comments</td><td colspan="3"></td></tr>
</table>

Figure 2.3. Field sheet

Channel Measurements

!

General Information

Name: ______________ Site Code: ______________ GPS Coordinates: ______________

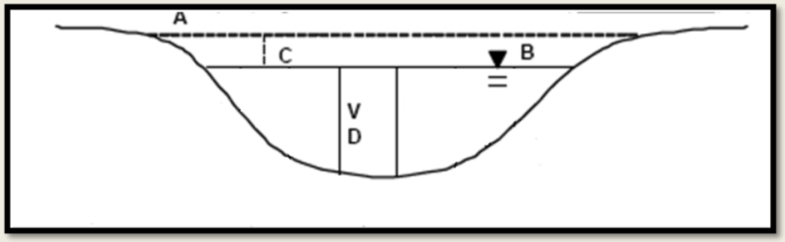
Sampling Date (DD/MM/YY): ______________ Weather Conditions: ______________

Local Land Use: ______________

Channel Data

A. Bankfull Width (m) ______________ B. Wetted Stream Width (m) ______________

C. Bankfull-Wetted Depth (cm) (height from water surface to bankfull) ______________



Velocity and Depth

i. Distance from Shore (m) ______________ ii. Depth (D) (cm) ______________

iii. Average Velocity (V) (m/s) ______________

!

Physical Water Monitoring Data

Sample Site #	Time Sampled (am/pm)	DO (mg/L)	DO (%)	Cond. (µS/cm)	Sp.Cond (µS/cm)	TDS (mg/L)	Temp. (^{o}C)	pH
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

Figure 2.4. Email correspondence

Good morning,

My name is Ashley Shelton and I am a graduate student of the Applied Science program at Saint Mary's University. As a part of my master's thesis, I am conducting research under the supervision of Dr. Cathy Conrad and Dr. Shannon Sterling.

As a follow-up to your previous involvement in the study evaluating accuracy of water quality monitoring data, I would like to invite you to participate in a brief survey.

The purpose of this survey is to determine your previous training and experience prior to your completion of the Wet-Pro online training course. This will be used as supplementary information to the main water quality study with the goal of examining if there is a connection with previous training and data accuracy and determining the effectiveness of the online Wet-Pro training course.

If you wish to participate please respond to environmental.network@smu.ca indicating your consent to participate and complete and attach the survey, which should take 15 minutes to complete.

I look forward to hearing from you.

Sincerely,
Ashley

Ashley Shelton
MSc. Candidate, **Saint Mary's University**

Saint Mary's University,
Department of Geography, Burke Bldg.
923 Robie Street,
Halifax, Nova Scotia, B3H 3C3
Phone: 902.491.6243
E-Mail: environmental.network@smu.ca

Figure 2.5. Informed consent form

INFORMATION/COVER LETTER

The Accuracy of Water Quality Monitoring Data: A Comparison Between Citizen Scientists and Professionals

REB File # 12-260

Ashley Shelton
Environmental Science
Saint Mary's University, 923 Robie Street, Halifax, NS B3H 3C3
Phone: 491-6243; Fax 496-8213;
Email address: environmental.network@smu.ca

Hello, my name is Ashley Shelton and I am a graduate student of the Applied Science program at Saint Mary's University. As a part of my master's thesis, I am conducting research under the supervision of Dr. Cathy Conrad and Dr. Shannon Sterling and through the funding of the Social Sciences and Humanities Research Council of Canada.

You are being invited to participate in a follow-up survey.

The purpose of this survey is determine your previous background and experience prior to the completion of the Wet-Pro online training course taken as a result of your involvement in the study evaluating accuracy of water quality monitoring data.

The goal of this survey is to determine any previous relevant work/volunteer experience, training and education. This will be used as supplementary information to the main water quality study with the goal of examining if there is a connection with previous training and data accuracy and determining the effectiveness of the online Wet-Pro training.

If you wish to participate please respond to environmental.network@smu.ca and attach your completed survey, which should take 15 minutes to complete. A response with a completed survey will indicate your consent to participate.

This survey will allow for a thorough analysis of the water quality data collected as a part of your involvement in the water study and can aid in an evaluation of the Wet-Pro online training. Overall the benefits of this research to participants can include potential for increased credibility to the data collected by volunteers and the potential for integration with government data. There is potential for increased knowledge on citizen scientists, and may lead to improvements to the Wet-Pro training course. The potential benefits to society include an increased awareness of environmental issues, and the potential of government agencies accepting volunteer data as credible, therefore leading to an increase fresh water quality data available to the general public and more informed decision making from government based on accurate scientific study.

There are no foreseeable risks associated with participation in this study and you are free to withdraw from the participating in the survey. Please contact environmental.network@smu.ca by November, 2012 if you wish to withdraw, to ensure that data analysis and reporting has not yet occurred. All survey results will be handled by the principal investigator, any reporting of data will maintain the privacy of the participants, and all names of participants will be kept confidential. The results of the study will be available to the public on the Community-Based Environmental Monitoring Network website (http://www.envnetwork.smu.ca/).

If you have any questions related to student research please contact:
Ashley Shelton, Principal Investigator: environmental.network@smu.ca

If you have any questions or concerns about ethical matters, you may contact the Chair of the Saint Mary's University Research Ethics Board at ethics@smu.ca or 420-5728.

Figure 2.6. Participant questionnaire

Participant Questionnaire:
Previous Education and Training

1) What is the name of your environmental community organization and your affiliation?

____

____

____

2) What was your experience with water quality monitoring programs prior to your involvement in this study? ____

____

____

3) Do you have any relevant education related to environmental monitoring? If yes, please list.

____

____

____

4) Do you any relevant training prior to the online Wet-Pro training? If yes, please list.

____

____

____

Appendix B

Table 3.6. Summary of treatment group participant survey responses

Site Code	Experience with Water Quality Monitoring (WQM) Programs, Relevant Education Related to WQM and Relevant Training
1	**Experience WQM Program:** Limited Experience **Education:** BSc. and diploma program **Training:** No training
2	**Experience WQM Program:** Limited Experience **Education:** BSc. **Training:** No training
3	N/A
4	N/A
5	N/A
6	**Experience WQM Program:** Work experience with maintaining equipment controlling water quality **Education:** No relevant education **Training:** No training
7	**Experience WQM Program:** Experience with CAMP program, main focus on fish. Some YSI temperature measurements. **Education:** BSc, some undergraduate research on lake water (field sampling and lab chemistry). Limnology, field ecology and aquatic invertebrates. **Training:** No formal training
8	**Experience WQM Program:** No experience with water quality monitoring programs. **Education:** MSc, with instruction in 1st year university course with topics on fresh water and environmental concerns. **Training:** No training in environmental water sampling
9	**Experience WQM Program:** No experience with water quality monitoring program. **Education:** PhD. No lab or field experience of any kind **Training:** No training
10	**Experience WQM Program:** No experience in water quality monitoring **Education:** MES, but not related to water quality monitoring **Training:** No training

Note: N/A responses refer to non-response of participant background survey

Table 3.6. (continued)

Site Code	Experience with Water Quality Monitoring (WQM) Programs, Relevant Education Related to WQM and Relevant Training
11	**Experience WQM Program:** Only 15 minute introduction to DO and E.coli sampling using a Van Dorn sampler. **Education:** No education related to environmental monitoring **Training:** No training
12	**Experience WQM Program:** Sampling for one sampling season **Education:** NSCC ENV Eng. Technology Diploma in Water Resources focusing on environmental monitoring and sampling for water quality and contamination. **Training:** Field and lab training
13	**Experience WQM Program:** Very little **Education**: BSc candidate in Environmental Science, theoretical education but very little hands-on experience. **Training:** No prior training
14	**Experience WQM Program:** Some training through college program and hands-on field training and lab experiences. **Education**: Some college and university training of env. monitoring, including hands on field training and in lab experiences. Course training in watershed ecology, riparian assessment, water quality techniques and survey equipment, dip netting and identifying invertebrates and other forms of monitoring (e.g. DFO CAMP). **Training:** YSI meter, electrofishing and field techniques (e.g. secchi disc)
15	**Experience WQM Program:** Not in a study this in-depth. Participated in Adopt-A-stream water temperature monitoring and used a YSI meter during the DFO CAMP program. Worked with community group program for 5 yrs. **Education:** BSc candidate in Environmental Science **Training:** No training in water sampling techniques
16	**Experience WQM Program:** Water quality monitoring and sampling with community group. **Education:** Chemical Engineering Candidate. Fundamentals of ENV Eng. course. **Training:** Education provided information on what was being monitored (pH, conductivity, turbidity) and previously used the water sampling equipment while water sampling with community group.
17	**Experience WQM Program:** Sporadic work exp. over the past 35 years. Currently monitoring water quality in lakes and streams for community group. **Education:** M.Sc. Biology **Training:** Some lab, university training and on-the-job practice.
18	**Experience WQM Program:** Five yrs. of community group water quality monitoring **Education:** MSc and further graduate study Environmental Physiology of Plants, though not in water quality **Training:** Periodic water quality testing

Table 3.7. Field and calibration observations

Site Code	Volunteer sampled first	Field Observations	Calibration Observations
1	No	• Troubleshooting: Volunteer handheld display froze. Volunteer requested aid from researcher to troubleshoot in the field. Removed batteries to re-start equipment.	• Participant asked many questions to researcher while performing calibration process. Researcher was not able to respond.
2	Yes	• Event #6: pH continued to slowly drop. Sample site was shallow, difficult to fully submerge sensors.	
3	No	• Volunteer was very thorough in checking channel measurements.	
4	Yes	• Troubleshooting: Volunteer re-measured at event #3 as a result of the large drop in pH measurement, no change in measurements was found.	• Troubleshooting: Performed the calibration process twice as participant thought an error occurred.
7	No		• Did not pour "used" solution into rinse bottles.
9	No	• Performed good bankfull measurement • Could not confirm the participant reset the velocity meter prior to taking measurement.	• Volunteer was very uncomfortable with researcher observing calibration. • Calibration process took approximately 1.5 hours. • Researcher could not observe the entire calibration process, as they needed personal space. • Researcher unable to confirm a complete calibration.

Table 3.7. (continued)

Site Code	Volunteer sampled first	Field Observations	Calibration Observations
10	Yes	• Had to move upstream from marker as depth was too shallow • Bankfull width measurement was estimated by participant, did not measure with tape.	
12	Yes		• Very confident with calibration procedure.
16	Yes	• When measuring bankfull width there was slack on the measuring tape. • Participant faced the wrong direction when taking velocity measurement but performed the correction and re-measured. • Took the velocity measurement 5 meters downstream from flagging tape.	
17	No	• Difficulty handling all gear necessary for channel measurements however performed field procedures well.	• Calibrated pH 7 twice
18	Yes	• Event #1: Probe not submerged enough to get full conductivity, specific conductivity and total dissolved oxygen measurements. Participant noted drop in field logbook but did not troubleshoot in field with a re-measurement. • Heavy rain by event #3.	• Did not press "Cal" to complete pH calibration.

Notes: Observations refer to actions performed by participants outside of standard procedures noted in training and site codes where no deviations from field and calibration procedures were noted were excluded.

Table 3.8. Wilcoxon signed-rank test results for water quality data divided by experience, education and training relevant to water quality monitoring compared to accuracy of volunteer

Water Quality Parameter (d_x)	No relevant experience		No relevant education		No relevant training	
	p-value	Significance *	*p*-value	Significance *	*p*-value	Significance *
d_T (°C)	• <0.001	• Yes	• 0.004	• Yes	• 0.001	• Yes
d_{pH}	• 0.202	• No	• <0.001	• Yes	• 0.001	• Yes
d_{DO} (mg L^{-1})	• 0.746	• No	• 0.001	• Yes	• 0.012	• Yes
d_{DO} (%)	• 0.884	• No	• 0.001	• Yes	• 0.006	• Yes
d_C (uS cm^{-1})	• 0.207	• No	• 0.021	• No	• 0.002	• Yes
d_{SPC} (uS cm^{-1})	• 0.050	• No	• 0.026	• No	• 0.001	• Yes
d_{TDS} (mg L^{-1})	• 0.798	• No	• 0.045	• No	• 0.005	• Yes
Water Quality Parameter (d_x)	**High relevant experience**		**High relevant education**		**High relevant training**	
	p-value	Significance *	*p*-value	Significance *	*p*-value	Significance *
d_T (°C)	• 0.315	• No	• 0.008	• No	• 0.702	• No
d_{pH}	• <0.001	• Yes	• 0.030	• No	• <0.001	• Yes
d_{DO} (mg L^{-1})	• 0.004	• Yes	• 0.011	• No	• 0.607	• No
d_{DO} (%)	• <0.001	• Yes	• 0.001	• Yes	• 0.690	• No
d_C (uS cm^{-1})	• 0.994	• No	• 0.001	• Yes	• 0.611	• No
d_{SPC} (uS cm^{-1})	• 0.788	• No	• <0.001	• Yes	• 0.368	• No
d_{TDS} (mg L^{-1})	• 0.798	• No	• 0.045	• No	• 0.005	• Yes

Note: (1) Values are omitted if equal to hypothesis median test value therefore “N for test” refers to sample size used for test
(2) * Using Bonforoni Correction for multiple testing, significance level α=0.0071

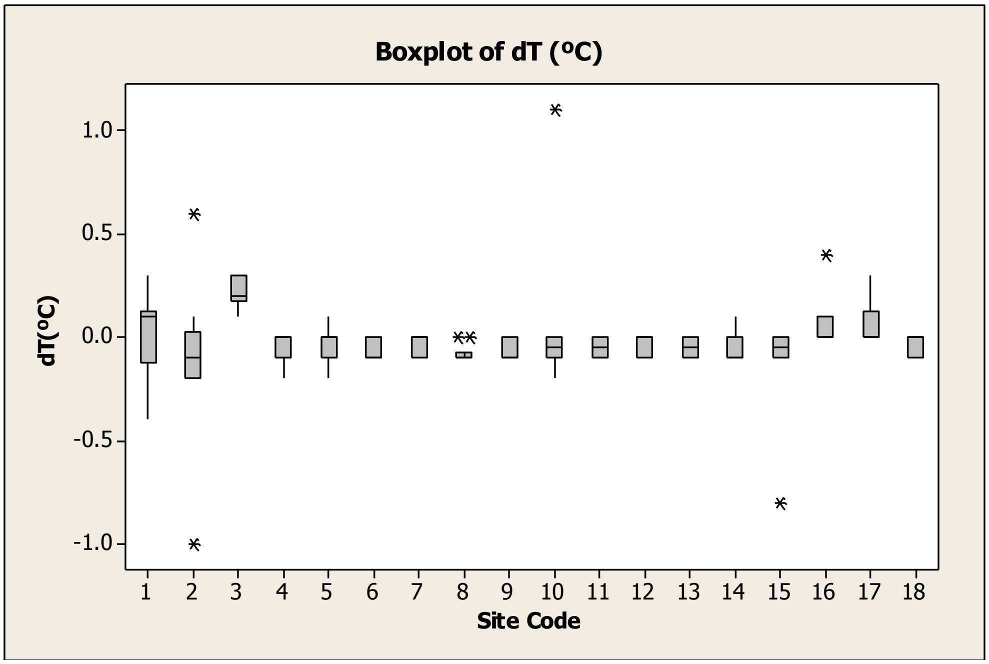


Figure 3.13. Difference in the water temperature (°C) (d_T) between the measurement collected from the treatment group and the measurement collected from the control group with whiskers marking the minimum and maximum data values for each sample site. The asterisks indicate outliers of the dataset.

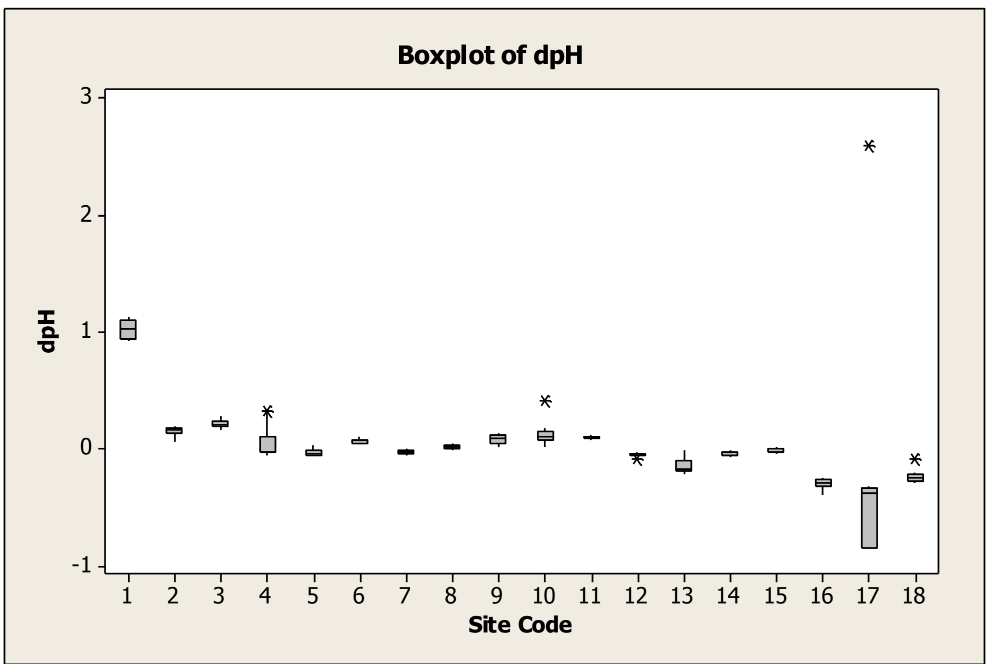


Figure 3.14. Difference in the pH data (d_{pH}) between the measurement collected from the treatment group and the measurement collected from the control group with whiskers marking the minimum and maximum data values for each sample site. The asterisks indicate outliers of the dataset.

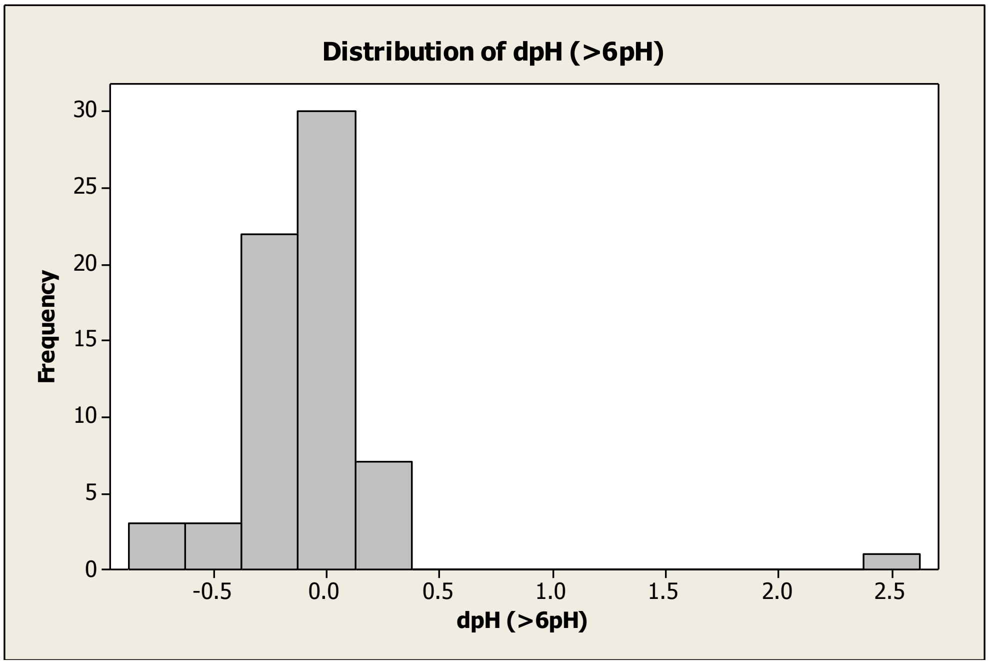


Figure 3.15. Difference in the pH data (d_{pH}) between the measurement collected from the treatment group and the measurement collected from the control group for all samples collected with the control group values less than 6 pH.

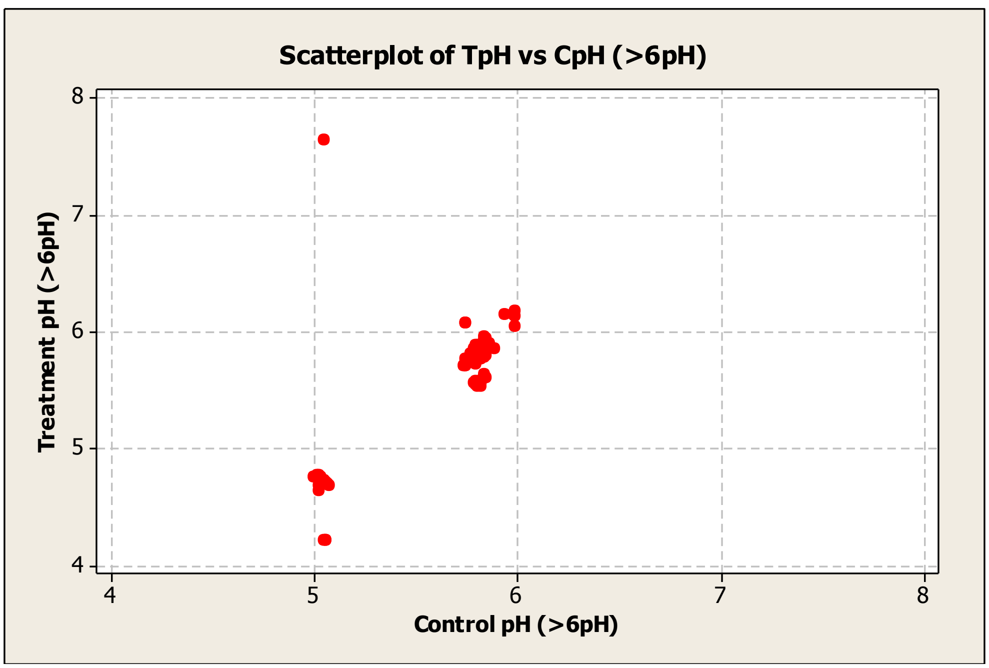


Figure 3.16. Scatterplot of control group pH measurements (C_{pH}) versus treatment group pH measurements (T_{pH}) for all samples collected with the control group values less than 6 pH.

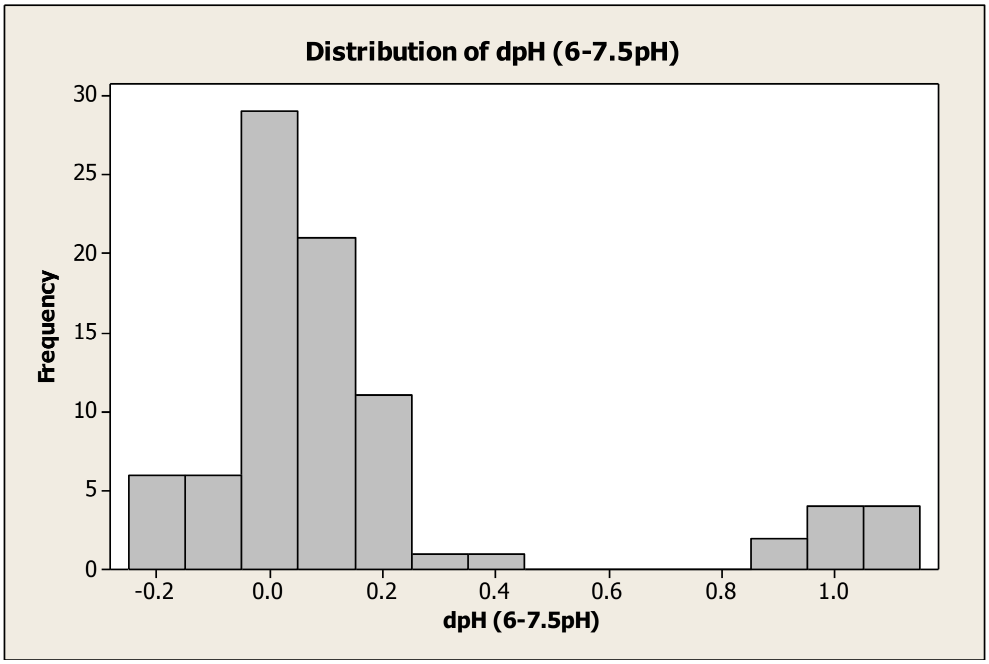


Figure 3.17. Difference in the pH data (d_{pH}) between the measurement collected from the treatment group and the measurement collected from the control group for all samples collected with the control group values between 6 and 7.5 pH.

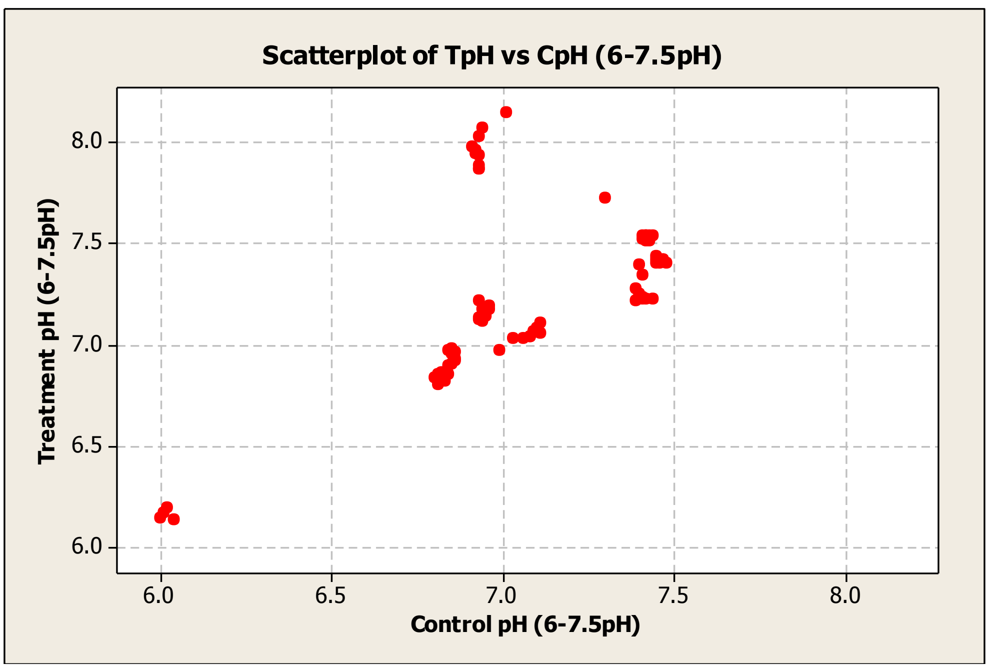


Figure 3.18. Scatterplot of control group pH measurements (C_{pH}) versus treatment group pH measurements (T_{pH}) for all samples collected with the control group values between 6 and 7.5 pH.

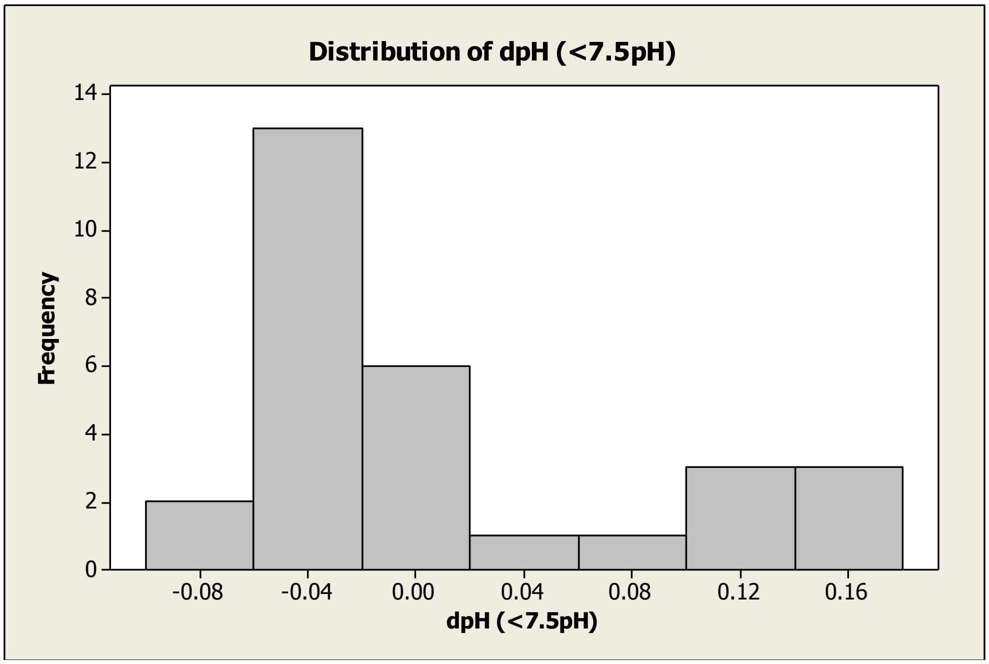


Figure 3.19. Difference in the pH data (d_{pH}) between the measurement collected from the treatment group and the measurement collected from the control group for all samples collected with the control group values greater than 7.5 pH.

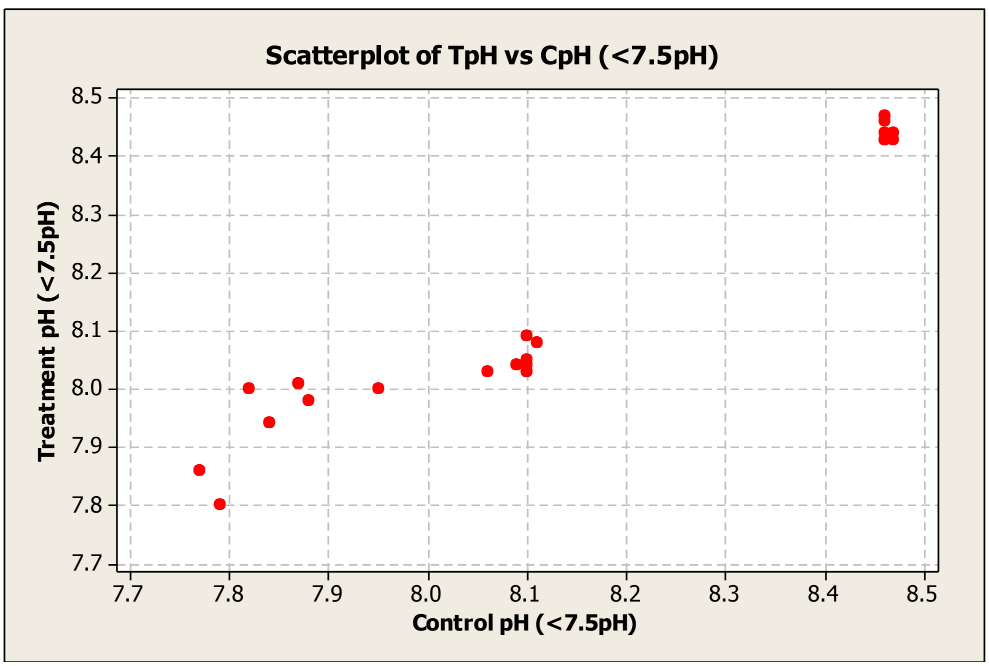


Figure 3.20. Scatterplot of control group pH measurements (C_{pH}) versus treatment group pH measurements (T_{pH}) for all samples collected with the control group values greater than 7.5 pH.

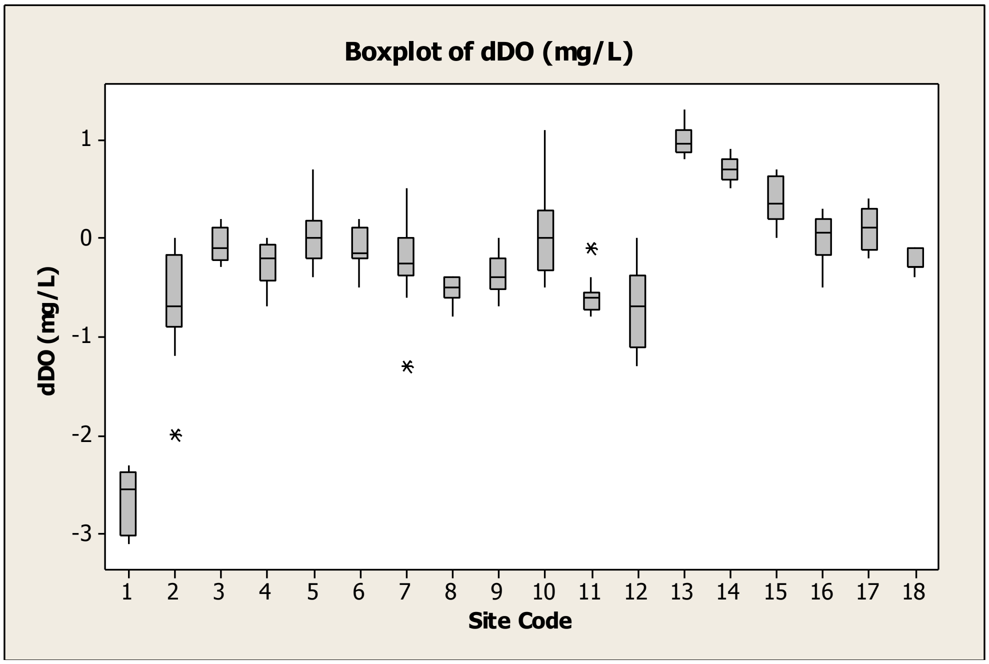


Figure 3.21. Difference in the DO (mg/L) data (d_{DO}) between the measurement collected from the treatment group and the measurement collected from the control group with whiskers marking the minimum and maximum data values for each sample site. The asterisks indicate outliers of the dataset.

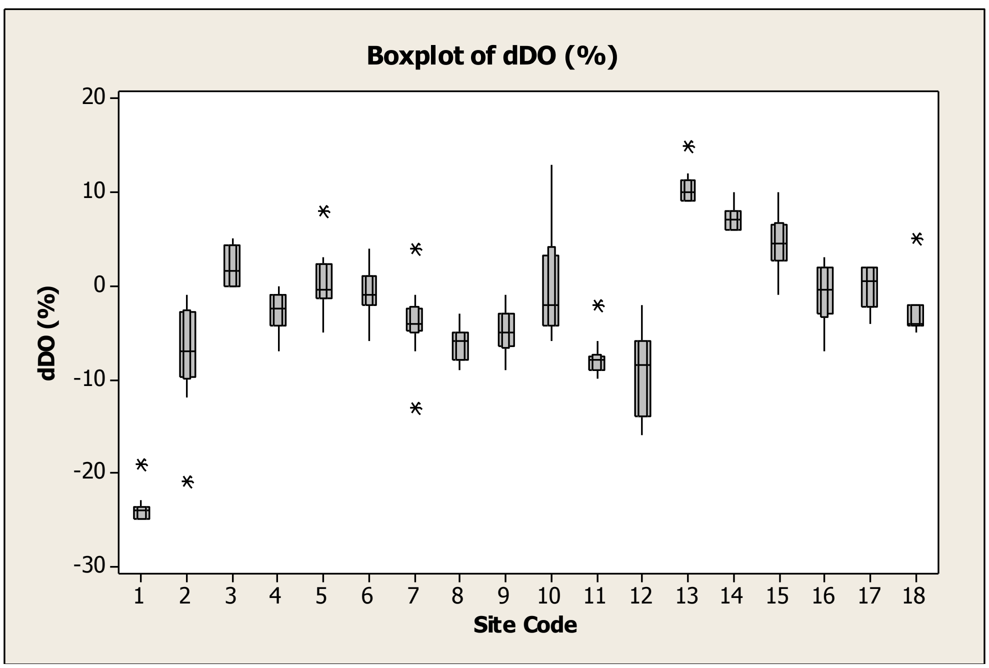


Figure 3.22. Difference in the DO (%) data (d_{DO}) between the measurement collected from the treatment group and the measurement collected from the control group with whiskers marking the minimum and maximum data values for each sample site. The asterisks indicate outliers of the dataset.

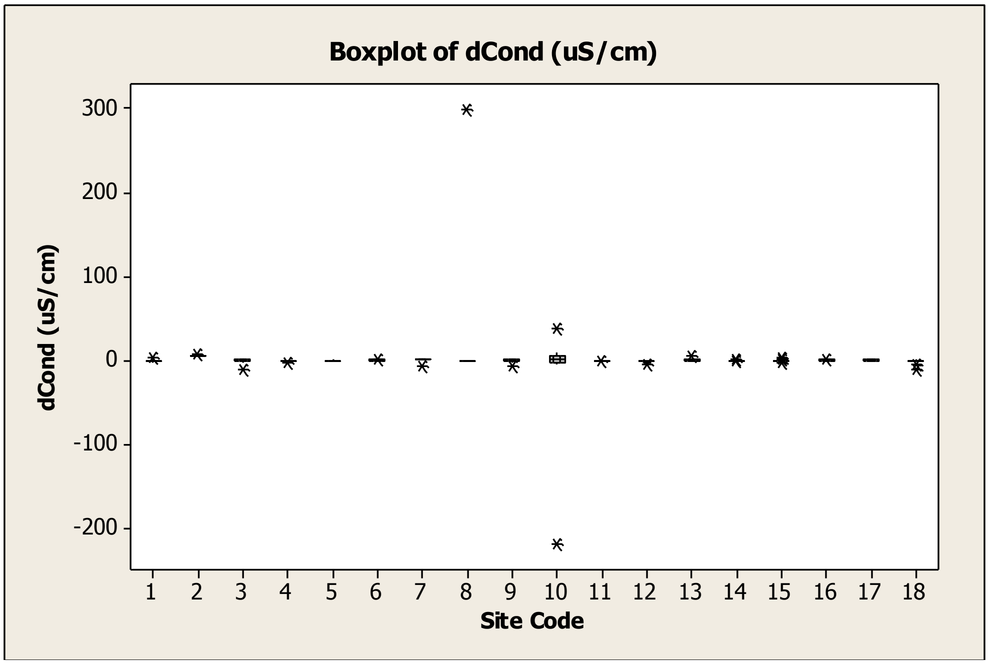


Figure 3.23. Difference in the conductivity (uS/cm) data (d_C) between the measurement collected from the treatment group and the measurement collected from the control group with whiskers marking the minimum and maximum data values for each sample site. The asterisks indicate outliers of the dataset.

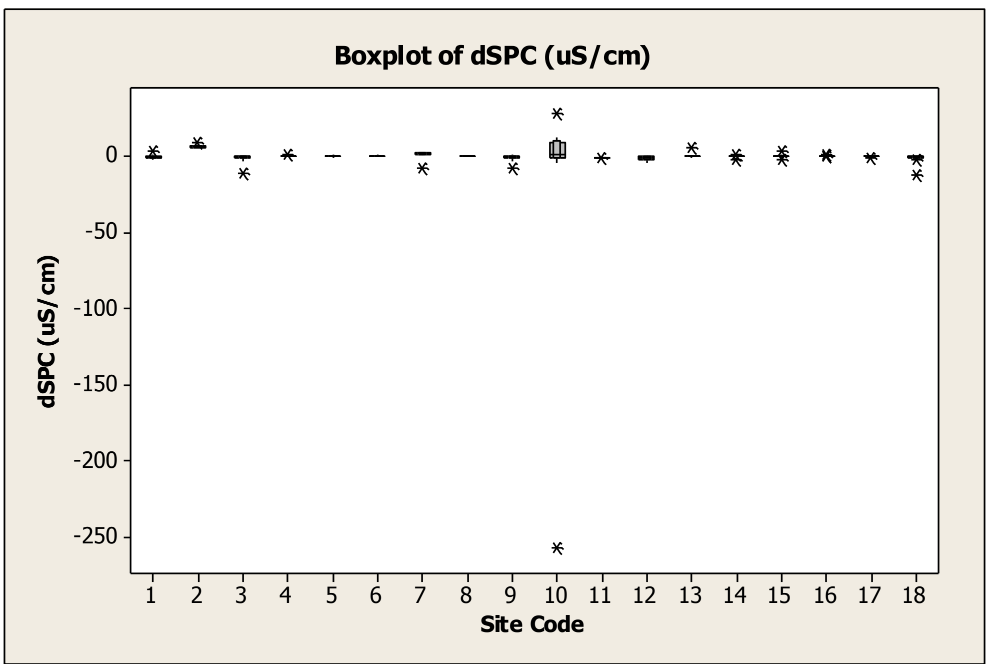


Figure 3.24. Difference in the SPC (uS/cm) data (d_{SPC}) between the measurement collected from the treatment group and the measurement collected from the control group with whiskers marking the minimum and maximum data values for each sample site. The asterisks indicate outliers of the dataset.

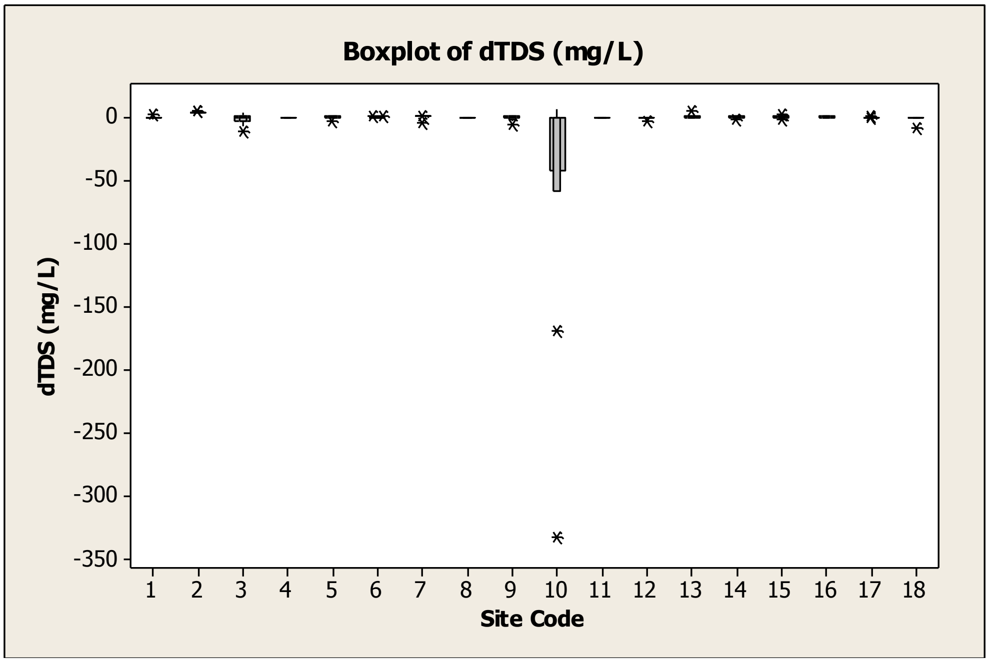


Figure 3.25. Difference in the TDS (mg/L) data (d_{TDS}) between the measurement collected from the treatment group and the measurement collected from the control group with whiskers marking the minimum and maximum data values for each sample site. The asterisks indicate outliers of the dataset.

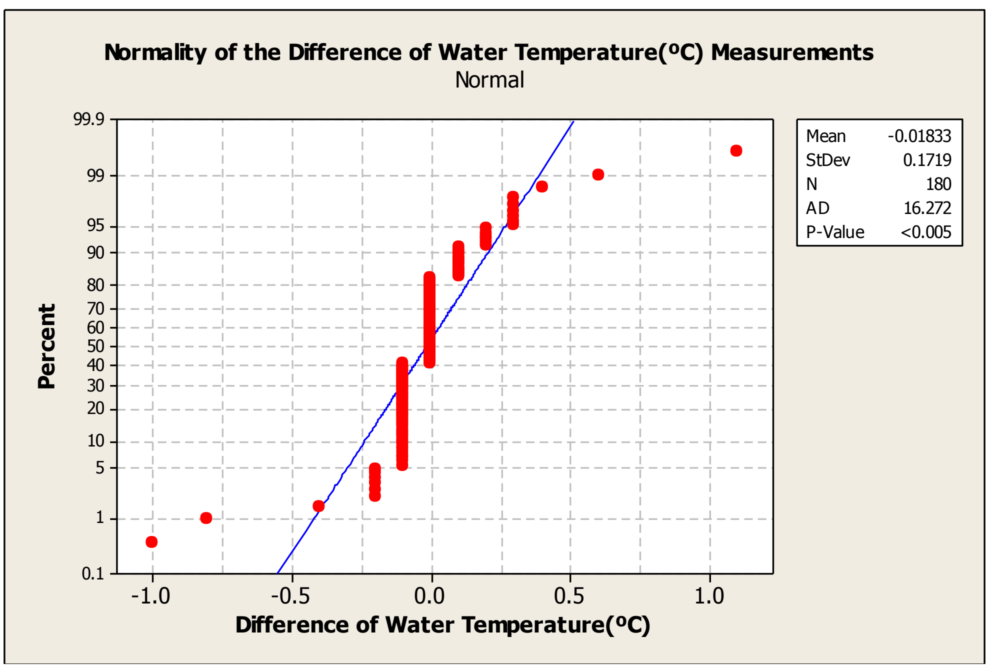


Figure 3.26. Normality of water temperature (°C) data

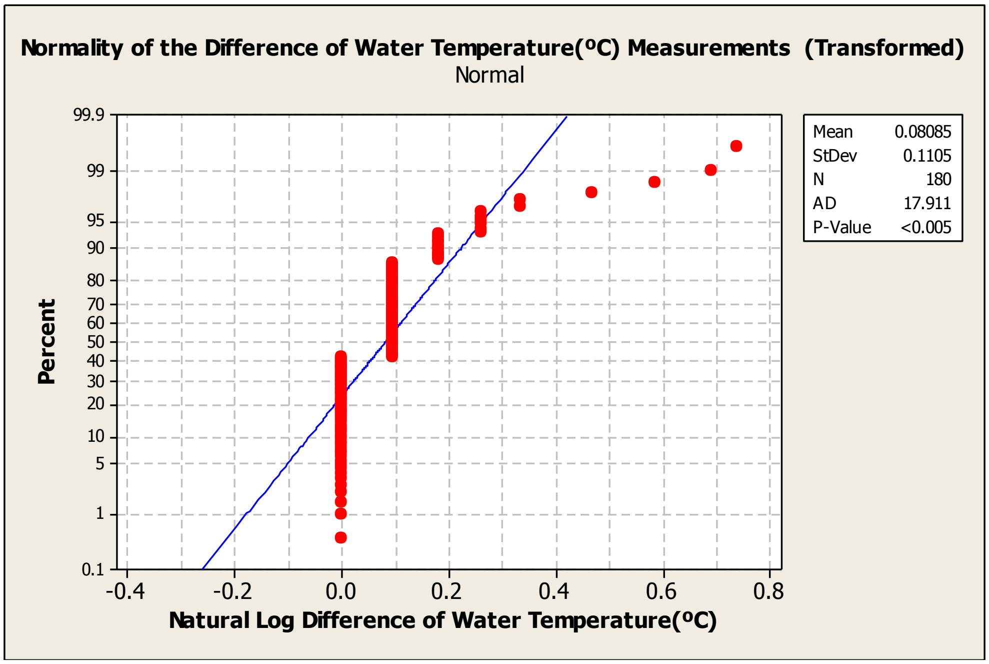


Figure 3.27. Normality of the transformed water temperature (°C) data

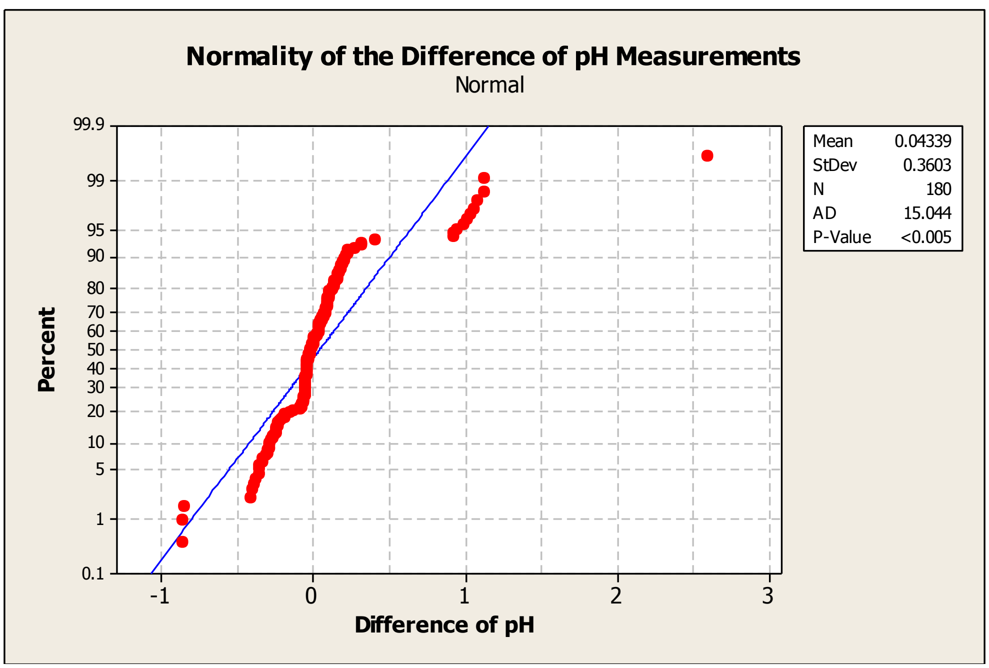


Figure 3.28. Normality of pH data

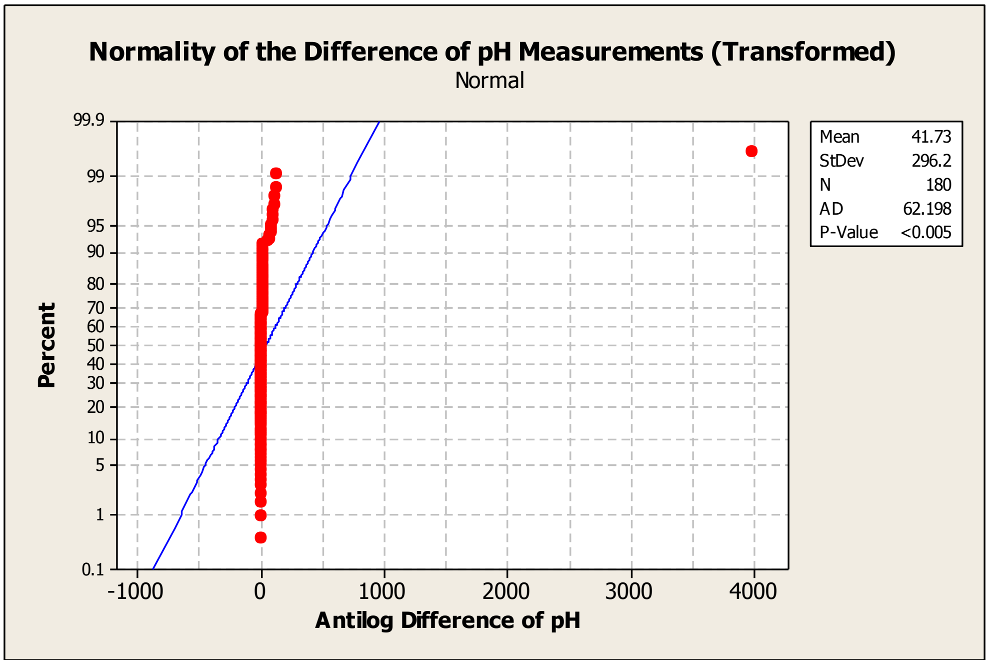


Figure 3.29. Normality of the transformed pH data

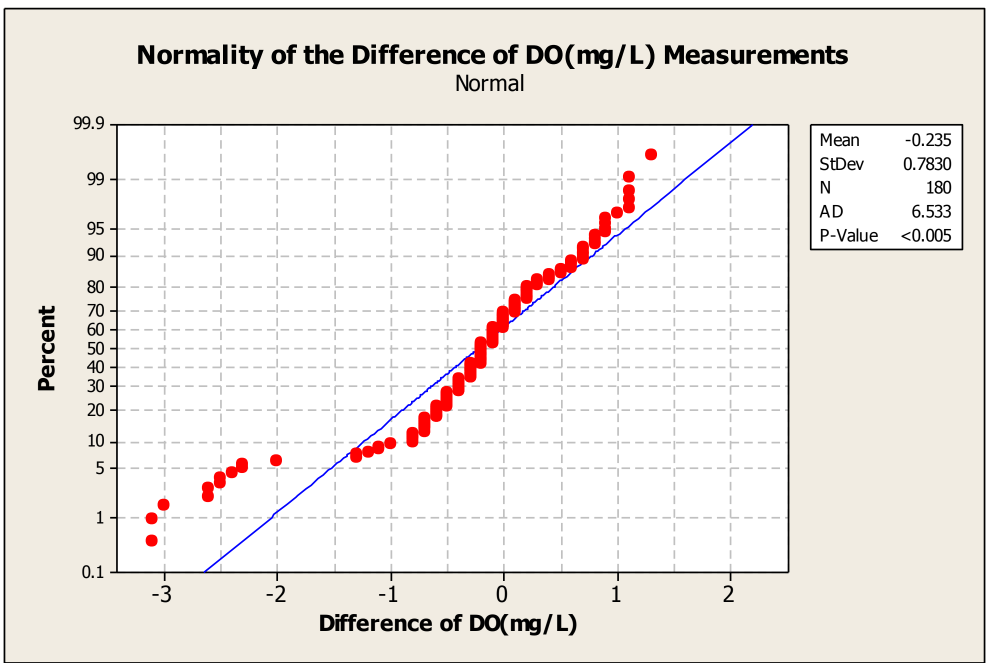


Figure 3.30. Normality of DO (mg/L) data

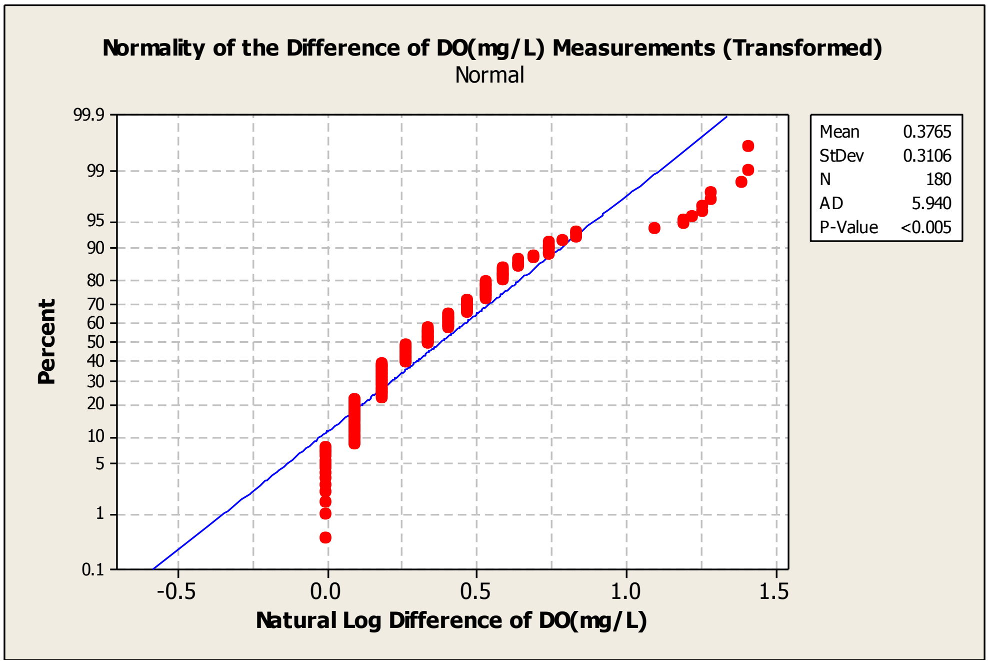


Figure 3.31. Normality of the transformed DO (mg/L) data

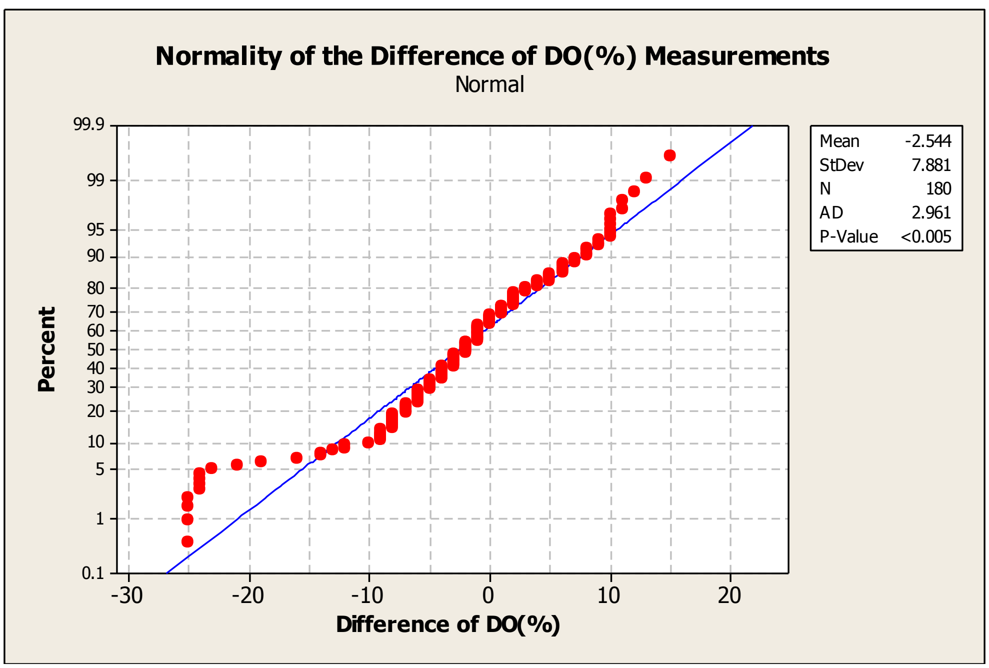


Figure 3.32. Normality test of DO (%) data

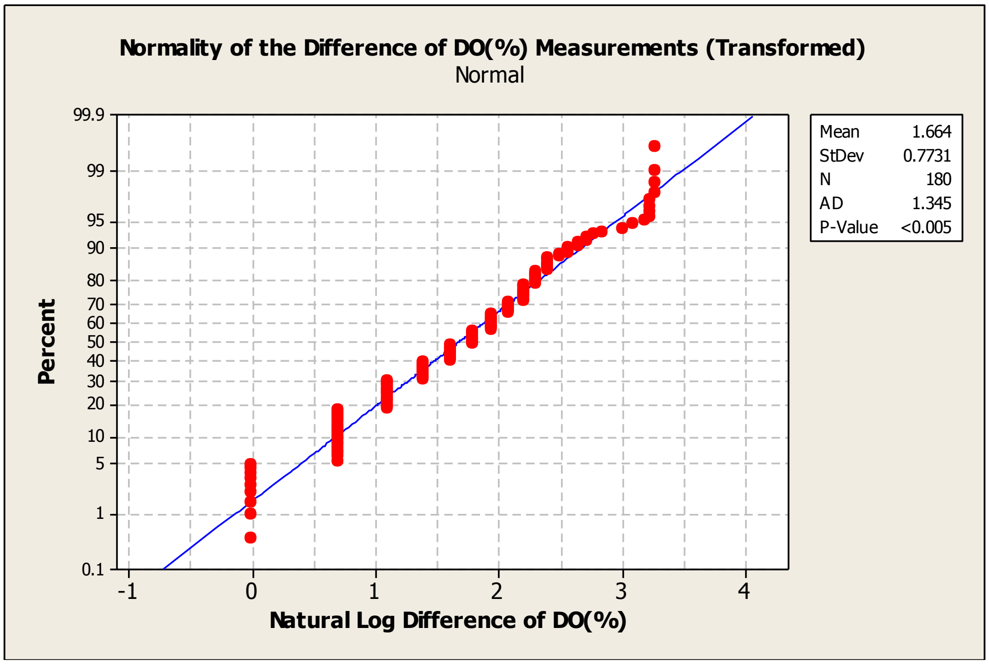


Figure 3.33. Normality of the transformed DO (%) data

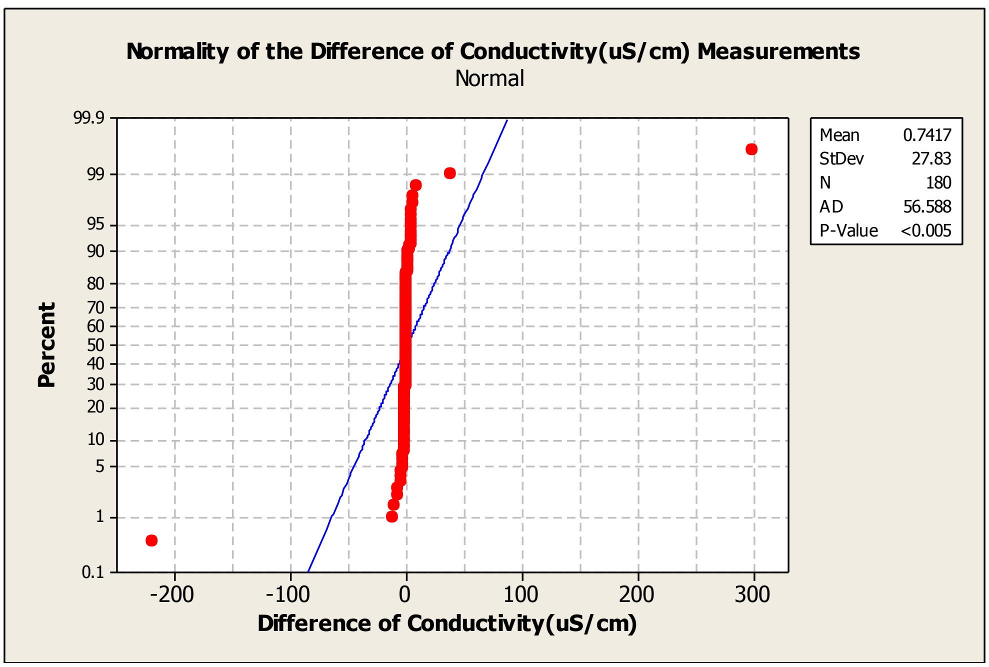


Figure 3.34. Normality of conductivity (uS/cm) data

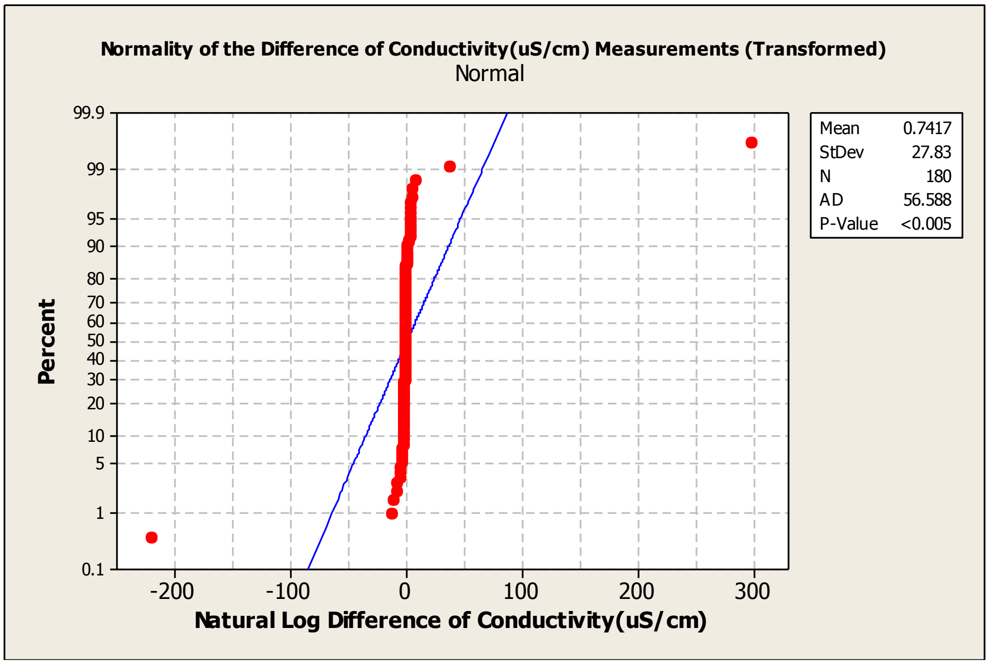


Figure 3.35. Normality of the transformed conductivity (uS/cm) data

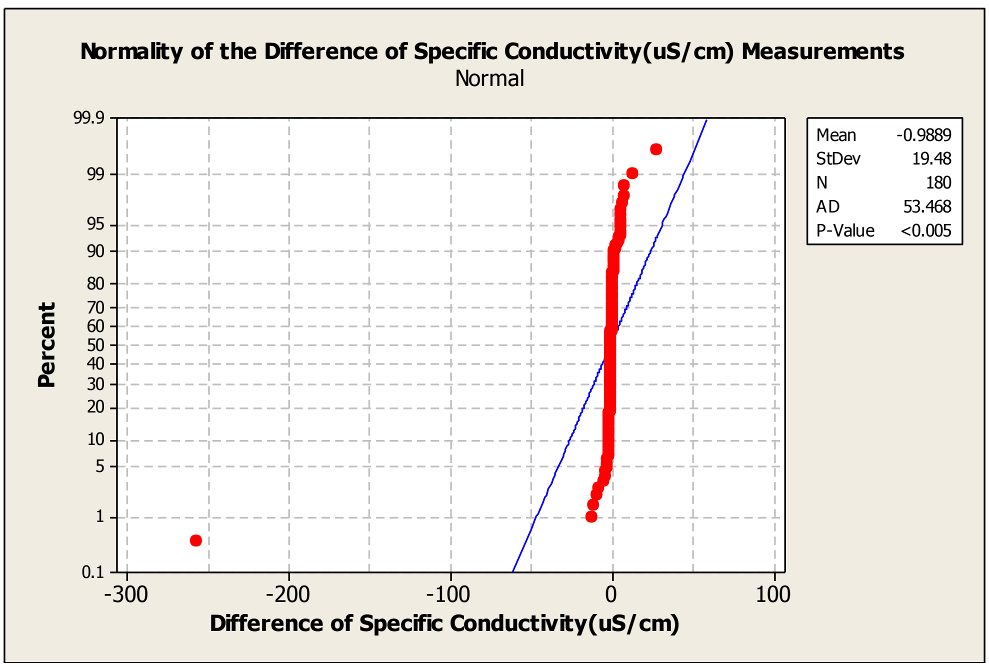


Figure 3.36. Normality of the specific conductivity (uS/cm) data

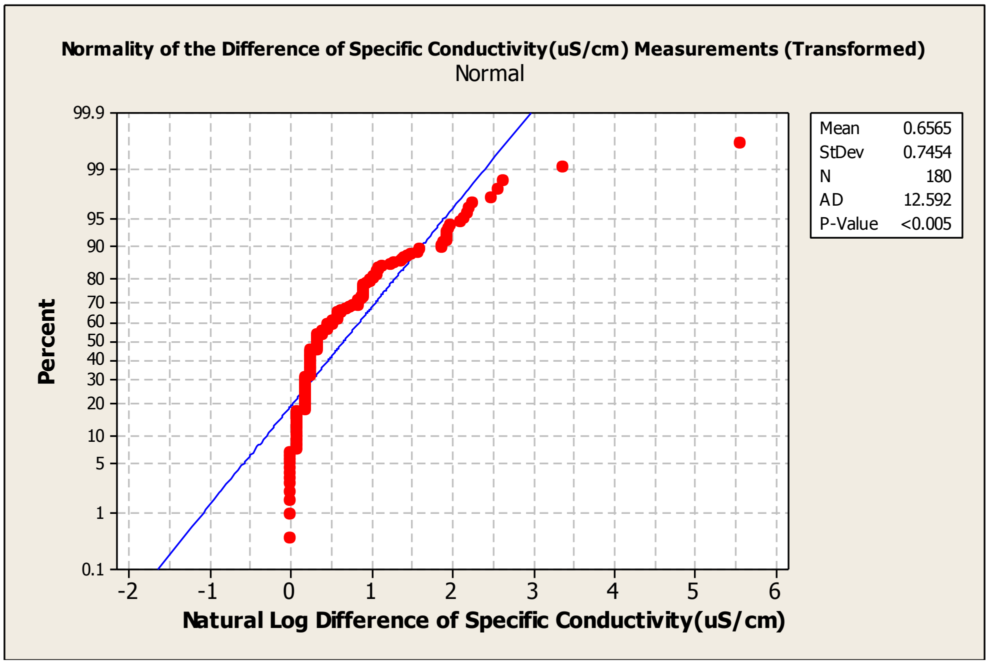


Figure 3.37. Normality of the transformed specific conductivity (uS/cm) data

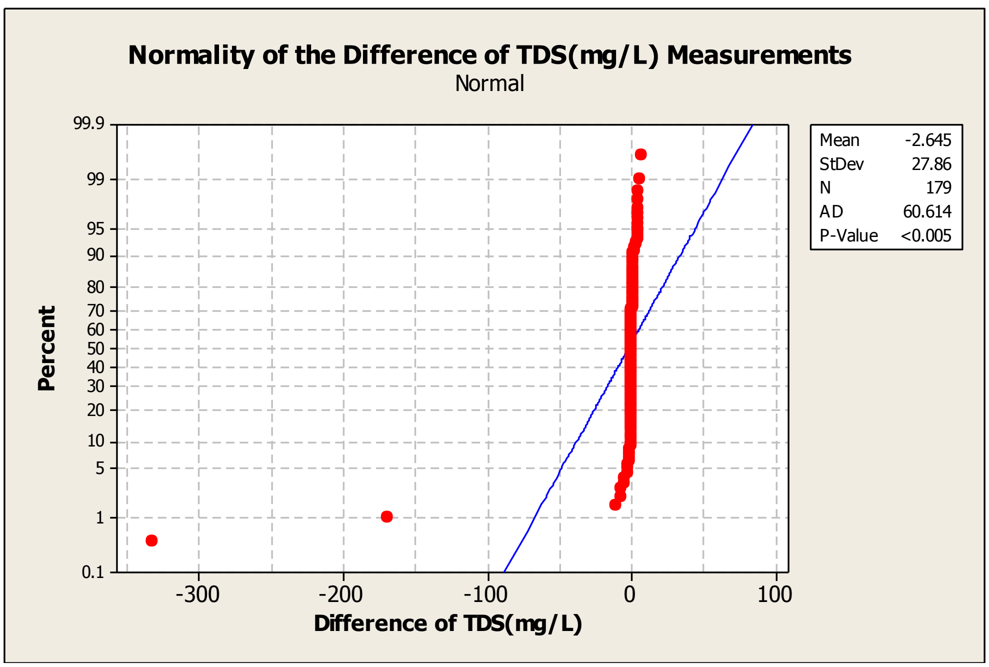


Figure 3.38. Normality of total dissolved solids (mg/L) data

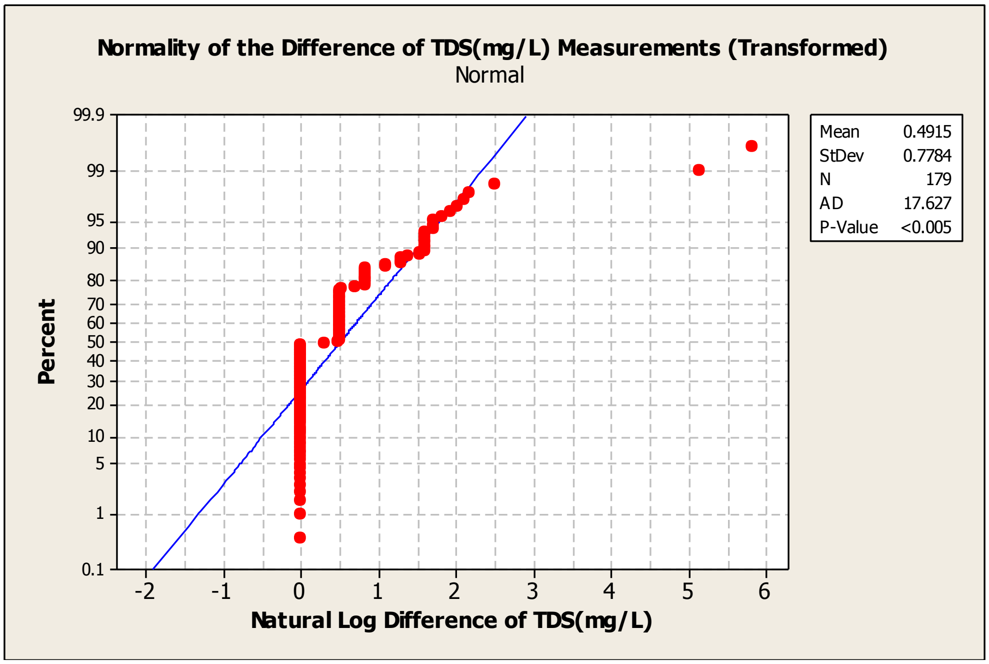


Figure 3.39. Normality of the transformed total dissolved solids (mg/L) data

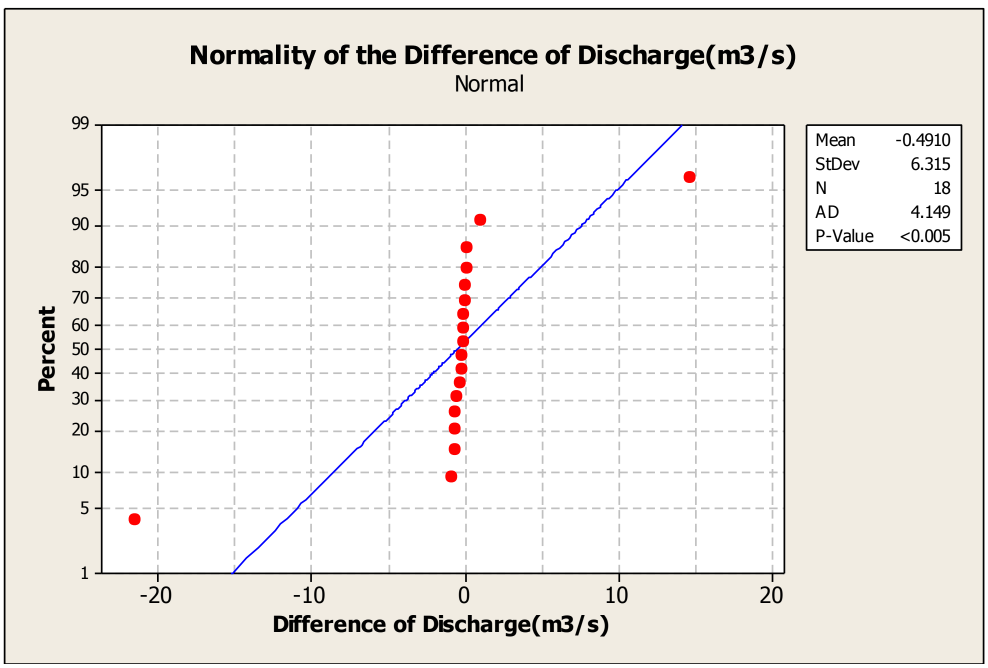


Figure 3.40. Normality of discharge (m3/s) data

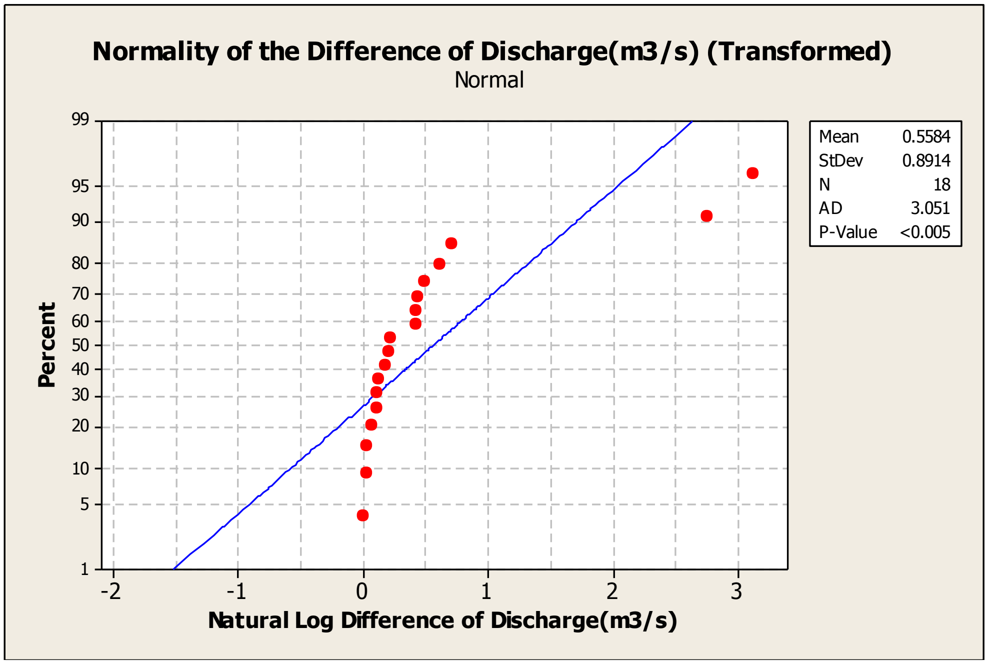


Figure 3.41. Normality of the transformed discharge (m3/s) data

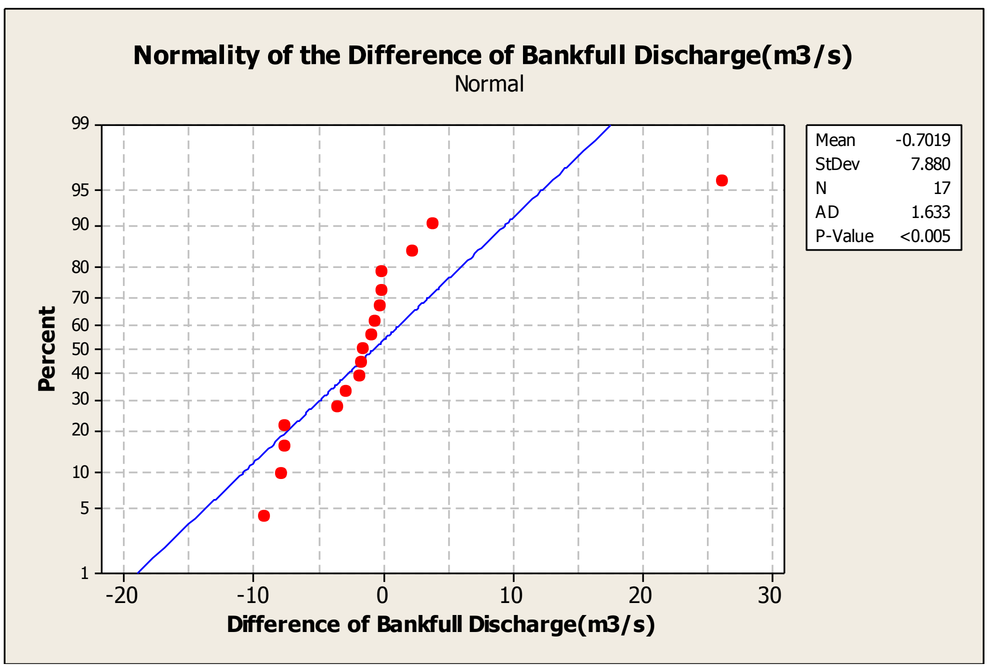


Figure 3.42. Normality of bankfull discharge (m3/s) data

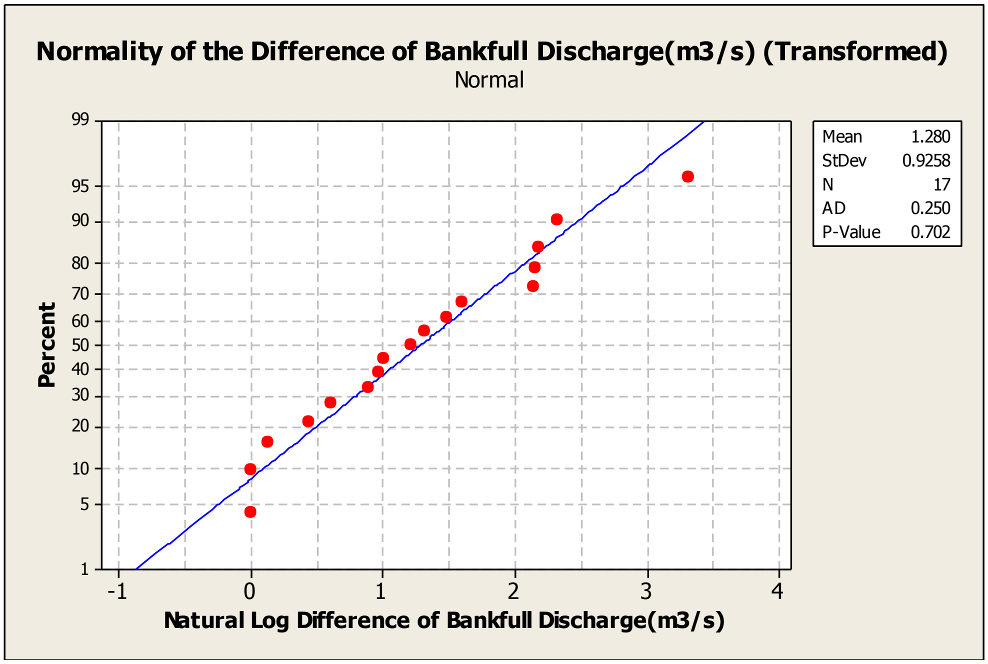


Figure 3.43. Normality of the transformed discharge (m3/s) data