Optimizing Water Quality Sampling Through Application Of Real Time Ionic Concentration Regression Models

Shibly Rahman, Renee Paterson

Water Resources Management Division, Department of Environment and Conservation, A1B4J6, St. John's, Canada

Abstract

The Water Resources Management Division of the Department of Environment and Conservation performs routine water sampling to measure the physical and chemical parameters of select water bodies in Newfoundland and Labrador. Ionic concentration parameter measurement is performed during routine water sampling to complement some of the key indicator parameters measured in real time at these select water bodies. The collection, laboratory analysis and measurement of water samples are a time consuming process. Some of the common conducting ions measured during routine sampling are sodium, calcium, chloride and sulphate. These conducting ions can be estimated using continuously measured specific conductance after observing the effect of flow. The estimated measurement will help identify the quality of water at a given point in time and hence save time and resources in performing routine sampling. It will also help estimate the quality of water in remote locations where routine sampling is not feasible. This paper compares four water bodies on the island part of Newfoundland and Labrador and estimates the ionic concentration using continuously measured specific conductance.

Keywords: Real Time Water Quality, Regression Analysis, Grab Sampling, Ionic Concentration

Introduction

The Water Resources Management Division (WRMD) of the Department of Environment and Conservation (ENVC) of the Province of Newfoundland and Labrador (NL) have established a near real time water quality (RTWQ) monitoring network throughout the province where key indicator water quality data is collected continuously. This water quality data can be used to monitor the health of aquatic ecosystems, establish trends and determine when specific water quality events occur. The information obtained from the network is needed by the WRMD to implement its mandate and allows managers and policy makers to make informed decisions on early warning of adverse water quality events. The general public, policy makers, government agencies and private sectors greatly benefit from such timely data and information.

The water quality parameters measured through the real time monitoring system are water temperature, pH, dissolved oxygen (DO), specific conductance (SC) and turbidity. Percent saturation and total dissolved solids are two additional parameters calculated from DO and SC. These key indicator parameters provide significant information to better understand the water quality of a particular water body. Routine water quality grab sampling is also performed in these select water bodies. The grab sampling is part of the Quality Assurance/Quality Control (QA/QC) protocol for NL RTWQ program which is used to measure ionic concentrations of water quality parameters. Some of these include sodium, calcium, chloride, and sulphate. The collection, shipping and analysis of grab samples in the lab require a significant period of time to measure the ionic concentration of the sample contents. This time lag can be greatly reduced if some of these ionic concentration parameters can be estimated in real time. As grab sample is collected at the same time as RTWQ parameter measurement, it is possible to correlate some of the grab sample parameters with the RTWQ parameters (see Figure 1) (Granato and Smith, 1999). Among all parameters, SC is more likely to correlate with the ion conducting parameters measured during grab sampling (Lind, 1970).

Water quality sampling sites across the island of Newfoundland were used to analyze site specific relationships between SC and sodium, calcium, chloride and sulphate ions. The sites chosen were Leary's Brook (LB), Waterford River (WR), Humber River (HR) and Rattling Brook below bridge (RB) (Figure 2). These sites have been sampled extensively for the last four to five years. The sites were selected based on the degree of anthropogenic activity taking place and the amount of dissolved solid material received by these water bodies in order that comparative analysis can be drawn from the data obtained.

This paper examines how the estimation of conducting ion concentration helps to optimize the resources and sampling time resulting in overall cost and time savings under the Water Quality Program. This approach would help the WRMD estimate the water quality in real time without waiting for lab analysis results. It will be very useful where tight timelines, budget constraints and human resources limitations are matters of concern. It can also be used to estimate water quality variables at remote sampling locations that are expensive and difficult to access. The results of this report will help to better understand how increased ionic concentration leads to elevated SC at impacted sites.

At first the methodology applied in this analysis is described followed by the description of the site locations. A brief discussion of the data collection and literature review is then performed. The effect of flow on each of the parameters is then shown. This is followed by statistical analysis on the dataset along with the regression models developed. Finally, the model verification and validation is performed along with concluding statements.

Methodology

Figure 1 shows the overall methodology of developing the ionic concentration model and applying the model to estimate ionic concentration in real time. The model is developed by using regression analysis on grab sample ionic concentration and real time specific conductance data.

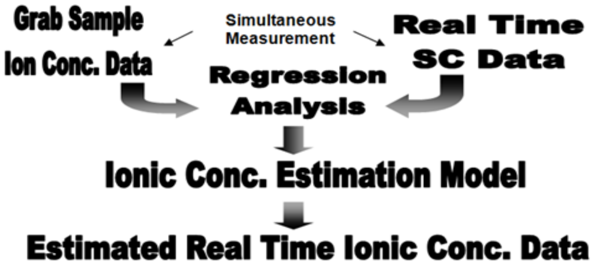

Figure 1. Methodology to estimate real time ionic concentration data

The model is then applied using real time specific conductance data on select stations to estimate the ionic concentration values for those stations. The model provides a measure of strength and variation of the relationship between real time and grab sample data. Site specific models for sodium, calcium, chloride and sulphate are developed in the four sampling locations across the island part of Newfoundland using SC data.

Description of site and sampling locations

Figure 2 shows the location of the four sites on the island part of Newfoundland from which the data is collected. The sites are: Leary's Brook, Waterford River, Humber River and Rattling Brook below bridge. These sites are chosen based on the degree of anthropogenic activity and availability of water quality data. Leary's Brook and Waterford River are located in an urban setting with a high level of anthropogenic impact from the surrounding areas. Rattling Brook is non-urban but in the middle of a construction site, while Humber River is non-urban with little impact from surrounding areas.

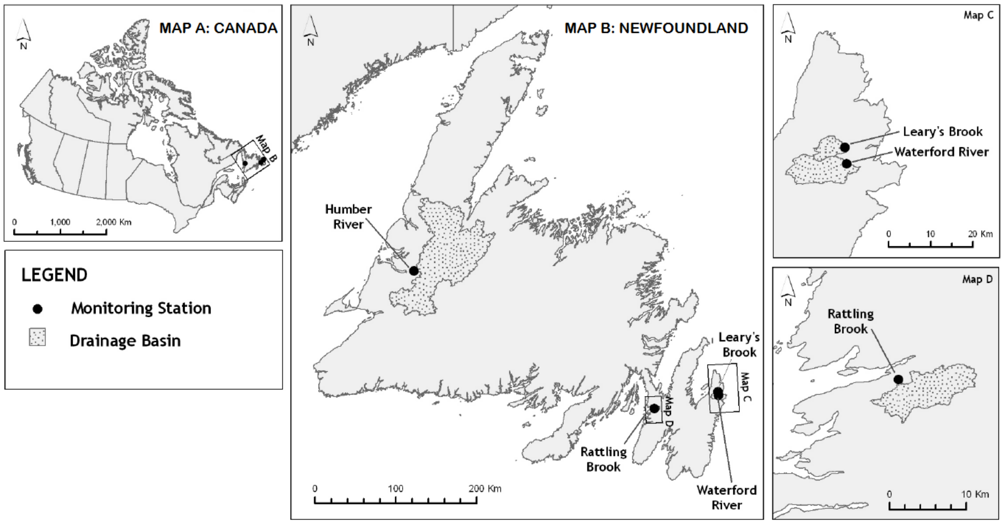

Figure 2. Geographic location of the four sites chosen for regression modelling

Leary's Brook at Clinch Crescent was the first RTWQ station established in 2001. The sampling site is located in a developed section of the City of St. John's close to Memorial University. One of the main shopping centres in the city is located immediately upstream of the sampling site where a portion of the river is culverted. The area is densely surrounded by houses, buildings, business facilities and major roads. Road salts are applied during the winter months which affect the water quality within the river. Significant urban runoff can be observed in the culvert area as a result of surrounding anthropogenic activities.

Waterford River at Kilbride station was established in 2005. The sampling site is situated near the downtown area of the City of St. John's. The area around the sampling site is densely surrounded by houses, buildings, roads and highways. Major industrial areas are also located within the drainage basin. Road salts are applied during the winter months which affect the water quality within the river. The river is highly impacted as a result of surrounding anthropogenic influence which affects the quality of water at the sampling site.

The Humber River is the second largest river on the island of Newfoundland. The sampling station was established in December 2003. It is classified as a non-urban station. There are a number of small communities located within the watershed but the overall population density is sparse. There are some transportation routes throughout the basin which are salted during the winter months. However, due to the large volume of water within the system, the ionic concentration is diluted.

Rattling Brook below bridge station was established in December 2006 on the south eastern Avalon Peninsula. It is within the construction zone of a commercial processing facility. Major work resulting from the construction of the processing facility is occurring along the river and access to the sampling sites is controlled due to security and safety concerns. The river is moderately impacted with ionic concentration due to sparse population and the presence of the processing plant and facilities.

Literature Review

Conductivity reveals the presence of dissolved materials in water (Williams, 1966) consisting of metallic ions, organic and inorganic materials. It is the ability of a fluid to conduct electricity. Specific conductance (SC) is the inverse of electrical resistivity, corrected at 25 °C, since fluids conduct more at higher temperatures. Hence SC is an indirect measure of the amount of dissolved substances (Hach, 2006). A detailed study by Granato and Smith (1999) in Northborough, Massachusetts applied regression analysis in their study to measure constituent calcium, sodium, and chloride on the basis of continuous records of SC of highway runoff. Christensen, Rasmussen and Ziegler (2002) and Ryberg (2006, 2007) also developed regression equations to estimate constituent concentration yields in water bodies in Kansas and North Dakota. Reham El-Korashey (2009) has applied regression analysis to estimate sodium and chloride in Bahr El Baqar Drain in Egypt using electrical conductivity as an explanatory variable. Recent study by Harvey, Lye and Khan (2011) shows some of the advances in RTWQ monitoring and how it can be applied in estimation of constituent parameter concentration. The Ordinary Least Square (OLS) has been applied in many studies (Granato & Smith, 1999, Christensen et al., 2002) and is a standard procedure to estimate water quality constituents (Hem, 1992).

Data Collection and Analysis

Conductivity data is collected through NL RTWQ monitoring network. The network consists of a series of monitoring stations with sensors collecting data across NL. Monitoring instruments as shown in figure 3 are deployed beneath the water's surface in a representative section of the stream which continuously measures RTWQ parameter data for each sampling station.

Figure 3. Water Quality Parameter Sensors

Grab sample data is collected using monthly grab sample results measured at an accredited lab. Daily flow data is collected using Water Survey of Canada's centrally-managed database HYDAT. The data retrieval and management for NL RTWQ network is shown in figure 4.

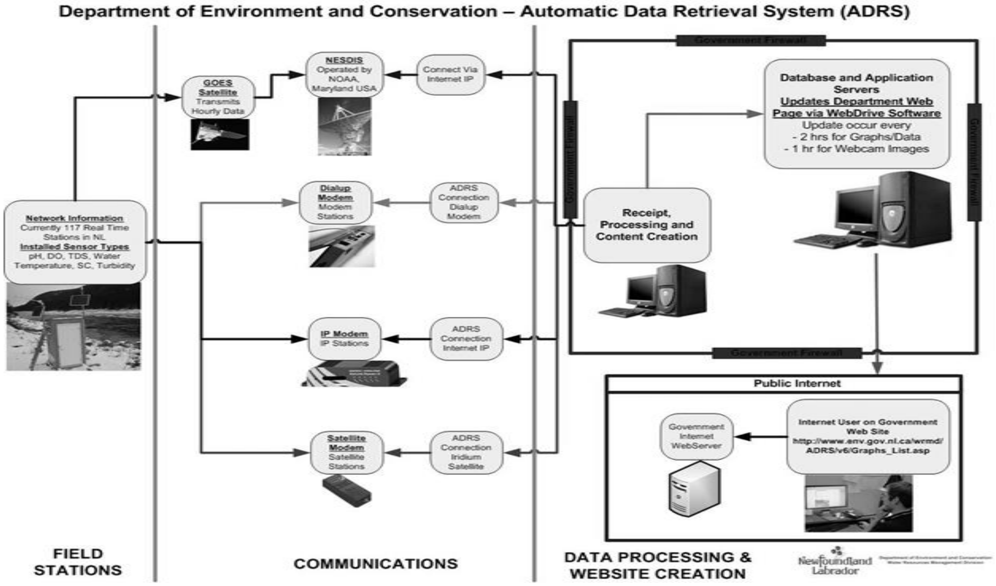

Figure 4. Real Time Data Retrieval and Management

The assumptions for Linear Regression are individually tested using Residual plots for the tested water quality parameters in Minitab^{TM}. The independence, homoscedasticity, and normality of the error distribution assumption for regression analysis for all parameters were fulfilled.

The box-plots in Figure 5 compare the variability of conductivity, sodium, calcium and chloride values across locations. Sulphate values were not tested due to the presence of less than detect values for two of the four stations. The median line was connected to all the stations to observe the variability and difference in parameter values. The urban stations showed more variability in parameter values in comparison to the rural stations. This can be noted by looking at the flatness of the box plots for the rural stations and the presence of outliers in the urban stations.

Table 1 shows the statistical measurements for real time Specific Conductance for all four stations. The variability of specific conductance measurement is much higher in the urban stations compared to the non-urban stations. The high variations can be due to increased snowmelt or storm runoff that takes place during seasonal weather changes.

Table1. Statistical analysis for real time data in all stations

Station	Sample Size	Range (µS/cm)	Mean	Median (Q2)	Q1	Q3
Leary's Brook	31	148.1 - 1346	450.6	360.2	287	505
Waterford River	30	235 - 1417	529.4	438.5	369.3	494.5
Humber River	29	25.5 - 43.4	34.94	35.6	31.75	38.6
Rattling Brook	29	27.2 - 41.5	34.13	35.1	31.5	36.3

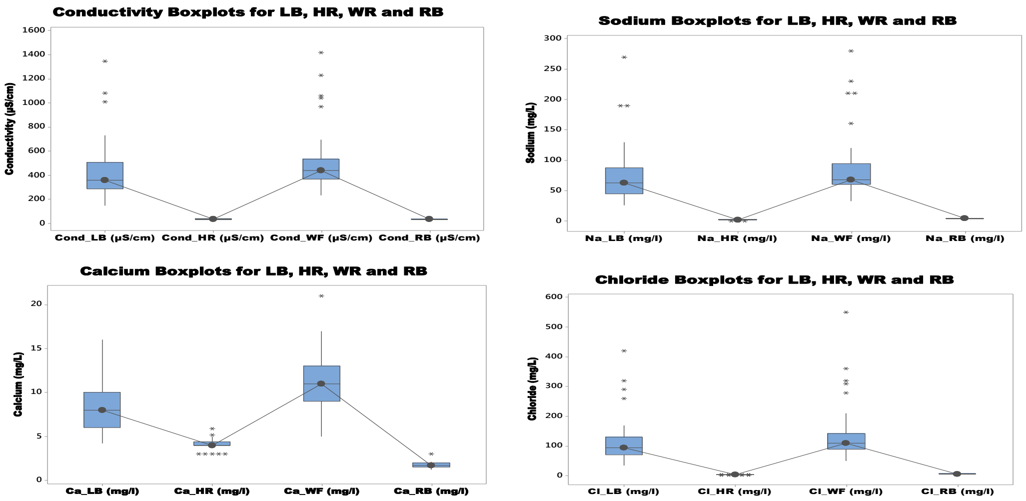

Figure 5. Comparisons of parameter values across locations

Effect of flow on parameter concentration

The interval plot for monthly flow for all stations is shown in figure 6. Four to five years of flow data obtained from HYDAT is used in this graph. The confidence interval for each month is also shown. The highest seasonal flow for most stations occurs during the month of April to May while the lowest seasonal flow occurs during July to October. The rise in flow corresponds to snow melt which immediately dips with a low flow season in the summer months.

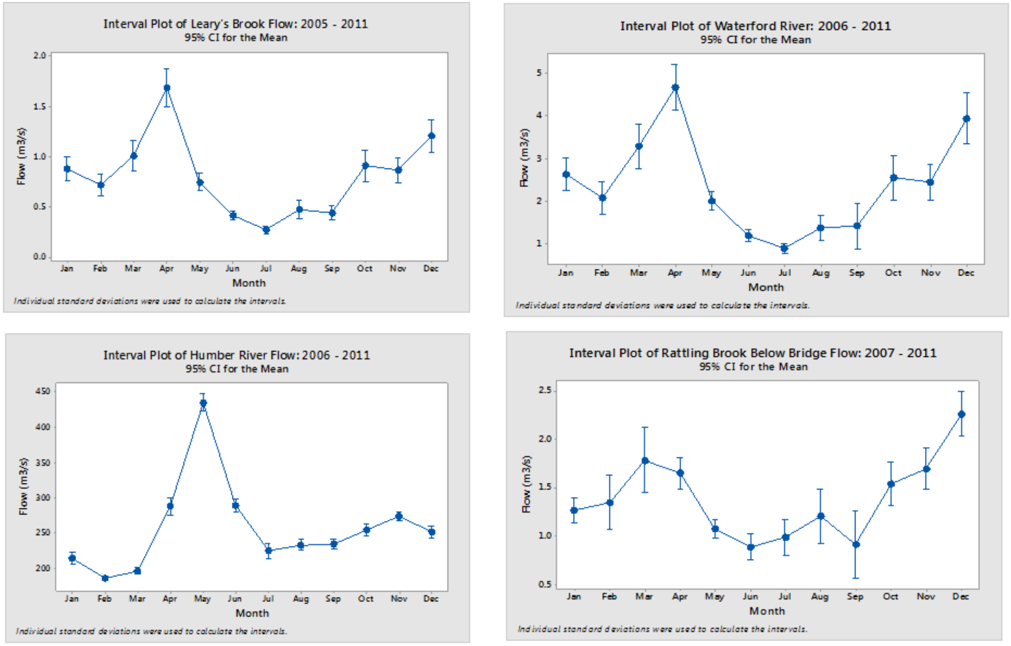

Figure 6: Seasonal Flow Interval Plot using four to five years of flow data for all stations

The effect of flow on parameter concentration is discussed in Clissie, Pollock and Cunjak (1996). Figure 7 show scatter plots with Lowess lines to see if relationships exists between parameter concentration and flow. Although flow plays a major role in controlling parameter concentration, the Lowess lines in this case displays lack of clear patterns between parameter concentration and flow. As instantaneous flow data was not available at the time of parameter concentration measurement, the average daily flow may dampen some of the effects of flow on parameter concentration.

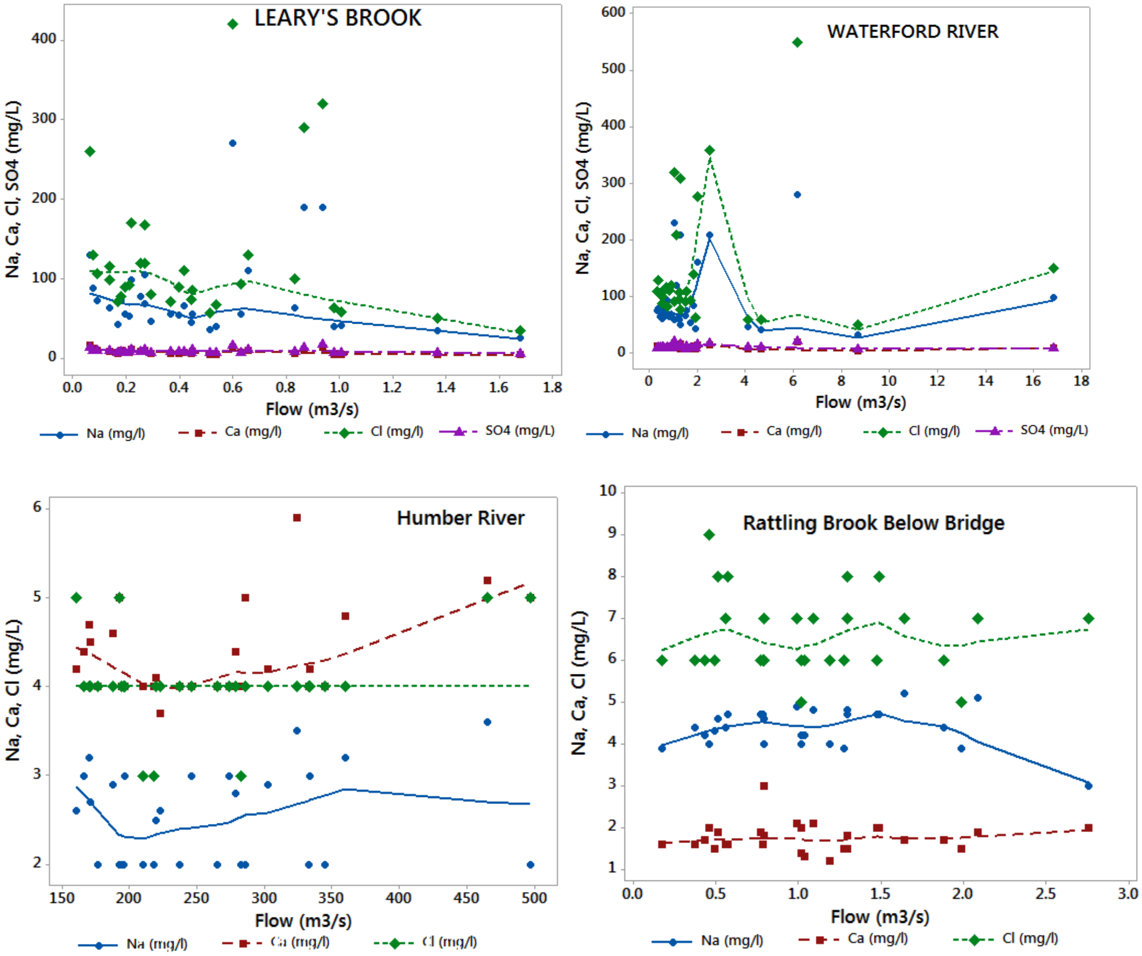

Figure 7. Effect of flow on parameter concentration

Regression analysis was performed between parameter concentration and flow is shown in table 2. All parameters for all stations show poor correlation between parameter concentration and flow. Transforming the model does not significantly improve this relationship.

Site Specific Parameter Model For All Stations

Figures 8 show Scatter plots with Lowess lines to check linear patterns for sodium, calcium, chloride and sulphate with respect to specific conductance (SC). The Lowess lines display linear patterns for Leary's Brook and Waterford River which is absent in Rattling Brook below Bridge and Humber River.

As shown in the previous section, the effect of flow on parameter concentration was minimal. Based on the Lowess lines in figure 8, the effect of specific conductance on parameter concentration was analyzed. The resulting models for all stations are shown in Table 3. Due to non-normality and the presence of outliers in most of the above parameter data values, log transformation was performed on the original data. Ordinary least square was applied on the log transformed data using Minitab™. Bias correction (Duan, 1983) was performed on the log transformed model.

Table 2. OLS Regression showing the effect of flow on parameter concentration

Variable	Regression Model	R-square	P-Value	Regression Model	R-square	P-Value
LEARY'S BROOK				**WATERFORD RIVER**		
Sodium	Na = 79.12 – 2.72 × Flow	0%	0.915	Na = 88.82 + 1.643 × Flow	0%	0.644
Calcium	Ca = 9.78 - 3.11 × Flow	13.3%	0.025	Ca = 11.49 - 0.1744 × Flow	0%	0.367
Chloride	Cl = 128.8 – 12.05 × Flow	0%	0.201	Cl = 137.5 + 4.056 × Flow	0%	0.523
Sulphate	SO4 = 10.07 - 0.54 × Flow	0%	0.672	SO4 = 12.11 - 0.047 × Flow	0%	0.816
HUMBER RIVER				**RATTLING BROOK**		
Sodium	Na = 2.29 + 0.0009 × Flow	0%	0.443	Na = 4.492 – 0.0913 × Flow	0%	0.548
Calcium	Ca = 3.734 + 0.0023 × Flow	0.03	0.03	Ca = 1.737 + .0283 × Flow	0%	0.802
Chloride	Cl = 3.598 + 0.0016 × Flow	5%	0.128	Cl = 6.722 – 0.106 × Flow	0%	0.735

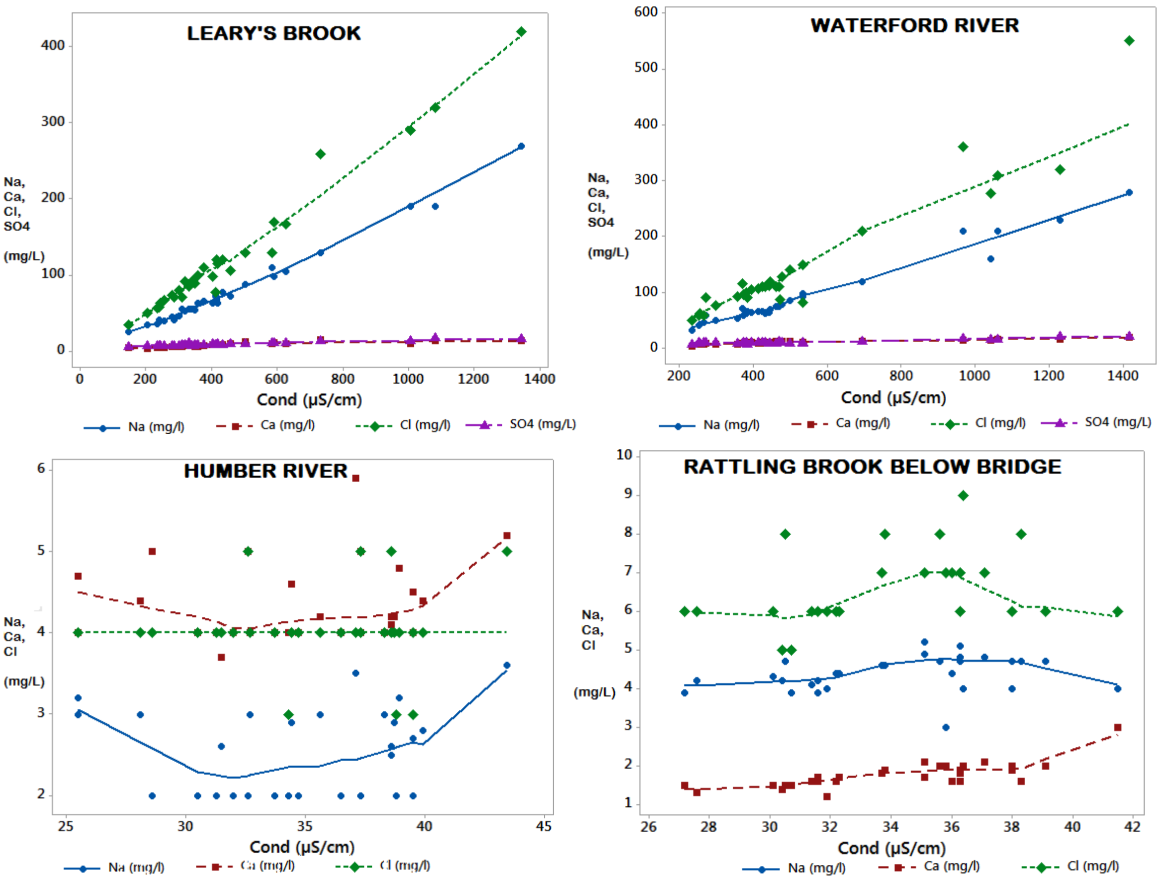

Figure 8. Scatterplot with Lowess Lines for all stations models

The models show strong correlation between ionic concentration parameters and specific conductance for Leary's Brook and Waterford River while a weaker correlation for Humber River and Rattling Brook. The strong correlation is indicated by high R-square value, low relative percentage difference (RPD) value, along with p-value < 0.01 indicating statistical significance for the correlation. The association is stronger in sodium and chloride in comparison with calcium and sulphate.

Table 3. Model for sodium, calcium, chloride and sulphate for all stations

Variable	Sample Size	Range (mg/L)	Median	Regression Model	RPD	R-square	P-Value
LEARY'S BROOK							
Sodium	31	26 - 270	63	$\log(Na) = -0.979 + 1.08 \times \log(Cond)$	0.78	98.70%	<0.01
Calcium	31	4.2 - 16	8	$\log(Ca) = -0.811 + 0.654 \times \log(Cond)$	1.56	80.60%	<0.01
Chloride	31	35 - 420	94	$\log(Cl) = -0.878 + 1.11 \times \log(Cond)$	2.88	96.70%	<0.01
Sulphate	31	6 - 18	9	$\log(SO4) = -0.220 + 0.461 \times \log(Cond)$	0.51	88.40%	<0.01
WATERFORD RIVER							
Sodium	30	33 - 280	68	$\log(Na) = -1.02 + 1.09 \times \log(Cond)$	2.25	96.50%	<0.01
Calcium	30	5 - 21	11	$\log(Ca) = -0.494 + 0.569\times\log(Cond)$	0.98	77.40%	<0.01
Chloride	30	51 - 550	110	$\log(Cl) = -0.990 + 1.15 \times \log(Cond)$	2.29	91.70%	<0.01
Sulphate	30	7 - 22	11	$\log(SO4) = -0.182+0.466\times\log(Cond)$	1.16	72.40%	<0.01
HUMBER RIVER							
Sodium	29	2-3.6	2.6	$\log(Na) = 0.54 + 0.103 \times \log(Cond)$	21.24	0.40%	0.734
Calcium	29	3.7-5.9	4.1	$\log(Ca) = 1.14 + 0.092 \times \log(Cond)$	15.5	1.3%	0.549
Chloride	29	3-5.0	4	$\log(Cl) = 1.32 + 0.019 \times \log(Cond)$	10	0%	0.916
RATTLING BROOK							
Sodium	29	3.0 - 5.2	4.4	$\log(Na) = 0.234 + 0.265 \times\log(Cond)$	7.93	6.20%	0.192
Calcium	29	1.2 - 3.0	1.7	$\log(Ca) = -1.76 + 1.31 \times \log(Cond)$	8.41	58.10%	<0.01
Chloride	29	5.0 - 9.0	6	$\log(Cl) = 0.092 + 0.472 \times \log(Cond)$	10.3	12.30%	0.103

Model Validation

The obtained results show that it is possible to predict ion concentration of sodium, calcium, chloride, and sulphate from real time specific conductance as long as there is enough variation within the parameter values of grab sample data. This would enable getting instantaneous estimations of parameter values thus saving wait times for grab sample results to be returned from a laboratory. The grab samples collected after the model development were used for model validation. The models obtained for Leary's Brook, Waterford River and Rattling Brook below Bridge were used for validation since Humber River parameters have shown poor correlation.

The graphs in Figure 9,10 and 11 shows the ionic concentration estimation and validation of parameters used in Leary's Brook, Waterford River and Rattling Brook models. The models from Table 3 were used to estimate ion concentration (sodium, calcium, chloride and sulphate) in real time. The model is represented by a line in the graph. The corresponding calibration grab sample values were placed as points within the graph to see how closely it fits to the model. As shown in figure 9, the calibration grab samples lie closely to the regression model line.

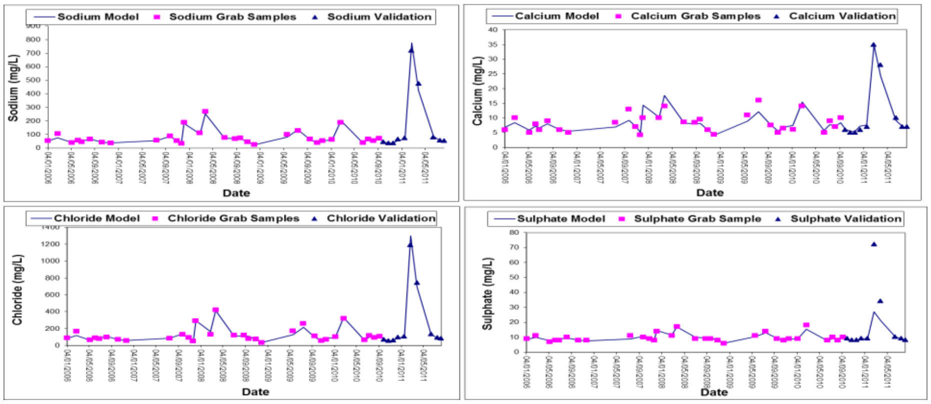

Figure 9. Model validation for Leary's Brook parameters

In order to validate the model, ten additional validation grab samples were used after the model development. The validation grab samples are represented as triangular points in order to distinguish between calibration and validation grab samples. As shown in figure 10 the validation grab samples lie closely to the model line.

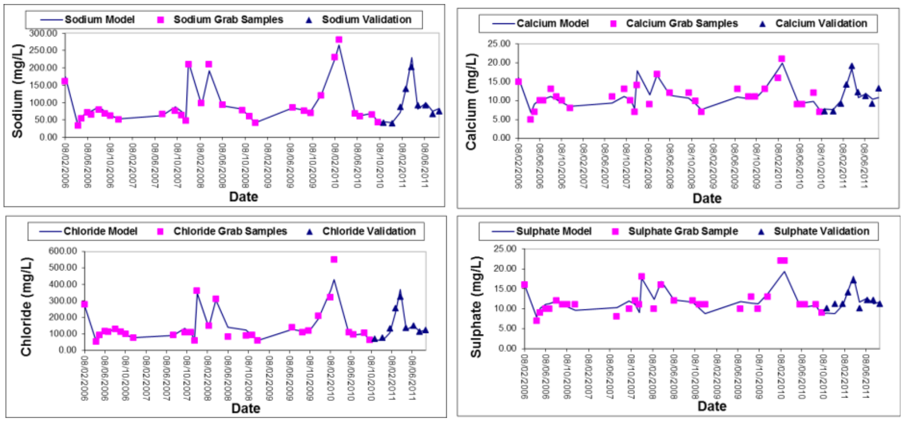

Figure 10. Model validation for Waterford River parameters

Calcium was the only parameter that showed a good fit for Rattling Brook. As shown in figure 11, the calibration and validation grab sample lies reasonably close to the regression model line.

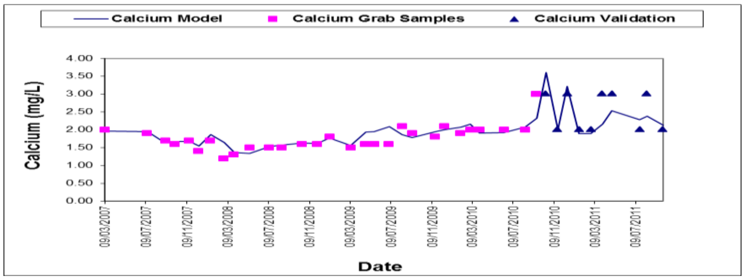

Figure 11. Model validation for Rattling Brook parameter

Conclusions

It is evident from the results presented that increased variation within grab sample measurements leads to a better regression model. This has been observed in the case of Leary’s Brook and Waterford River as well as for calcium in Rattling Brook below bridge. The variation in the level of ionic concentration is largely due to the presence of anthropogenic influence within these rivers. In the case of Humber River with little anthropogenic influence, the ionic concentration of most parameter measurements were below the detection limit, and hence it was difficult to apply any statistical tests to identify if a relationship exists between real time parameters and grab samples. The high flow of water in that river diluted most of the parameter concentrations which is represented in the low measurements of parameter values.

This study will aid in estimating ionic concentration in real time for the sites where a good fit for regression was obtained. It will also reduce the time delay required to measure water quality constituents at the laboratory by estimating ionic concentration instantaneously. Using the real time parameter specific conductance, the model will help predict the surrogate parameters (sodium, chloride, calcium and sulphate) in real time which can be viewed graphically. In order to maintain the accuracy of the model, it must be calibrated every year when newer grab samples are available. This will adjust the model accuracies based on the updated grab sample values.

Potential parameters of interest can be estimated in emerging real time sites using real time parameters as predictors by applying the methodological analysis applied in this study. One such parameter can be total suspended solids (TSS) which can be estimated using real time turbidity. This would be beneficial to industries monitoring real time water quality parameters who would like to ensure that the TSS values are in compliance with the current regulations. Another area of application of this model is to identify the impact of water quality due to the application of road salts. Operational decisions can be made in a proactive manner with the available estimated data.

Acknowledgement

Feedback on statistical analysis was provided by Dr Leonard Lye, Professor of Engineering and Applied Science from Memorial University. The paper was thoroughly reviewed by colleagues Kyla Brake and Rob Wight (Environmental Scientists at WRMD) with valuable insights. Technical advice relating to the "Data Collection and Management" portion of the report was provided by Leona Hyde (Water Resources Technician). Finally, assistance was provided by colleague Keith Abbott (Environmental Scientist at WRMD) in the generation of the station location maps.

References

Christensen, V.G., Rasmussen, P.P., Ziegler, A.C., 2002. Real time water quality monitoring and regression analysis to estimate nutrient and bacteria concentrations in Kansas streams, Water Science and Technology, Vol. 45, No. 9, pp. 205-219.

Church, S.E., von Guerard, Paul, and Finger, S.E., 2007, Integrated investigations of environmental effects of historical mining in the Animas River watershed, San Juan County, Colorado: U.S. Geological Survey.

Clissie, D., T. L. Pollock, and R. A. Cunjak, 1996, Variation in stream water chemistry and hydrograph separation in a small drainage basin, Journal of Hydrology, 178, 137– 157.

Duan, N., 1983, Smearing estimate - a nonparametric retransformation method. Journal of the American Statistical Association, Vol. 78, No. 383, p. 605-610.

Granato, G. E, & Smith, K. P. 1999. Estimating concentrations of road-salt constituents in highway-runoff from measurements of specific conductance. Northborough, Mass.: U.S. Dept. of the Interior, U.S. Geological Survey.

Hach Company, 2006. Hydrolab DS5X, DS5, and MS5 Water Quality Multiprobes USER MANUAL, Edition 3, Catalog Number 003078HY (http://s.campbellsci.com/documents/ca/manuals/series_5_man.pdf)

Harvey, R., Lye, L. and Khan, A., 2011, Recent Advances in the Analysis of Real-time Water Quality Data Collected in Newfoundland and Labrador, Canadian Water Resources Journal, 36:4, 349-361.

Hem, J.D., 1992. Study and interpretation of the chemical characteristics of natural water (3rd ed): U.S.Geological Survey Water-Supply Paper 2254, 263 p.

Lind, C.J., 1970. Specific conductance as a means of estimating ionic strength: U.S. Geological Survey Professional Paper 700-D, p. D272-280

Reham El-Korashey, 2009. Using Regression Analysis to Estimate Water Quality Constituents in Bahr El Baqar Drain, Journal of Applied Sciences Research, 5(8): 1067-1076.

Ryberg, K.R., 2006. Continuous Water-Quality Monitoring and Regression Analysis to Estimate Constituent Concentrations and Loads in The Red River of the North, Fargo, North Dakota, 2003-05, U.S. Geological Survey Water Resources Investigations Report 2006-5241, pp: 35.

Ryberg, K.R., 2007. Continuous water-quality monitoring and regression analysis to estimate constituent concentrations and loads in the Sheyenne River, North Dakota, 1980–2006: U.S. Geological Survey Scientific Investigations Report 2007–5153, pp: 22.

Williams, W. D. 1966, Conductivity and the concentration of total dissolved solids in Australian lakes. Australian Journal of Marine and Freshwater Research 17, 169-76.