TECHNICAL REPORT NO. 64

# REGRESSION ANALYSIS OF SPECIFIC CONDUCTANCE AND CHLORIDE CONCENTRATIONS

SOUTHEASTERN WISCONSIN REGIONAL PLANNING COMMISSION

## SOUTHEASTERN WISCONSIN REGIONAL PLANNING COMMISSION

**KENOSHA COUNTY**
John Holloway
Amy Maurer
Robert W. Pitts

**RACINE COUNTY**
Jonathan Delagrave
Peggy L. Shumway
Vacant

**MILWAUKEE COUNTY**
Donna Brown-Martin
Priscilla Coggs-Jones
Michael Maistelman

**WALWORTH COUNTY**
Charles L. Colman, *Chairman*
Brian E. Holt
Mary Knipper

**OZAUKEE COUNTY**
Thomas H. Buestrin
Natalia Minkel-Dumit
Eric Stelter

**WASHINGTON COUNTY**
Katrina Hanson
Jeffrey D. Schleif
David L. Stroik, *Treasurer*

**WAUKESHA COUNTY**
Michael A. Crowley, *Secretary*
James T. Dwyer, *Vice-Chairman*
Dewayne J. Johnson

## SOUTHEASTERN WISCONSIN REGIONAL PLANNING COMMISSION STAFF

Stephanie Hacker, AICP, LEED AP ..... Executive Director
Benjamin McKay, AICP ..... Deputy Director
Christopher Hiebert, PE ..... MPO Director
Elizabeth Larsen, SPHR, SHRM-SCP ..... Director of Administration
Joel Dietl, AICP ..... Chief Land Use Planner
Laura Herrick, PE, CFM ..... Chief Environmental Engineer
Ryan Hoel, PE ..... Chief Transportation Engineer
Eric Lynde ..... Chief Special Projects Planner
Rob Merry, PLS ..... Chief Surveyor
Nakeisha Payne ..... Public Involvement and Outreach Manager
Thomas Slawski, PhD ..... Chief Biologist

Special acknowledgement is due to Joseph Boxhorn, PhD, Principal Planner; Justin Poinsatte, PhD, Principal Specialist; Karin Hollister, PE, Principal Engineer; Nicklaus Neureuther, Specialist, and Megan Deau, Senior Graphic Designer for their efforts in the preparation of this report.

## REGIONAL CHLORIDE IMPACT STUDY TECHNICAL ADVISORY COMMITTEE

Thomas Grisa, *Chairman* ..... Department of Public Works, City of Brookfield
Laura Herrick, *Secretary* ..... Southeastern Wisconsin Regional Planning Commission
Mandy Bonneville ..... Walworth County Land Use and Resource Management Department
Patrice Brunette[1] ..... Racine County Highway Department
Karl Buck ..... Federal Highway Administration, Wisconsin Division
Brian Cater ..... Department of Public Works, City of Kenosha
Cody Churchill ..... Wisconsin Department of Transportation
Steven Corsi ..... U.S. Geological Survey
Timothy Grundl ..... University of Wisconsin Milwaukee
David Hart ..... Wisconsin Geological and Natural History Survey
Bryan Hartsook ..... Wisconsin Department of Natural Resources
Craig Helker ..... Wisconsin Department of Natural Resources
Richard Hough ..... Department of Public Works, Walworth County
Samantha Katt[2] ..... Wisconsin Department of Natural Resources
Scott Kroeger ..... Public Works and Development, City of Muskego
Matthew Magruder ..... Milwaukee Metropolitan Sewerage District
Max Marechal ..... Engineering Department, City of West Bend
Daniel Moudry ..... Department of Public Works, Waukesha County
Cheryl Nenn ..... Milwaukee Riverkeeper
Neal O'Reilly ..... University of Wisconsin Milwaukee
Charles Paradis ..... University of Wisconsin Milwaukee
Steven Peterson[3] ..... U.S. Geological Survey
Timothy Popanda ..... Village of Paddock Lake
Scott Schmidt ..... Washington County Highway Department
Anthony Spears ..... Department of Public Works, City of Cudahy
Kurt Sprangers ..... Department of Public Works, City of Milwaukee
David Strifling ..... Marquette University Law School
Michael Wieser ..... Engineering and Public Works, City of Cedarburg

[1] *Prior to May 2023, Roley Behm, Director of Public Works, served on the Technical Advisory Committee.*

[2] *Prior to June 2023, Benjamin Benninghoff, Natural Resource Basin Supervisor, served on the Technical Advisory Committee.*

[3] *Prior to October 2023, John Walker, Director, Midwest Science Center, USGS, served on the Technical Advisory Committee*

TECHNICAL REPORT
NUMBER 64

# REGRESSION ANALYSIS OF SPECIFIC CONDUCTANCE AND CHLORIDE CONCENTRATIONS

Prepared by the
Southeastern Wisconsin Regional Planning Commission
W239 N1812 Rockwood Drive
P.O. Box 1607
Waukesha, Wisconsin 53187-1607
www.sewrpc.org

The preparation of this publication was financed in part through project funds provided by the U.S. Department of Transportation Federal Highway Administration, Wisconsin Department of Natural Resources, Milwaukee Metropolitan Sewerage District, Fund for Lake Michigan, and the Southeastern Wisconsin Regional Planning Commission.

May 2024

TABLE OF CONTENTS

TABLE OF CONTENTS



## LIST OF FIGURES

**Chapter 2**



**Chapter 3**



**Appendix B**



## LIST OF MAPS

**Chapter 2**

## LIST OF TABLES

# INTRODUCTION 1

## 1.1 PURPOSE OF THIS REPORT

This report documents the development of regression models that can be used to estimate the concentration of chloride in surface water from measurements of the water's specific conductance.

Conductivity measures the ability of water to conduct an electric current. Because this ability is affected by water temperature, conductivity values are corrected to a standard temperature of 25 degrees Celsius (°C)[1] or 77 degrees Fahrenheit (°F). This corrected value is referred to as specific conductance. Pure water is a poor conductor of electrical currents and exhibits low values of specific conductance. For example, distilled water produced in a laboratory has a specific conductance in the range of 0.5 to 3.0 microSiemens per centimeter ($\mu$S/cm), a very low value. The ability of water to carry a current depends upon the presence of ions in the water, and on their chemical identities, total concentration, mobility, and electrical charge. Solutions of many inorganic compounds, such as salts, are relatively good conductors. As a result, specific conductance gives a measure of the concentration of dissolved solids in water, with higher values of specific conductance indicating higher concentrations of dissolved solids.

Under certain circumstances, measurements of specific conductance may act as a useful surrogate for measurements of the concentrations of individual dissolved substances. For example, measurements of specific conductance may be able to give indications of chloride concentrations in receiving waters. Specific conductance can be measured in the field using a hand-held meter while measuring chloride concentration requires laboratory analysis. In addition, specific conductance in a waterbody can be measured continuously using dedicated data logging equipment. The use of specific conductance as a surrogate would allow the substitution of specific conductance measurements for chloride measurements in water quality monitoring. Such a substitution would allow for monitoring of chloride concentrations at substantially lower costs than could be achieved through direct monitoring of chloride using less robust equipment such as chloride probes or chemical analysis. In addition, the availability of automated monitoring equipment for specific conductance enables monitoring of chloride concentrations at much finer time scales than would be possible through sampling for chemical analysis. This potentially provides a more detailed understanding of the dynamics of chloride in stream and river systems, especially in relation to events such as winter storms and snowmelts.

Data analysis by the U.S. Geological Survey (USGS) found a linear relationship between specific conductance and chloride concentration at higher values of conductance and chloride concentration.[2] This suggests that during periods when chloride is being carried into receiving waters by discharges of stormwater or snowmelt, ambient chloride concentrations could be estimated using specific conductance.

The USGS regression model was developed using data from 17 Wisconsin streams, including streams located in the urban and rural areas. Comparison of the chloride concentrations predicted by the USGS regression model to actual historical chloride concentrations in samples collected from the Root River in Milwaukee and Racine Counties, Wisconsin found that the regression model was a poor predictor of chloride concentrations in the Root River. The model usually predicted higher chloride concentrations based on specific conductance than were observed in the River, with an average difference between predicted and observed concentrations of about 25 percent of the observed concentrations and a maximum difference of about 277 percent of observed concentrations.[3]

---

[1] *Acronyms and abbreviations used in this report are defined in Appendix A.*

[2] *S.R. Corsi, D.J. Graczyk, S.W. Geis, N.L. Boot, and K.D. Richards, "A Fresh Look at Road Salt: Aquatic Toxicity and Water Quality Impacts on Local, Regional, and National Scales,"* Environmental Science and Technology, *44: 7,376-7,382, 2010.*

[3] *SEWRPC Community Assistance Planning Report No. 316,* A Restoration Plan for the Root River Watershed, *July 2014.*

As part of the Chloride Impact Study, Commission staff developed and refined regression models for predicting chloride concentration from specific conductance for portions of the Region. These models were developed using simultaneous specific conductance and chloride data collected by Commission staff at 41 stream and river sampling stations throughout the Southeastern Wisconsin Region over the period October 2018 through May 2021.[4] These models will be used throughout the Chloride Impact Study to estimate chloride concentrations from measurements of specific conductance.

## 1.2 RELATIONSHIP OF THIS REPORT TO THE CHLORIDE IMPACT STUDY

This technical report presents some of the findings from the Commission's Chloride Impact Study.[5] This study was initiated due to heightened public concern over the growing use of road salt and evidence of increasing chloride concentrations in surface water and groundwater within the Southeastern Wisconsin Region. The findings of this study are being presented in a series of reports.

Major objectives of the chloride impact study include:

1. Documenting historical and existing conditions and trends in chloride concentrations in surface and groundwater in the Southeastern Wisconsin Region

2. Evaluating the potential for increased amounts of chloride in the environment to cause impacts to surface water, groundwater, and the natural and built environment in the Region

3. Identifying the major sources of chloride to the environment in the Region

4. Investigating and defining the relationship between the introduction of chloride into the environment and the chloride content of surface and groundwater

5. Developing estimates of chloride loads introduced into the environment under existing conditions and forecasts of such loads under planned land use conditions

6. Evaluating the potential effects of climate change on the major sources of chloride under planned land use conditions

7. Reviewing the state-of-the-art of technologies and best management practices affecting chloride inputs to the environment and developing performance and cost information for such practices and technologies

8. Exploring legal and policy options for addressing chloride contributions to the environment

9. Developing and evaluating alternative chloride management scenarios for minimizing impacts to the environment from chloride use while meeting public safety objectives

10. Present recommendations for the management of chloride and mitigation of impacts of chloride on the natural and built environment

This report presents regression models that were developed to estimate chloride concentrations in surface water from measurements of specific conductance. These models were applied to develop estimates of chloride concentrations that were used in analyses in other stages of the Study. These analyses will be presented in other technical reports.

[4] *Data collection is described in detail in SEWRPC Technical Report No. 61,* Field Analyses for the Chloride Impact Study, *September 2023.*

[5] *SEWRPC Planning Report No. 57,* A Chloride Impact Study for Southeastern Wisconsin, *in preparation.*

## 1.3 REPORT FORMAT AND ORGANIZATION

This report documents the development of regression models for estimating concentrations of chloride from specific conductance. It is organized into three chapters.

Following this initial chapter, Chapter 2 summarizes the methods used to develop the regression models. This summary begins with a discussion of the statistical background regarding regression techniques. This is followed by descriptions of the data and statistical methods used to develop the regressions. The chapter also discusses methods used to evaluate the resulting regression models.

Chapter 3 shows the results of the regression analysis, presenting two models for estimating the concentration of chloride from specific conductance. The chapter then provides an evaluation of each model. Finally, the chapter provides guidance for the use of the models.

This report includes two appendices. Appendix A defines acronyms and abbreviations used in this report. Appendix B discusses development of site-specific linear regressions to estimate chloride concentrations from specific conductance for the six lakes sampled by Commission staff as part of the Chloride Impact Study. It also provides some guidance for developing estimates of chloride concentration in other lakes of the Region.

# METHODS 2

## 2.1 INTRODUCTION

As part of the Chloride Impact Study, Commission staff developed and refined two regression models for estimating chloride concentration from specific conductance for streams in portions of the Region. These models were developed using simultaneous specific conductance and chloride data collected at 41 stream and river sampling sites throughout the Southeastern Wisconsin Region.[6] This chapter describes the methods used to develop and evaluate the regression models. The resulting models are described and assessed in Chapter 3 of this report.

## 2.2 STATISTICAL BACKGROUND

### Linear Equations

The simplest form of a linear equation is one with two variables whose graph is a line. Figure 2.1 summarizes some of the characteristics of a simple linear equation. Such an equation takes the form:

$$y = mx + b$$

Where:

*y* and *x* are variables and can take different numerical values

- *y* is the dependent variable
- *x* is the independent variable

*m* and *b* are parameters or constants in the equation

- *m* is the slope of the line or the number of units that y changes when x changes by one unit
- *b* is the y-intercept or the value of y when x = 0

Simple linear equations can be used to describe many processes and relationships in the physical world. For example, they are often used to calculate rates of change that biological processes follow.

### Regression Analysis

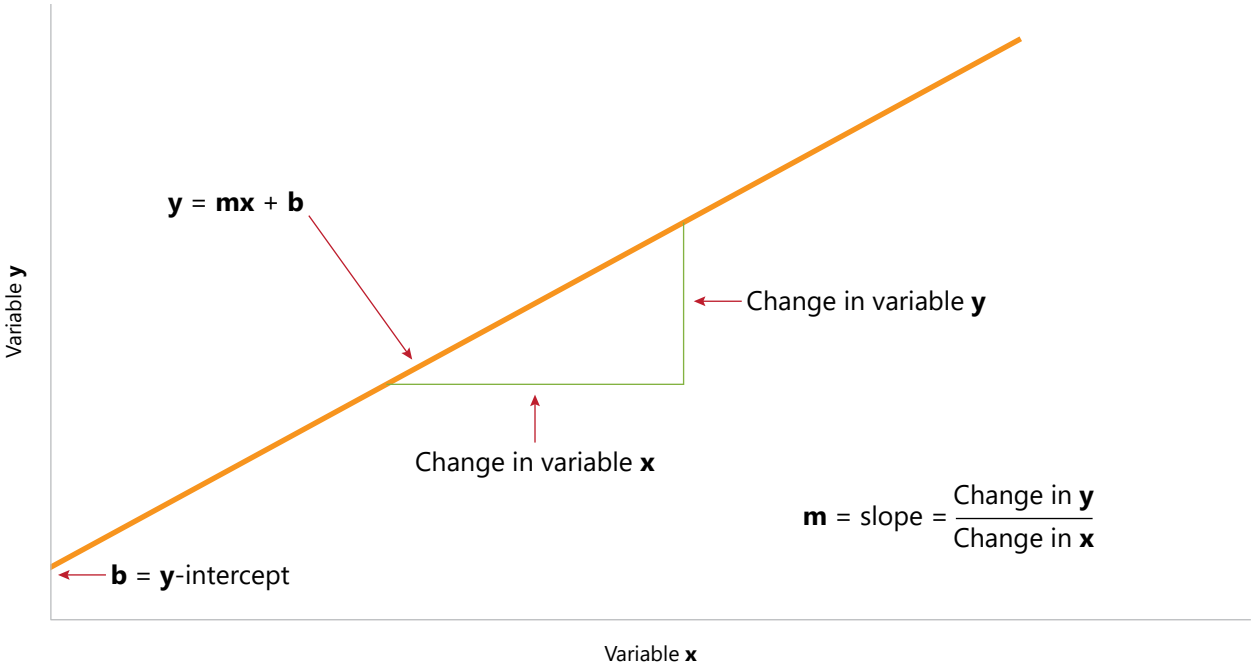
Regression analysis consists of a collection of statistical techniques for assessing and estimating the relationship between a dependent variable and one or more independent variables. Regression analysis can provide:

- A determination of whether a relationship exists between the dependent variable and one or more independent variables
- A mathematical equation that describes the relationship between the variables, if a relationship exists
- An assessment of the accuracy of the description and any predictions achieved by the equation

In the case of multiple regression, which looks at more than one independent variable, it can also assess the relative contribution of each independent variable to variations in the dependent variable.

[6] *Data collection is described in detail in SEWRPC Technical Report No. 61,* Field Analyses for the Chloride Impact Study, *September 2023.*

**Figure 2.1**
**Characteristics of a Linear Equation**




*Source: SEWRPC*

Regression analysis is used for two major purposes. First, the development of a regression model allows the average value of the dependent variable to be predicted or estimated from values of the independent variable or variables. Second, in an experimental context, regression analysis can be used to infer a causal relationship between the independent variables or variables and the dependent variable. The use of regression methods in this study will focus on the first use.

There are several forms of regression analysis. Simple linear regression is the most commonly used form. This method examines the relationship between two variables and finds a line that most closely fits the data according to a specific mathematical condition. This line can be used to predict the average value of the dependent variable when the independent variable takes a given value. Multiple linear regression is a similar method that develops a linear equation relating the dependent variable to more than one independent variable. Piecewise linear regression is a method that partitions the independent variable into intervals and develops separate linear equations for each interval. Linear mixed effects models examine the effects of both fixed and random factors on the dependent variable. Nonlinear regression is a method used when the relationship between the dependent variable and independent variable or variables is not linear.

### *Assumptions of Linear Regression*

Linear regression makes several assumptions about the data used to develop the regression model. Regression models developed from data that do not meet these assumptions may give poor estimates of the average value of the dependent variable. To some extent, these assumptions depend on the mathematical method used to develop the regression model. This discussion will focus on the assumptions of simple linear regression using the method of ordinary least squares, which is a commonly used method of conducting simple linear regression.

#### Data Are Representative of Population of Interest

Linear regression assumes that the sample data used to develop the regression model are representative of the data of interest. Representative data are those in which the sample used to develop the model matches the characteristics of the larger population of interest. These include the ranges of both variables and the shape and variability of the underlying distribution of data. If this assumption is not satisfied, the resulting regression model may give poor estimates of the dependent variable and its variance for a given value of

the independent variable and may be unsuitable for developing confidence and prediction intervals around the regression equation. In such a case, interpretation of the results of the model should be restricted to the population that the data actually represent.

Linear Relationship Between Variables

Linear regression assumes that the dependent variable is linearly related to the independent variable. If this assumption is not satisfied, the resulting regression model will give poor estimates of the dependent variable and its variance for a given value of the independent variable. Therefore, it will be unsuitable for developing confidence and prediction intervals around the regression equation. Using some other functional form will most likely produce a better model of the data. In some cases, this may be done by mathematically transforming one or both variables. Such a transformation can linearize some relationships between variables. For example, the growth of unconstrained populations follows a curvilinear relationship over time. Because of this, models of unconstrained population growth over time generally use time as the independent variable and the logarithm of population size as the dependent variable. This logarithmic transformation of population size linearizes the relationship. Not all relationships between variables can be linearized through transforming the data. More advanced statistical methods such as piecewise regression or nonlinear regression may be more appropriate in these instances.

Independent Variables Are Fixed Values

The underlying statistical model for linear regression assumes that the independent variables are chosen or fixed values rather than random variables, that is that they are assessed without measurement error. For most environmental data, this assumption is generally unrealistic. If this assumption is not satisfied, the resulting regression model could give poor estimates of the variance around estimates of the dependent variable for a given value of the independent variable and may be unsuitable for developing confidence intervals and prediction intervals around the regression equation. These problems can be minimized if the if the independent variable is measured with high accuracy. In any case, the regression model may still be applicable for predicting the average value of the dependent variable for a given value of the independent variable.

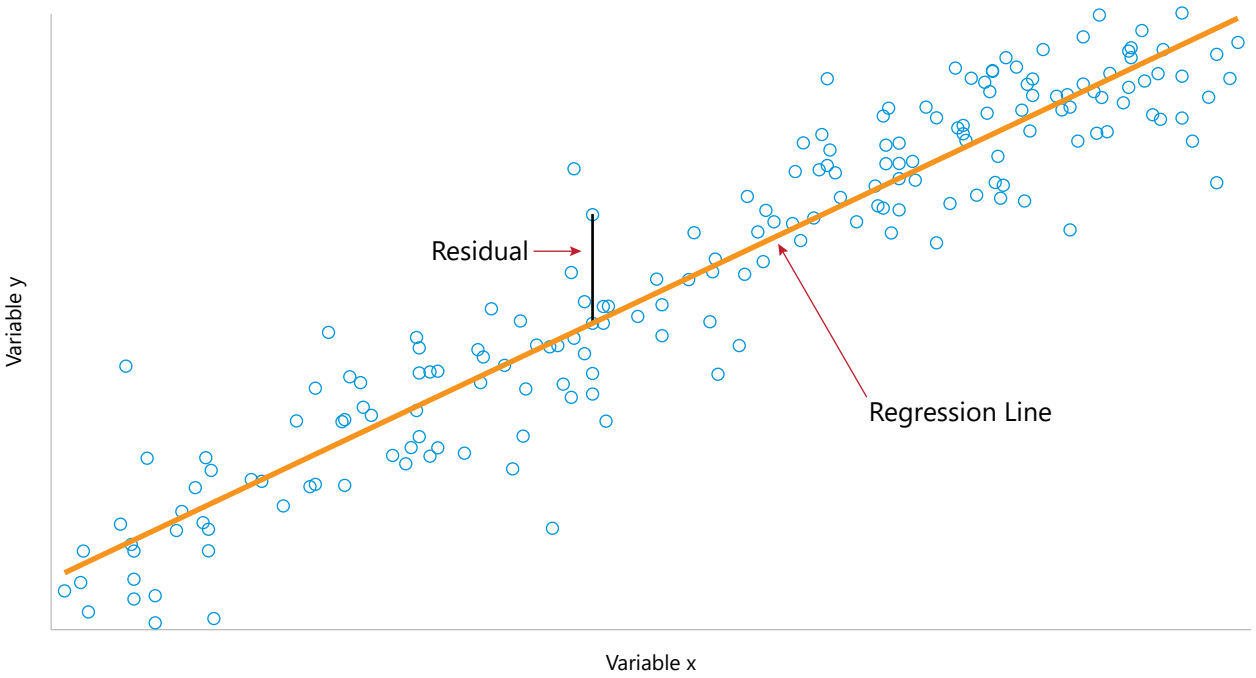
Homoscedasticity

The line developed in a linear regression model represents a mean value of the dependent variable at given values of the independent variable. In this type of model, residuals consist of the deviations from the line fitted by the regression to the observed value. For any data point in a regression, the residual is the difference between the value of the dependent variable in the observed data and the value of the dependent variable predicted by the regression equation for the corresponding value of the independent variable. An example of a residual in a linear regression model is shown in Figure 2.2. The residuals in a linear regression represent the variation in the data that is not accounted for by the relationship between the independent and dependent variables, including both variation due to other factors and random variation.

Linear regression assumes that the variance of the residuals along the regression line is constant. This condition is referred to as homoscedasticity. An example of this condition is shown in Figure 2.3. In the figure, the amount of variation around the regression line does not change as the independent variable changes. Figure 2.4 shows an instance in which the assumption of constant variance of residuals along the regression line is not met. This condition is referred to as heteroscedasticity. In this case, the amount of variation around the line increases as the value of the independent variable increases. Other examples of heteroscedasticity include the amount of variation around the line decreasing as the independent variable increases and more complex relationships.

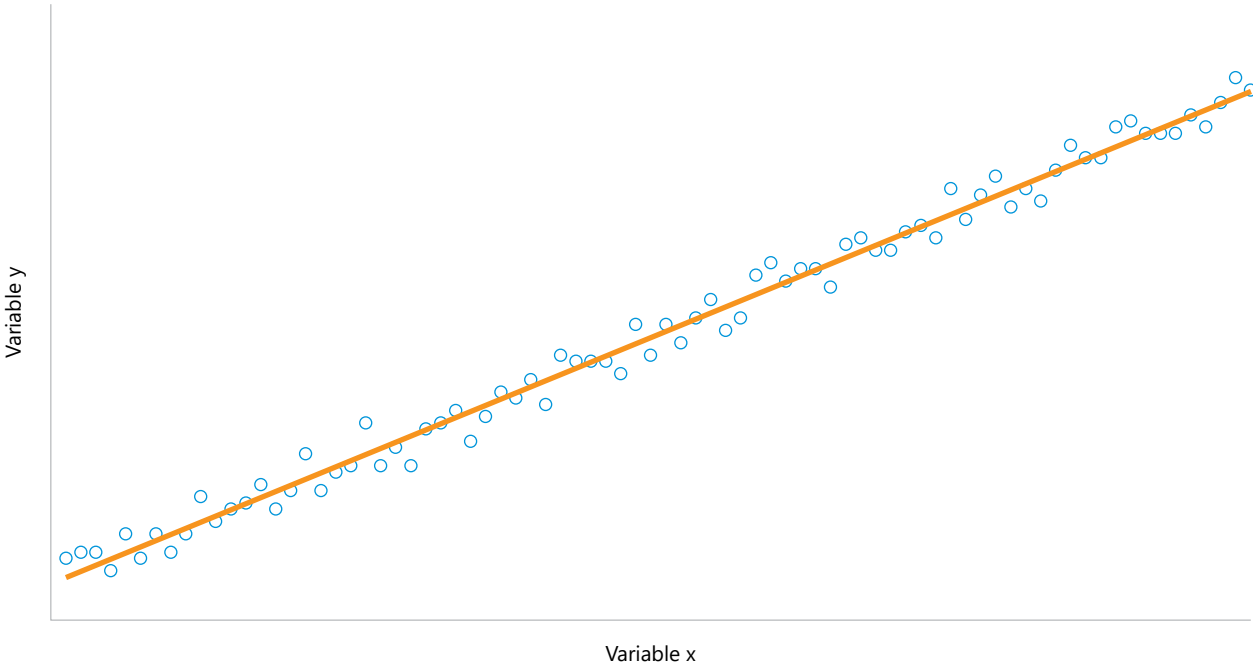
If the assumption of homoscedasticity is not satisfied, the resulting regression model will give poor estimates of the variance around estimates of the dependent variable for a given value of the independent variable and will be unsuitable for developing confidence intervals and prediction intervals around the regression equation. In many cases where this assumption is not met, mathematical transformation of one or both variables may allow a model to better meet the homoscedasticity assumption. Although, such transformation will change the underlying statistical model.

**Figure 2.2**
**Residuals in a Linear Equation**



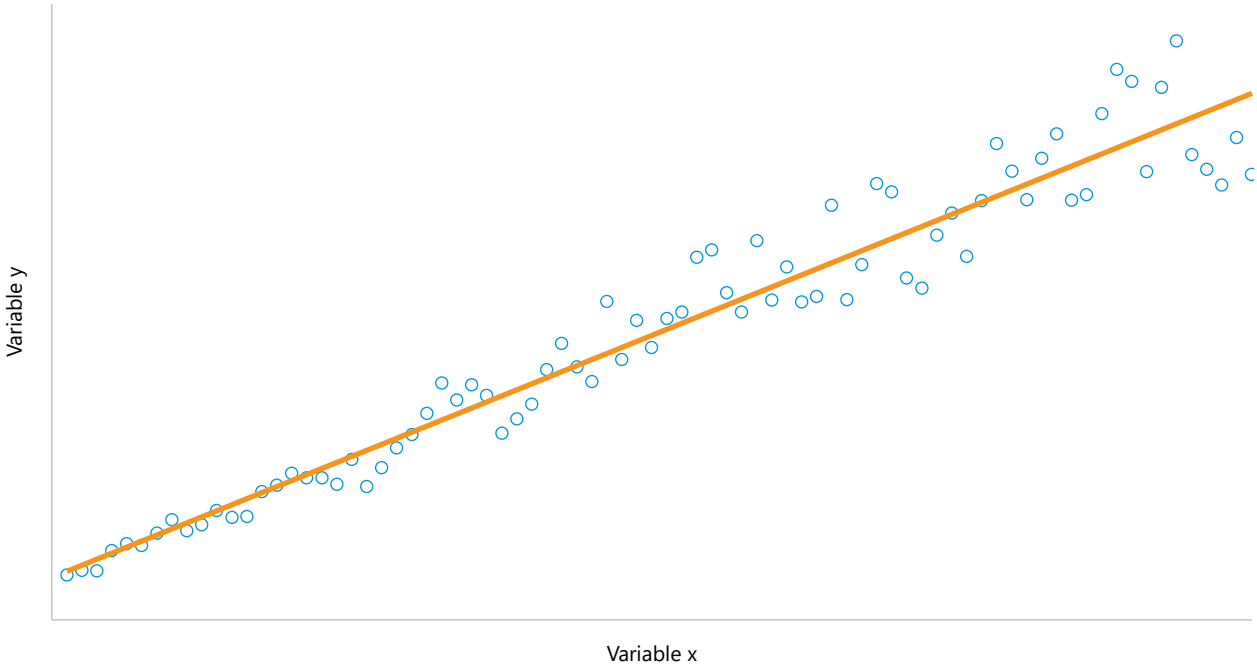

*Source: SEWRPC*

**Figure 2.3**
**Homoscedastic Data**




*Source: SEWRPC*

**Figure 2.4**
**Heteroscedastic Data**




*Source: SEWRPC*

Normal Distribution of Residuals

Linear regression assumes that for any value of the independent variable the residuals are normally distributed about the average value of the dependent variable. If this assumption is not satisfied, the resulting regression model will give poor estimates of the variance around estimates of the dependent variable for a given value of the independent variable. Therefore, the regression model will be unsuitable for developing confidence intervals and prediction intervals around the regression equation. In many cases where this assumption is not met, mathematical transformation of one or both variables may allow a model to better meet the normal distribution of residuals assumption. Such transformation will change the underlying statistical model.

While the residuals of linear regression models must be distributed normally, linear regression does not require that the independent or dependent variables be normally distributed. It should be noted that because the regression line represents a mean, it is sensitive to the presence of outliers in the data much like other means are.

Independence of Residuals

Linear regression assumes that the residuals are independent from one another. The residuals should not show a pattern such as increasing or decreasing in relation to one another or any other factor. This is especially important when time series data are used to construct the model. If this assumption is not satisfied, the resulting regression model will give poor estimates of the variance around estimates of the dependent variable for a given value of the independent variable and will be unsuitable for developing confidence intervals and prediction intervals for the regression equation. In some instances where independence of residuals is not met, there are more sophisticated statistical methods, such as repeated measures linear regression, that can be used to develop a model.

Assumptions that Data Need to Satisfy for Linear Regression

The number of assumptions that a regression model needs to satisfy depends on the purpose for which the regression is to be used.[7] Some uses require that the model satisfies more assumptions than would need to be met for other purposes. If the regression is going to be used to test hypotheses or if confidence or prediction intervals are to be developed around the resulting equation, the regression will need to satisfy all the assumptions previously discussed. If the regression is only to be used to predict values of the dependent variable from values of the independent variable, the regression will need to satisfy only the assumptions that the variables are linearly related, and that the data are representative of the population of interest. While satisfying other assumptions may result in better predictions, the regression equation will still be usable for predictive purposes if they are not met. Other uses of the regression model may require that more assumptions be met.

## 2.3 DATA USED TO DEVELOP REGRESSIONS

**Site Selection**

Regression models to estimate instream chloride concentration were developed using data gathered as part of the Chloride Impact Study. Commission staff conducted water quality monitoring to collect chloride concentrations and specific conductance data. These data were collected at 41 sampling sites on streams within the Southeastern Wisconsin Region. These sites are shown on Map 2.1 and described in Table 2.1. Sites were selected to ensure a balanced geographic distribution of monitoring locations among the counties and watersheds in the Region. In addition, monitoring sites were selected to provide a set of locations that are representative of the variety of stream conditions within the Region. Factors considered in the selection of sites included:

- Land use in the areas draining to the monitoring sites
- The presence and absence of wastewater treatment facilities and stormwater management systems discharging into streams
- The size of streams, including both stream order and observed or modeled stream discharge
- The locations of U.S. Geological Survey (USGS)[8] stream gages
- Sources of water supply in areas draining to monitoring sites
- Presence of chloride-related impaired water designations pursuant to Section 303(d) of the Federal Clean Water Act
- Availability of historical water quality monitoring data
- Conditions within the stream and riparian area
- Legal and safe access to monitoring sites

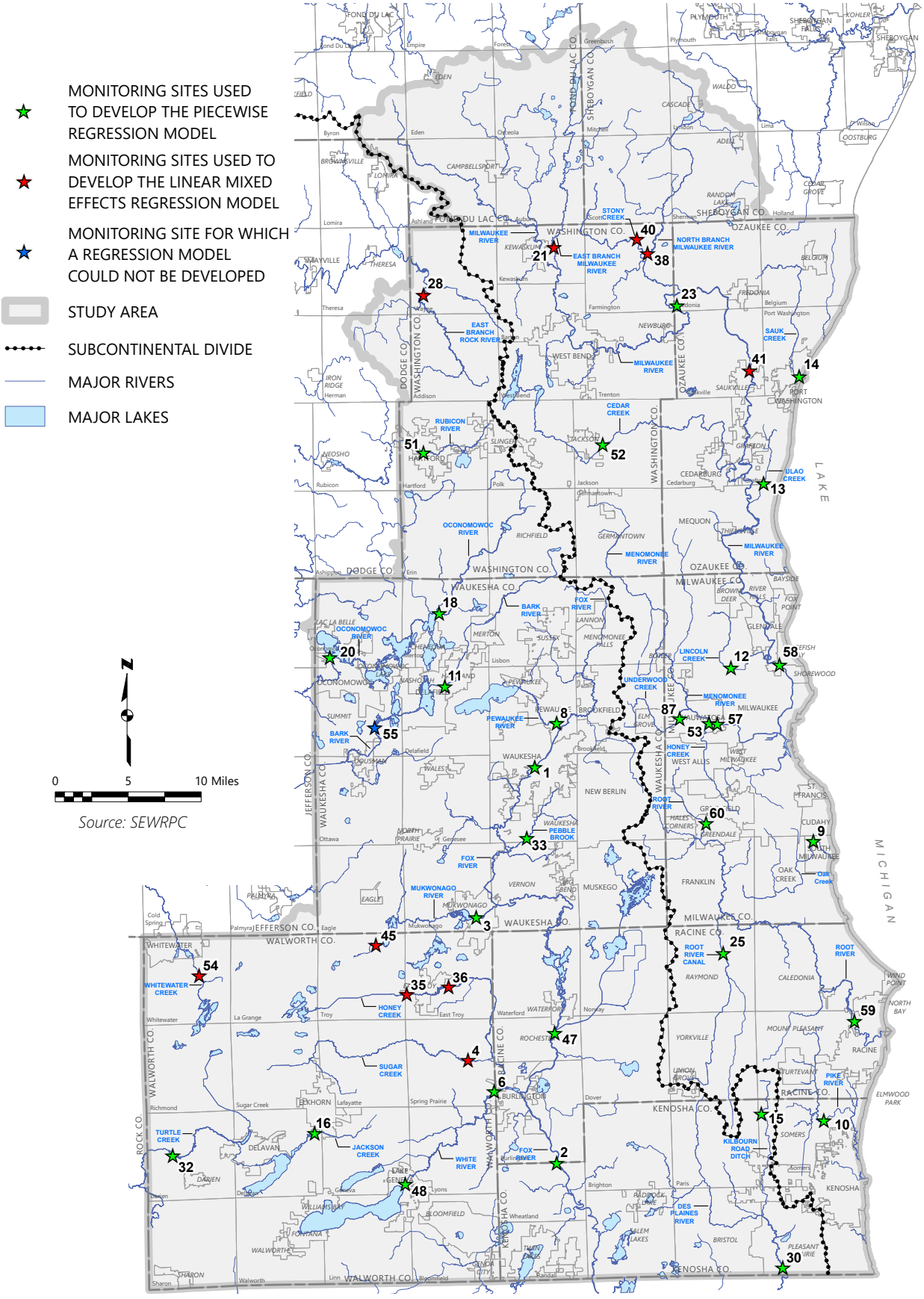
A description of the process used to select sites for monitoring for the Chloride Impact Study is given in a separate technical report.[9] That report also provides descriptions of the monitoring sites and their contributing drainage areas.

[7] *D.R. Helsel et al.,* Statistical Methods in Water Resources*, U.S. Geological Survey Techniques and Methods, Book 4, Hydrological Analysis and Interpretation, Section A, Statistical Analysis, Chapter 3, 2020.*

[8] *Acronyms and abbreviations used in this report are defined in Appendix A.*

[9] *SEWRPC Technical Report No. 61,* Field Monitoring and Data Collection for the Chloride Impact Study*, September 2023.*

## Map 2.1
## Stream Monitoring Sites Used for Developing Regression Models for the Chloride Impact Study



## Table 2.1
## Stream Monitoring Sites for the Chloride Impact Study

| Site ID[a] | Site Name | Major Watershed | Site County | Counties Within Drainage Area[b] | Drainage Area Size (sq mi) | SWIMS Station ID | Nearest USGS Streamgage | Latitude | Longitude | Site Location |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Fox River at Waukesha | Fox River | Waukesha | Waukesha, Washington | 126.3 | 683310 | 05543830 | 43.00501682 | -88.24428955 | Fox River about 100 feet downstream of Prairie Avenue near USGS Gage 05543830 at Waukesha |
| 2 | Fox River at New Munster | Fox River | Kenosha | Waukesha, Walworth, Racine, Kenosha, Jefferson, Milwaukee, Washington | 807.1 | 523093 | 05545750 | 42.61102994 | -88.22575534 | Fox River about 30 feet downstream of CTH JB near USGS Gage 05545750 at New Munster |
| 3 | Mukwonago River at Mukwonago | Fox River | Waukesha | Waukesha, Walworth, Jefferson | 85.4 | 10032435 | 05544200 | 42.85698382 | -88.32736057 | Mukwonago River 35 feet downstream of STH 83 and 200 feet downstream of USGS Gage 05544200 at Mukwonago |
| 4 | Sugar Creek | Fox River | Walworth | Walworth | 60.5 | 10029083 | -- | 42.71494642 | -88.34238151 | Sugar Creek about 60 feet upstream of Potter Road near Spring Prairie |
| 6 | White River near Burlington | Fox River | Walworth | Walworth, Racine, Kenosha | 112.2 | 653104 | -- | 42.68340253 | -88.30797773 | White River 40 feet downstream of CTH JS near Burlington |
| 8 | Pewaukee River | Fox River | Waukesha | Waukesha | 38.1 | 10051685 | -- | 43.04793066 | -88.21308887 | Pewaukee River at Steinhafels about 1,000 feet downstream of Busse Road |
| 9 | Oak Creek | Oak Creek | Milwaukee | Milwaukee | 25.8 | 413913 | 04087204 | 42.92486133 | -87.86938351 | Oak Creek 385 feet downstream of 15th Avenue and USGS Gage 04087204 at Oak Creek |
| 10 | Pike River | Pike River | Kenosha | Kenosha, Racine | 36.6 | 10034961 | 04087257 | 42.64700492 | -87.86516338 | Pike River at Petrifying Springs Park about 1,500 feet upstream of USGS Gage 04087257 |
| 11 | Bark River Upstream | Rock River | Waukesha | Waukesha, Washington | 35.0 | 683427 | 05426067 | 43.15954154 | -88.36944299 | Bark River about 100 feet downstream of STH 83 and about 3,950 feet upstream of USGS Gage 05426067 at Nagawicka Road |
| 12 | Lincoln Creek | Milwaukee River | Milwaukee | Milwaukee | 11.0 | 10047562 | 040869416 | 43.09927104 | -87.97527082 | Lincoln Creek about 400 feet downstream of 51st Blvd and about 2,500 feet upstream of USGS 040869416 Gage at Sherman Boulevard |
| 13 | Ulao Creek | Milwaukee River | Ozaukee | Ozaukee | 9.2 | 10050932 | -- | 43.28115708 | -87.92473975 | Ulao Creek about 40 feet downstream of CTH W |
| 14 | Sauk Creek | Sauk Creek | Ozaukee | Ozaukee, Sheboygan | 31.7 | 10030655 | -- | 43.38648777 | -87.87253643 | Sauk Creek about 400 feet upstream of Wisconsin Street |
| 15 | Kilbourn Road Ditch | Des Plaines River | Kenosha | Racine, Kenosha | 8.5 | 10051686 | -- | 42.65507120 | -87.94899341 | Kilbourn Road Ditch at CTH A |
| 16 | Jackson Creek | Rock River | Walworth | Walworth | 9.8 | 10051687 | 05431016 | 42.64536095 | -88.55068624 | Jackson Creek about 3,000 feet downstream of STH 67 and about 4,400 feet upstream of USGS Gage 05431016 at Mound Road |
| 18 | Oconomowoc River Upstream | Rock River | Waukesha | Washington, Waukesha | 41.3 | 683245 | -- | 43.11796620 | -88.51890233 | Oconomowoc River about 325 feet upstream of STH 83 |

**Table continued on next page.**

## Table 2.1 (Continued)

| Site ID[a] | Site Name | Major Watershed | Site County | Counties Within Drainage Area[b] | Drainage Area Size (sq mi) | SWIMS Station ID | Nearest USGS Streamgage | Latitude | Longitude | Site Location |
|---|---|---|---|---|---|---|---|---|---|---|
| 20 | Oconomowoc River Downstream | Rock River | Waukesha | Waukesha, Washington, Dodge, Jefferson | 100.4 | 10051688 | -- | 43.47604420 | -88.38240756 | Oconomowoc River near Lac La Belle Outlet about 75 feet downstream of STH 16 |
| 21 | East Branch Milwaukee River | Milwaukee River | Washington | Sheboygan, Fond Du Lac, Washington | 49.4 | 10051139 | -- | 43.52109322 | -88.20310120 | East Branch Milwaukee River at STH 28 |
| 23 | Milwaukee River Downstream of Newburg | Milwaukee River | Ozaukee | Fond Du Lac, Washington, Sheboygan, Ozaukee, Dodge | 264.6 | 10051689 | -- | 43.46025398 | -88.03691368 | Milwaukee River about 1,000 feet upstream of Hickory Drive (extended) and Washington/Ozaukee County line |
| 25 | Root River Canal | Root River | Racine | Racine, Kenosha | 58.8 | 10016596 | 04087233 | 42.81548800 | -87.99495284 | Root River Canal at USGS Gage 04087233 at 6 Mile Road (CTH G) |
| 28 | East Branch Rock River | Rock River | Washington | Washington, Dodge | 54.7 | 10032027 | -- | 42.62553785 | -88.74234642 | East Branch Rock River about 80 feet downstream of CTH D |
| 30 | Des Plaines River | Des Plaines River | Kenosha | Kenosha, Racine | 114.6 | 303054 | 05527800 | 42.50164176 | -87.92539857 | Des Plaines River at 122nd St (CTH ML) about 7,800 feet upstream of USGS Gage 05527800 at Russel Road (Illinois) |
| 32 | Turtle Creek | Rock River | Walworth | Walworth | 94.0 | 10051690 | -- | 43.31952281 | -88.38667623 | Turtle Creek about 230 feet upstream of USH 14 |
| 33 | Pebble Brook | Fox River | Waukesha | Waukesha | 16.0 | 10008183 | -- | 42.93472331 | -88.25683580 | Pebble Brook about 300 feet upstream of CTH XX |
| 35 | Honey Creek Upstream of East Troy | Fox River | Walworth | Walworth | 37.7 | 10032440 | -- | 42.78177625 | -88.42317446 | Honey Creek about 800 feet downstream of Townline Road at Michael Fields Agricultural Institute |
| 36 | Honey Creek Downstream of East Troy | Fox River | Walworth | Walworth | 44.6 | 653244 | -- | 42.78823546 | -88.36653679 | Honey Creek at Carver School Road |
| 38 | North Branch Milwaukee River | Milwaukee River | Washington | Sheboygan, Ozaukee, Washington | 105.8 | 10029089 | -- | 43.51262786 | -88.07534337 | North Branch Milwaukee River about 25 feet downstream of CTH XX |
| 40 | Stony Creek | Milwaukee River | Washington | Washington, Sheboygan, Fond Du Lac | 17.8 | 673267 | -- | 43.52741053 | -88.08937392 | Stony Creek at CTH X |
| 41 | Milwaukee River near Saukville | Milwaukee River | Ozaukee | Fond Du Lac, Washington, Sheboygan, Ozaukee, Dodge | 448.3 | 10051691 | -- | 43.39366252 | -87.94024145 | Milwaukee River near Friendship Lane (extended) near Saukville |
| 45 | Mukwonago River at Nature Road | Fox River | Walworth | Walworth, Waukesha, Jefferson | 24.4 | 10029287 | -- | 42.83108888 | -88.46375625 | Mukwonago River about 150 feet downstream of Nature Road and upstream of Lulu Lake |
| 47 | Fox River at Rochester | Fox River | Racine | Waukesha, Racine, Walworth, Jefferson, Milwaukee, Washington | 455.6 | 10032438 | 05544475[c] | 42.74014301 | -88.22477829 | Fox River about 1,700 feet upstream of Rochester Dam near USGS Gage 05544475 at Rochester |
| 48 | White River at Lake Geneva | Fox River | Walworth | Walworth | 29.1 | 10051692 | 055451345 | 42.59328722 | -88.43008313 | White River about 1,430 feet downstream of Geneva Lake outlet and USGS Gage 055451345 |
| 51 | Rubicon River | Rock River | Washington | Washington, Dodge | 27.5 | 10051693 | -- | 42.80382218 | -88.70293308 | Rubicon River at West Side Park about 250 feet upstream of Grant Street |
| 52 | Cedar Creek | Milwaukee River | Washington | Washington, Ozaukee | 53.6 | 673048 | -- | 43.32350934 | -88.14256630 | Cedar Creek about 150 feet upstream of STH 60, east of Jackson |

**Table continued on next page.**

## Table 2.1 (Continued)

| Site ID[a] | Site Name | Major Watershed | Site County | Counties Within Drainage Area[b] | Drainage Area Size (sq mi) | SWIMS Station ID | Nearest USGS Streamgage | Latitude | Longitude | Site Location |
|---|---|---|---|---|---|---|---|---|---|---|
| 53 | Honey Creek at Wauwatosa | Menomonee River | Milwaukee | Milwaukee | 10.7 | 10030407 | 04087119 | 43.04426929 | -88.00683244 | Honey Creek about 1,500 feet upstream of the confluence with the Menomonee River and about 600 feet upstream of USGS Gage 04087119 |
| 54 | Whitewater Creek | Rock River | Walworth | Walworth | 18.8 | 653291 | -- | 43.04745799 | -88.45981016 | Whitewater Creek about 30 feet upstream of Millis Road near Whitewater |
| 55 | Bark River Downstream | Rock River | Waukesha | Waukesha, Washington | 53.2 | 683424 | -- | 43.15954154 | -88.36944299 | Bark River about 50 feet upstream of Genesee Lake Road |
| 57 | Menomonee River at Wauwatosa | Menomonee River | Milwaukee | Milwaukee, Waukesha, Washington, Ozaukee | 124.5 | 10012584 | 04087120 | 43.04348983 | -87.99543034 | Menomonee River near Jacobus Park and about 1,500 feet downstream of USGS Gage 04087120 at 70th Street in Wauwatosa |
| 58 | Milwaukee River at Estabrook Park | Milwaukee River | Milwaukee | Washington, Ozaukee, Fond Du Lac, Sheboygan, Milwaukee, Dodge | 684.7 | 413640 | 04087000 | 43.10080823 | -87.90949931 | Milwaukee River at Estabrook Park about 2,100 feet downstream of Port Washington Road and 330 feet upstream of USGS Gage 04087000 |
| 59 | Root River near Horlick Dam | Root River | Racine | Racine, Milwaukee, Waukesha, Kenosha | 189.7 | 10044817 | 04087240 | 42.74522748 | -87.82038887 | Root River at Racine Country Club Golf Course Bridge and about 2,600 feet downstream USGS Gage 04087240 at STH 38 |
| 60 | Root River at Grange Avenue | Root River | Milwaukee | Milwaukee, Waukesha | 15.0 | 413716 | 04087214 | 42.94500273 | -88.01399744 | Root River near USGS Gage 04087214 |
| 87 | Underwood Creek | Menomonee River | Milwaukee | Waukesha, Milwaukee | 19.0 | 10031613 | 04087088 | 43.05008628 | -88.04639671 | Underwood Creek at Gravel Sholes Park about 870 feet downstream of STH 100 at USGS Gage 04087088 |

[a] *See Map 2.1 for locations of each monitored site.*

[b] *Counties are listed in the order of largest proportion of the drainage area.*

[c] *The USGS gage on the Fox River at Rochester only measures water level and does not measure streamflow discharge.*

*Source: SEWRPC*

### Sample Collection

The data used to develop the regression models consisted of paired values of chloride concentration and specific conductance. Chloride values came from water samples collected at stream monitoring sites. In most cases, the water sample was collected at the location of the in-stream CTD-10 sensor that was used to continuously monitor specific conductance, water temperature, and water level. In some instances, this was not possible because the site of the sensor could not be safely accessed due to the presence of high water or ice. In those cases, the sample was collected as close to the sensor as was possible to do safely. These samples were sent to the Wisconsin State Laboratory of Hygiene for chemical analysis.

Samples included those collected during regular monthly sampling throughout the study period as well as event-based samples collected at some sites predominantly during winter snow events and snowmelt events. The event-based sampling was conducted to capture the full range of chloride concentrations and specific conductance levels that occur at monitoring sites within the Region. Figure 2.5 shows the levels of chloride and specific conductance from monthly and event sampling at the monitoring station on the Fox River at Waukesha (Site 1). The figure shows that regular monthly sampling often fails to capture the higher values of chloride concentration and specific conductance levels that can be associated with winter storms and snowmelt events. A description of sampling methods and monitoring techniques used in the Chloride Impact Study is provided in a separate technical report.[10]

As part of developing the regression models the chloride concentrations in water quality samples were paired with the specific conductance level in the stream at the time of sample collection. In most cases, chloride concentrations were paired with the specific conductance reported by the CTD-10 sensor that was installed at the monitoring site. In a few instances, the data from the CTD-10 sensor were adjusted due to dampening from fouling by sediment, algae, bacteria, or macroinvertebrates. In other cases, chloride concentrations were paired with specific conductance measured at the site of water sample collection using a handheld Aqua TROLL 500 multiparameter sonde. Data from the sonde were used when they were judged to be more representative of the water sample than data from the CTD-10 sensor. This occurred when the presence of ice or safety considerations required collection of the water sample at some location other than the site of the CTD-10 sensor or when data from the sensor showed evidence of dampening for which the data could not be adjusted. Information on installation and operation of the CTD-10 sensors, adjustment of specific conductance data from the CTD-10 sensors for dampening, and calibration and use of the Aqua TROLL 500 sonde are presented in a separate technical report.[11]

### Regression Dataset

Figure 2.6 shows the full stream dataset used to develop the Chloride Impact Study regression models. Monthly water samples were collected at 41 stream monitoring sites over 25 months from October 2018 through October 2020 (see Map 2.1 and Table 2.1). In addition, targeted sampling was conducted during winter storm and snowmelt events at 27 selected sites throughout the study period and through the 2020-2021 winter season. The dataset consisted of 1,104 samples and included 998 monthly samples and 106 event samples.

## 2.4 STATISTICAL METHODS

### Statistical Models

All statistical analyses used to develop the regression models for the Chloride Impact Study were conducted using the R statistical programing language.[12] Three methods were used. First, simple linear regressions were developed for each site to examine whether there were significant relationships between specific conductance and chloride. Next, piecewise linear regression was used to develop a general model that could be applied to data from most of the stream monitoring sites. Finally, a linear mixed effects model was developed for those monitoring sites at which the piecewise regression could not be applied. These techniques are described in the following three subsections.

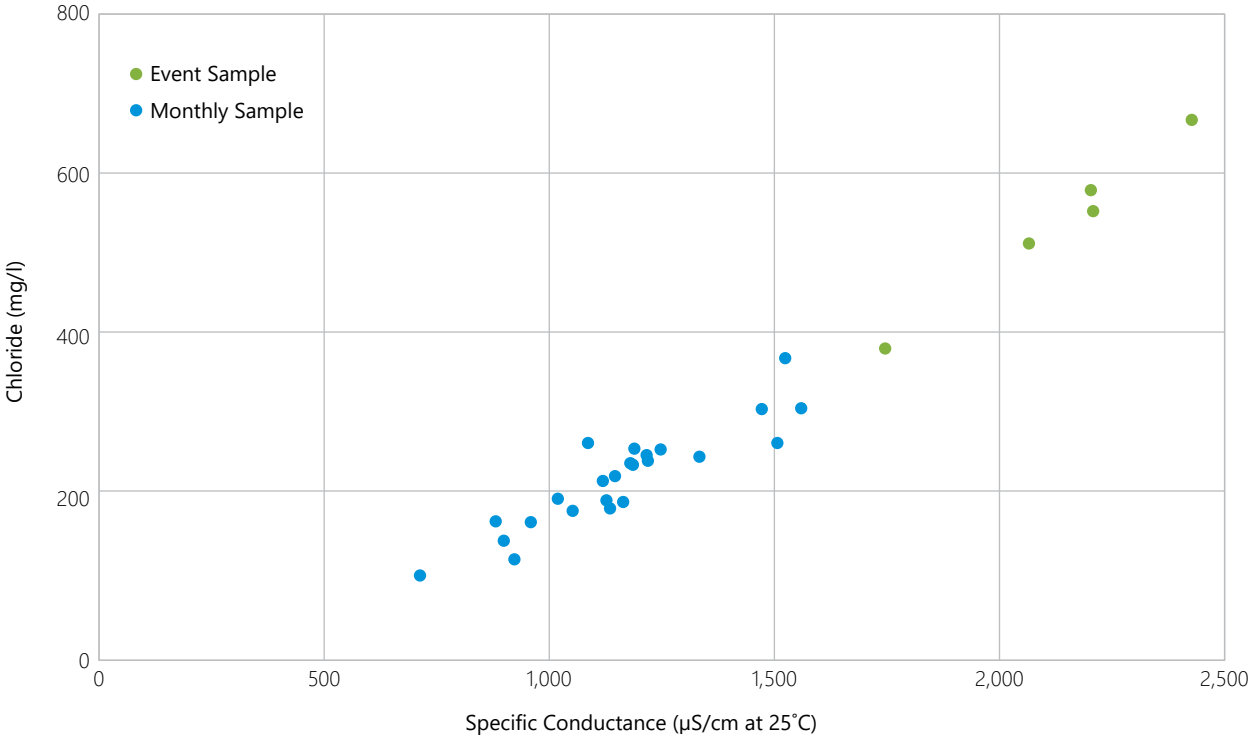
---

[10] Ibid.

[11] Ibid.

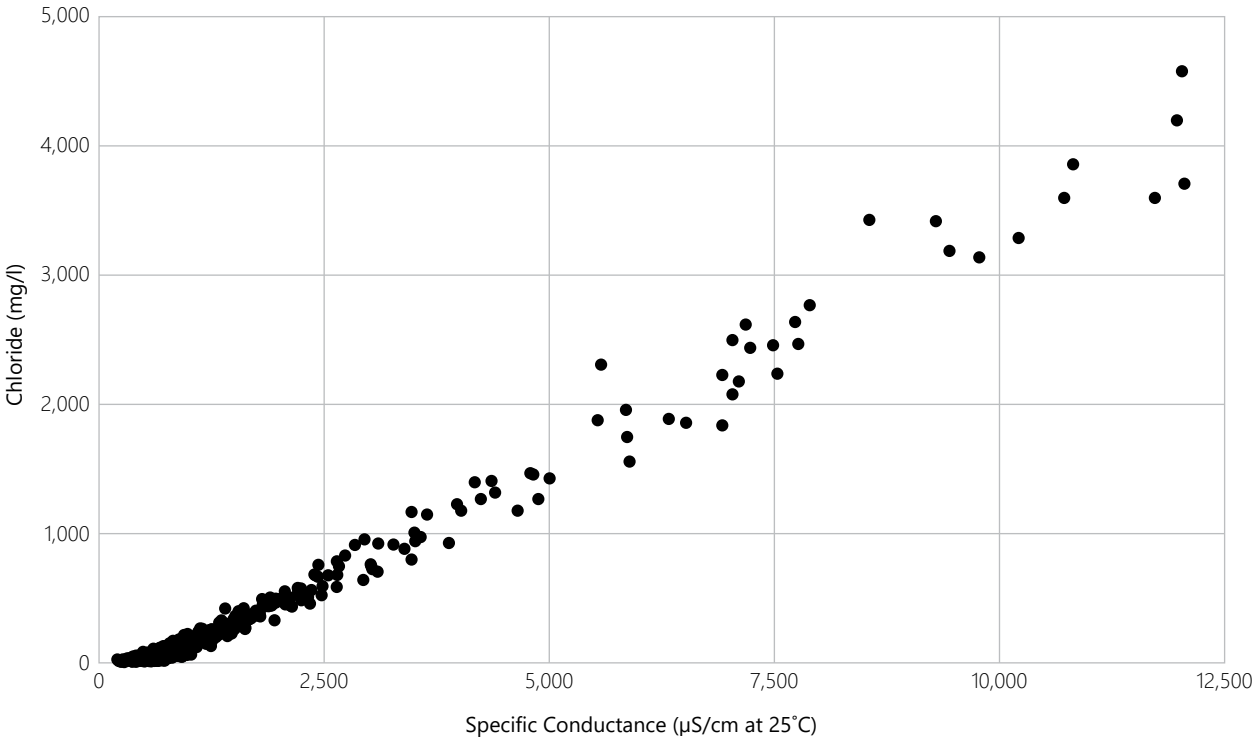
[12] *R Core Team,* R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, cran.r-project.org, 2021.*

## Figure 2.5
## Chloride Concentrations and Specific Conductance Levels from Monthly and Event Samples at Site 1, the Fox River in Waukesha




*Source: SEWRPC*

## Figure 2.6
## Paired Specific Conductance and Chloride for Chloride Study Sites




*Source: SEWRPC*

### *Simple Linear Regression*

Simple linear regression examines the relationship between two variables and finds a line that most closely fits the data according to a specific mathematical condition. While several mathematical criteria can be used for choosing an equation that best fits the data, simple linear regression often uses the method of ordinary least squares. This method calculates an equation for the line that minimizes the sum of squares of the residuals, or deviations from the line.

Regression equations developed using the method of ordinary least squares have four properties. First, the mean of the residuals in a regression model developed using this method equals zero. Second, the mean of the predicted values of the dependent variable resulting from the regression equals the mean of the corresponding observed values in the data used to develop the regression. Third, the regression line passes through the centroid of the data used to develop the regression. And fourth, the variance of the predicted values of the dependent variable is less than the variance of the observed responses, unless the dependent and independent variables are perfectly correlated.[13]

Linear regression models were developed in R using the "lm" function from the "stats" package.[14] Summary statistics for each linear regression model were calculated using the "summary" function from base R as well as the "tidy" and "glance" functions from the "broom" package.[15,16] Functions from the "tidyverse" package were also utilized to format and graph data.[17]

### *Piecewise Linear Regression*

Piecewise linear regression is a method that partitions the independent variable into intervals and develops separate linear equations for each interval.[18] This method is useful when the relationship between the independent variable and the dependent variable in different regions of the data follows different linear trends. The piecewise method fits separate line segments to each interval of the data using simple linear regression.

There are three important considerations in developing a piecewise regression model. First, the model should use the fewest number of segments necessary to sufficiently model the data. Including too many segments could lead to overfitting by the model which can reduce its performance in making predictions. Second, developing a piecewise regression model requires that breakpoints be identified. These are the points at which the slope of the regression model changes. In some cases, these points may be known before analysis. More often, they need to be determined as part of the development of the model. This is generally done using an iterative algorithm which varies potential breakpoints and seeks to minimize the sum of the squares of the residuals across the entire model. Third, the regression model should be continuous at the breakpoints. This means that the two equations used for the line segments on either side of a breakpoint should predict equal values of the dependent variable at the breakpoint.

Piecewise regression produces a model consisting of a system of linear equations. Each equation models a single line segment and is valid only within its specified range.

---

[13] *Perfect correlation of these variables requires that the $R^2$ of the regression relationship equals 1.0. The $R^2$ of a regression equation is discussed in the section on evaluating regression models later in this Chapter.*

[14] *R Core Team, 2021,* op. cit.

[15] Ibid.

[16] *D. Robinson, A. Hayes, and S. Couch,* broom: Convert Statistical Objects into Tidy Tibbles, *R package version 1.0.3, 2023.*

[17] *H. Wickham et al., "Welcome to the Tidyverse,"* Journal of Open Source Software, *4(43): 1686, 2019*.

[18] *Piecewise regression is also known as segmented regression and broken-stick regression.*

For the Chloride Impact Study the piecewise regression models were developed in R by first creating a linear regression model using the "lm" function from the "stats" package and then using "segmented" function from the "segmented" package on this linear regression model.[19,20,21] Starting specific conductance breakpoint values were provided to the "segmented" function to help estimate optimal breakpoints. Summary statistics for the piecewise regression model were calculated using the "summary.segmented" function from the "segmented" package. Functions from the "tidyverse" package were also utilized to format and graph data.[22]

***Linear Mixed Effects Model***

Linear mixed effects regression models are an extension of simple linear regression that examines the effects of both fixed factors and random factors on the dependent variable. This method is often used when data are not independent from one another. Such non-independence can arise from hierarchical structure within the data. An example of this would be data from sites on several streams that have drainage areas with different bedrock or soil chemistry that contribute to the baseline or background water chemistry at the site. The data in this example are structured into two levels: a site level consisting of samples collected at single sites and a group level consisting of the sampling of several different sites.

When there are multiple levels in the data, the variation in the dependent variable can occur either within each group or between the groups. The group-level data are independent of one another. The within-group data are not, in the stream example they consist of repeated measurements at the same site. There are several ways to deal with non-independence within hierarchical data. For example, a regression could be developed using the mean values of the independent and dependent variables from each site. While this approach will yield estimates, it does not take advantage of all the information in the data. Important differences may be obscured or lost by averaging the data. In addition, the model developed from mean values would not be able to estimate individual measurements at a single site. It could only be used to estimate average concentrations at different sites. Alternatively, the data could be analyzed by developing separate linear regressions for each site. This approach results in the production of multiple models that are not closely related to one another. In addition, none of these models take advantage of data from all the sites. Finally, the results from these individual models can be noisy because each model is based on a relatively small amount of data.

Linear mixed effects models explore the differences of effects within groups and effects between groups of data. They incorporate both fixed and random effects within the statistical model. The fixed effects come from the fixed nature of the independent variable and are reflected in a fixed parameter value for the relationship between the dependent and independent variables. The random effects account for the variation among sites. These are assumed to follow a distribution such as the normal distribution. The random effects are reflected in a random parameter value for the relationship between the dependent and independent variables.

For example, a simple linear regression has two parameters: the slope (m) and the y-intercept (b). A linear mixed effects regression model for samples collected at multiple sites could be designed with the slope as a fixed parameter and the y-intercept as a random parameter. The result of this model would be a set of equations with one equation for each site. Each equation would have the same slope but could potentially have different y-intercepts. This would reflect variation between sites in the relationship between the dependent and independent variables. The model would also produce a consensus equation with the same slope and an average y-intercept.

[19] *R Core Team, 2021,* op. cit.

[20] *V.M.R. Muggeo, "Estimating Regression Models with Unknown Break-points,"* Statistics in Medicine, *22: 3055-3071, 2003.*

[21] *V.M.R. Muggeo*, *"segmented: an R Package to Fit Regression Models with Broken-Line Relationships,"* R News 8(1): 202-25, *2008*.

[22] *Wickham et al., 2019,* op. cit.

For the Chloride Impact Study the linear mixed effects models were developed using the "lme" function from the "nlme" package.[23] Conditional and marginal $R^2$ values for the linear mixed effects regression models were calculated using the "r2_nakagawa" function from the "performance" package.[24] Summary statistics for the linear mixed effects model were calculated using the "summary.lme" function from the "nlme" package. Functions from the "tidyverse" package were also utilized to format and graph data.[25]

## 2.5 DEVELOPMENT OF REGRESSION MODELS

A series of steps was used to develop regression models to predict chloride concentration from specific conductance for the Chloride Impact Study. Each step was based on examination of the data and the results of the previous step.

### Preliminary Simple Linear Regression

First, a preliminary simple linear regression was developed for the entire stream dataset. This model is shown in Figure 2.7. Examination of the residuals from this model showed that it does not meet the assumptions of linear regression. The residuals were not distributed normally about the average value of chloride concentration. This was especially the case at lower values of specific conductance. Below specific conductance values of about 1,000 microSiemens per centimeter (µS/cm), the relationship between chloride and specific conductance appears less steep than at higher specific conductance values. At specific conductance values below about 440 µS/cm, this regression model consistently predicted lower concentrations of chloride than were observed in the data. Between specific conductance values of about 440 µS/cm and about 1,200 µS/cm, this model generally predicted higher concentrations of chloride than were observed in the data. Above about 1,200 µS/cm, the model did not systematically underpredict or overpredict chloride concentrations in the data; however, heteroscedasticity was present in the data with the variation about the regression line increasing with increasing values of specific conductance. Comparing the cloud of data points to the line developed by the model suggests that the relationship between chloride and specific conductance is not strictly linear. Because of these issues, Commission staff concluded that an adequate model of the relationship between chloride concentration and specific conductance could not be developed through simple linear regression using these two variables.

### Simple Linear Regression with Transformed Data

Attempts were made to improve the performance of the simple linear regression model by transforming the data. For example, logarithmic transformation was examined. Chloride concentration was converted to the common logarithm of chloride concentration using the "log" function in base R.[26] A simple linear regression model was developed for predicting the logarithm of chloride concentration from specific conductance. Transformation did not linearize the relationship between chloride concentration and specific conductance or produce a better fit to the data. Examination of the residuals showed that this model deviated from the assumptions of linear regression more than did the model that used untransformed data.

### Preliminary Piecewise Regression

A preliminary examination of the Chloride Impact Study entire stream dataset was conducted using piecewise linear regression. This model is shown in Figure 2.8. This model included two segments: one representing the relationship between chloride concentration and specific conductance at lower levels of both variables and another representing the relationship at higher levels of the variables. The breakpoint between the two segments was determined algorithmically by minimizing the sum of the squares of the residuals across the entire model.[27] The breakpoint in this model was at a specific conductance of 725 µS/cm. This corresponded to a predicted chloride concentration of 63 milligrams per liter (mg/l).

---

[23] *J. Pinheiro, D. Bates, S. DebRoy, and D. Sarkar,* nlme: Linear and Nonlinear Mixed Effects Models, *R package version 3.1-152, 2021*.

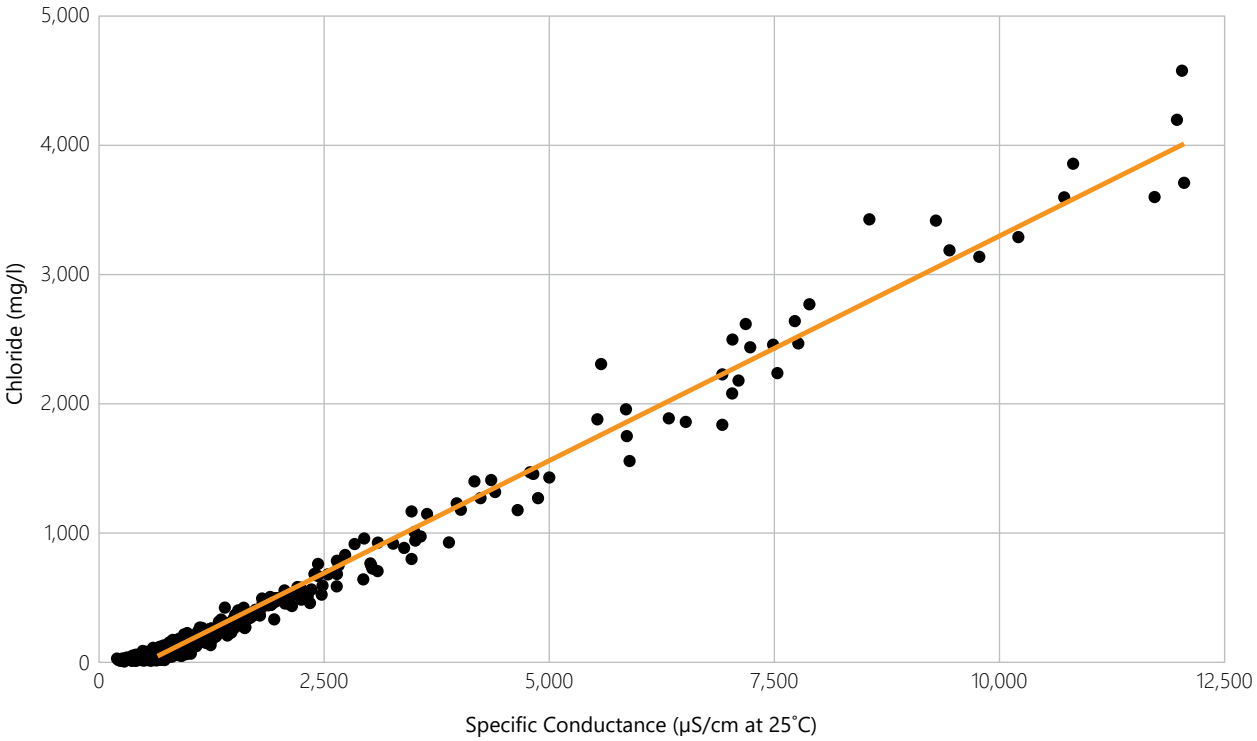
[24] *D. Lüdecke et al., "performance: An R Package for Assessment, Comparison, and Testing of Statistical Models,"* Journal of Open Source Software, *6(60): 3139, 2021*.

[25] *Wickham et al., 2019,* op. cit.

[26] *R Core Team, 2021,* op. cit.

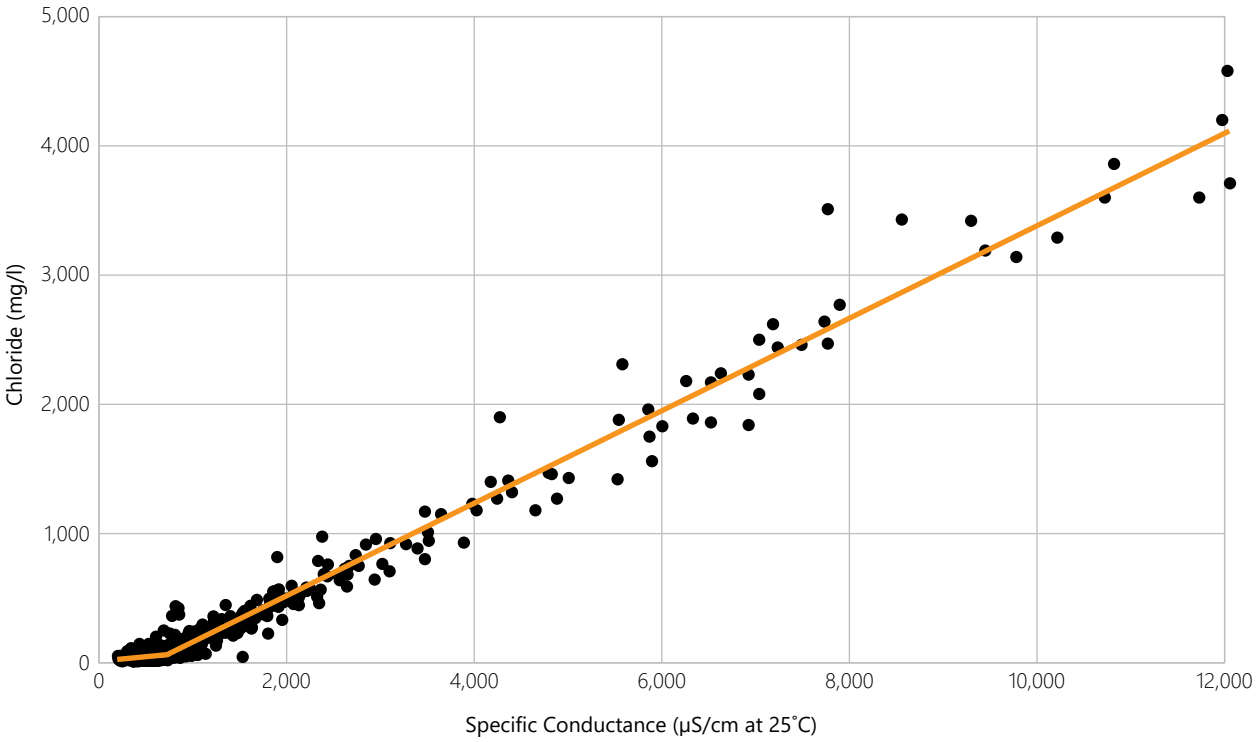
[27] *A specific conductance of 800 uS/cm was provided as a starting value to estimate the piecewise regression model breakpoint using the "segmented" function in R.*

**Figure 2.7**
**Preliminary Simple Linear Regresssion of Chloride Versus Specific Conductance**




*Source: SEWRPC*

**Figure 2.8**
**Preliminary Piecewise Regression of Chloride Versus Specific Conductance**




*Source: SEWRPC*

Examination of this preliminary piecewise linear regression revealed that the model gave a poor representation of the relationship between chloride concentration and specific conductance, especially at low levels of specific conductance. This was due in part to high variability in the relationship between the variables at low levels of specific conductance.

**Grouping of Sample Sites for Model Development**

Site-specific linear regression models were developed for each of the 41 monitoring sites to examine how potential sources of variability in the relationship between chloride concentration and specific conductance might affect the development of models representing the general relationship between these two water quality constituents. These individual models were developed using simple linear regression. Examination of the models revealed patterns in the data. In particular, several sites had maximum specific conductance observations that were lower than the preliminary piecewise regression breakpoint of 725 µS/cm. At these sites, most of the chloride concentrations and the site-specific regression line lay below the piecewise regression. This indicated that the preliminary piecewise regression model was systematically overpredicting chloride concentrations relative to the observed concentrations at these sites (see Figure 2.9). In addition, the specific conductance value at which the preliminary piecewise regression relationship crossed the x-axis (x-intercept) was greater than zero. The x-intercept of a linear equation is greater than zero any time that the y-intercept is less than zero. An x-intercept that is greater than zero indicates that the model predicts negative values of chloride concentration at lower levels of specific conductance. This is a nonsensical result that is not physically possible.

Based on the results of the examination of the individual site-specific regressions, the dataset was divided into three groups. This was done as an iterative process in which sites that deviated too much from the preliminary piecewise regression were removed from the first group and treated separately. The first group consisted of data from those sites at which the ranges of specific conductance and chloride concentration extended beyond the breakpoint of the preliminary piecewise regression (725 µS/cm, 63 mg/l). Data from 30 sampling sites were included in this group (see Map 2.2).[28] A model for these data was developed using piecewise regression. The second group consisted of data from those sites at which the ranges of specific conductance and chloride concentration did not extend beyond the breakpoint from the preliminary piecewise regression, and for which the piecewise regression was not a good fit. Data from 10 sampling sites were included in this group. A model for these data was developed using a linear mixed effects regression (see Map 2.3). The third group consisted of sites at which the site-specific linear regression was not statistically significant, and whose data were not well-fit by either the piecewise regression nor the linear mixed effects models. Data from one site was included in this group and no model was developed for these data (see Map 2.4). Assignment of sampling sites to regression model groups is shown in Table 2.2.

**Final Regression Models**

***Piecewise Linear Regression Model***

A piecewise linear regression model was developed using data from 30 sampling sites from the Chloride Impact Study. The locations of these stations and their drainage areas are shown on Map 2.2. This model included three segments. The locations of the breakpoints between segments were determined algorithmically by minimizing the sum of the squares of the residuals across the entire model.[29] The model used three segments in order to produce a better fit to the data than was seen in the preliminary piecewise linear regression, which was composed of two segments. In addition, the use of three segments reduced the magnitude of the x-intercept. This model is described in Chapter 3 of this report.

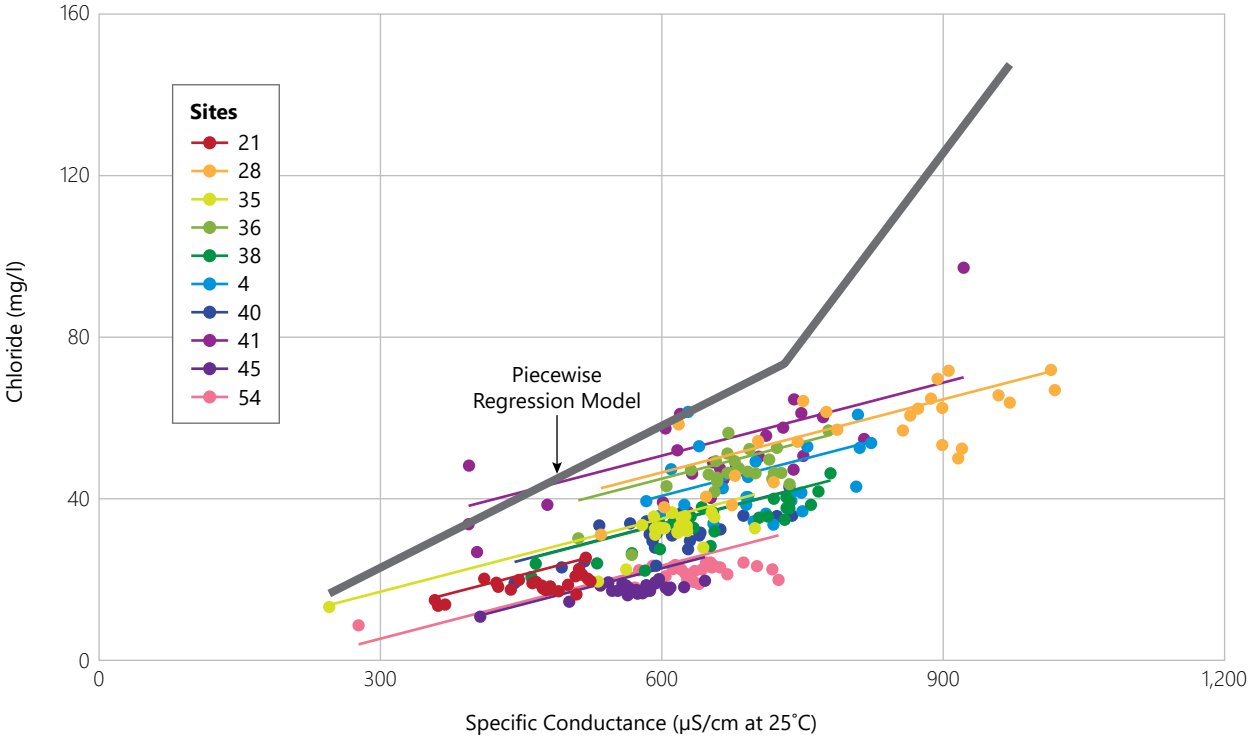
***Linear Mixed Effects Regression Model***

A linear mixed effects regression model was developed using data from 10 Study sampling stations. The locations of these stations and their drainage areas are shown on Map 2.3. In developing this model, the slopes of the regression lines were treated as a fixed effect and the y-intercepts were treated as a random effect. This reflects the assumption that the differences among the site-specific regressions for these sites are due to differences in the concentration of ions other than chloride and sodium and results mostly from

[28] *See SEWRPC TR-61,* op. cit. *for detailed maps of the drainage areas.*

[29] *Specific conductance values of 800 and 1,300 µS/cm were provided as starting values to estimate the breakpoints using the "segmented" function in R.*

**Figure 2.9**
**Site-Specific Regressions of Chloride Versus Specific Conductance for Sites with Overpredictions of Chloride by the Preliminary Piecewise Regression**




*Source: SEWRPC*

differences in the background water chemistry at these sites. This is equivalent to assuming that the slope of the relationship between specific conductance level and chloride concentration is the same at all 10 sites and the differences in slopes observed in the site-specific regressions represent random variations in the data. The differences in the y-intercepts in the site-specific regressions for these sites represent random effects due to differences in background water chemistry.

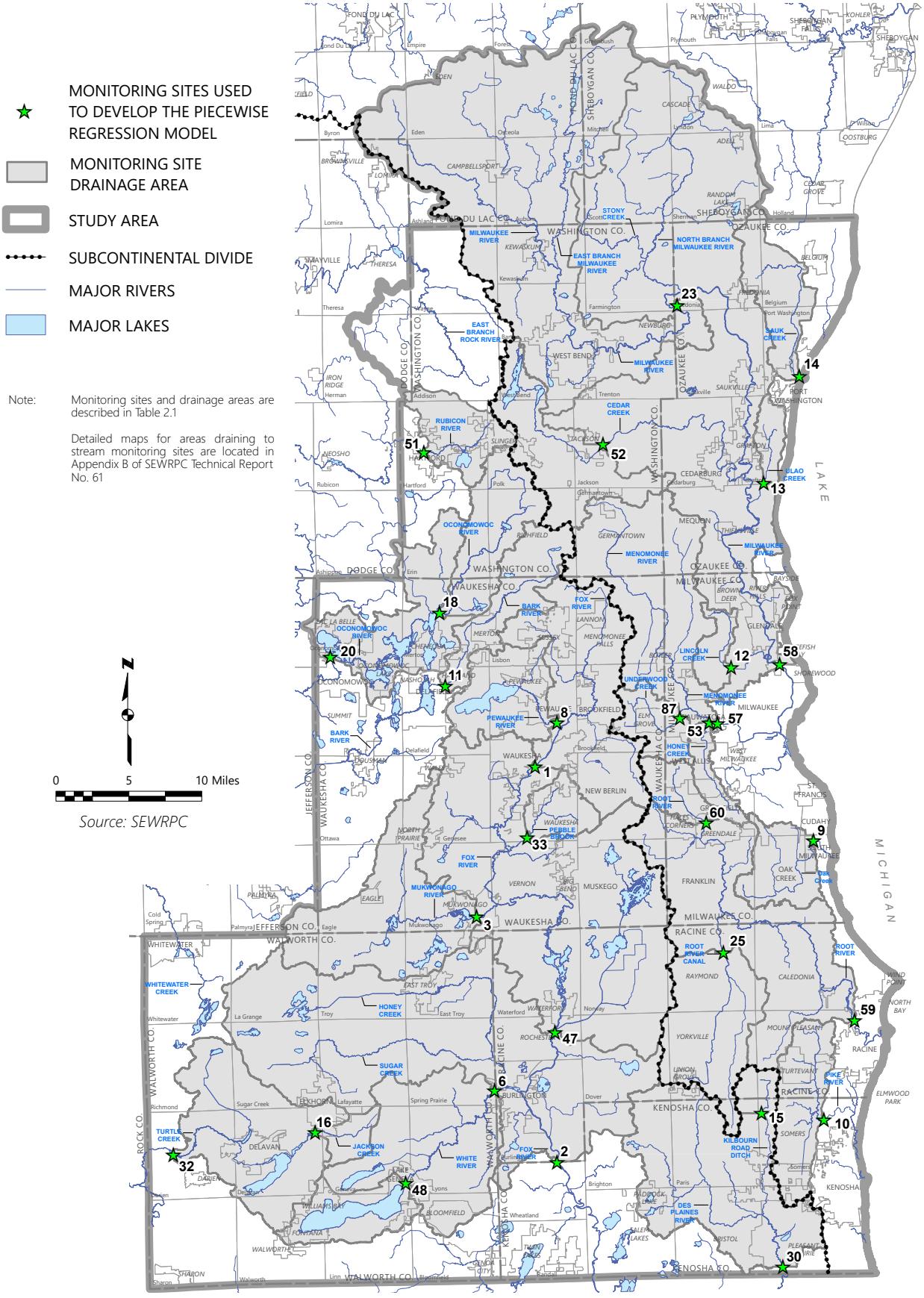
The linear mixed effects model produced an equation for each of the sites used in its development. These equations have the same slope but differ in their y-intercepts. In addition, the model produced a single equation with the same slope and an average y-intercept, as a representative value for all 10 sites. This model is described in Chapter 3 of this report.

## 2.6 EVALUATION OF REGRESSION MODELS

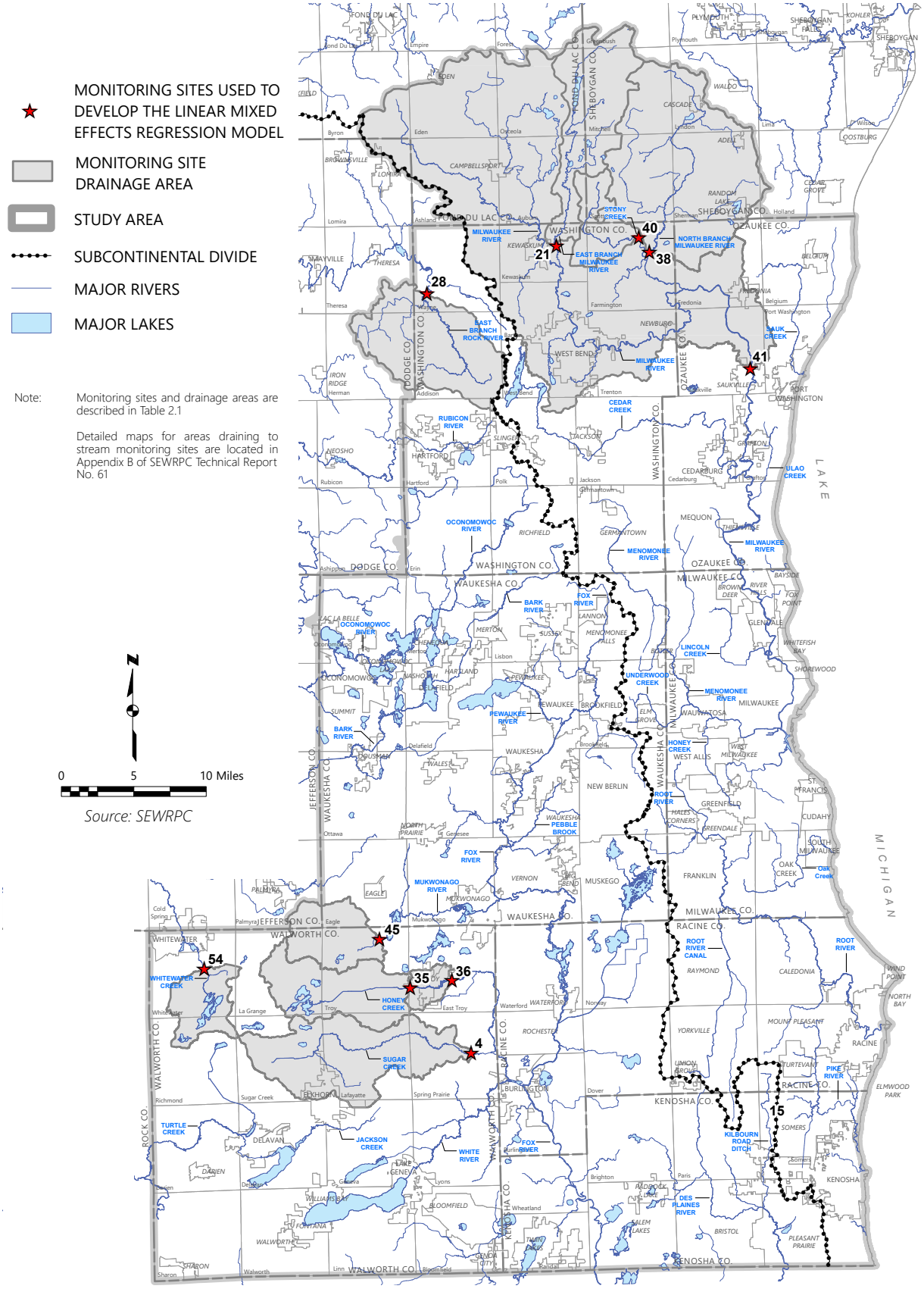
An important aspect of a developing regression model is to evaluate how well the model fits the data and how well it performs at estimating the values of the dependent variable from values of the independent variable. Several methods were used to evaluate the regression models that were developed for the Chloride Impact Study. These include examination of regression statistics, examination of the regression residuals, and examination of the continuity of the model. This section describes methods that were used to evaluate the final regression models. A discussion of the actual results for the regression models is included in Chapter 3.

For use as a predictive device, it is important that the results of any regression model make sense relative to the physical world. For example, a regression model that is meant to predict concentrations of a substance should not predict negative concentrations at reasonable values of the independent variable. A couple of different strategies can be used to address the production of nonsensical results by a regression model. In some cases, the nonsensical result may occur in a part of the range of the independent variable that is rarely encountered. In such a case setting the prediction to zero may be an adequate solution to this problem. In other instances, it may be necessary to further refine the model.

## Map 2.2
## Stream Monitoring Sites and Associated Drainage Areas Used to Develop the Piecewise Regression Model



## Map 2.3
## Stream Monitoring Sites and Associated Drainage Areas Used to Develop the Linear Mixed Effects Regression Model



## Map 2.4
## Stream Monitoring Sites and Associated Drainage Area for Which a Regression Model Could Not Be Developed

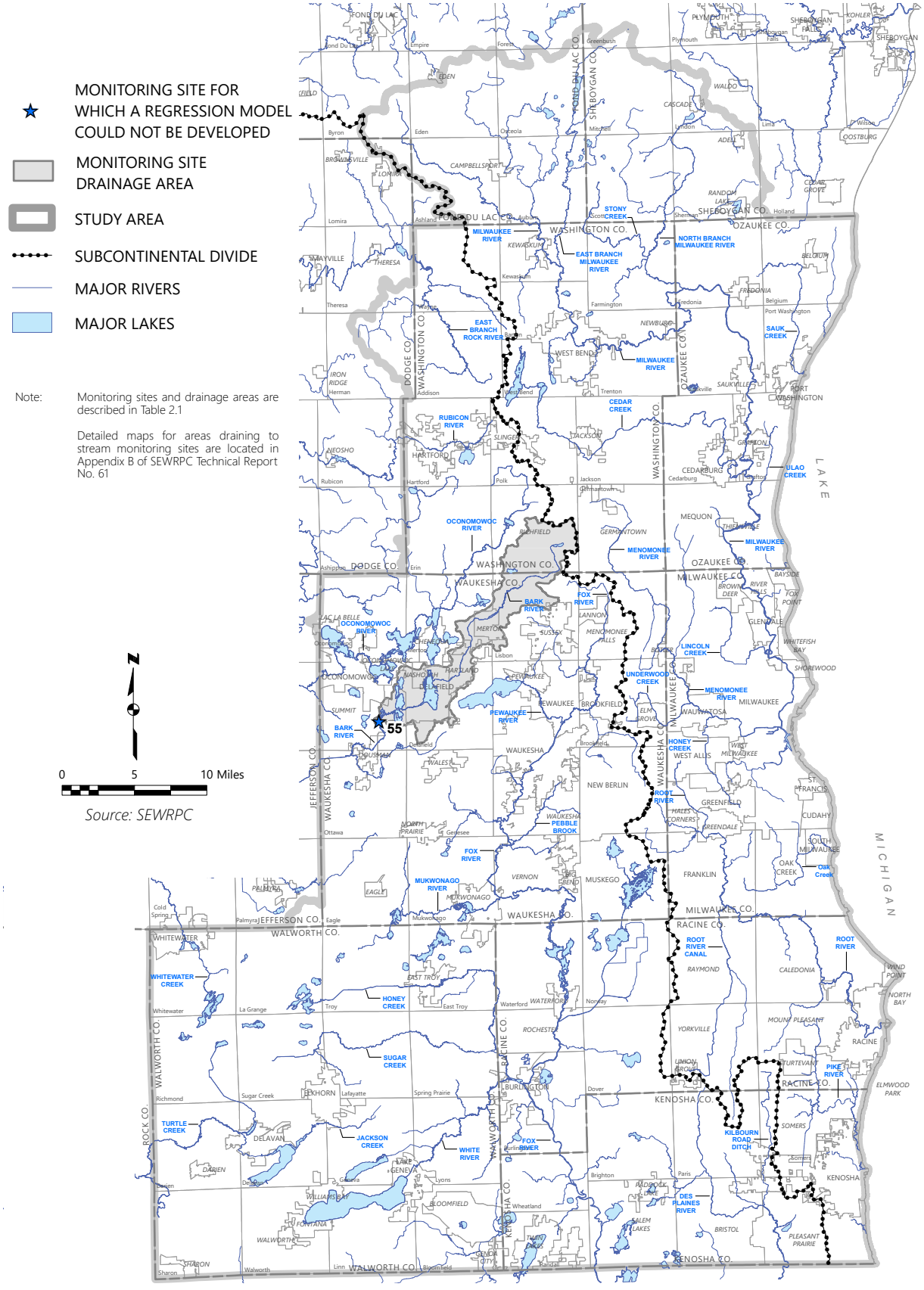


**Table 2.2**
**Regression Models Applicable to SEWRPC Water Quality Monitoring Sites**

| Site Number | Site Name | Regression Model |
|---|---|---|
| 1 | Fox River at Waukesha | Piecewise regression |
| 2 | Fox River at New Munster | Piecewise regression |
| 3 | Mukwonago River at Mukwonago | Piecewise regression |
| 4 | Sugar Creek | Linear mixed effects |
| 6 | White River near Burlington | Piecewise regression |
| 8 | Pewaukee River | Piecewise regression |
| 9 | Oak Creek | Piecewise regression |
| 10 | Pike River | Piecewise regression |
| 11 | Bark River Upstream | Piecewise regression |
| 12 | Lincoln Creek | Piecewise regression |
| 13 | Ulao Creek | Piecewise regression |
| 14 | Sauk Creek | Piecewise regression |
| 15 | Kilbourn Road Ditch | Piecewise regression |
| 16 | Jackson Creek | Piecewise regression |
| 18 | Oconomowoc River Upstream | Piecewise regression |
| 20 | Oconomowoc River Downstream | Piecewise regression |
| 21 | East Branch Milwaukee River | Linear mixed effects |
| 23 | Milwaukee River Downstream of Newburg | Piecewise regression |
| 25 | Root River Canal | Piecewise regression |
| 28 | East Branch Rock River | Linear mixed effects |
| 30 | Des Plaines River | Piecewise regression |
| 32 | Turtle Creek | Piecewise regression |
| 33 | Pebble Brook | Piecewise regression |
| 35 | Honey Creek Upstream of East Troy | Linear mixed effects |
| 36 | Honey Creek Downstream of East Troy | Linear mixed effects |
| 38 | North Branch Milwaukee River | Linear mixed effects |
| 40 | Stony Creek | Linear mixed effects |
| 41 | Milwaukee River near Saukville | Linear mixed effects |
| 45 | Mukwonago River at Nature Road | Linear mixed effects |
| 47 | Fox River at Rochester | Piecewise regression |
| 48 | White River at Lake Geneva | Piecewise regression |
| 51 | Rubicon River | Piecewise regression |
| 52 | Cedar Creek | Piecewise regression |
| 53 | Honey Creek at Wauwatosa | Piecewise regression |
| 54 | Whitewater Creek | Linear mixed effects |
| 55 | Bark River Downstream | No model |
| 57 | Menomonee River at Wauwatosa | Piecewise regression |
| 58 | Milwaukee River at Estabrook Park | Piecewise regression |
| 59 | Root River near Horlick Dam | Piecewise regression |
| 60 | Root River at Grange Avenue | Piecewise regression |
| 87 | Underwood Creek | Piecewise regression |

*Source: SEWRPC*

### Coefficient of Determination ($R^2$)

The coefficient of determination ($R^2$) measures how well the model fits the data. It indicates how much of the variability in the dependent variable can be accounted for by the model. Normally the value of $R^2$ ranges between 0.0 and 1.0. An $R^2$ of 0.0 indicates that the variability in the dependent variable is independent of variability in the independent variable, and the model does not provide a good fit to the data. An $R^2$ of 1.0 indicates that all the variability in the dependent variable is accounted for by the variability in the independent variable. This would constitute a perfect fit or correlation in which all the data points used to develop the model lie on the regression line.

It is important to recognize that there are several aspects of a model that are not reflected in the $R^2$ statistic. $R^2$ provides no information as to whether variation in the independent variable is the cause of variation in the dependent variable. $R^2$ cannot indicate whether the best mathematical function was used to develop the regression equation. Whether the model might be improved by transforming the data is not reflected in the $R^2$ value. Finally, $R^2$ does not indicate whether enough data were used in developing the model to allow users to draw a solid conclusion.

In practice, the calculation of $R^2$ can be subject to bias. Because of this, an adjusted $R^2$ is commonly calculated, in which a correction is made to $R^2$ to reduce bias. It should be noted that this correction may not eliminate all bias in the adjusted $R^2$.

**Residual Analysis**

Residual analysis is a graphical technique used to evaluate regression models by examining the differences between the observed values and the values predicted by the model. In this analysis, the residuals from the regression are plotted against the predicted values of the dependent variable (see Figure 2.10). A good residual plot has several characteristics:

- There is a high density of points close to zero and a lower density of points farther away from zero
- The distribution of points is symmetric about zero
- The distribution of points has a random pattern

Residual analysis provides a means of evaluating whether the model meets some of the assumptions of linear regression. For example, a "U" shape to the residual plot can indicate that the data violates the assumption of a linear relationship between the variables. Similarly, an increase in the width of the band of points around zero from right to left in the plot indicates that the variability in the data increases as the values of the dependent variable increases, which violates the assumption of homoscedasticity. Residual analysis can also reveal violations of the assumptions that the residuals are normally distributed and that they are independent of one another.

**Cross-Validation**

Cross-validation is a method used to examine the robustness of a regression model. This procedure examines whether a model overfits the data used to develop it. An overfit of the data occurs when the model fits the data so tightly that it may account for random noise and unwanted trends in the development data that will not be present in other datasets to which the model is applied. Cross-validation examines a model for overfitting by removing some of the development data, recalculating the model with the remaining development data, and comparing the results to the original model. Comparing the results of the recalculated model with the original model can reveal whether the model overfits the data.

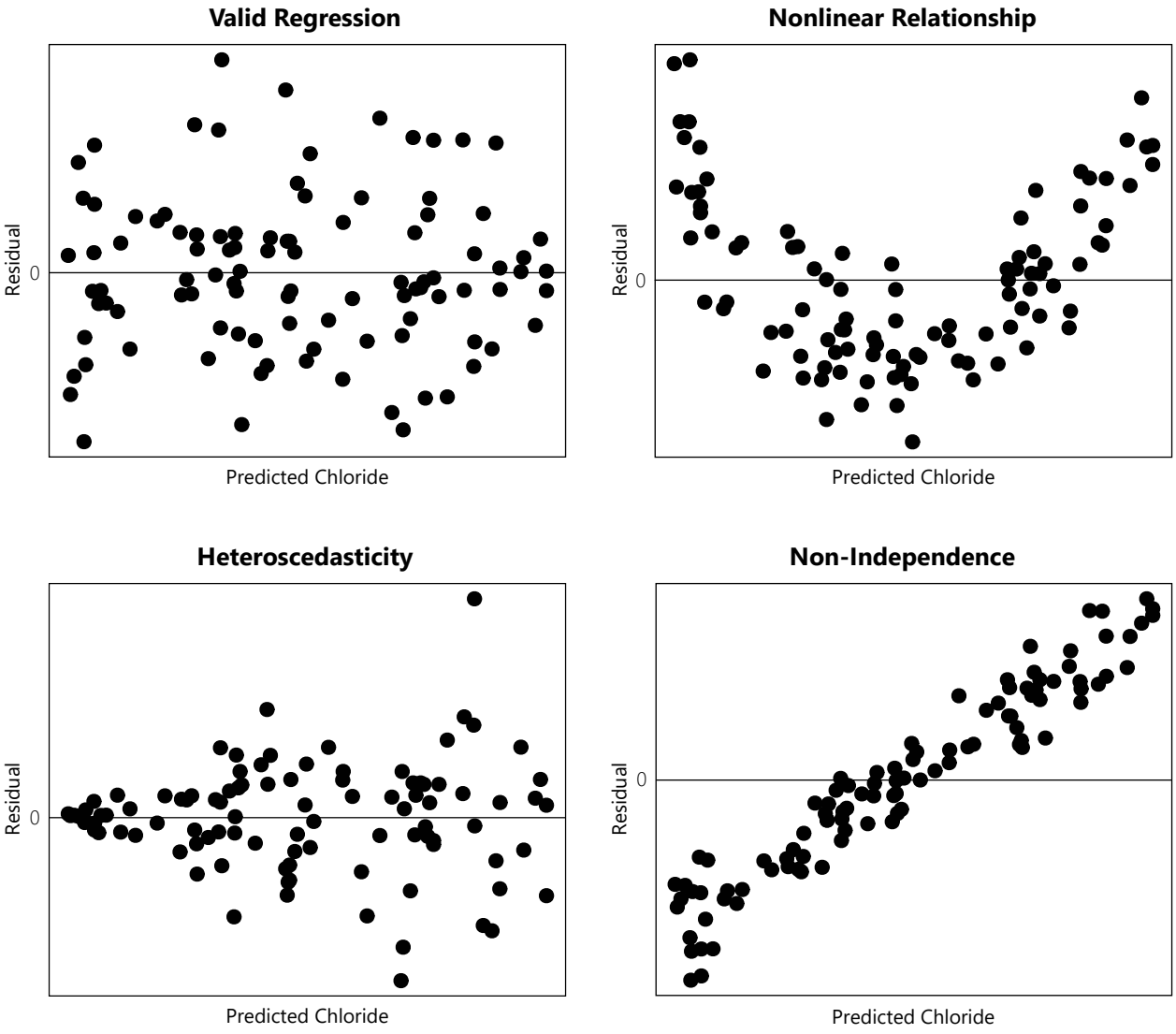
Several methods can be used for cross-validation. K-fold cross-validation was used to examine the Chloride Impact Study models developed for estimating chloride from specific conductance. In this application, a 10-fold cross-validation was performed.

The following procedure was used to cross-validate the regression models developed for the Chloride Impact Study:

1. The data used to develop the model were randomly divided into 10 groups of equal size
    a. For the piecewise regression model, data were randomly divided irrespective of sites[30]

[30] *Data were assigned to groups by randomly assigning each observation a number between one and the total number of observations in the entire dataset using the "sample" function from base R. These assigned numbers were ranked into ten groups using the "ntile" function from the "tidyverse" package.*

**Figure 2.10**
**Analysis of Regression Residuals**




*Source: Wikimedia Commons and SEWRPC*

b. For the linear mixed effects model, data were randomly divided into ten subsets within each site and the subsets from the individual sites were combined to form 10 data groups, such that each group had roughly equal amounts of data from each site[31]

2. One of the data groups was removed and the model was recreated using data from the remaining nine groups

   a. The model was then used to predict the chloride concentration from each specific conductance observation in the removed data group and the mean squared error of the predicted versus actual chloride concentrations was calculated

[31] *Data were assigned to groups within each site following a modified approach known as a stratified K-fold cross-validation, where the random number assignment and ranking into ten groups occurred within each site as opposed to across the entire dataset. This division within site was included to account for treating the site as a random effect in the linear mixed effects regression. All sites included in the linear mixed effects model had either 25 or 26 observations, resulting in either two or three observations from each site in each cross-validation group.*

3. Step 2 was performed 10 times, each time leaving a different group of data out

4. Parameters from the 10 cross-validation models were compared to those from the original model. If the final model overfits the data, the values of the parameters developed by the regression analysis should be sensitive to the removal of data

Two metrics were used to evaluate the results of the cross-validation. The first was the mean squared error (MSE), which is a measure of the dispersion of data points around a regression model. A smaller MSE indicates that the points are more closely grouped around the regression line than a larger MSE. For different models applied to the same data, a lower MSE indicates a better fit of the model to the data. The second was the mean absolute error (MAE) which is the average distance along the y-axis of points from the regression model. A lower MAE indicates a better fit to the data.

**Analyses Specific to the Piecewise Regression**

***Equation Simplification and Examination of Continuity at Breakpoints***

The piecewise regression equations developed using R were presented as complex, multiple-term expressions. While the equation for the line segment representing the lowest values of the independent variable was presented in the form $y = mx + b$, the equations of the other line segments were presented in a form in which terms modifying the parameters were added to the initial equation. These equations were simplified algebraically to put them into the form $y = mx + b$.

In addition, the parameters (m, b) of the simplified piecewise regression equations were rounded from those in the original equations generated using R. This was done in order to have them conform to the number of significant digits in the data used to develop them. Rounding can affect the model results, creating differences between the original regression model output and the results computed using the simplified equations. Therefore, additional testing was performed on the simplified equations to verify their performance. The results of the simplified equations were validated against the original regression model results to ensure the equations produced reasonably close results. Further testing was performed on the piecewise regression equations near the equation breakpoints to evaluate how the simplified equations performed at and around the junction of two segments of the piecewise regression.

To verify the performance of the simplified equations after rounding the parameters and breakpoints, chloride concentrations were computed for a variety of specific conductance values using both the original equations and the simplified equations. The adjacent segments of the original piecewise regression equations converged at the computed breakpoints, but the breakpoints were rounded to whole numbers for use with the simplified equations, resulting in minor differences in the chloride concentrations computed around these points. The simplified equations were used to predict chloride concentrations from specific conductance values at and adjacent to the piecewise regression equation breakpoints. The paired data were plotted to confirm continuity of the piecewise regression equation between the adjacent segments.

***Comparison of Estimated Chloride Concentration to Measured Values***

The performance of the piecewise regression model was examined by comparing chloride concentrations estimated from specific conductance to measured chloride concentrations using historical datasets. This was done using paired specific conductance and chloride data from water quality sampling in four mostly urban watersheds in the Region: the Kinnickinnic River watershed, the Menomonee River watershed, the Oak Creek watershed, and the Root River watershed. The historical data were collected by four agencies: the Milwaukee Metropolitan Sewerage District (MMSD), the Southeastern Wisconsin Regional Planning Commission (SEWRPC), the USGS, and the Wisconsin Department of Natural Resources (WDNR) and were downloaded from the MMSD Corridor Study Database, the WDNR SWIMS database, and the USGS NWIS database.[32] The data included over 23,307 paired samples collected between 1964 and 2022.

Two groups of comparisons between estimated and measured concentrations of chloride were made. First, the estimated concentrations were directly compared to the corresponding measured concentrations. Second, the performance of the piecewise regression in predicting exceedances of regulatory thresholds in comparison to measured data was examined.

[32] *WDNR data included volunteer monitoring data from the University of Wisconsin-Madison Division of Extension/WDNR Water Action Volunteer program.*

Direct Comparison of Estimated Chloride Concentration to Measured Values

The comparison of estimated chloride concentration to measured values was made using several methods. In each watershed, the difference between the value estimated by the piecewise regression model and the corresponding measured value was calculated. The percentage by which the estimated chloride concentration differed from the measured concentration was also computed. These measures were summarized by calculating mean and median values for the watershed, both on a raw difference basis and an absolute value basis. Estimated chloride concentrations were also graphed against the corresponding measured value. Graphs were drawn for all data from each watershed as well as subsets of the data. Subsets included separate graphs for data from each agency that conducted monitoring, each stream and major site in the watershed, and several different time periods. This method was not used to evaluate the linear mixed effects regression model because Commission staff were unable to find adequate datasets containing paired samples of specific conductance and chloride concentration from streams for which this model would likely be applied.

Evaluation of Model Performance in Predicting Exceedance of Regulatory Thresholds

The ability of the piecewise regression model to correctly predict exceedances of regulatory thresholds was examined using paired specific conductance and chloride data from historical water quality sampling in four mostly urban watersheds in the Region: the Kinnickinnic River watershed, the Menomonee River watershed, the Oak Creek watershed, and the Root River watershed. For each sample, the concentration of chloride estimated from specific conductance using the piecewise regression and the measured concentration of chloride were compared to chloride regulatory thresholds to examine whether the two values were consistent with one another in their assessments of whether a threshold had been exceeded. Comparisons were made to three Wisconsin regulatory thresholds that address chloride: the acute toxicity criterion for aquatic life, the chronic toxicity criterion for aquatic life, and the secondary drinking water maximum contaminant limit. The values of these criteria are given in Table 2.3. Three measures were used to evaluate the consistency of the estimated concentrations with the measured values. These included:

**Table 2.3**
**Regulatory Thresholds for Chloride Used to Assess Performance of the Piecewise Regression Model**

| Criterion | Value (mg/l) |
|---|---|
| Acute toxicity for aquatic life | 757 |
| Chronic toxicity for aquatic life | 395 |
| Drinking water secondary maximum contaminant limit[a] | 250 |

[a] *This is also the value of the Wisconsin groundwater enforcement standard for chloride.*

*Source: Wisconsin Department of Natural Resources*

- The percentage of samples in which both the concentration estimated by the piecewise regression and the measured concentrations were either above or below the threshold was calculated
- The rate of false positives for exceedances among values estimated by the piecewise regression[33]
- The rate of false negatives for exceedances among values estimated by the piecewise regression[34]

It should be noted that this method is not appropriate for evaluating the linear mixed effects regression model because that model is not meant for estimating higher concentrations of chloride. This qualification and the application of the regression models will be discussed further in Chapter 3.

[33] *False positives occur when the concentration of chloride estimated by the model exceeds the regulatory threshold while the measured value is below the threshold.*

[34] *False negatives occur when the concentration of chloride estimated by the model is below the regulatory threshold while the measured value exceeds the threshold.*

# RESULTS 3

## 3.1 INTRODUCTION

As previously discussed in this Report, Commission staff developed two regression models for estimating average concentrations of chloride from specific conductance in streams of the Southeastern Wisconsin Region—a piecewise regression model and a linear mixed effects regression model. Chapter 2 of this Report describes the methods used to develop these models. This chapter presents the models, evaluates their performance, and provides guidance regarding the situations in which each model can be applied.

## 3.2 PIECEWISE REGRESSION MODEL

### Model Equation

Figure 3.1 shows the piecewise regression model that was developed for estimating average chloride concentration from specific conductance. The equations for this model are given in Table 3.1. The table also shows the ranges of specific conductance and chloride concentration in the Study data used to develop the piecewise regression model. This model consists of four equations, each applicable over a distinct range of specific conductance and these ranges do not overlap with one another. The orange dots on the graph show the breakpoints between the line segments comprising this model. These breakpoints occur at specific conductance levels equal to 732 microSiemens per centimeter (µS/cm)[35,36] and 2,123 µS/cm.

The three equations for specific conductance above 103 µS/cm were generated by the piecewise regression algorithm. The equation covering the lowest portion of the specific conductance range was added because the equation estimating chloride over the range of 103 µS/cm to 732 µS/cm crosses the x-axis at a specific conductance of 103 µS/cm. As a result, this equation would predict a negative concentration of chloride for specific conductance values at and below 103 µS/cm. Because this result makes no physical sense, estimates of chloride concentrations at specific conductance values at and below 103 µS/cm were set to zero milligrams per liter (mg/l).

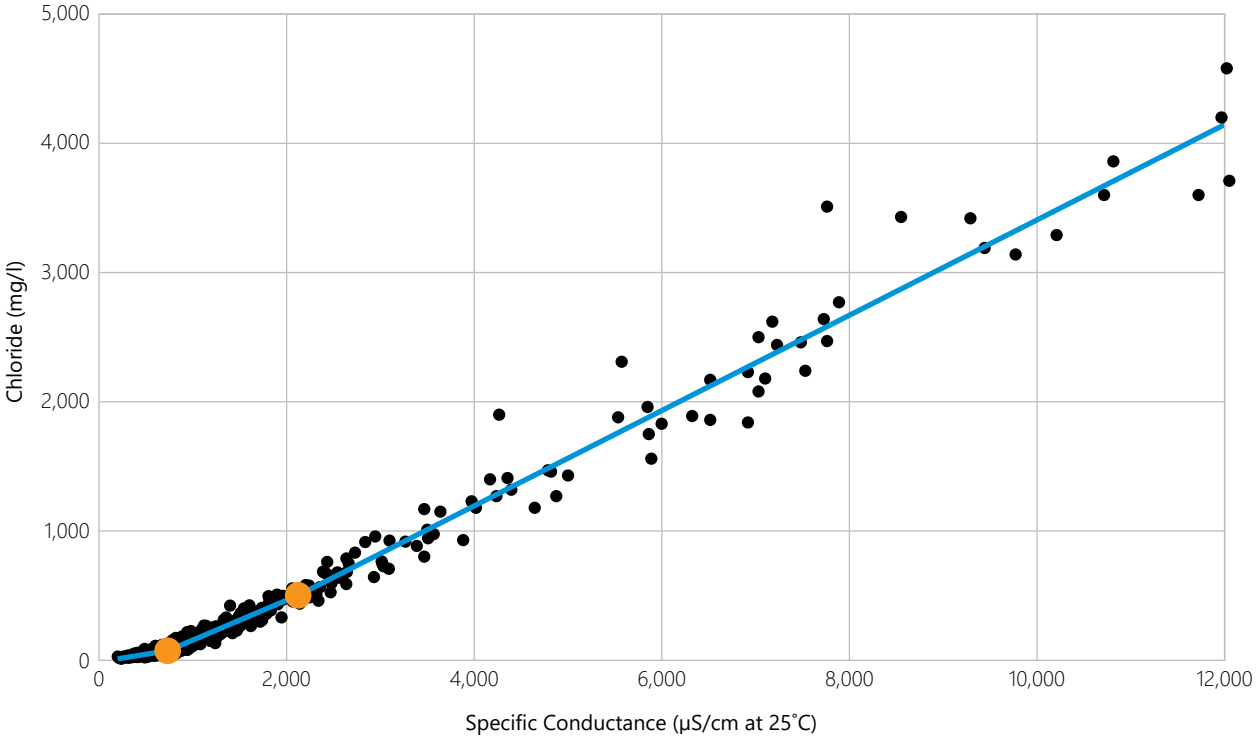
Two facts support the use of zero mg/l in this model as an estimate for chloride concentration when specific conductance was at or below 103 µS/cm. First, numerous ions beside chloride and sodium contribute to specific conductance. Some of these ions are likely to be present in stream water even when chloride is absent. Second, specific conductance values as low as or lower than 103 µS/cm were very rare in the continuous specific conductance data collected as part of the Chloride Impact Study. Values of specific conductance that were this low were observed only in a few very flashy urban streams for brief periods during large rain events. Because of this, it is expected that only a small number of specific conductance readings from the 30 sites used to develop this model will have chloride concentrations estimated to be zero mg/l.

Table 3.1 and Figure 3.1 show that at higher ranges of specific conductance, the slope of the equation for the line segment representing the relationship between specific conductance and chloride is greater than at lower ranges. This likely reflects the fact that when chloride concentrations are high, chloride and its associated cations account for a much greater share of specific conductance than they do when their concentrations are low. Thus, when large amounts of chloride salts enter streams, chloride accounts for much of the specific conductance that is present.

[35] *Acronyms and abbreviations used in this report are defined in Appendix A*

[36] *Specific conductance is actual conductivity normalized to the equivalent conductivity at 25°C.*

## Figure 3.1
## Piecewise Regression Model for Estimating Chloride from Specific Conductance in Southeastern Wisconsin Streams




Note: Black dots indicate the data used to develop the model. Orange dots indicated the breakpoints in the model.

*Source: SEWRPC*

## Table 3.1
## Piecewise Regression Model for Estimating Chloride Concentration from Specific Conductance

| Specific Conductance Range (µS/cm at 25°C)[a] | Equation to Estimate Chloride Concentration (mg/l) |
|---|---|
| SC ≤ 103 | $[Cl^-] = 0$ |
| 103 < SC ≤ 732 | $[Cl^-] = 0.1171 \times SC - 12.0$ |
| 732 < SC ≤ 2,123 | $[Cl^-] = 0.3084 \times SC - 151.9$ |
| SC > 2,123 | $[Cl^-] = 0.3687 \times SC - 280.0$ |

| Range of Values[b] | |
|---|---|
| **Specific Conductance (µS/cm at 25°C)** | **Chloride (mg/l)** |
| 200 – 12,050 | 11 – 4,163 |

Note: SC indicates specific conductance. [Cl-] indicates chloride concentration.

[a] *The standard temperature used for adjusting conductivity to calculate specific conductance is 25°C, which is equivalent to 77°F.*

[b] *Estimates outside the range of data used to develop the regression model should be treated as less reliable due to extrapolating beyond the range of values used to develop the model.*

*Source: SEWRPC*

**Evaluation of Piecewise Regression Model**

The piecewise regression model is based on 818 paired samples of specific conductance and chloride from 30 sampling sites. Levels of specific conductance in this dataset spanned a range of almost 12,000 µS/cm. Similarly, concentrations of chloride spanned a range of over 4,150 mg/l (see Table 3.1).

The piecewise regression model has a residual standard error of 73.4 based on 812 degrees of freedom. The piecewise regression model was found to be significant at a probability less than 0.01. The piecewise regression equations have an adjusted $R^2$ value of 0.9833. This means that variation in specific conductance accounts for over 98 percent of the variation in chloride concentration.

***Residual Analysis of the Piecewise Regression***

The examination of residuals shows that the piecewise regression model has marked heteroscedasticity, with variability in chloride concentrations increasing at higher levels of specific conductance. As discussed in Chapter 2 of this Report, this should not cause any problem in using the model for estimating average chloride concentrations from levels of specific conductance; however, it does preclude computation of valid confidence intervals around the model.

***Continuity of the Piecewise Regression Model at Breakpoints***

As discussed in Chapter 2 of this Report, the parameters of the line segments making up the piecewise regression model and the breakpoints were rounded from those in the original equations generated using R to have them conform to the number of significant digits in the development data. Slopes were rounded to four significant digits, y-intercepts were rounded to the nearest tenth of a milligram per liter, and breakpoints were rounded to the nearest whole number. Examination of continuity of the model at the breakpoints showed only minor discrepancies that were on the order of 0.1 mg/l. Discrepancies this small were judged to be acceptable.

***Cross-Validation of Piecewise Regression Model***

Cross-validation was performed on the piecewise regression model. This procedure consisted of ten iterations. In each iteration, 10 percent of the samples were removed from the data and a piecewise regression was developed using the remaining data. A different 10 percent was removed for each iteration. The results from the cross-validation were compared to the original piecewise regression using several metrics. These results are summarized in Table 3.2. More information on the methods used for cross-validation is given in Chapter 2 of this Report.

<u>Comparison of Slopes</u>

The slopes generated by the iterations of the cross-validation were compared to the slopes of three segments of the piecewise regression model. The piecewise regression model had three segments, each with its own slope (see Table 3.1). The first segment covered the lowest portion of the regression, the specific conductance range extending from 103 µS/cm to 732 µS/cm. The second and third segments covered progressively higher portions of the specific conductance range.

All the slopes from the cross-validation for the first segment of the model were within 15 percent of the slope for this segment from the piecewise regression model. Most were within 5 percent (see Table 3.2). All the slopes from the cross-validation for the second segment of the model were within 4 percent of the slope for this segment from the piecewise regression model. Most were within 3 percent. All the slopes from the cross-validation for the third segment of the model were within 8 percent of the slope for this segment from the piecewise regression model, while most were within 5 percent.

This comparison of slopes indicates that the piecewise regression model is not sensitive to individual samples. This stability of the model relative to slope is greatest in the second segment, which constitutes the middle of the range of data used to develop the model. The higher variation in cross-validation iteration slopes in the first segment is not surprising. This portion of the model represents the range of data in which ions other than chloride have the greatest influence on specific conductance. Similarly, the slightly higher variation in slopes in the third segment of the model may reflect the heteroscedasticity of the paired specific conductance-chloride concentration dataset.

**Table 3.2**
**Piecewise Regression Versus Cross-Validation Model Parameters**

| Model Parameter | Piecewise Model | Cross-Validation Results | | |
|---|---|---|---|---|
| | | Minimum | Mean | Maximum |
| 1st Segment Slope | 0.1171 | 0.0985 | 0.1180 | 0.1775 |
| 2nd Segment Slope | 0.3084 | 0.2942 | 0.3018 | 0.3163 |
| 3rd Segment Slope | 0.3687 | 0.3619 | 0.3729 | 0.4225 |
| X-Intercept | 103 | 27 | 81 | 130 |
| Y-Intercept (1st Segment)[a] | -12.0 | -16.1 | -9.3 | -2.6 |
| Y-Intercept (2nd Segment) | -151.9 | -152.5 | -147.7 | -139.8 |
| Y-Intercept (3rd Segment) | -280.0 | -300.8 | -278.6 | -259.6 |
| 1st Breakpoint | 732 | 688 | 717 | 742 |
| 2nd Breakpoint | 2,123 | 1,717 | 2,000 | 3,031 |
| $R^2$ | 0.983 | 0.982 | 0.983 | 0.986 |
| Mean Squared Error | 5,348 | 4,492 | 5,333 | 5,757 |

[a] *The first segment y-intercept constitutes the y-intercept for the piecewise regression model.*

*Source: SEWRPC*

Comparison of y-intercepts

The y-intercepts generated by the iterations of the cross-validation were compared to those from the piecewise regression model. Note that the analysis generated y-intercepts for all three segments of the model. While the y-intercepts associated with the second and third segments are not part of the model, they are important for determining both the slopes of the associated segments of the model and continuity of the model at break points.

All the y-intercepts from the cross-validation for the first segment of the model were within 70 percent of the y-intercept for this segment from the piecewise regression model (see Table 3.2). Most were within 15 percent. While these percentages are high, they represent small differences in the magnitude of chloride concentration. For the second segment of the model all the y-intercepts from the cross-validation were within 8 percent of the y-intercept for this segment from the piecewise regression model, and most were within 5 percent. All the y-intercepts from the cross-validation for the third segment of the model were within 8 percent of the y-intercept for this segment from the piecewise regression model. Most were within 5 percent.

The magnitudes of the differences in the y-intercepts generated by the cross-validation from those of the piecewise regression model were small. All the y-intercepts for the first segment of the model from the cross-validation were within 9 mg/l of the y-intercept from the piecewise regression. Most were within 2 mg/l. For the second segment of the model from the cross-validation all the y-intercepts were within 12 mg/l of the y-intercept from the piecewise regression, and most were within 5 mg/l. All the y-intercepts for the third segment of the model from the cross-validation were within 20 mg/l of the y-intercept from the piecewise regression. Most were within 15 mg/l. Given the magnitudes of chloride concentrations estimated from the ranges of these segments, these differences in y-intercepts between iterations of the cross-validation and the piecewise regression are quite small.

Comparison of x-intercepts of the First Segment

The x-intercept of the first segment of the piecewise regression model indicates the level of specific conductance that is observed when no chloride is present in the water. The magnitude of this quantity is determined by the slope and y-intercept of the equation for the first segment of the model.

All the x-intercepts generated by the iterations of the cross-validation were within 74 percent of the xintercept of the piecewise regression model (see Table 3.2). Most of the x-intercepts were within 10 percent. The magnitude of the x-intercepts varied between 27 µS/cm and 130 µS/cm. As previously noted, specific conductance measurements this low were very rare in both the grab sample and continuous specific conductance datasets.[37]

[37] *For more information on this see SEWRPC Technical Report No. 63,* Chloride Conditions and Trends in Southeastern Wisconsin*, in preparation.*

Comparison of Breakpoints

The breakpoints in the piecewise regression model determine the levels of specific conductance at which the model changes from one segment to another. The first breakpoint indicated the location of the shift from the first segment of the model to the second segment. The second breakpoint indicated the location of the shift from the second segment to the third segment.

All the first breakpoints generated by the iterations of the cross-validation were within 6 percent of the first breakpoint of the piecewise regression (see Table 3.2). Most of the values were within 1 percent. All the second breakpoints generated by the iterations of the cross-validation were within 43 percent of the second breakpoint of the piecewise regression, and most values were within 1 percent.

Comparison of $R^2$

The $R^2$ generated by each iteration of the cross-validation was greater than 0.98 (see Table 3.2). These all differed from the $R^2$ of the piecewise regression by less than 1 percent. The high $R^2$ values in both the piecewise regression model and the cross-validation iterations partially reflect the large number of samples used to develop the model. As previously noted, the piecewise regression was based on 818 paired samples. Depending on the iteration, the cross-validation iterations were based on either 736 or 737 samples.

Comparison of Mean Squared Errors

The mean squared error (MSE) is a measure of the dispersion of data points around a regression model. A smaller MSE indicates that the points are more closely grouped around the regression line than a larger MSE. For different models applied to the same data, a lower MSE indicates a better fit of the model to the data.

The MSE of the piecewise regression model was 5,348. The MSEs from the iterations of the cross-validation ranged between 4,492 and 5,757, with a mean value of 5,333. This average was very close to the MSE from the piecewise regression (see Table 3.2). The MSEs from five cross-validations iterations were higher than the MSE of the piecewise regression, while the MSEs from the remaining five iterations were lower. All cross-validation MSEs were within 16 percent of the MSE from the piecewise regression, and most MSE were within 10 percent.

The iterations of the cross-validation also estimated chloride concentration from each specific conductance observation in the removed data group and calculated MSEs of the predicted versus actual chloride concentrations. These MSEs varied from 1,600 to 13,443, with a mean value of 5,615. These differences in MSE represent relatively small differences in averaged mean absolute error (MAE). The MAE across the 10 iterations of the cross-validation ranged between 26.1 mg/l and 40.2 mg/l with an average of 33.1 mg/l. This average was about 5.0 mg/l lower than the MAE of chloride estimates generated from specific conductance using the piecewise regression when compared to measured data.

These MSE and MAE results are not surprising given both the high variability in the relationship between chloride concentration and specific conductance at low values of specific conductance and the heteroscedasticity in the dataset (see Figure 3.1) Random removal of a few outlying points in some iterations of the cross-validation likely resulted in a considerable reduction of the associated MSEs.

Summary of Cross-Validation Results

The results of the cross-validation of the piecewise regression model show that model parameters and other metrics are reasonably stable to removal of 10 percent of the data used to develop the model. The slopes of the segments of the model and the breakpoints are quite stable. While the x-intercepts of the first model segment varied among iterations, the values were all at levels of specific conductance that were rarely observed. The $R^2$ values of the cross-validation iterations were almost identical to the $R^2$ of the piecewise regression. While some cross-validation MSEs were lower than the MSE from the piecewise regression, this is likely due to the heteroscedasticity of the dataset and high variability of the specific conductance-chloride relationship at low values of specific conductance. Taken together, these results suggest that the piecewise regression model is not overfitting the dataset.

***Comparison of Estimates to Measured Values of Chloride***

The performance of the piecewise regression was evaluated by comparing concentrations of chloride estimated from specific conductance to historical measured values of chloride in paired specific conductance-chloride samples. This analysis was conducted using 23,371 paired samples from the Kinnickinnic River, Menomonee River, Oak Creek, and Root River watersheds collected between 1964 and 2022. Specific conductance in these samples ranged between 78 µS/cm and 19,800 µS/cm. Chloride concentrations ranged between 0.5 mg/l and 6,470 mg/l. These ranges were wider than those of the dataset used to develop the piecewise regression. These paired data were collected by several agencies including the Milwaukee Metropolitan Sewerage District (MMSD), the U.S. Geological Survey (USGS), the Wisconsin Department of Natural Resources (WDNR), and the Southeastern Wisconsin Regional Planning Commission (Commission). None of these data were used to develop the piecewise regression model.

Figure 3.2 compares the chloride concentrations estimated from specific conductance using the piecewise regression model to the corresponding measured concentrations for samples collected between 1964 and 2022. In each graph, the x-axis shows the measured concentration, and the y-axis shows the estimated concentration. Points above the blue line indicate samples in which the estimated concentration was higher than the measured concentration, or model overprediction. Points below the blue line indicate samples in which the measured concentration was higher than the estimated concentration, or underprediction by the regression model.

Figure 3.2 shows that there is considerable variation in how close the estimates of chloride concentration were to the measured value. The average difference for all four watersheds between the estimated and measured concentrations was slightly less than 38 mg/l. There could be several sources for these differences. Some differences between the estimated and measured concentrations would be expected given that a regression model estimates average values of a dependent variable for a given value of the independent variable. In addition, some of the differences between the estimated and measured values of chloride concentration may reflect differences in the concentrations of other ions in the samples. Since all the ions that are present in a sample contribute to specific conductance, differences in the concentrations of other ions will affect the relationship between specific conductance and chloride. Finally, the data shown in the figure were collected over a period of about 60 years. Changes in measurement and laboratory analytical techniques over that period could account for some of the differences.

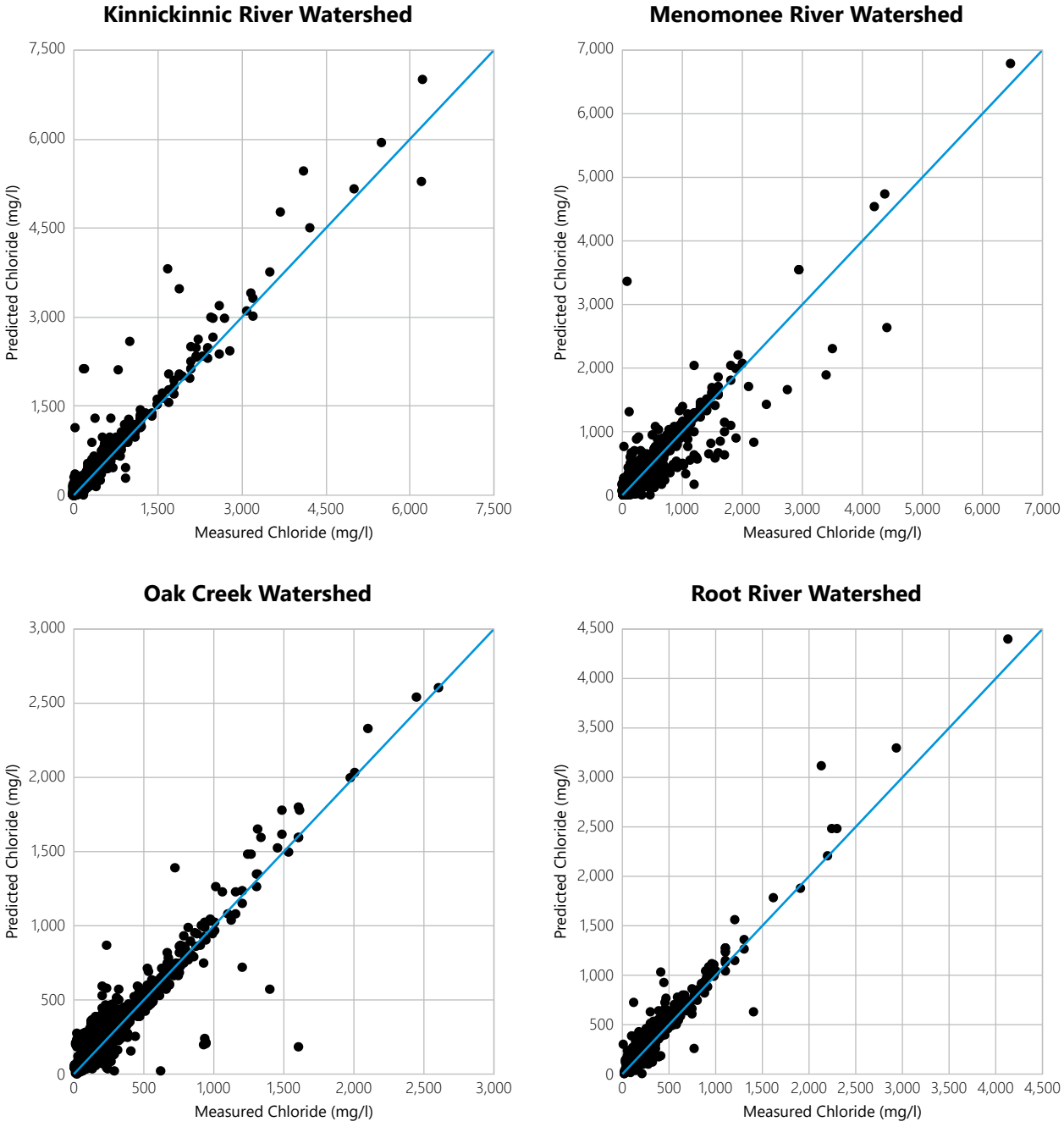
Figure 3.2 also shows that the piecewise regression tends to estimate higher chloride concentrations than the measured concentration more often than it estimates lower concentrations. While the extent to which this overprediction occurs varies, results from all four watersheds used to examine the performance of the piecewise regression show evidence of this overprediction effect. In addition, the model substantially underpredicts chloride concentrations for a subset of samples in the Menomonee River watershed.

The variation in how closely the piecewise regression model estimates measured chloride concentrations and the tendency of the model to overpredict chloride concentrations as compared to measured values led to additional examination of the performance of the model using the test dataset. Three potential sources of performance differences were examined: differences among data from different water quality monitoring agencies, differences among data from different streams within watersheds, and differences among data collected during different time periods.

Few differences were seen among monitoring agencies in comparisons of chloride concentrations estimated by the piecewise regression model to the corresponding measured concentrations. The major exception to this generalization is that many of the paired samples in the Menomonee River watershed in which the model underpredicted chloride concentration consisted of historical samples collected by the WDNR. It should also be noted that there were not enough data collected by USGS in the Kinnickinnic River watershed to make a valid comparison.

Only a few differences were seen among streams within watersheds when comparing chloride concentrations estimated by the piecewise regression model to the corresponding measured concentrations. Most of the paired samples in the Menomonee River watershed in which the model underpredicted chloride concentration consisted of samples from Honey Creek and Noyes Creek. Fewer samples showing underprediction by the model were collected from the mainstem of the Menomonee River, Underwood Creek, or the Little Menomonee River. Many streams in each watershed lacked sufficient data to make valid comparisons.

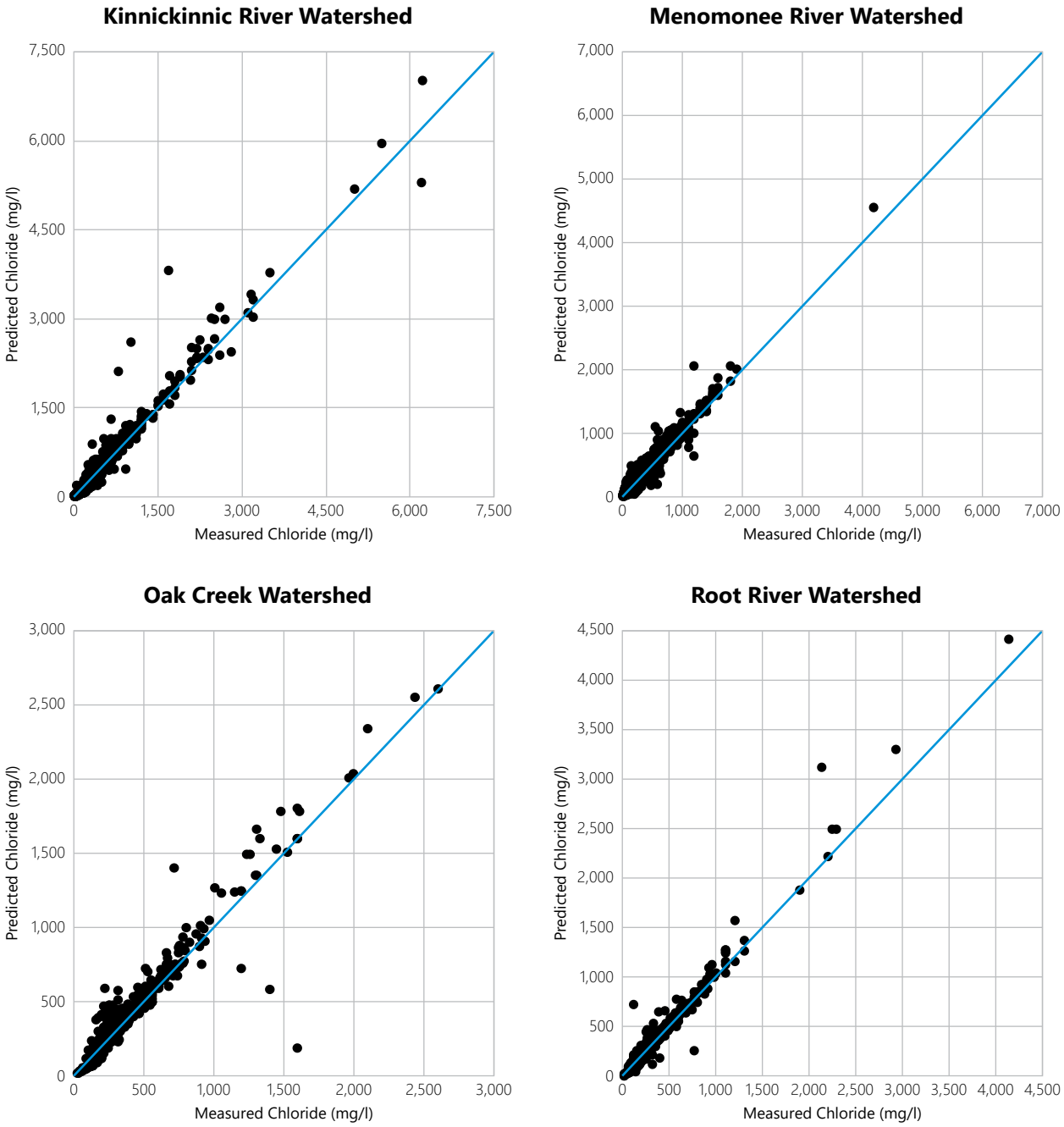
**Figure 3.2**
**Comparison of Chloride Concentrations Estimated by the Piecewise Regression Model to Measured Chloride Concentrations: 1964-2022**




*Source: SEWRPC*

The piecewise regression model performed better at estimating chloride concentrations from more recent specific conductance data than from older data. There was less variation in how close the estimates of chloride concentration were to the measured value in data collected in recent years than in data collected in earlier years. Figure 3.3 compares the chloride concentrations estimated from specific conductance using the piecewise regression model to the corresponding measured concentrations for 2011-2022. There is less variation in how close the estimates of chloride concentration were to the measured value in the Menomonee River, Oak Creek, and Root River watersheds than in the 1964-2022 data shown in Figure 3.2.

**Figure 3.3**
**Comparison of Chloride Concentrations Estimated by the Piecewise Regression Model to Measured Chloride Concentrations: 2011-2022**




*Source: SEWRPC*

Two factors could account for the better performance of the piecewise regression model using more recent data. As previously mentioned, changes in measurement and laboratory analytical techniques over that period could account for this. There have been improvements in conductivity sensor technology over the last 60 years and this might be reflected in the data used to evaluate the model.

The better performance of the model using more recent data could also reflect changes in the concentrations of other ions in streams of these watersheds over time. The relationship between specific conductance and chloride is influenced by the presence of other ions because they also contribute specific conductance. Long-term trends toward changes in concentrations of ions other than chloride have been documented for

these four watersheds.[38] Between 1975 and 2001, statistically significant decreases were observed at some sampling stations in these watersheds for ammonia; pH; and some dissolved metals and metalloids such as arsenic, cadmium, chromium, lead, and nickel. Significant increases were also observed in concentrations of some nutrients such as nitrate and dissolved phosphate and other metals such as copper and zinc. Some of these changes may have been complex. For example, the pH in Oak Creek decreased between 1975 and 2006 and increased between 2007 and 2016.[39]

Similarly, the concentration of chloride has increased over time in many streams of the Southeastern Wisconsin Region.[40] This has likely increased the contribution of chloride to specific conductance leading to better performance of the model when applied to more recent data.

The results of the comparison of chloride concentration estimates produced by the model to corresponding historically measured values have implications for the use of the piecewise regression model. The greater variability in the difference between estimates produced by the model and measured values observed in historical data than in recent data suggests that the model is best applied to recent data. In addition, the results of this evaluation suggest that the relationship between specific conductance and chloride concentration in individual streams may change over time, perhaps either as techniques for measuring these two water quality constituents change or improve or as ambient concentrations of other ionic water quality constituents change. It would be prudent to periodically conduct additional validation of the piecewise regression model in the future using paired specific conductance and chloride samples. If future comparisons of estimates of chloride concentrations generated by the model to measured values suggest that model performance is decreasing, the model may need to be revised.

Utility of the Piecewise Regression Model at Identifying Exceedances of Regulatory Thresholds

Estimates of chloride concentrations produced by the piecewise regression model were also compared to concurrent measured values by evaluating whether the model correctly predicted exceedances of regulatory thresholds. Comparisons were made to three Wisconsin regulatory thresholds that address chloride: the acute toxicity criterion for aquatic life of 757 mg/l, the chronic toxicity criterion for aquatic life of 395 mg/l, and the secondary drinking water maximum contaminant limit (MCL) of 250 mg/l (see Table 2.3). Three types of comparisons were made. The percentage of samples in which the model correctly identified whether each regulatory threshold was exceeded was calculated. The percentages of false positives and false negatives were also calculated. These comparisons were made for over 23,000 paired specific conductance-chloride samples from the four watersheds previously discussed from 1964-2022.

Table 3.3 shows the results of the calculation examining whether the piecewise regression model correctly identified exceedances of regulatory thresholds. With one exception, the model correctly identified whether each regulatory threshold was exceeded for 90 percent or more of the samples. The exception is that the model correctly identified exceedance of the secondary drinking water MCL for about 81 percent of samples in the Oak Creek watershed. In all watersheds the percentage of correct identifications was higher at higher concentration regulatory thresholds. For example, the model correctly identified whether the acute toxicity criterion was exceeded for over 99 percent of samples.

Table 3.4 shows the percentage of samples for which the piecewise regression model incorrectly identified that an exceedance of a regulatory threshold had occurred. In general, the rates at which these false positives occurred were very low, usually under 5 percent. There were two exceptions to this generalization. Higher percentages of false positives were seen for identification of exceedances of the secondary drinking water MCL in the Oak Creek and Root River watersheds. In general, lower rates of false positives were seen at higher regulatory thresholds than at lower thresholds.

[38] *SEWRPC Technical Report No. 39,* Water Quality Conditions and Sources of Pollution in the Greater Milwaukee Watersheds, *November 2007.*

[39] *SEWRPC Community Assistance Planning Report No. 330,* A Restoration Plan for the Oak Creek Watershed, *December 2021.*

[40] *See, for example, SEWRPC Technical Report No. 39,* op. cit.

**Table 3.3**
**Percent of Samples from 1964-2022 with Regulatory Threshold Exceedance Correctly Predicted by Piecewise Regression Model**

| Watershed | Number of Samples | Regulatory Threshold Exceedance Correctly Predicted | | |
|---|---|---|---|---|
| | | Acute toxicity 757 mg/l (percent) | Chronic Toxicity 395 mg/l (percent) | Drinking Water Secondary MCL 250 mg/l (percent) |
| Kinnickinnic River | 6,077 | 99.4 | 98.2 | 96.6 |
| Menomonee River | 12,016 | 99.5 | 97.4 | 94.4 |
| Oak Creek | 3,387 | 99.4 | 97.1 | 80.8 |
| Root River | 1,901 | 99.4 | 96.9 | 90.0 |
| Total | 23,371 | -- | -- | -- |
| Weighted Average[a] | -- | 99.5 | 97.5 | 92.7 |

[a] *Average weighted by number of samples.*

*Source: SEWRPC*

**Table 3.4**
**Percent of Samples from 1964-2022 with False Positives Predicted by Piecewise Regression Model[a]**

| Watershed | Number of Samples | Regulatory Threshold False Positives Predicted | | |
|---|---|---|---|---|
| | | Acute toxicity 757 mg/l (percent) | Chronic Toxicity 395 mg/l (percent) | Drinking Water Secondary MCL 250 mg/l (percent) |
| Kinnickinnic River | 6,077 | 0.5 | 1.5 | 2.8 |
| Menomonee River | 12,016 | 0.3 | 3.6 | 4.8 |
| Oak Creek | 3,387 | 0.2 | 2.7 | 23.4 |
| Root River | 1,901 | 0.4 | 3.6 | 12.0 |
| Weighted Average[b] | 23,371 | 0.3 | 2.1 | 7.6 |

[a] *False positives occur when the concentration of chloride estimated by the model exceeds the regulatory threshold while the measured value is below the threshold.*

[b] *Average weighted by number of samples.*

*Source: SEWRPC*

Table 3.5 shows the percentage of samples for which the piecewise regression model incorrectly identified that an exceedance of a regulatory threshold had not occurred. The rates at which these false negatives occurred were higher than the rates at which false positives occurred (see Table 3.4). In contrast to the pattern seen with false positives, higher rates of false negatives were seen at higher regulatory thresholds than at lower thresholds.

When examined using paired samples, the piecewise regression model does a better job identifying exceedances of higher regulatory thresholds than lower thresholds. When tested against either the acute toxicity criterion for aquatic life or the chronic toxicity criterion for aquatic life, the percentage of false positives generated by the model was quite low (see Table 3.4). This suggests that when chloride concentrations estimated using this model are higher than these thresholds, high confidence can be assigned to the idea that an exceedance has occurred. The model does not perform as well using a lower threshold such as the secondary drinking water MCL. Because of this, when chloride concentrations estimated using this model are higher than this threshold, less confidence should be placed in the likelihood that an exceedance has occurred. The application of the piecewise regression model is discussed in Section 3.4.

## 3.3 LINEAR MIXED EFFECTS REGRESSION MODEL

### Model Equations

Figure 3.4 shows the linear mixed effects regression model. The colored lines show the regression lines for individual sampling sites and the black line shows the consensus regression line. The gray line on the figure shows the lower portion of the piecewise regression model.

## Table 3.5
## Percent of Samples from 1964-2020 with False Negatives Predicted by Piecewise Regression Model[a]

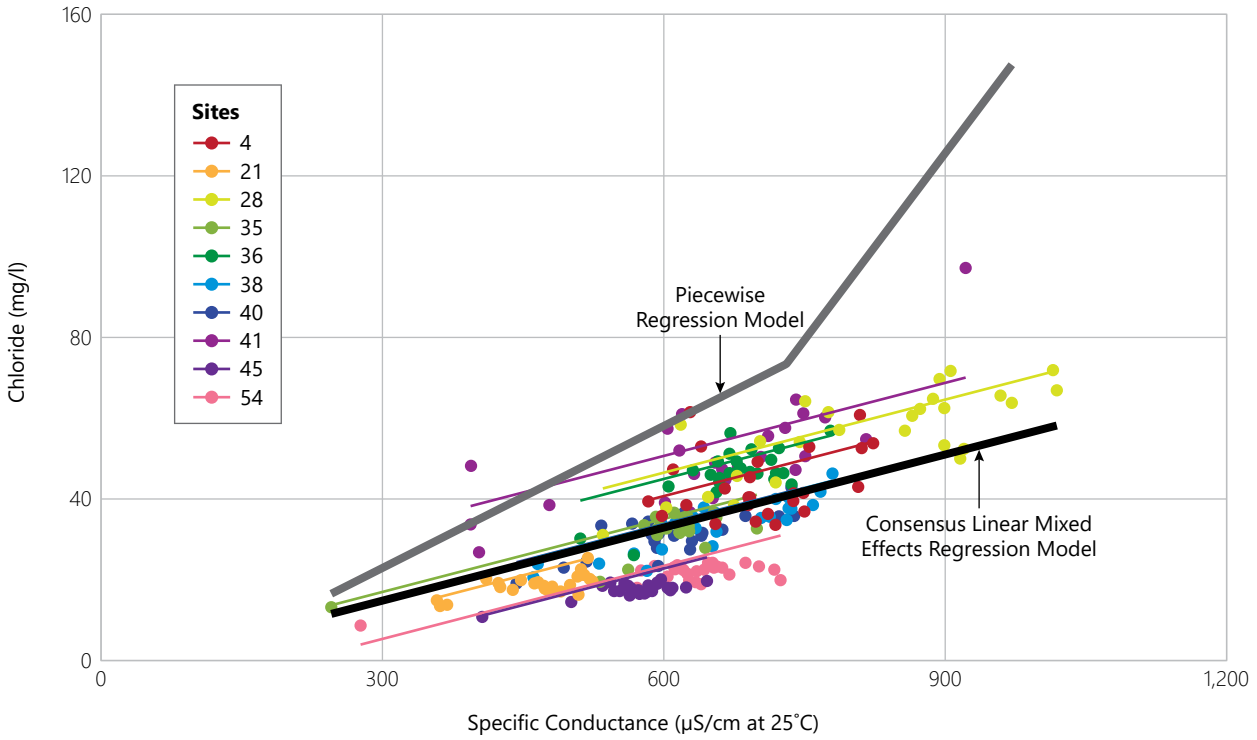
| Watershed | Number of Samples | Regulatory Threshold False Negatives Predicted | | |
|---|---|---|---|---|
| | | Acute toxicity 757 mg/l (percent) | Chronic Toxicity 395 mg/l (percent) | Drinking Water Secondary MCL 250 mg/l (percent) |
| Kinnickinnic River | 6,077 | 5.3 | 5.4 | 6.1 |
| Menomonee River | 12,016 | 18.3 | 11.8 | 10.2 |
| Oak Creek | 3,387 | 18.3 | 5.1 | 5.9 |
| Root River | 1,901 | 8.5 | 2.2 | 3.5 |
| Weighted Average[b] | 23,371 | 14.1 | 8.4 | 7.8 |

[a] *False negatives occur when the concentration of chloride estimated by the model is below the regulatory threshold while the measured value exceeds the threshold.*

[b] *Average weighted by number of samples.*

*Source: SEWRPC*

## Figure 3.4
## Linear Mixed Effects Model for Estimating Chloride from Specific Conductance in Southeastern Wisconsin Streams




*Source: SEWRPC*

The equations for the linear mixed effects model are given in Table 3.6. This model consists of eleven sets of equations with each set consisting of one or two equations. Ten of the equation sets consist of models for individual sample sites. The eleventh consists of a consensus model that may be used for other sites for which this linear mixed effects regression may be applicable. The table also includes the ranges of specific conductance and chloride concentration in the data that were used to develop the linear mixed effects regression model.

In seven sets of equations, the equation covering the lowest portion of the specific conductance range was added because the main equation estimating chloride crosses the x-axis at a positive specific conductance. The magnitude of the x-intercept in these models varied, ranging between 19 µS/cm and 223 µS/cm. As a

result, these equations would predict negative concentrations of chloride for specific conductance values at and below the x-intercept. Because this result makes no physical sense, estimates of chloride concentrations at specific conductance values at and below the x-intercepts were set to zero mg/l.

Two facts support this use of zero mg/l by this model as an estimate for chloride concentration when specific conductance was at or below the x-intercept. First, numerous ions beside chloride and sodium contribute to specific conductance. Some of these ions are likely to be present in stream water even when chloride is absent. Second, specific conductance values for each equation generated by the linear mixed effects model never got as low as or lower than the value of the respective x-intercept.

**Evaluation of Linear Mixed Effects Regression Model**

The linear mixed effects regression model is based on 253 paired samples of specific conductance and chloride from 10 sampling sites. The linear mixed effects regression model was developed from a dataset that spanned a specific conductance range of about 775 µS/cm and a chloride range of about 47 mg/l (see Table 3.6).

The model examined 10 sites based on 242 degrees of freedom. The linear mixed effect regression model was found to be significant at a probability less than 0.0001. Two $R^2$ values are associated with the linear mixed effects model. The marginal $R^2$ describes the proportion of the variance accounted for by the fixed factor alone, in this case the effect of specific conductance. The conditional $R^2$ describes the proportion of the variance accounted for by both the fixed and random factors, in this case effects of both specific conductance and differences among sampling sites in background water chemistry. The marginal $R^2$ of the consensus linear mixed effects model is 0.306. This means that variation in specific conductance accounts for about 31 percent of the variation in chloride concentration. The conditional $R^2$ of the consensus model is 0.838, indicating that the variation in specific conductance and the variation among sites in background water chemistry accounts for about 84 percent of the variation in chloride concentration.

The fact that the conditional $R^2$ is so much higher than the marginal $R^2$ is not surprising considering that this model examines the relationship between specific conductance and chloride concentration at sites where the values of both these water quality constituents are quite low. At low levels of specific conductance and chloride concentration, other ions that are present in the water have much greater impact on specific conductance than they would at higher chloride concentrations.

Differences among sample sites in concentrations of chloride, specific conductance, and some ionic components of water quality are shown in Table 3.7. Given the low levels of specific conductance seen at these sites, these differences may account for some of the variation among the equations for individual sample sites shown in Table 3.6.

***Cross-Validation of Linear Mixed Effects Model***

A stratified cross-validation was performed on the linear mixed effects regression model. This procedure consisted of 10 iterations. The data were organized into 10 groups containing paired data from each of the 10 sites, with the goal of each group having equal numbers of samples from each site. In each iteration, one group was removed from the data. A linear mixed effects regression was developed for the remaining data. This regression produced separate equations for each site and a consensus equation. A different subset was removed for each iteration. The results from the cross-validation were compared to the original linear mixed effects regression model using several metrics. The results of the cross-validation for the linear mixed effects regression model are summarized in Table 3.8. More information on the methods used for cross-validation is given in Chapter 2 of this Report.

Comparison of Slopes

The consensus equation slopes generated by iterations of the cross-validation were compared to the slope of the consensus of the linear mixed effects regression model. Since the slope of the equation for each site was the same as the consensus equation slope, this comparison for the site-specific equations was not needed (see Table 3.6).

## Table 3.6
## Linear Mixed Model for Estimating Chloride Concentration from Specific Conductance

| Site Number | Site | Specific Conductance Range (µS/cm at 25°C)[a] | Equation to Estimate Chloride Concentration (mg/l) |
|---|---|---|---|
| 4 | Sugar Creek | SC > 0 | $[Cl^-] = 0.0603 \times SC + 4.5$ |
| 21 | East Branch Milwaukee River | SC ≤ 100<br>SC > 100 | $[Cl^-] = 0$<br>$[Cl^-] = 0.0603 \times SC - 6.0$ |
| 28 | East Branch Rock River | SC > 0 | $[Cl^-] = 0.0603 \times SC + 10.3$ |
| 35 | Honey Creek Upstream of East Troy | SC ≤ 19<br>SC > 19 | $[Cl^-] = 0$<br>$[Cl^-] = 0.0603 \times SC - 1.1$ |
| 36 | Honey Creek Downstream of East Troy | SC > 0 | $[Cl^-] = 0.0603 \times SC + 8.8$ |
| 38 | North Branch Milwaukee River | SC ≤ 42<br>SC > 42 | $[Cl^-] = 0$<br>$[Cl^-] = 0.0603 \times SC - 2.5$ |
| 40 | Stony Creek | SC ≤ 40<br>SC > 40 | $[Cl^-] = 0$<br>$[Cl^-] = 0.0603 \times SC - 2.4$ |
| 41 | Milwaukee River near Saukville | SC > 0 | $[Cl^-] = 0.0603 \times SC + 14.5$ |
| 45 | Mukwonago River at Nature Road | SC ≤ 223<br>SC > 223 | $[Cl^-] = 0$<br>$[Cl^-] = 0.0603 \times SC - 13.4$ |
| 54 | Whitewater Creek | SC ≤ 211<br>SC > 211 | $[Cl^-] = 0$<br>$[Cl^-] = 0.0603 \times SC - 12.7$ |
| -- | Consensus Model[b] | SC ≤ 55<br>SC > 55 | $[Cl^-] = 0$<br>$[Cl^-] = 0.0603 \times SC - 3.3$ |

| | Range of Values | |
|---|---|---|
| | Specific Conductance (µS/cm at 25°C) | Chloride (mg/l) |
| | 245 – 1,020 | 11.5 – 97.2 |

Note: SC indicates specific conductance. $[Cl^-]$ indicates chloride concentration.

[a] *The standard temperature used for adjusting conductivity to calculate specific conductance is 25°C, which is equivalent to 77°F.*

[b] *The consensus model is to be used to estimate chloride concentration at other sites for which this model is applicable.*

[c] *Estimates outside the range of data used to develop the regression model should be treated as less reliable due to extrapolation beyond the range of values used to develop the model.*

*Source: SEWRPC*

All the slopes from iterations of the cross-validation were within 8 percent of the slope from the consensus equation in the linear mixed effects regression model (see Table 3.8). Most slopes were within 2 percent. The two iterations with the highest differences in slope from the consensus model contained the two highest chloride concentrations in the dataset.[41] The slopes from the iterations of the cross-validation and the linear mixed effects regression model were all less than 0.07 mg/l per µS/cm. Because these are such low slopes, the magnitude in the variations in slopes among the iterations of the cross-validation are quite low. This comparison of slopes indicates that linear mixed effects model is not sensitive to individual samples across most of its applicable range of values but does show some sensitivity at the upper end of the chloride range.

Comparison of y-intercepts

The y-intercepts generated by the iterations of the cross-validation were compared to those from the linear mixed effects regression model. This evaluation included comparing the y-intercepts for the consensus model from each cross-validation iteration to that of the linear mixed effects regression and the y-intercepts generated for each individual sample site from each cross-validation iteration to that of the linear mixed effects regression.

All the y-intercepts for the consensus equations from iterations of the cross-validation were within 90 percent of the y-intercept for the linear mixed effects model consensus equation (see Table 3.8). Most yintercepts were within 12 percent of the consensus model y-intercept. While these percentages are high,

[41] *The highest chloride concentration included in the linear mixed effects regression dataset was nearly three times higher than the median chloride concentration of the dataset.*

## Table 3.7
## Summary Statistics for Water Quality Constituents at Sampling Sites Used to Develop the Linear Mixed Effects Regression Model: 2018-2020

| | Sampling Site[a] | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Statistic** | **4** | **21** | **28** | **35** | **36** | **38** | **40** | **41** | **45** | **54** |
| Calcium | | | | | | | | | | |
| Samples | 25 | 25 | 26 | 26 | 25 | 25 | 25 | 25 | 25 | 26 |
| Minimum (mg/l) | 65.3 | 41.7 | 47.3 | 30.2 | 66.9 | 49.0 | 44.2 | 49.7 | 49.6 | 34.3 |
| Mean (mg/l) | 82.0 | 55.6 | 81.0 | 76.4 | 77.3 | 73.6 | 69.9 | 68.5 | 70.8 | 75.7 |
| Maximum (mg/l) | 92.6 | 66.4 | 95.8 | 92.6 | 94.4 | 87.4 | 81.3 | 82.7 | 84.7 | 89.8 |
| Standard Deviation | 7.2 | 6.2 | 13.0 | 11.0 | 6.1 | 10.1 | 8.5 | 8.6 | 6.9 | 9.9 |
| Chloride | | | | | | | | | | |
| Samples | 25 | 25 | 26 | 26 | 25 | 25 | 25 | 25 | 25 | 26 |
| Minimum (mg/l) | 33.6 | 13.5 | 31.1 | 13.2 | 26.1 | 20.6 | 19.1 | 26.8 | 10.8 | 8.7 |
| Mean (mg/l) | 43.7 | 18.7 | 56.1 | 31.7 | 46.2 | 33.7 | 31.2 | 50.7 | 17.6 | 21.6 |
| Maximum (mg/l) | 61.5 | 25.4 | 71.9 | 36.6 | 56.9 | 46.3 | 42.9 | 97.2 | 20.1 | 24.2 |
| Standard Deviation | 8.1 | 2.6 | 11.0 | 5.4 | 6.6 | 6.6 | 5.1 | 13.4 | 1.9 | 3.1 |
| Hardness | | | | | | | | | | |
| Samples | 25 | 25 | 26 | 26 | 25 | 25 | 25 | 25 | 25 | 26 |
| Minimum (mg/l as $CaCO_3$) | 281 | 200 | 222 | 130 | 278 | 228 | 205 | 231 | 218 | 150 |
| Mean (mg/l as $CaCO_3$) | 359 | 267 | 371 | 341 | 346 | 342 | 326 | 322 | 316 | 348 |
| Maximum (mg/l as $CaCO_3$) | 403 | 317 | 437 | 388 | 391 | 407 | 379 | 381 | 360 | 402 |
| Standard Deviation | 31 | 30 | 59 | 48 | 25 | 47 | 41 | 42 | 30 | 46 |
| Magnesium | | | | | | | | | | |
| Samples | 25 | 25 | 26 | 26 | 25 | 25 | 25 | 25 | 25 | 26 |
| Minimum (mg/l) | 28.7 | 23.2 | 25.1 | 13.2 | 26.2 | 25.6 | 23.0 | 26.0 | 22.8 | 15.6 |
| Mean (mg/l) | 37.4 | 31.2 | 41.0 | 36.5 | 37.2 | 38.4 | 36.9 | 36.7 | 33.9 | 38.5 |
| Maximum (mg/l) | 41.6 | 36.7 | 48.5 | 42.0 | 41.4 | 45.8 | 42.7 | 43.0 | 38.2 | 43.5 |
| Standard Deviation | 3.3 | 3.6 | 6.7 | 5.8 | 3.6 | 5.4 | 4.9 | 5.1 | 3.2 | 5.3 |
| Potassium | | | | | | | | | | |
| Samples | 25 | 25 | 26 | 26 | 25 | 25 | 25 | 25 | 25 | 26 |
| Minimum (mg/l) | 1.8 | 0.7 | 1.9 | 1.8 | 1.9 | 2.4 | 1.9 | 2.3 | 0.8 | 1.4 |
| Mean (mg/l) | 2.4 | 1.8 | 3.5 | 2.3 | 2.3 | 3.6 | 2.8 | 3.4 | 1.4 | 2.3 |
| Maximum (mg/l) | 3.7 | 2.9 | 7.6 | 4.6 | 3.9 | 5.4 | 4.1 | 6.1 | 1.7 | 6.1 |
| Standard Deviation | 0.5 | 0.5 | 1.5 | 0.6 | 0.4 | 0.7 | 0.6 | 0.9 | 0.2 | 0.9 |

**Table continued on next page.**

## Table 3.7 (Continued)

| Statistic | Sampling Site[a] 4 | 21 | 28 | 35 | 36 | 38 | 40 | 41 | 45 | 54 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Sodium** | | | | | | | | | | |
| Samples | 25 | 25 | 26 | 26 | 25 | 25 | 25 | 25 | 25 | 26 |
| Minimum (mg/l) | 15.8 | 6.9 | 15.9 | 6.9 | 12.1 | 10.7 | 10.5 | 14.6 | 4.8 | 4.1 |
| Mean (mg/l) | 21.2 | 10.4 | 28.5 | 13.7 | 21.9 | 17.7 | 16.6 | 27.9 | 7.6 | 8.3 |
| Maximum (mg/l) | 30.1 | 12.5 | 36.4 | 16.2 | 27.6 | 23.4 | 21.5 | 55.1 | 8.9 | 9.5 |
| Standard Deviation | 3.9 | 1.3 | 5.2 | 2.3 | 3.2 | 3.3 | 2.2 | 7.7 | 0.8 | 1.0 |
| **Specific Conductance** | | | | | | | | | | |
| Samples | 25 | 25 | 26 | 26 | 25 | 25 | 25 | 25 | 25 | 26 |
| Minimum (µS/cm at 25°C) | 444 | 349 | 535 | 245 | 511 | 461 | 248 | 394 | 406 | 362 |
| Mean (µS/cm at 25°C) | 661 | 452 | 876 | 593 | 661 | 652 | 478 | 572 | 552 | 621 |
| Maximum (µS/cm at 25°C) | 811 | 524 | 1,130 | 656 | 736 | 780 | 687 | 741 | 624 | 724 |
| Standard Deviation | 94 | 57 | 143 | 76 | 49 | 89 | 138 | 97 | 56 | 70 |
| **Sulfate** | | | | | | | | | | |
| Samples | 25 | 25 | 26 | 26 | 25 | 25 | 25 | 25 | 25 | 26 |
| Minimum (mg/l) | 17.1 | 2.4 | 19.5 | 11.4 | 28.0 | 13.0 | 11.2 | 11.4 | 13.2 | 12.9 |
| Mean (mg/l) | 23.0 | 8.4 | 28.3 | 43.9 | 40.3 | 21.5 | 17.9 | 18.8 | 19.6 | 31.7 |
| Maximum (mg/l) | 27.9 | 14.4 | 39.7 | 71.8 | 68.6 | 36.7 | 34.7 | 26.9 | 26.6 | 39.0 |
| Standard Deviation | 3.1 | 2.3 | 6.1 | 12.6 | 8.7 | 6.1 | 4.9 | 4.0 | 3.2 | 5.3 |

[a] *Locations of sample sites are shown on Map 2.3 and described in Table 2.1.*

*Source: SEWRPC*

**Table 3.8**
**Linear Mixed Effects Versus Consensus Cross-Validation Model Parameters**

| Model Parameter | Linear Mixed Effects | Consensus Cross-Validation | | |
|---|---|---|---|---|
| | | Minimum | Mean | Maximum |
| Slope | 0.0603 | 0.0557 | 0.0603 | 0.0648 |
| X-Intercept | 54.8 | 7.4 | 52.9 | 96.7 |
| Y-Intercept | -3.3 | -6.3 | -3.3 | -0.4 |
| Marginal $R^2$ | 0.306 | 0.261 | 0.307 | 0.353 |
| Conditional $R^2$ | 0.838 | 0.830 | 0.839 | 0.853 |
| Mean Squared Error | 67.7 | 62.9 | 67.6 | 70.4 |

Note: The marginal $R^2$ describes the proportion of the variance accounted for by the effect of specific conductance. The conditional $R^2$ describes the proportion of the variance accounted for by both specific conductance and differences among sampling sites in background water chemistry.

*Source: SEWRPC*

they represent small differences in the magnitude of chloride concentration. The cross-validation consensus y-intercepts were all within 3 mg/l of the consensus y-intercept from the linear mixed effects regression model. As with the comparison of slopes, the iterations with the largest differences in y-intercepts from the consensus model were iterations that contained the highest chloride concentrations in the dataset.

All the y-intercepts for the site-specific equations from the cross-validation were within 2 mg/l of the corresponding y-intercept for that site from the linear mixed effects regression. Most of the y-intercepts were within 1 mg/l. Expressed as a percent, all the y-intercepts for the site-specific equations from the iterations were within 34 percent of their corresponding site-specific y-intercept from the linear mixed effects model. Most y-intercepts were within 10 percent. Site 35, Honey Creek Upstream in Walworth County, had the largest differences in its y-intercepts between the cross-validation iterations and the linear mixed effects model while Site 45, Mukwonago River at Nature Road, had the smallest differences. These results imply that the y-intercepts are generally not sensitive to individual samples, although some sites, such as Site 35, are more sensitive than others.

Comparison of x-intercepts

The x-intercept of the linear mixed effects regression model indicates the level of specific conductance that is observed when no chloride is present in the water. This specific conductance is determined by the slope and y-intercept of the equation for the model. Separate x-intercepts were found for the consensus equation and each of the site-specific equations.

All the x-intercepts generated by the iterations of the cross-validation for the consensus equation were within 87 percent of the xintercept of the consensus equation from the linear mixed effects regression model (see Table 3.8). Most consensus equation x-intercepts were within 11 percent. The magnitude of the x-intercepts varied between 7 µS/cm and 97 µS/cm. All the x-intercepts generated by iterations of the cross-validation for site-specific equations were within 45 percent of the x-intercept from the corresponding equation in the linear mixed effects regression model. Most of these x-intercepts were within 10 percent of the x-intercept from the corresponding equation in the linear mixed effects regression model. As with the y-intercept, Site 35 had the largest differences in x-intercepts between the cross-validation iterations and the linear mixed effects regression, while Site 45 had the smallest differences.

Comparison of $R^2$

The marginal $R^2$ values generated by the iterations of the cross-validation for the linear mixed effects model varied from 0.261 to 0.353, with a mean of 0.307 (see Table 3.8). All of the marginal $R^2$s were within 16 percent of that of the consensus equation of the linear mixed effects regression model. This indicates that there are not considerable differences among the iterations in how much of the variation in chloride concentration is accounted for by variation in specific conductance alone. The iterations with the greatest differences in marginal $R^2$ values were the iterations where the highest chloride concentrations were not included in the cross-validation dataset.

The conditional $R^2$ values generated by the iterations of the cross-validation varied from 0.830 to 0.853, with a mean of 0.839. All of them were within 2 percent of the conditional $R^2$ of the linear mixed effects regression model. This indicated that in all iterations, most of the variation in chloride concentrations is accounted for by the combination of the variation in specific conductance and the variation among sampling sites in background water chemistry.

Comparison of Mean Squared Error

The MSE of the consensus linear mixed effect model was 67.7 (see Table 3.8). The MSEs from the iterations of the cross-validation ranged between 62.9 and 70.4, with a mean value of 67.6. This average was very close to the MSE from the consensus linear mixed effects regression model. The MSEs from six cross-validations iterations were higher than the MSE of the linear mixed effects regression. The MSEs from the other cross-validation iteration were lower. All cross-validation MSEs were within 7 percent of the MSE from the linear mixed effects regression and most MSEs were within 3 percent.

Ability of Cross-Validation Equations to Estimate Chloride Concentrations for Removed Data

The iterations of the cross-validation also estimated chloride concentration from each specific conductance observation in the removed data group with the resultant linear mixed effects equation from the cross-validation. Differences between predicted versus actual chloride concentrations were determined, and the differences ranged from 2.7 to 4.8 mg/l, with a mean of 3.7 mg/l.

Summary of Cross-Validation Results

The results of the cross-validation of the linear mixed effects regression model indicate that model parameters and other metrics are reasonably stable to removal of 10 percent of the data used to develop the model. However, some sites are less stable than others and removal of the highest chloride concentrations in the model dataset seemingly had a larger effect on the resultant equations than lower concentration samples. When compared to the parameters of the consensus equation from the linear mixed effects regression model on a percentage basis, the slopes and y-intercepts of the iterations of the cross-validation show limited variability. Given that magnitudes of the variations of slopes in the iterations were small, some of this variability reflects the rather shallow slopes in the model. Individual iterations showed greater variability in the parameters of the equations for individual sites than for the consensus model.

A few factors may contribute to the variability in the parameters of the linear mixed effects regression model during cross validation. First, the number of samples used to develop this model was smaller than that used to develop the piecewise regression model. This means that the relative contribution of each sample to development of the model was greater. This is especially important given that outliers were present at a few sampling sites (see Figure 3.4). Removal of one of more outliers during an iteration may have had an outsized effect on the parameter of the iteration. In addition, some lack of balance in removal of samples from data for individual sites may have affected the results. Ideally, exactly 10 percent of samples from each site would be removed during each iteration. Since there were about 25 samples taken at each site, in practice this meant that for each iteration two samples were removed from the data for some sites and three samples were removed from the data for other sites. For each iteration, this likely resulted in some sites having greater influence on the parameters of the consensus model than other sites had.

Commission staff could not identify a large, comparable dataset with low chloride concentrations within Southeastern Wisconsin to use for validating the linear mixed effects regression model as was done for the piecewise regression model. Consequently, the comparison of predicted versus measured chloride concentrations from the 10 percent of data removed in each cross-validation iteration is the only external validation for this model. While these results indicate that the model performs fairly well, it should be noted that the validation dataset is still only drawn from the same ten sites that were used to develop the model, so the model may perform more favorably for samples from these sites than for completely external sites.

Taken together, these results suggest that the linear mixed effects model should be used with caution. The site-specific equations seem adequate for estimating chloride concentrations at the sites that were used for developing the model. The variability shown by parameters for the consensus equation in the iterations of the cross-validation, the low marginal $R^2$, and the inability to compare against a completely external validation dataset suggests that the consensus equation might not produce acceptable estimates of chloride at other sites as this equation seems more sensitive to differences in background water chemistry than the piecewise regression. The application of the linear mixed effects model is discussed in the next section of the Chapter.

## 3.4 APPLICATION OF THE MODELS

**Application of the Piecewise Regression Model**

For applications at most sites, the piecewise regression model is preferred over the linear mixed effects regression model. The piecewise regression model covers more of the specific conductance levels and chloride concentrations that are likely to be encountered in streams of southeastern Wisconsin compared to the linear mixed effects model. When compared to the entire Study dataset, the piecewise regression better captures the departures of the data from linearity (see Figure 3.1). In addition, the linear mixed effects model consists of lines developed from data at sites with low values of specific conductance and low concentrations of chloride over narrow ranges. The linear mixed effects model is designed for sites with low values of each and will not give estimates of chloride concentration that approach regulatory limits. Such limits are beyond the range of the model.

The evaluation of the piecewise regression model given in this Chapter indicates that it should be appropriate to use for estimating average concentrations of chloride from specific conductance at many stream sites in the Southeastern Wisconsin Region. A few important considerations should be taken into account when applying it. First, an estimate generated by the model represents the average chloride concentration that is expected to be observed at a particular level of specific conductance. There will be variation around this average, so some deviation of estimated concentrations from measured values in paired specific conductance-chloride samples should be expected.

Second, comparisons of chloride concentrations estimated using the piecewise regression model to historical measured chloride values from paired samples in four watersheds showed that the model tended to overestimate chloride concentration in these watersheds more often than it underestimated it. The samples used for this comparison came mostly from sites that have moderate to large amounts of urban land use in their drainage areas. This was due to the lack of independent datasets with paired specific conductance-chloride data from sites located in more rural areas. It is not known whether the piecewise regression model will tend to overestimate or underestimate chloride concentrations in areas with less urban development. The small percentage of false positives found when examining exceedance of water quality criteria suggests that this tendency to overpredict chloride concentration should not pose a major problem for determining whether regulatory thresholds have been exceeded; however, it could lead to overestimates of chloride loads at some sites.

A third consideration for using the piecewise regression model is that it may not perform well in streams immediately downstream of lakes. This is because the large volume of water in a lake will tend to buffer concentrations of chloride in water flowing out of the lake. This acts to attenuate the concentrations of chloride downstream of the lake and dampen the hydrological, anthropogenic, and seasonal patterns observed as compared to those in streams upstream of the lake.[42] It is not certain how far downstream the impact of this lake buffering effect may extend. It is likely to vary among stream sites depending on site-specific details. Factors that may affect whether the model performs poorly downstream of a lake include:

- The residence time of water in the lake, with longer residence times leading to poorer model performance
- Land use in the contributing drainage area to the lake
- Land use in the contributing drainage area downstream of the lake but upstream of the sampling site
- The presence, density, or relative proximity of transportation infrastructure in the contributing drainage area downstream of the lake but upstream of the sampling site
- The difference in relative amounts of flow to the sampling site originating from the lake and from the portion of the drainage area downstream from the lake

[42] *L.A. Rock and H.A. Dugan, "Lakes Protect Downstream Riverine Habitats from Chloride Toxicity,"* Limnology and Oceanography, *68:1,216-1,231, 2023*.

The final consideration for applying the piecewise regression model is that at low levels of specific conductance there is much variability in the relationship between specific conductance and chloride concentration (see Figure 2.6 in Chapter 2). This variability reflects the influence of other ions on specific conductance. At low chloride concentrations, there is simply not enough chloride in the water to dominate specific conductance. Thus, specific conductance may be a poor predictor of chloride concentration at low levels of specific conductance, and the piecewise regression model may not perform as well at low levels of specific conductance as it does at intermediate and higher levels.

These considerations have implications for the use of the piecewise regression model. When using the model to estimate chloride concentrations at a stream site that was not used in its development, it would be prudent to compare the estimates generated by the model to measured chloride concentrations from paired specific conductance-chloride samples. This comparison can be made using recent paired samples from the site, if they are available. If no such samples are available, sample collection is recommended. If the sample comparison shows that the piecewise regression model produces reasonable estimates of chloride concentration for the stream site, the model should be appropriate for application at the site. If the estimates produced by the model are not reasonable, then use of an alternative model may be warranted. Under circumstances that are discussed in the following section, it may be possible to apply the linear mixed effects regression model. Otherwise, it may be necessary to develop a site-specific regression model for that stream location.

**Application of the Linear Mixed Effects Regression Model**

The site-specific linear equations from the linear mixed effects regression model can be used to estimate chloride concentrations at the sites for which they were developed (see Table 3.6). As previously noted, for other stream locations using the piecewise regression model is preferred over the consensus equation from the linear mixed effects model.

Despite this, there may be some situations in which it is appropriate to apply the consensus equation from the linear mixed effects model. In such cases, some considerations should be taken into account. As previously noted, this model was developed for application to sites with low levels of specific conductance and low chloride concentrations. The linear mixed effects regression model will not give estimates of chloride concentration that approach regulatory limits and thus is not suitable for evaluating whether streams meet water quality criteria for chloride. As with the piecewise regression model, an estimate generated by the linear mixed effects regression model represents the average chloride concentration that is expected to be observed at a particular level of specific conductance and there will be variation around this average. This model may not perform well at stream sites immediately downstream of lakes due to the buffering effects of lakes on downstream concentrations as was previously discussed. Finally, the linear mixed effects model is likely to perform best at sites with drainage areas that are similar to the sites used in its development.[43]

Similar drainage area characteristics to those used for developing the linear mixed effects model suggests that it would be best to apply the linear mixed effects model only to data that meet several conditions. First, sampling at a stream site should show that the piecewise regression model systematically overestimates chloride concentrations at the site. Additionally, sampling at the site should show that specific conductance never exceeds 1,000 µS/cm. Third, the percentage of urban land uses in the drainage area contributing to the site should be less than 20 percent. The WDNR has two online models that can be used to estimate the amount of urban land uses in a drainage area to a site, PrestoLite[44] and the Wisconsin Water Explorer.[45] Finally, no major highway or transportation network areas should drain to the waterbody in close proximity upstream of the stream site. It is important to recognize that the linear mixed effects regression model may not produce good estimates for all sites that meet these criteria. There may also be other circumstances in which it performs poorly.

[43] *Site drainage areas are characterized in SEWRPC Technical Report No. 61,* Field Monitoring and Data Collection for the Chloride Impact Study*, September 2023.*

[44] *PrestoLite can be accessed at dnr.wisconsin.gov/topic/SurfaceWater/PRESTO.HTML, accessed January 3, 2024.*

[45] *The Wisconsin Water Explorer can be accessed at dnr.wisconsin.gov/topic/SurfaceWater/WEx.html, accessed January 3, 2024.*

As with the piecewise regression model, it would be prudent to compare the estimates generated by the linear mixed effects regression model to measured chloride concentrations from paired specific conductance-chloride samples. This comparison can be made using recent paired samples from the site, if they are available. If no such samples are available, it would be prudent to collect some. If the comparison to sampling shows that the linear mixed effects regression model produces reasonable estimates of chloride concentrations for the site, the model should be appropriate for application at the stream site.

**Regression Models for Lake Data**

The piecewise regression and linear mixed effects regression models were designed specifically for estimating chloride concentrations from specific conductance at stream and river sites. These models should not be applied to lake data.

Site specific linear regression models were developed for four of the lakes sampled as part of the Chloride Impact Study. These models are described in Appendix B of this report. Appendix B also discusses strategies for estimating chloride concentrations from specific conductance for other lakes in the Southeastern Wisconsin Region.

# APPENDICES

# ACRONYMS AND ABBREVIATIONS

# APPENDIX A

| | |
|---|---|
| °C | Degrees Celsius |
| °F | Degrees Fahrenheit |

**B**

| | |
|---|---|
| b | y-intercept of a linear equation |

**M**

| | |
|---|---|
| m | Slope of a linear equation |
| MCL | Maximum contaminant limit |
| mg/l | Milligrams per liter |
| MMSD | Milwaukee Metropolitan Sewerage District |
| MAE | Mean absolute error |
| MSE | Mean squared error |

**R**

| | |
|---|---|
| $R^2$ | Coefficient of determination |

**S**

| | |
|---|---|
| SEWRPC | Southeastern Wisconsin Regional Planning Commission |

**U**

| | |
|---|---|
| USGS | U.S. Geological Survey |

**W**

| | |
|---|---|
| WDNR | Wisconsin Department of Natural Resources |

**µ**

| | |
|---|---|
| µS/cm | MicroSiemens per centimeter |

# APPENDIX B

# LAKE-SPECIFIC REGRESSIONS FOR LAKES SAMPLED AS PART OF THE CHLORIDE IMPACT STUDY

## INTRODUCTION

As part of the Chloride Impact Study, Commission staff collected paired samples of specific conductance and chloride concentration from six lakes in the Southeastern Wisconsin Region: Big Cedar Lake, Geneva Lake, Little Muskego Lake, Moose Lake, Silver Lake (Washington County), and Voltz Lake. These lakes were chosen to provide balanced geographic representation of the Region and to include a variety of lake types including drainage, drained, seepage, and spring lakes. Sampling data was taken quarterly for 2018-2021 at multiple depths in each lake. More information on these lakes is available in a separate technical report.[46]

The piecewise and linear mixed effects regression models described in this report were developed for estimating chloride concentrations in rivers and streams of the Region. The datasets used to develop these models did not include any samples from lakes. In addition, there are substantial differences between running water systems such as streams and rivers and standing water systems such as lakes and ponds. For example, flowing water systems have physical mixing that allows more of the water volume to interact with the atmosphere as well as short water residence times. New water replaces old water more frequently in a flowing water system than in a standing water system. At any one time, only a small portion of water near the surface of a lake or pond is in direct contact with the atmosphere. Lakes and ponds that are deep enough will undergo seasonal thermal stratification. During stratification, deeper portions of a lake or pond are physically prevented from interacting with the atmosphere. Mixing in lakes and ponds that stratify may be restricted to short periods during the spring and fall. Such stratification is less common in streams and rivers and occurs mostly in dammed impoundments.

The physical differences between running water systems and standing water systems can lead to differences in their water chemistry. Because of this, models designed to estimate chloride concentrations from specific conductance in stream and river systems may perform poorly when applied to lake and pond systems.

[46] *SEWRPC Technical Report No. 61,* Field Monitoring and Data Collection for the Chloride Impact Study, *September 2023*.

Figure B.1 shows comparisons of chloride concentrations estimated from specific conductance using the piecewise regression model developed for streams described earlier in this report to measured concentrations of chloride in the six lakes sampled as part of the Chloride Impact Study. When the chloride concentration estimated by the model was the same as the measured concentration, the point lies along the one-to-one line which is indicated by the blue line on the graph. Points above the blue line indicate that the concentration estimates generated by the model were higher than the measured concentrations. Points below the blue line indicate that the piecewise regression model estimated concentrations were lower than the measured concentrations.

Figure B.1 shows that the piecewise regression model performs poorly for most of the lakes examined. The model systematically underestimates measured concentrations of chloride in Big Cedar, Little Muskego, and Moose Lakes and overestimates measured concentrations in Silver and Voltz Lakes. The model does not appear to systematically overestimate or underestimate chloride concentrations in Geneva Lake; however, it should be noted that this lake had very narrow ranges of levels of specific conductance and concentrations of chloride. In addition, chloride concentrations and specific conductance levels in Geneva Lake are quite close to some of the data used to develop the piecewise regression model since one stream monitoring site was located on the White River (site number 48) a short distance downstream from the outlet of Geneva Lake. Because of this, samples collected at this river site often effectively measured chloride concentrations and specific conductance level within the lake.

These results suggest that the piecewise regression model performs poorly when applied to data from lakes and should not be used to estimate chloride concentrations from specific conductance data collected from lakes. As an alternative approach, Commission staff attempted to develop lake-specific regression models to use for estimating chloride concentrations from specific conductance in these six lakes.

## LAKE SPECIFIC REGRESSION DEVELOPMENT

Separate models were developed for each lake using simple linear regression on untransformed data. Statistical computations and significance testing were conducted using SYSTAT version 10.[47]

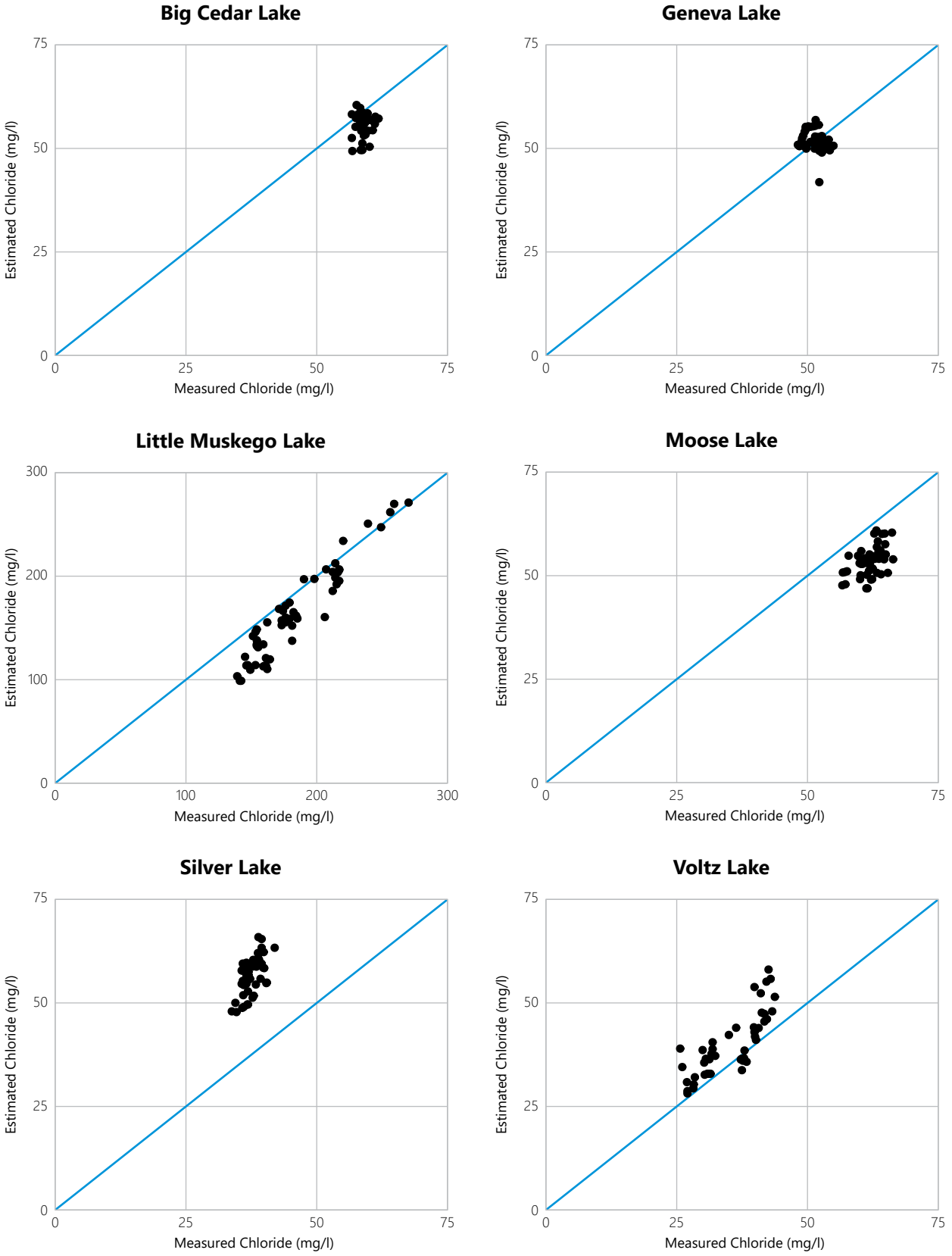
Figure B.2 shows the relationships between the level of specific conductance and chloride concentration in the six lakes. The green line indicates the linear regression model relationship between the two variables. The results of the linear regression analysis are given in Table B.1. The P-values given in the table indicate that statistically significant relationships between specific conductance and chloride concentration were found for four lakes: Little Muskego, Moose, Silver, and Voltz Lakes. The relationships between specific conductance and chloride concentration for Big Cedar and Geneva Lakes were not statistically significant. This suggests that specific conductance cannot be used to estimate chloride concentration for these two lakes using only the Chloride Impact Study dataset.

For the four lakes with significant regressions, the adjusted $R^2$ values shown in Table B.1 provide information on how well the level of specific conductance estimates chloride concentration. For two lakes, variability in specific conductance levels accounts for most of the variability in chloride concentration. The regression model for Little Muskego Lake has an $R^2$ of 0.924, indicating that variability in specific conductance accounts for over 92 percent of the variability in chloride concentration. This is a tight relationship and suggests that specific conductance is a good predictor of chloride concentration in Little Muskego Lake. The regression model for Voltz Lake has an $R^2$ of 0.647, indicating that variability in specific conductance accounts for almost two-thirds of the variability in chloride concentration. While the relationship in this lake is not as tight as the one in Little Muskego Lake, it does indicate that specific conductance is a reasonable predictor of chloride concentration in Voltz Lake.

By contrast, the adjusted $R^2$ values for the regression models for Moose and Silver Lakes indicate that specific conductance accounts for less than half the variability in chloride concentration in these lakes: 36 percent for Silver Lake and less than 20 percent for Moose Lake. This suggests that specific conductance is not as good a predictor of chloride concentration in these lakes as it is in Little Muskego Lake and Voltz Lake.

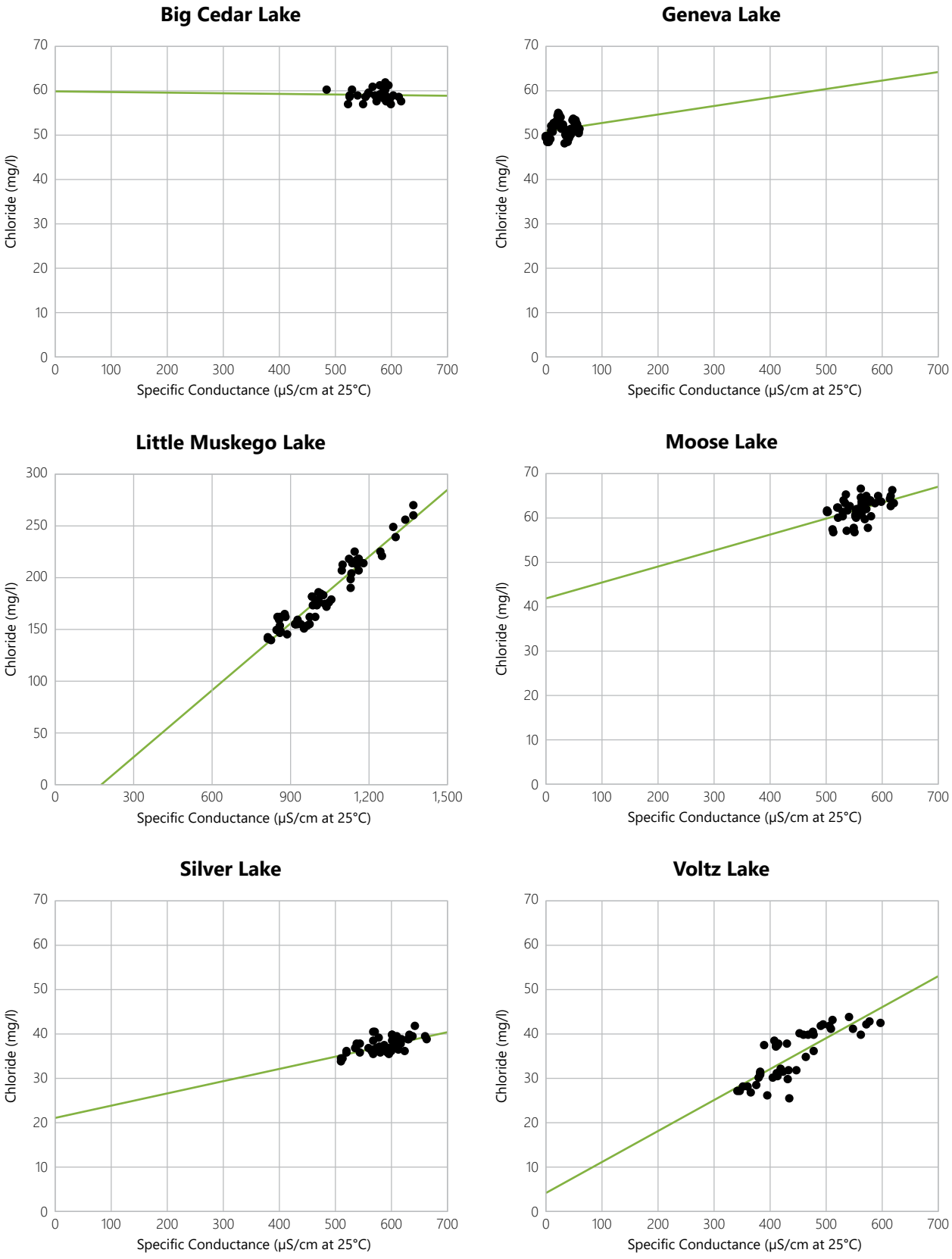
[47] *SYSTAT Software, Inc.,* SYSTAT version 10.2, *Richmond, California, 2002.*

**Figure B.1**
**Comparison of Chloride Concentrations Estimated from Specific Conductance to Measured Values in Six Lakes**




*Source: SEWRPC*

## Figure B.2
## Relationship Between Chloride Concentration and Specific Conductance in Six Lakes




*Source: SEWRPC*

**Table B.1**
**Site-Specific Regression Results for Lakes Sampled as a Part of the Chloride Impact Study[a]**

| Lake | Samples | Slope | y-Intercept | P-value[b] | Adjusted $R^2$ |
|---|---|---|---|---|---|
| Big Cedar Lake | 43 | -0.001 | 59.8 | 0.839 | 0.000 |
| Geneva Lake | 55 | -0.022 | 63.1 | 0.073 | 0.042 |
| Little Muskego Lake | 62 | 0.215 | -37.9 | <0.001 | 0.924 |
| Moose Lake | 52 | 0.036 | 41.7 | <0.001 | 0.195 |
| Silver Lake (Washington County) | 56 | 0.028 | 21.1 | <0.001 | 0.360 |
| Voltz Lake | 45 | 0.070 | 4.4 | <0.001 | 0.647 |

[a] *The regression equations take the form $\{Cl^-\} = m \times SC + b$, where $\{Cl^-\}$ is the concentration of chloride in mg/l, m is the slope, SC is the value of specific conductance in µS/cm, and b is the y-intercept.*

[b] *A regression equation is considered significant at a P-value less than 0.050.*

*Source: SEWRPC*

### Potential Sources of Differences in Regression Results Among Lakes

Differences in the ranges of specific conductance and chloride concentration observed in these lakes during data collection for the Chloride Impact Study might account for the differences in the significance and $R^2$ values of the regression models. Table B.2 shows the minimum and maximum values of specific conductance and chloride observed in these lakes during the quarterly sampling period from October 2018 through February 2021. The lakes where significant regression models could not be developed had narrow specific conductance ranges, about 134 microSiemens per centimeter (µS/cm) for Big Cedar Lake and 128 µS/cm for Geneva Lake. Similarly, the ranges over which chloride concentration varied in these lakes were also quite narrow: 5 milligrams per liter (mg/l) for Big Cedar Lake and 7 mg/l for Geneva Lake. For the most part, the two lakes with significant regression models in which specific conductance accounted for less than half the variability in chloride concentration had slightly wider ranges. The ranges of specific conductance in Moose Lake and Silver Lake were 119 µS/cm and 153 µS/cm, respectively, and the ranges of chloride were 9 mg/l and 8 mg/l, respectively. These ranges of specific conductance and chloride concentration might simply be too narrow for developing good regression models for estimating chloride concentrations.

By contrast, the ranges of specific conductance and chloride observed in the two lakes in which specific conductance accounted for more than half of the variability in chloride concentration were broader. The ranges of specific conductance in Little Muskego Lake and Voltz Lake were 558 µS/cm and 256 µS/cm, respectively, and the ranges of chloride were 131 mg/l and 18 mg/l, respectively. These wider ranges likely contributed to the greater amount of variation in chloride concentration accounted for by specific conductance in the respective regression models for these lakes.

The differences in the significance of and the amount of variability in chloride accounted for by the regression models may reflect differences in the characteristics of these six lakes. Specifically, the results of regression analysis correspond with differences in average water residence time in these lakes.[48] The lakes for which better regression models were developed have relative short residence times, 0.9 years for Little

Muskego Lake[49] and 2.2 years for Voltz Lake.[50] The lakes for which no statistically significant regression model could be developed had longer residence times—5.5 years for Big Cedar Lake[51] and 13.9 years for

[48] *Residence time is the number of years required for natural water sources under typical weather conditions to fill the lake one time. Natural water sources include runoff from surrounding areas, precipitation falling directly upon a lake, water entering from tributary streams, and water contributed to a lake by groundwater.*

[49] *SEWRPC Memorandum Report No. 155, 3rd Edition,* An Aquatic Plant Management Plan for Little Muskego Lake, Waukesha County, Wisconsin*, June 2019*.

[50] *SEWRPC Memorandum Report No. 159,* An Aquatic Plant Management Plan for Voltz Lake, Kenosha County, Wisconsin*, January 2005*.

[51] *SEWRPC Memorandum Report No. 137,* A Water Quality and Stormwater Management Plan for Big Cedar Lake, Washington County, Wisconsin*, August 2001*.

**Table B.2**
**Ranges of Specific Conductance and Chloride Data Used to Develop Site-Specific Regressions for Lakes**

| Lake | Minimum Specific Conductance (µS/cm) | Maximum Specific Conductance (µS/cm) | Minimum Chloride Concentration (mg/l) | Maximum Chloride Concentration (mg/l) |
|---|---|---|---|---|
| Big Cedar Lake | 485 | 619 | 57 | 62 |
| Geneva Lake | 460 | 588 | 48 | 55 |
| Little Muskego Lake | 813 | 1,371 | 139 | 270 |
| Moose Lake | 503 | 622 | 57 | 66 |
| Silver Lake (Washington County) | 511 | 664 | 34 | 42 |
| Voltz Lake | 342 | 598 | 26 | 44 |

*Source: SEWRPC*

Geneva Lake.[52] Silver Lake, for which specific conductance accounted for less than half of the variability in chloride concentration, has an intermediate residence time of 3.2 years.[53]

Water residence time has not been calculated for Moose Lake. Its residence time might be estimated by examining those of other seepage lakes in the Region. Lake Wandawega, a shallow seepage lake, has an average residence time of about 2.0 years.[54] Residence time for Lake Denoon, a deeper seepage lake, has been estimated to range between 3.1 and 4.2 years.[55] This suggests that a reasonable estimate for the residence time in Moose Lake might be between about two and four years. Since the volume of water contained in Lake Denoon is slightly less than that of Moose Lake,[56] it suggests that the residence time of water in Moose Lake may be toward the higher end of this estimated range. If the actual residence time is near the higher end of this range, this lake would also have a residence time in between those for which better regression models were developed and those for which no statistically significant model could be developed.

The results of these lake-specific regression analyses suggest that that the ease or difficulty of developing a lake-specific regression model for a lake will be influenced by its water residence time. Specific conductance levels and chloride concentrations are likely to be more stable in lakes with longer residence times than those with shorter residence times. In order to develop statistically significant regression models to estimate chloride concentration from specific conductance in lakes with longer residence times, it may be necessary to compile larger datasets based on sampling conducted over longer time periods. Even when this is done, the resulting models may not adequately estimate chloride concentrations because the percentage of variability in chloride concentration accounted for by specific conductance may be quite small.

[52] *SEWRPC Community Assistance Planning Report No. 60, 2nd Edition,* A Lake Management Plan for Geneva Lake, Walworth County, Wisconsin*, May 2008.*

[53] *SEWRPC Memorandum Report No. 123, 2nd Edition,* A Lake Protection and Recreational Use Plan for Silver Lake, Washington County, Wisconsin*, December 2005.*

[54] *SEWRPC Memorandum Report No. 175,* An Aquatic Plant Management Plan for Lake Wandawega, Walworth County, Wisconsin*, April 2009.*

[55] *SEWRPC Community Assistance Planning Report No. 155,* A Lake Management Plan for Lake Denoon, Racine and Waukesha Counties, Wisconsin*, December 2017.*

[56] *Lake Denoon has a volume of 2,940 acre-feet. Based on surface area of 87 acres and a mean depth of 40 feet, Moose Lake has a volume of about 3,480 acre-feet.*

## APPLICATIONS OF LAKE-SPECIFIC CONDUCTANCE/ CHLORIDE REGRESSIONS TO LAKE DATA

### Lakes Sampled for the Chloride Study

Based on the results of the regression analysis described in this appendix, two strategies should be applied for estimating chloride concentrations in the six lakes sampled as part of the Chloride Impact Study. Chloride concentrations for Little Muskego Lake and Voltz Lake can be estimated from specific conductance using the lake-specific regression equations presented in Table B.1. These models should give adequate estimates of chloride in these two lakes.

The regression models developed for Big Cedar, Geneva, Moose, and Silver Lakes are not adequate for estimating chloride concentrations from specific conductance, either because they are not statistically significant or because they account for a small portion of the variation in the data. The narrow ranges of chloride concentrations observed in these lakes over the period October 2018 through February 2021 indicate that chloride concentrations in these lakes are fairly stable over the short term (see Table B.2). This suggests that chloride concentrations at any point during the year could be estimated using the average concentration for that year. The adequacy of an estimate made this way could be checked by comparing specific conductance from a sample to the range for the lake that is shown in Table B.2. If specific conductance is within the given range, using the annual average chloride concentration will likely give an adequate estimate.

Because there is a general trend toward chloride concentrations increasing in southeastern Wisconsin lakes over time,[57] using the current average chloride concentration to estimate chloride concentration in these lakes will not remain a viable approach over the long term. Two approaches could be taken to address this problem. Because these methods can be applied to other lakes in addition to the six sampled as part of the Chloride Impact Study, they are described in the next section.

### General Methods for Estimating Chloride Concentration in Lakes

The discussion in the previous section suggests two approaches that could be used to develop estimates of chloride concentration for an individual lake. One approach would be to use recent and historical data that include paired specific conductance and chloride samples to develop a regression model for estimating chloride concentration from specific conductance. There are several issues that should be kept in mind when taking this approach. First, developing a good regression model will require that there be sufficient variability in the data. For lakes with longer residence times, this will likely require having a data record that spans several to many years. Second, to fully encompass the relationship between specific conductance and chloride, it would be best to have samples from several locations in the water column. Anoxia develops in the hypolimnia of many lakes in southeastern Wisconsin during summer stratification. This causes changes in the chemistry of sediment and hypolimnetic water. Such changes include the solubilization of some ions which are sequestered in sediment under oxic conditions. Because they contribute to the electrical conductivity of water, the entry of these ions into hypolimnetic water could potentially affect the relationship between specific conductance and chloride. Thus, having data that include samples from both surface water and deeper water would allow development of a more representative model of the relationship between specific conductance and chloride in a lake. Third, this approach assumes that the relationship between specific conductance and chloride concentration in a given lake does not change over time. Because the concentrations of other ions may also be changing over time, this may not always be a valid assumption.

An alternative approach would be to periodically sample chloride concentration in a lake throughout the year to establish an average concentration. This average could be used for a number of years to represent chloride concentration in the lake. While a better average might be derived by collecting water at several points of the water column, such sampling should at a minimum include collection of water from both the surface and near the lake bottom. The interval over which an estimate developed this way will remain valid will likely vary with the residence time of water in the lake. For lakes with short residence times such an estimate may be valid for only a couple years. On the other hand, an estimate developed for a lake with a longer residence time may remain valid for several years. When using this method to develop an estimate, it would be prudent to collect at least one sample annually to check the continued validity of the chloride concentration estimate.

[57] *See, for example, SEWRPC Technical Report No. 63,* Chloride Conditions and Trends in Southeastern Wisconsin*, in preparation.*