# Data 603 – Statistical Modelling with Data

# Final Report

## Calgary's Surface Water pH Level and Related Factors

December 1, 2024

by:
Arthur Sumague (UCID: 30121834)
Binh Nguyen (UCID: 302630057)
David Errington (UCID: 30269897)
Lalith Nandakumar (UCID: 30262992)
Tsz-Chuen Hui (Alvin) (UCID: 30240301)

**TABLE OF CONTENTS**

**TABLE OF FIGURES**



**TABLE OF TABLES**

# 1 INTRODUCTION

## 1.1 MOTIVATION

### *1.1.1 Overview & Context*

It is no secret that the health and well-being of our planet's ecosystems, and of all of us here on Earth, is greatly dependent on the quality of our water resources. Over the past 30 years or so, places like Canada have made great strides in improving their water quality. However, although Canada is ranked fourth among the top 17 OECD countries in the world, according to the Conference Board of Canada (2024), there are still environmental and human health problems related to water quality evident across the country.

Here in Calgary, we chose to focus our research on some of the main factors that affect water quality—specifically pH, water temperature, turbidity, conductivity, and dissolved oxygen—and explore the potential relationships that might exist between them. For those unfamiliar with these terms:

- **pH** is a measure of the acidity or alkalinity of a water body, ranging from 0 to 14.
- **Turbidity** is a measure of the level of particles such as sediment, plankton, or other organic by-products.
- **Conductivity** is a measure of water's ability to carry an electric current, which depends on the presence of ions.
- **Dissolved oxygen** is a measure of free (i.e., not chemically combined) oxygen dissolved in water.
- **Water Temperature** is a measure of the water's overall thermal energy.

Each of these factors is important and makes up what is often referred to as the Water Quality Index which was developed by the Yale Center for Environmental Law & Policy. Among these parameters, pH is especially important because it determines how safe our drinking water is. Just like our body needs to maintain the right pH balance, our drinking water needs the correct pH level to be safe for consumption and to protect our health.

## 1.2 OBJECTIVES

### *1.2.1 Goals & Research Questions*

The primary objective of our project's data and visual analytics is to examine the relationship between various water quality parameters and pH levels, and specifically address two key research questions:

i. How can we create a predictive model for pH levels in Calgary's surface water, and what key factors (i.e., dissolved oxygen, turbidity, and/or conductivity) significantly correlate with changes in pH?
ii. Are there any seasonal variations in pH levels?

## 1.3 SCIENTIFIC DISCUSSION ON DEPENDENT & INDEPENDENT VARIABLES

Let's first introduce the relationship between pH (dependent variable) and other independent variables.

### *1.3.1 pH and Oxygen*

The scientific literature has determined a correlation between water pH levels and oxygen concentrations.

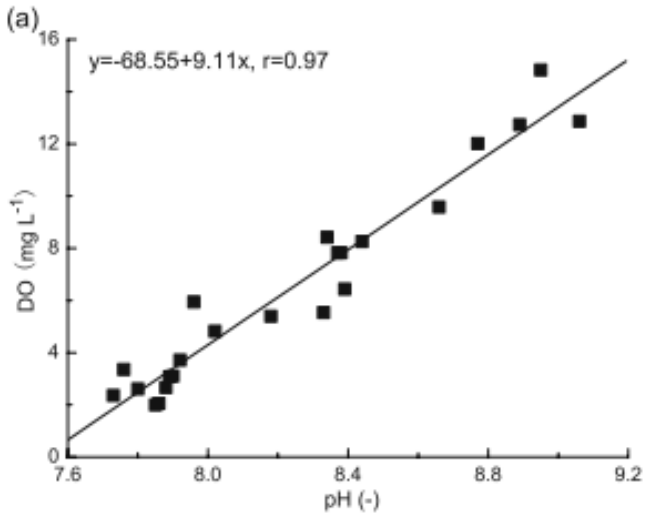


*Figure 1. Regression curves for pH and dissolved oxygen (DO) for enclosure 1. (C. Zang, 2011)*

The literature describes the water chemistry as pH decreases, the increased H+ ions react with the oxygen in the water, resulting in a decrease in DO. Therefore, simply by the chemical equilibrium between pH and DO, pH can generally have a positive linear relationship with DO (C. Zang, 2011).

### *1.3.2 pH and Conductivity*

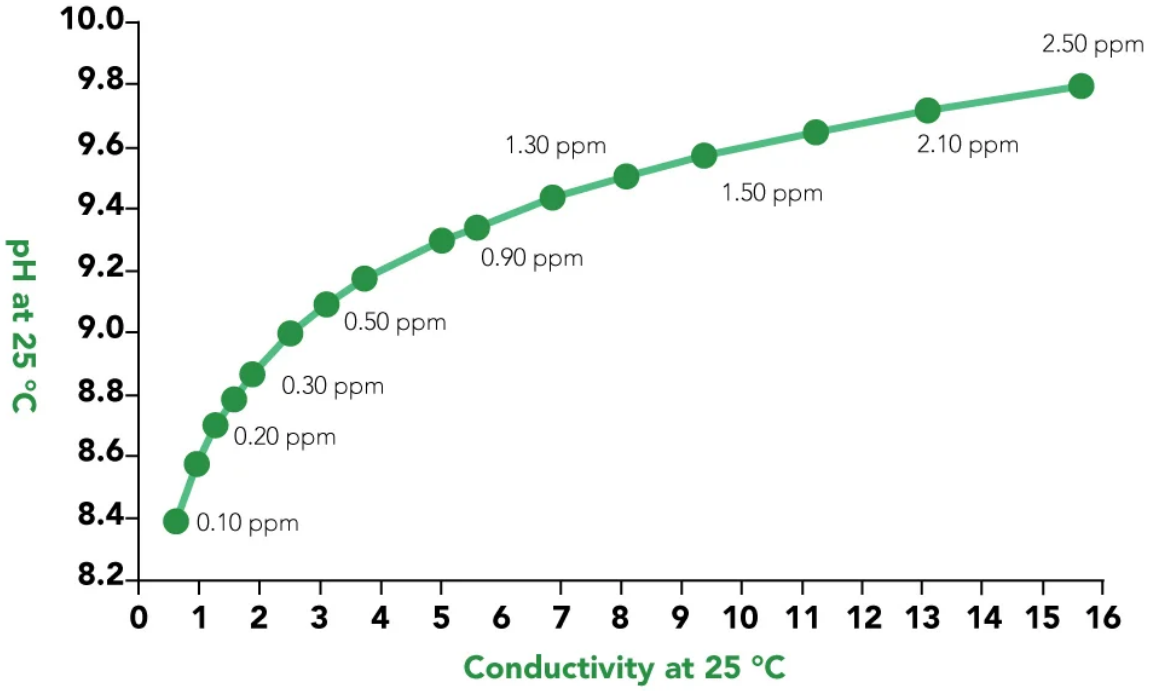


*Figure 2. What is the Relationship Between EC (electrical conductivity) & pH? (Atlas-Scientific, 2024)*

The above graph, from Atlas Scientific (2024), has determined a positive correlation between pH and conductivity. Their scientific reasoning behind this is that while at lower pH, they find a higher concentration of H+ ions, which takes up the space for other conductive cations (Ca2+, Ni2+, Mg2+..., etc.). In contrast to a higher pH, they find a higher concentration of OH–, and subsequently more conductive aforementioned cations to accompany the negative charge, thus increasing an aqueous solution's conductivity.

Additionally, water quality within Calgary is considered to be hardness due to Ca(2+) ions and Mg(2+) ions, which is the result of dissolved limestone in our water (City of Calgary, 2024).

### *1.3.3 pH and Water Temperature*

The formation of hydrogen ions (hydroxonium ions) and hydroxide ions from water is an endothermic process (absorbs energy). Because energy is being absorbed; they should find that with higher pH, they will see a decrease in temperature (temperature and pH are negatively correlated). This relationship is further supported by the open-source textbook LibreTexts Chemistry. Their theoretical calculations of pH of differing pure water temperatures are as published and are summarized here (Chem Libre Texts, 2024).

*Table 1. Temperature Dependence of the pH of pure Water (Chem Libre Texts, 2024).*

| T (°C) | $K_w$ ($mol^2$ $dm^{-6}$) | pH | pOH |
|---|---|---|---|
| 0 | $0.114 \times 10^{-14}$ | 7.47 | 7.47 |
| 10 | $0.293 \times 10^{-14}$ | 7.27 | 7.27 |
| 20 | $0.681 \times 10^{-14}$ | 7.08 | 7.08 |
| 25 | $1.008 \times 10^{-14}$ | 7.00 | 7.00 |
| 30 | $1.471 \times 10^{-14}$ | 6.92 | 6.92 |
| 40 | $2.916 \times 10^{-14}$ | 6.77 | 6.77 |
| 50 | $5.476 \times 10^{-14}$ | 6.63 | 6.63 |
| 100 | $51.3 \times 10^{-14}$ | 6.14 | 6.14 |

In natural environments, pH can be affected due to thermal stratification (colder water at the bottom, hotter water at the top due to fluidic densities). The stratification results in lower depths of the water body due to the lack of photosynthesis by plants and algae and respiration by organisms decomposing organic matter (because organisms don't want to be cold) (EPA United States Environmental Protection Agency, 2024).

### *1.3.4 pH and Seasons*

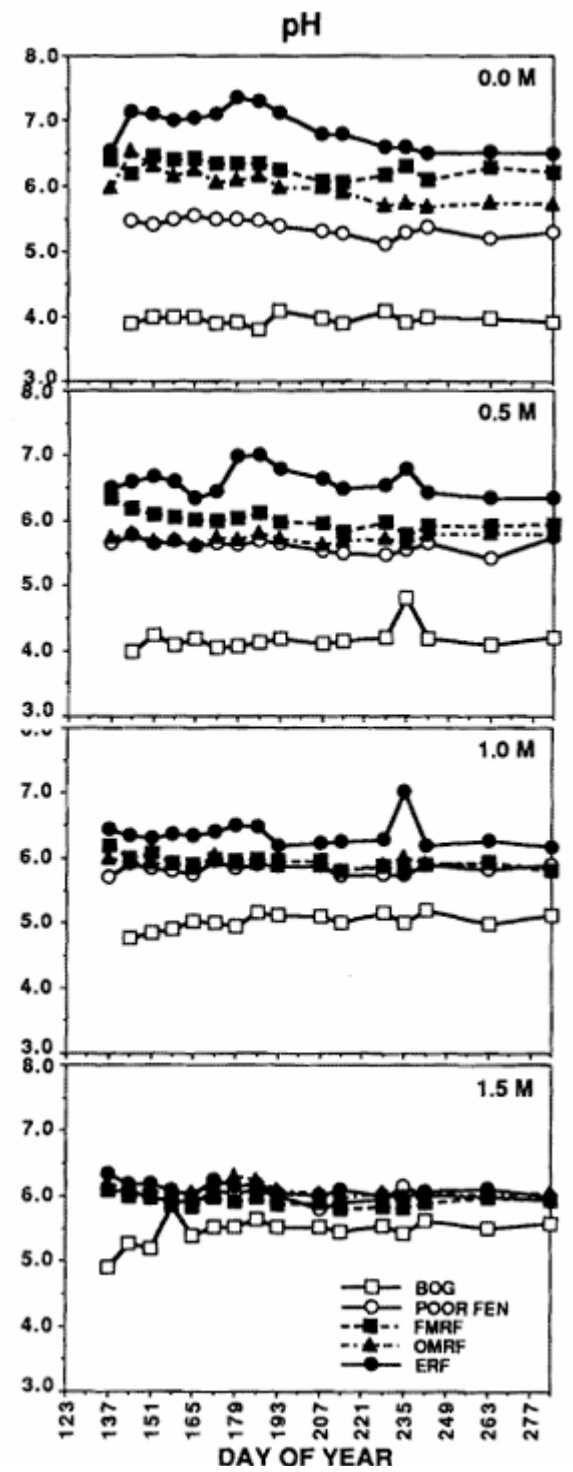


*Figure 3. Differences in pH in the various peatland types during the 1989 sampling period from (Dale H et. al, n.d.).*

A study on seasonal changes on differing surface bodies of water in Alberta concluded that seasonality shows minimal changes on pH across the seasons. This suggests that the pH levels are stable throughout the year, showing no significant fluctuations between seasons. The slight dip in pH during winter might be due to reduced biological activity, such as less decomposition or organic matter processing, and potentially lower photosynthetic activity, which influences carbon dioxide levels in water. The slight rise in pH in spring could result from increased biological activity, including algae blooms or increased photosynthesis, which consume carbon dioxide and temporarily elevate pH levels. Additionally, seasonality in the Alberta bodies of water is known to influence other factors such as conductivity because as the snow melts, water becomes diluted, (City of Calgary, 2024). lowering the conductivity. Seasonality may be a significant influential factor to other independent variables, which garners its importance in this study.

In summary, the scientific research discussed above highlights the following relationship:

- pH and Oxygen: A positive linear relationship exists between pH and dissolved oxygen.
- pH and Conductivity: pH and conductivity are positively correlated.
- pH and Water Temperature: Temperature and pH are negatively correlated.
- pH and Seasons: Seasonal changes minimally impact pH, which remains stable year-round.

# 2 METHODOLOGY

## 2.1 DATASET

This dataset was provided by the City of Calgary's in situ (on-site) water quality monitoring program, which collects water parameters in real-time using multi-parameter sondes in surface waters. The dataset analyzed in this project spans from 2019-11-26 11:00:00 to 2024-09-05 10:15:00. The raw data comprises approximately 1.8 GB, containing over 15 million records.

### *2.1.1 Collection Method*

Data collection employed two primary methodologies: continuous and discrete. In the continuous method, data is logged by sondes at 15-minute intervals over extended periods. The discrete method involves individual measurements manually recorded by field technicians. This study focuses on data collected by sondes in surface waters only (i.e. the continuous method) to minimize human error during the process.

### *2.1.2 Details of the Data*

The monitored parameters include water temperature, specific conductance, pH, dissolved oxygen, and turbidity. Details are as follows:

| Column | Description | Measures |
|---|---|---|
| ID | Unique identifier for each data point | Numeric |
| Sample Site | Location of water sample collection | • Bow River<br>• Elbow River<br>• Fish Creek<br>• Nose Creek<br>• Pine Creek<br>• Shepard Wetland<br>• West Nose Creek |

| Column | Description | Measures |
|---|---|---|
| Sample Date | Date and time of sample collection | Format: `YYYY-MM-DD HH:MM:SS` |
| Parameter | Physical characteristic measured | • Conductivity<br>• Dissolved Oxygen Concentration<br>• pH<br>• Temperature<br>• Turbidity |
| Result Units | Measurement units for each parameter | • Conductivity : µS/cm<br>• Dissolved Oxygen Concentration : mg/L<br>• pH : pH units<br>• Temperature : °C<br>• Turbidity Units : NTU |
| Field Data Description | Method of data collection | • Continuous: Automated logging at 15-minute intervals<br>• Discrete: Manual collection by field technicians |
| Numeric Result | Measured value of the parameter | Numeric (floating point) |

### *2.1.3 License*

This dataset is publicly accessible through the City of Calgary's Open Data Portal in multiple formats (CSV, RDF, RSS, etc.) Under the Open Government License for the City of Calgary, users are permitted to "Copy, modify, publish, translate, adapt, distribute or otherwise use [this] Information in any medium, mode or format for any lawful purpose" (Open Calgary Terms of Use, 2024).

## 2.2 APPROACH

The main programming for this project is `R` and most of the development was done in `RStudio`, which is a free, open-source programming language well-suited for statistical analysis and data visualization.

In addition to `R` and `RStudio`, the following libraries were used in our analytic process and model building:

| Library | Purpose |
|---|---|
| `olsrr` | Used to perform stepwise regression analysis. |
| `ggplot2` | Used to create a scatter plot for correlation analysis. |
| `stringr` | Used on data cleaning and wrangling |
| `lubridate` | Used on data cleaning and wrangling on data-time objects. |
| `mctest` | Used to perform multicollinearity check using VIF |
| `lmtest` | Used to perform equal variance Assumption check using `bptest` |

| Library | Purpose |
|---|---|
| `psych` | Used to create correlation charts. |

### 2.3 WORKFLOW

- Data cleaning and wrangling process
    - Filter out records that were collected manually. This study focuses on data collected by devices only to minimize human error during the process.
    - Filter out non-informative columns to reduce the overall data size.
    - Pivot the dataset so that all predictors are placed in different columns of the same rows.
    - Standardize column names to facilitate subsequent data processing.
    - Filter out rows if NA values are found in the predictors.
    - Remove outliers in the pH measurement.
    - Create a new column called “season” based on the date field. This new column will be used as a categorical predictor.
- Simple linear regression
- Pre-modelling testing
    - Checking multicollinearity (VIF)
    - Plotting to check the linearity between the response (pH) and predictors (water_temp, conductivity, oxygen, turbidity)
- Multiple linear regression
    - First-order model
    - The predictor selection (stepwise selection and all-possible-regressions selection procedure)
    - Interaction model
    - Correlation matrix
    - Polynomial model
- Checking assumptions
    - Linearity assumption (residuals plot)
    - Independence assumption (no date & time data used in the project)
    - Equal variance assumption (the Breusch-Pagan test)
    - Normality assumption (Kolmogorov-Smirnov test used as the sample size larger than 5,000 observations)
    - Transformations for nonnormality and heteroscedasticity (log-transformation)
- Prediction
    - The model was trained on data from November 2019 to September 2024, and the model was tested on data from September to October 2024.

### 2.4 CONTRIBUTIONS

The project work was distributed among team members based on their expertise and project requirements. While each member had primary responsibilities, our team maintained continuous collaboration through regular meetings to make sure the project was on the right track. The following table details each member's key responsibilities:

| Group Member | Roles and Responsibilities |
|---|---|
| Tsz Chuen Hui (Alvin) | • Led data cleaning, filtering, and pivoting.<br>• Performed model validation to improve data quality for better predictions.<br>• Contributed to report writing and performed quality assurance reviews. |
| Binh Nguyen | • Initiated model development<br>• Carried out model validation to realize the best performance and reliability.<br>• Contributed to report writing and performed quality assurance reviews. |
| David Errington | • Validated the model for overfitting and accuracy, interpreted the model<br>• Visualized results through plots and graphs.<br>• Contributed to report writing and drafted initial documentation |
| Lalith Nandakumar | • Took part in data cleaning tasks like standardization, performed model validation to ensure correctness of the model<br>• Assessed model's performance via ability to predict test data.<br>• Contributed to report writing and drafted initial documentation |
| Arthur Sumague | • Performed EDA on the dataset, provided models' validation contributions<br>• Fine-tuned data insights toward model accuracy.<br>• Contributed to report writing and drafted initial documentation |

# 3 MAIN RESULTS OF THE ANALYSIS

## 3.1 DEPENDENT & INDEPENDENT VARIABLES

In this study, the multiple linear regression was used as a predictive model for pH levels in Calgary's surface water. Details of the dependent and independent variables are listed below.

*Table 2. Variable description of the water dataset. Indicates our dependent (response) and independent variables, and descriptions of each.*

| Column | Description | Measures |
|---|---|---|
| *Dependent Variable* | | |
| pH | Unique identifier for each data point | Numeric |
| *Independent Variable(s)* | | |
| Season | Season of the year | Categorical |
| Temperature | Temperature of the water | Numeric |
| Conductivity | Electrical conductivity of the water for ion detection | Numeric |
| Turbidity | Transparency of the water for any suspended particles | Numeric |
| Oxygen | Dissolved Oxygen Concentration | Numeric |

### *3.1.1 Exploratory Data Analysis (EDA) of our Variables*

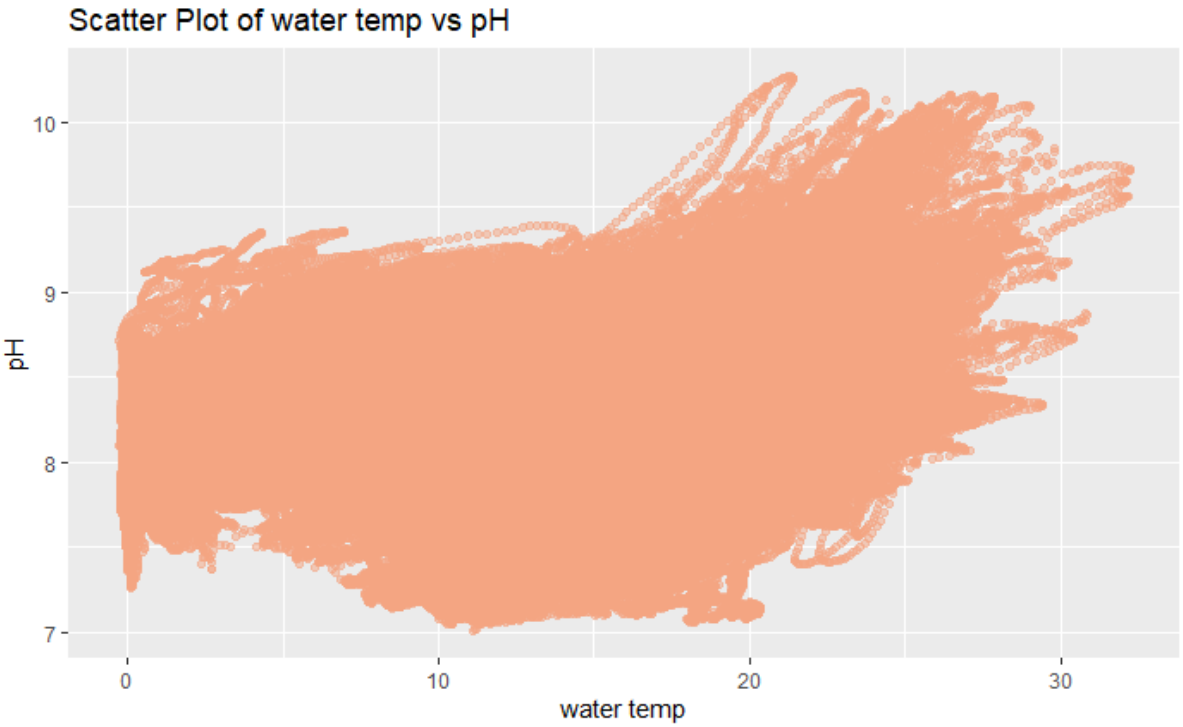


*Figure 4. Scatter plot of Water Temperature vs. pH of water dataset.*

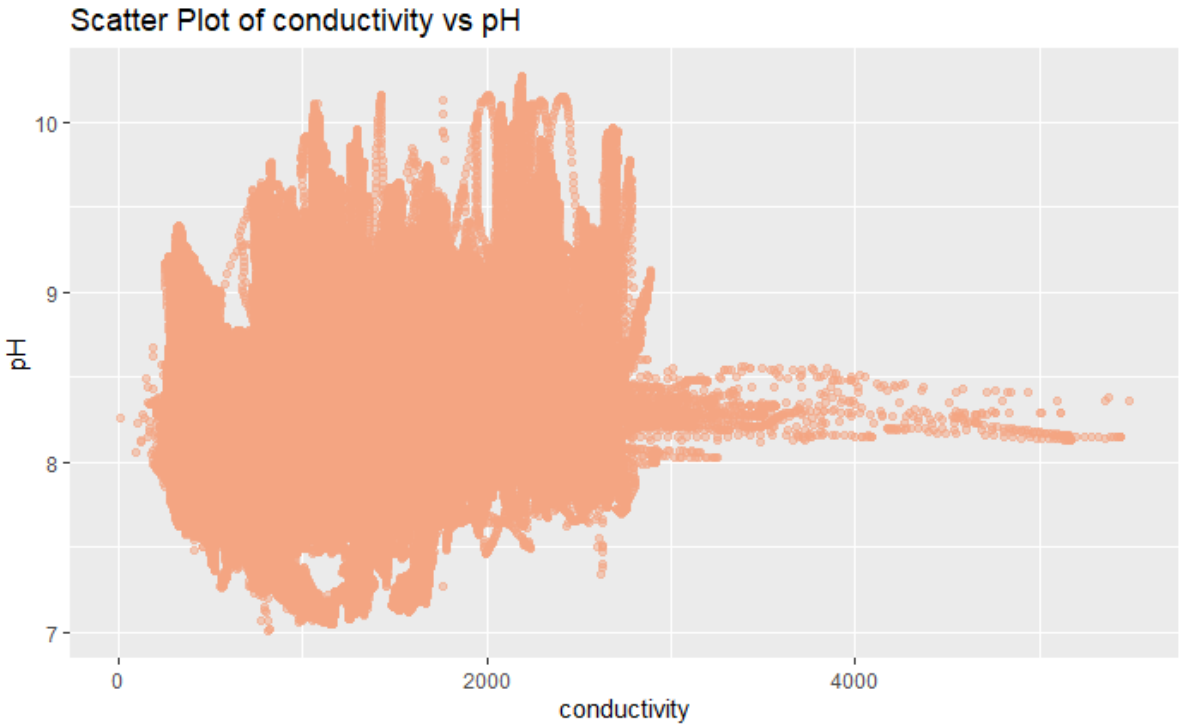


*Figure 5. Scatter plot of Conductivity vs. pH of water dataset.*

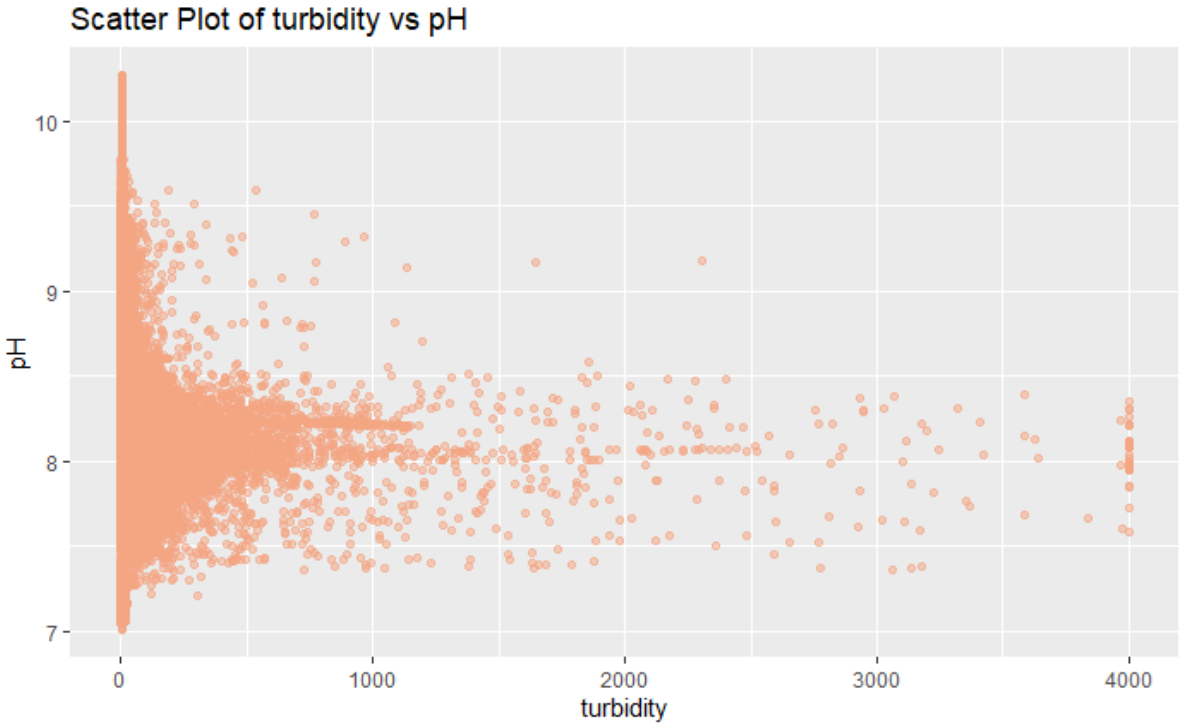


*Figure 6. Scatter plot of turbidity vs. pH of water dataset.*

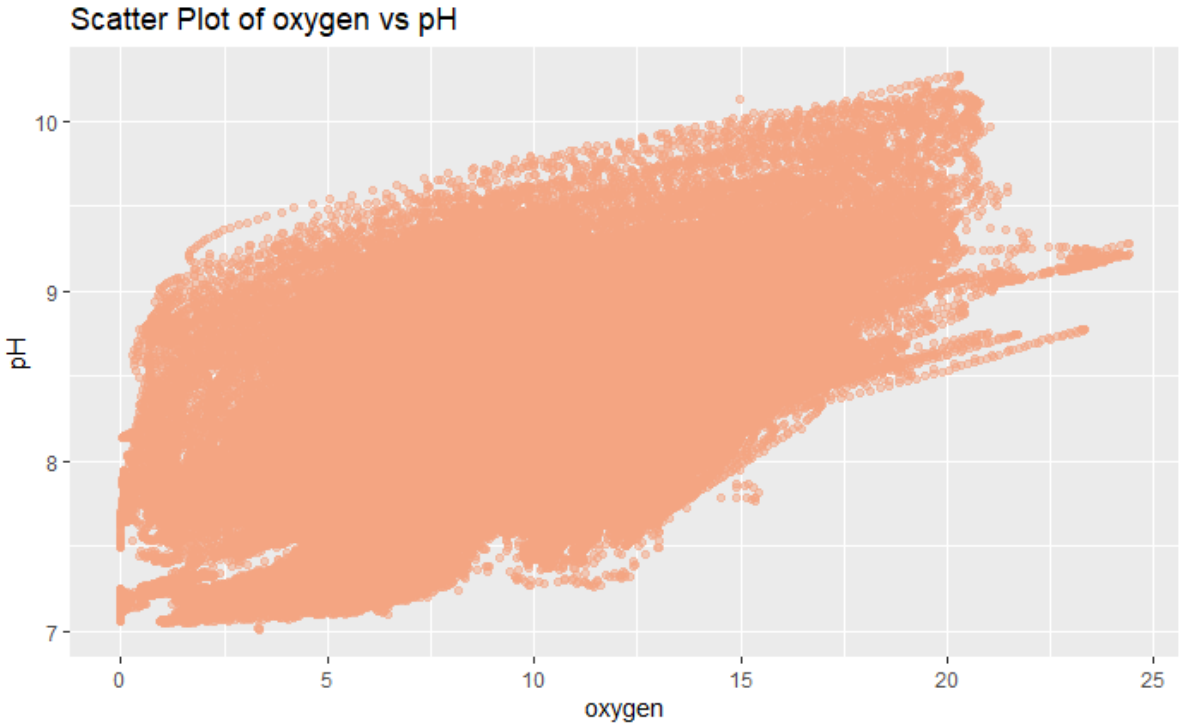


*Figure 7. Scatter plot of oxygen vs. pH of water dataset.*

## 3.2 FIRST ORDER MODEL

### *3.2.1 Univariate Linear Regression Model*

The project was initiated with a single independent variable, water temperature.

$$\widehat{\text{pH}} = 8.178 + 9.799 \times 10^{-3} \cdot \text{water\_temp}$$

The resulting model has an adjusted R-squared of 0.0517 (5.17%) and a residual standard error of 0.2907. This model is insubstantial for modelling because only 5.17% of the variance of pH is explained by the model. So, we moved forward with multivariate linear regression modelling.

### *3.2.2 Multivariate Linear Regression Model*

We initially modelled our data using the first-order additive model. The resulting model:

$$\widehat{\text{pH}}_{\text{Season}} = \begin{cases} 6.938 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Autumn} \\ 6.938 + 0.07250 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Spring} \\ 6.938 - 0.002923 + 0.02813 \cdot \text{wate\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Summer} \\ 6.938 - 0.07844 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Winter} \end{cases}$$

$$= \begin{cases} 6.938 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Autumn} \\ 7.0105 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Spring} \\ 6.935077 + 0.02813 \cdot \text{wate\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Summer} \\ 6.85956 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Winter} \end{cases}$$

### *3.2.3 Interpretation*

Provided us with coefficients that are significantly different from 0 ($\alpha = 0.05 >$ p-value) using individual t-tests. The following model provided us with an adjusted R-squared 0.509 with 0.2092 residual standard error. Thus far, our model explains 50.9% of the variation we observe within our dependent variable, pH. The explanatory variables included are water temperature, conductivity, turbidity, oxygen, and season as a dummy variable. Therefore, we will continue onwards with these variables for further modelling methods.

## 3.3 STEPWISE REGRESSION PROCEDURE

To confirm our resulting model from t-tests method, we modelled using stepwise modelling as well from `olsrr` package; `ols_step_both_p()`. We opted for this function specifically because we thought it would be the most stringent and strict method to ensure that all the independent variables included in our model are significant contributors. This method also ensures that for each added independent variable for each step, it checks the previous step's independent variable for significance. This method upholds the standard of significance which we favoured to ensure significance in our model, despite the higher rate of Type I error. The resulting model we achieved through this method is:

$$\widehat{\text{pH}}_{\text{Season}} = \begin{cases} 6.938 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Autumn} \\ 6.938 + 0.07250 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Spring} \\ 6.938 - 0.002923 + 0.02813 \cdot \text{wate\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Summer} \\ 6.938 - 0.07844 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Winter} \end{cases}$$

$$= \begin{cases} 6.938 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Autumn} \\ 7.0105 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Spring} \\ 6.935077 + 0.02813 \cdot \text{wate\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Summer} \\ 6.85956 + 0.02813 \cdot \text{water\_temp} + 7.323 \times 10^{-5} \cdot \text{conductivity} - 2.137 \times 10^{-4} \cdot \text{turbidity} + 0.09707 \cdot \text{oxygen} & \text{if Winter} \end{cases}$$

which is identical from the t-tests method.

### *3.3.1 Interpretation*

Additionally, with the stepwise method for modelling, the function provides for us a hierarchy of independent variables towards their significance on contributing to the effect of the dependent variable. We find that oxygen is the most contributing, with a t-value of 1049.963 and turbidity the least with a t-value of -53.85. The full output can be viewed, with the most significant at the top, descending.

*Table 3. Stepwise output and assessment of each independent variable as it was sequentially added to a full additive model using* `ols_step_both_p()` *function.*

```
                 Estimate  Std. Error    t-value       p-Value
(Intercept)     6.938e+00   1.235e-03   5619.871   < 2e-16 ***
oxygen          9.707e-02   9.245e-05   1049.963   < 2e-16 ***
water_temp      2.813e-02   4.572e-05    615.257   < 2e-16 ***
conductivity    7.323e-05   3.877e-07    188.899   < 2e-16 ***
seasonSpring    7.250e-02   5.357e-04    135.327   < 2e-16 ***
seasonSummer   -2.923e-03   7.087e-04     -4.125   < 2e-16 ***
seasonWinter   -7.844e-02   8.498e-04    -92.298   < 2e-16 ***
turbidity      -2.137e-04   3.968e-06    -53.850   < 2e-16 ***
---

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 3.4 ALL-POSSIBLE-REGRESSIONS SELECTION

Another method of ensuring that we correctly chose the additive first-order model was to assess via the Akaike information criterion (AIC), Mallow's Cp Criterion (Cp) and Adjusted R-squared.

### *3.4.1 Interpretation*

The screening criteria for all possible regressions are outlined below.

*Table 4. Consolidated output of R-squared, adjusted R-squared, Mallow's (Cp) Criterion & Akaike information criterion (AIC).*

| Num. of Variables | R-squared | Adjusted R-squared | Cp | AIC |
|---|---|---|---|---|
| 1 | 0.1595740 | 0.1595733 | 942473.750 | 326299.5 |
| 2 | 0.4703773 | 0.4703765 | 104110.300 | -285282.1 |
| 3 | 0.4948982 | 0.4948963 | 37973.263 | -348064.2 |
| 4 | 0.5078993 | 0.5078970 | 2905.853 | -382601.0 |
| 5 | 0.5089743 | 0.5089717 | 8.000 | -385495.7 |

As observed, with 5 variables within our model; we maximize Adjusted R-squared, Cp is closest to $k + 1$ terms ($5 + 1 = 6$), and AIC is the least.

### 3.5 INTERACTION MODEL

The resulting interaction is as follows:

$$\widehat{\text{pH}_i} = \begin{cases} 6.234 + 0.04633X_{\text{water temp}_i} + 6.274 \times 10^{-4}X_{\text{conductivity}_i} + 3.201 \times 10^{-3}X_{\text{turbidity}_i} + 0.1515X_{\text{oxygen}_i} \\ \quad -1.278 \times 10^{-5}X_{\text{water temp}_i}X_{\text{conductivity}_i} + 2.818 \times 10^{-4}X_{\text{water temp}_i}X_{\text{oxygen}_i} \\ \quad -5.052 \times 10^{-8}X_{\text{conductivity}_i}X_{\text{turbidity}_i} - 4.617 \times 10^{-5}X_{\text{conductivity}_i}X_{\text{oxygen}_i} \\ \quad +1.500 \times 10^{-5}X_{\text{turbidity}_i}X_{\text{oxygen}_i} & \text{(Fall)} \\ \\ 6.30181 + 0.04232X_{\text{water temp}_i} + 7.1419 \times 10^{-4}X_{\text{conductivity}_i} - 4.42 \times 10^{-4}X_{\text{turbidity}_i} \\ \quad +0.13921X_{\text{oxygen}_i} - 1.278 \times 10^{-5}X_{\text{water temp}_i}X_{\text{conductivity}_i} \\ \quad +2.818 \times 10^{-4}X_{\text{water temp}_i}X_{\text{oxygen}_i} + 8.679 \times 10^{-5}X_{\text{conductivity}_i}X_{\text{turbidity}_i} \\ \quad -3.643 \times 10^{-3}X_{\text{turbidity}_i}X_{\text{oxygen}_i} & \text{(Spring)} \\ \\ 6.31933 + 0.03581X_{\text{water temp}_i} + 7.1544 \times 10^{-4}X_{\text{conductivity}_i} + 2.88 \times 10^{-4}X_{\text{turbidity}_i} \\ \quad +0.142165X_{\text{oxygen}_i} - 1.278 \times 10^{-5}X_{\text{water temp}_i}X_{\text{conductivity}_i} \\ \quad +2.818 \times 10^{-4}X_{\text{water temp}_i}X_{\text{oxygen}_i} - 2.913 \times 10^{-3}X_{\text{turbidity}_i}X_{\text{oxygen}_i} & \text{(Summer)} \\ \\ 5.934 + 0.14693X_{\text{water temp}_i} - 4.80 \times 10^{-4}X_{\text{conductivity}_i} - 2.916 \times 10^{-3}X_{\text{turbidity}_i} \\ \quad +0.19873X_{\text{oxygen}_i} - 1.278 \times 10^{-5}X_{\text{water temp}_i}X_{\text{conductivity}_i} \\ \quad +2.818 \times 10^{-4}X_{\text{water temp}_i}X_{\text{oxygen}_i} + 1.500 \times 10^{-5}X_{\text{turbidity}_i}X_{\text{oxygen}_i} & \text{(Winter)} \end{cases}$$

With a resulting adjusted R-squared of 0.6022 and a residual standard error of 0.1555.

#### *3.5.1 Interpretation*

All interaction terms were found to be significant ($\alpha = 0.05 >$ p-value) and are included in the model. The model improved slightly (~4% increase in adjusted R-squared from the additive model). The model has an adjusted R-squared 0.5369 with 0.2032 residual standard error. The model explains 53.69% of the variation found in the dependent variable, pH. Although not a substantial improvement, we continued with this model as our best attempt to get an accurate model for the data.

### 3.6 HIGHER ORDER MODEL

Based on the results of the correlation analysis (Table 5), oxygen showed the highest potential for a higher-order relationship with the dependent variable, pH. Therefore, oxygen was selected to build the higher-order model.
Due to computational limitations and the size of our dataset, we could not provide graphs using `ggally` however, we were able to provide correlation coefficients.

*Table 5. Correlation Matrix of our five variables.*

| | pH | water temperature | conductivity | turbidity | oxygen |
|---|---|---|---|---|---|
| **pH** | 1.0000000 | 0.22738713 | 0.14974224 | -0.05816150 | 0.39946711 |
| **water temp** | 0.2273871 | 1.00000000 | 0.15858746 | 0.01151997 | -0.57378283 |
| **conductivity** | 0.1497422 | 0.15858746 | 1.00000000 | 0.03881448 | -0.11869078 |
| **turbidity** | -0.0581615 | 0.01151997 | 0.03881448 | 1.00000000 | -0.05882457 |
| **oxygen** | 0.3994671 | -0.57378283 | -0.11869078 | -0.05882457 | 1.00000000 |

Therefore, the resulting second-order model for oxygen to the power of 2 ($X_{\text{oxygen}}^2$) is:

$$\widehat{pH} = \begin{cases} 5.974 + 0.04906X_{\text{water temp}} + 4.800 \times 10^{-4}X_{\text{conductivity}} + 1.895 \times 10^{-3}X_{\text{turbidity}} + 0.1982X_{\text{oxygen}} \\ -1.816 \times 10^{-3}X_{\text{oxygen}}^2 - 7.812 \times 10^{-6}X_{\text{water temp}}X_{\text{conductivity}} - 2.555 \times 10^{-5}X_{\text{water temp}}X_{\text{turbidity}} \\ -4.983 \times 10^{-4}X_{\text{water temp}}X_{\text{oxygen}} + 1.045 \times 10^{-7}X_{\text{conductivity}}X_{\text{turbidity}} - 3.510 \times 10^{-5}X_{\text{conductivity}}X_{\text{oxygen}} \\ -1.594 \times 10^{-4}X_{\text{turbidity}}X_{\text{oxygen}} \quad \text{(Fall)} \\ \\ 6.236 + 0.04906X_{\text{water temp}} + 4.800 \times 10^{-4}X_{\text{conductivity}} + 1.895 \times 10^{-3}X_{\text{turbidity}} + 0.1982X_{\text{oxygen}} \\ -1.816 \times 10^{-3}X_{\text{oxygen}}^2 - 7.812 \times 10^{-6}X_{\text{water temp}}X_{\text{conductivity}} - 2.555 \times 10^{-5}X_{\text{water temp}}X_{\text{turbidity}} \\ -4.983 \times 10^{-4}X_{\text{water temp}}X_{\text{oxygen}} + 1.045 \times 10^{-7}X_{\text{conductivity}}X_{\text{turbidity}} - 3.510 \times 10^{-5}X_{\text{conductivity}}X_{\text{oxygen}} \\ -1.594 \times 10^{-4}X_{\text{turbidity}}X_{\text{oxygen}} + 0.2624 - 8.893 \times 10^{-3}X_{\text{water temp}} \\ +6.680 \times 10^{-5}X_{\text{conductivity}} - 6.210 \times 10^{-4}X_{\text{turbidity}} - 1.578 \times 10^{-2}X_{\text{oxygen}} \quad \text{(Spring)} \\ \\ 6.130 + 0.04906X_{\text{water temp}} + 4.800 \times 10^{-4}X_{\text{conductivity}} + 1.895 \times 10^{-3}X_{\text{turbidity}} + 0.1982X_{\text{oxygen}} \\ -1.816 \times 10^{-3}X_{\text{oxygen}}^2 - 7.812 \times 10^{-6}X_{\text{water temp}}X_{\text{conductivity}} - 2.555 \times 10^{-5}X_{\text{water temp}}X_{\text{turbidity}} \\ -4.983 \times 10^{-4}X_{\text{water temp}}X_{\text{oxygen}} + 1.045 \times 10^{-7}X_{\text{conductivity}}X_{\text{turbidity}} - 3.510 \times 10^{-5}X_{\text{conductivity}}X_{\text{oxygen}} \\ -1.594 \times 10^{-4}X_{\text{turbidity}}X_{\text{oxygen}} + 0.1568 - 3.416 \times 10^{-3}X_{\text{water temp}} \\ +6.754 \times 10^{-5}X_{\text{conductivity}} - 4.797 \times 10^{-4}X_{\text{turbidity}} - 2.061 \times 10^{-2}X_{\text{oxygen}} \quad \text{(Summer)} \\ \\ 5.559 + 0.04906X_{\text{water temp}} + 4.800 \times 10^{-4}X_{\text{conductivity}} + 1.895 \times 10^{-3}X_{\text{turbidity}} + 0.1982X_{\text{oxygen}} \\ -1.816 \times 10^{-3}X_{\text{oxygen}}^2 - 7.812 \times 10^{-6}X_{\text{water temp}}X_{\text{conductivity}} - 2.555 \times 10^{-5}X_{\text{water temp}}X_{\text{turbidity}} \\ -4.983 \times 10^{-4}X_{\text{water temp}}X_{\text{oxygen}} + 1.045 \times 10^{-7}X_{\text{conductivity}}X_{\text{turbidity}} - 3.510 \times 10^{-5}X_{\text{conductivity}}X_{\text{oxygen}} \\ -1.594 \times 10^{-4}X_{\text{turbidity}}X_{\text{oxygen}} - 0.4149 + 2.363 \times 10^{-2}X_{\text{water temp}} \\ -3.393 \times 10^{-5}X_{\text{conductivity}} - 2.229 \times 10^{-4}X_{\text{turbidity}} + 2.777 \times 10^{-2}X_{\text{oxygen}} \quad \text{(Winter)} \end{cases}$$

### *3.6.1 Interpretation*

The second-order model does show the significance for the second order of oxygen ($\alpha$ = 0.05 > p-value), however, the adjusted R-squared is 0.5397, an improvement of 0.0307 (3.07%).

## 3.7 FINAL MODEL

The final model we determined contained the following independent variables: water temperature, conductivity, turbidity, oxygen and seasons.

These independent variables were verified via significance from individual t-tests method and through stepwise modelling. Additionally, we found significant interaction between all the independent variables, except for water temperature and turbidity. We decided to omit higher-order variables because the improvement of the model was minimal (Adjusted R-Squared value improved by less than 5%) and would thus result in overfitting the model.

Here is a generalized equation for our model:

$$\begin{aligned}\widehat{\text{pH}_i} &= \beta_0 + \beta_1 X_{\text{water temp}_i} + \beta_2 X_{\text{conductivity}_i} + \beta_3 X_{\text{turbidity}_i} + \beta_4 X_{\text{oxygen}_i} \\ &+\beta_5 X_{\text{water temp}_i} X_{\text{conductivity}_i} + \beta_6 X_{\text{water temp}_i} X_{\text{turbidity}_i} + \beta_7 X_{\text{water temp}_i} X_{\text{oxygen}_i} \\ &+\beta_8 X_{\text{conductivity}_i} X_{\text{turbidity}_i} + \beta_9 X_{\text{conductivity}_i} X_{\text{oxygen}_i} + \beta_{10} X_{\text{turbidity}_i} X_{\text{oxygen}_i} \\ &+\beta_{11} \text{Spring}_i + \beta_{12} \text{Summer}_i + \beta_{13} \text{Winter}_i \\ &+\beta_{14} \text{Spring}_i X_{\text{water temp}_i} + \beta_{15} \text{Spring}_i X_{\text{conductivity}_i} + \beta_{16} \text{Spring}_i X_{\text{turbidity}_i} + \beta_{17} \text{Spring}_i X_{\text{oxygen}_i} \\ &+\beta_{18} \text{Summer}_i X_{\text{water temp}_i} + \beta_{19} \text{Summer}_i X_{\text{conductivity}_i} + \beta_{20} \text{Summer}_i X_{\text{turbidity}_i} + \beta_{21} \text{Summer}_i X_{\text{oxygen}_i} \\ &+\beta_{22} \text{Winter}_i X_{\text{water temp}_i} + \beta_{23} \text{Winter}_i X_{\text{conductivity}_i} + \beta_{24} \text{Winter}_i X_{\text{turbidity}_i} + \beta_{25} \text{Winter}_i X_{\text{oxygen}_i}\end{aligned}$$

And the following equation with sub-models is:

$$\widehat{pH} = \begin{cases} 6.200 + 0.04852 X_{\text{water temp}} + 4.797 \times 10^{-4} X_{\text{conductivity}} + 1.696 \times 10^{-3} X_{\text{turbidity}} + 0.1574 X_{\text{oxygen}} \\ -8.619 \times 10^{-6} X_{\text{water temp}} X_{\text{conductivity}} - 2.564 \times 10^{-5} X_{\text{water temp}} X_{\text{turbidity}} \\ -4.332 \times 10^{-4} X_{\text{water temp}} X_{\text{oxygen}} + 1.552 \times 10^{-7} X_{\text{conductivity}} X_{\text{turbidity}} \\ -3.528 \times 10^{-5} X_{\text{conductivity}} X_{\text{oxygen}} - 1.444 \times 10^{-4} X_{\text{turbidity}} X_{\text{oxygen}} \quad \text{(Fall)} \\ \\ 6.461 + 0.04852 X_{\text{water temp}} + 4.797 \times 10^{-4} X_{\text{conductivity}} + 1.696 \times 10^{-3} X_{\text{turbidity}} + 0.1574 X_{\text{oxygen}} \\ -8.619 \times 10^{-6} X_{\text{water temp}} X_{\text{conductivity}} - 2.564 \times 10^{-5} X_{\text{water temp}} X_{\text{turbidity}} \\ -4.332 \times 10^{-4} X_{\text{water temp}} X_{\text{oxygen}} + 1.552 \times 10^{-7} X_{\text{conductivity}} X_{\text{turbidity}} \\ -3.528 \times 10^{-5} X_{\text{conductivity}} X_{\text{oxygen}} - 1.444 \times 10^{-4} X_{\text{turbidity}} X_{\text{oxygen}} + 0.2619 \\ -9.021 \times 10^{-3} X_{\text{water temp}} + 6.832 \times 10^{-5} X_{\text{conductivity}} \\ -5.994 \times 10^{-4} X_{\text{turbidity}} - 1.566 \times 10^{-2} X_{\text{oxygen}} \quad \text{(Spring)} \\ \\ 6.292 + 0.04852 X_{\text{water temp}} + 4.797 \times 10^{-4} X_{\text{conductivity}} + 1.696 \times 10^{-3} X_{\text{turbidity}} + 0.1574 X_{\text{oxygen}} \\ -8.619 \times 10^{-6} X_{\text{water temp}} X_{\text{conductivity}} - 2.564 \times 10^{-5} X_{\text{water temp}} X_{\text{turbidity}} \\ -4.332 \times 10^{-4} X_{\text{water temp}} X_{\text{oxygen}} + 1.552 \times 10^{-7} X_{\text{conductivity}} X_{\text{turbidity}} \\ -3.528 \times 10^{-5} X_{\text{conductivity}} X_{\text{oxygen}} - 1.444 \times 10^{-4} X_{\text{turbidity}} X_{\text{oxygen}} + 0.09127 \\ -3.629 \times 10^{-3} X_{\text{water temp}} + 6.592 \times 10^{-5} X_{\text{conductivity}} \\ -4.038 \times 10^{-4} X_{\text{turbidity}} - 1.332 \times 10^{-2} X_{\text{oxygen}} \quad \text{(Summer)} \\ \\ 5.854 + 0.04852 X_{\text{water temp}} + 4.797 \times 10^{-4} X_{\text{conductivity}} + 1.696 \times 10^{-3} X_{\text{turbidity}} + 0.1574 X_{\text{oxygen}} \\ -8.619 \times 10^{-6} X_{\text{water temp}} X_{\text{conductivity}} - 2.564 \times 10^{-5} X_{\text{water temp}} X_{\text{turbidity}} \\ -4.332 \times 10^{-4} X_{\text{water temp}} X_{\text{oxygen}} + 1.552 \times 10^{-7} X_{\text{conductivity}} X_{\text{turbidity}} \\ -3.528 \times 10^{-5} X_{\text{conductivity}} X_{\text{oxygen}} - 1.444 \times 10^{-4} X_{\text{turbidity}} X_{\text{oxygen}} - 0.3460 \\ +2.362 \times 10^{-2} X_{\text{water temp}} - 3.739 \times 10^{-5} X_{\text{conductivity}} \\ -2.327 \times 10^{-4} X_{\text{turbidity}} + 2.237 \times 10^{-2} X_{\text{oxygen}} \quad \text{(Winter)} \end{cases}$$

We find that this model produces an adjusted R-squared of 0.5369 This tells us that 53.69% of the variation of pH is explained by the model. Additionally, the residual standard error is 0.2032.

## 3.8 ASSUMPTION TESTING

The following is an extensive investigation into the linear model assumptions of our model.

### *3.8.1 Linearity Assumption*

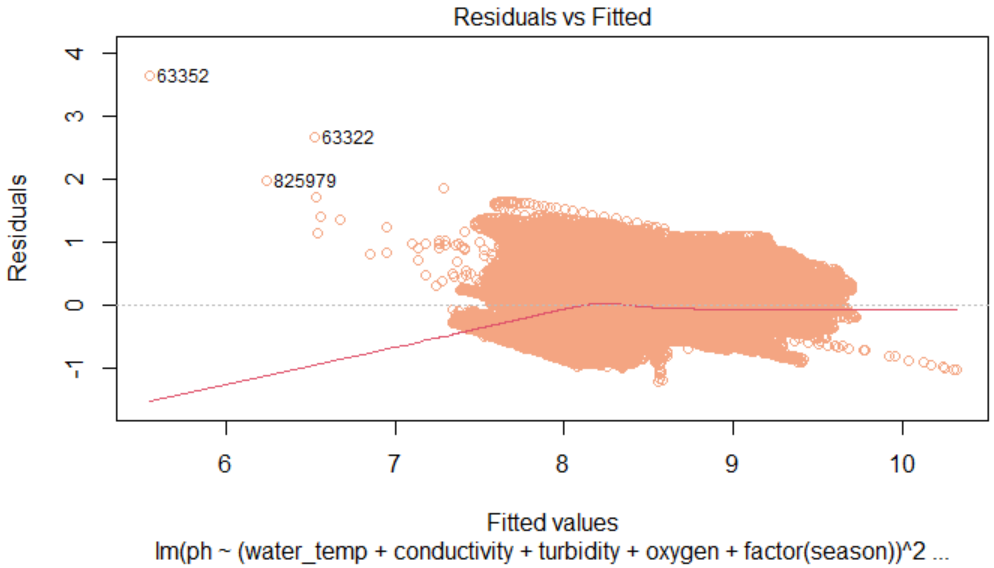


*Figure 8. Residual plot (Residuals vs. Fitted values of the model) of the proposed linear model.*

The trend line does center closely around 0, with fitted values from < 8 residual values ~2 to ~4. The plot is not perfectly horizontal as per the assumptions for linearity.

### *3.8.2 Independence Assumption*

The multi-linear regression model also carries an assumption that error terms are uncorrelated (must be mutually independent). Our data is highly sensitive to this assumption, because our data is gathered over time and is geographically categorized (as data is observed from locations across Calgary's bodies of water).

### *3.8.3 Equal Variance Assumption*

For the model to be valid, we assume that all data points have equal variance (homoscedastic). From Figure 4, we can see directly the variance of the data. Although there is no "funneling" we do see that the data is not randomly scattered, indicating that the variability of the data (residuals) depend on the predicted values (fitted values).

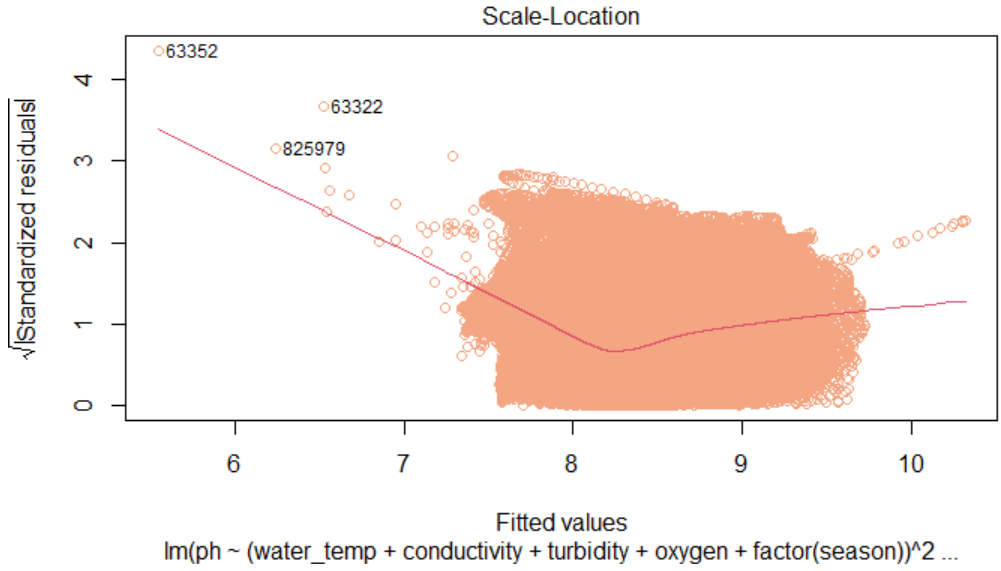


*Figure 9. Scale-location plot of the model.*

We can see from Figure 9 that the model displays heteroscedasticity because the line-of-best-fit is not perfectly horizontal, and the points are not randomly varying from the line. The line-of-best-fit actually shows a trend, decreasing until ~8 fitted values, it begins to increase.

To formally assess the variance of the data points, we used the Breush-Pagan test.
Hypotheses:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 \ldots = \sigma_n^2$$
$$H_a : \text{atleast one } \sigma_n^2 \text{ is different from the others} i = 1, 2, \ldots, n$$

The test produced a `BP = 174145`, `df = 25`, p-value `< 2.2e-16`. Therefore, we reject the null hypothesis and conclude that the data is not homoscedastic (null) and is therefore heteroscedastic (alternative).

### *3.8.4 Normality Assumption*

In order for us to build a valid and functional multi-linear regression model from our data, we must meet the assumption that the errors between observed and predicted values are normally distributed. We formally test for this via a Kolmogorov-Smirnov test.

$$H_0 : \text{the sample data are significantly normally distributed}$$
$$H_a : \text{the sample data are not significantly normally distributed}$$

The Kolmogorov-Smirnov resulted in a `D = 0.064542`, p-value `< 2.2e-16` with a two-sided alternative hypothesis. Because of this, we reject the null hypothesis and conclude that the data is not significantly normally distributed (alternative).

### *3.8.5 Multicollinearity*

To test for multicollinearity amongst our independent variables, we used the *Variance of Inflation Factor (VIF)* test. VIF testing of our independent variables results in no detection of multicollinearity between our independent numerical variables: water temperature, conductivity, turbidity and oxygen. VIF output is as follows:

*Table 6. VIF output.*

| | VIF | Detection |
|---|---|---|
| water temperature | 1.5110 | 0 |
| conductivity | 1.0283 | 0 |
| turbidity | 1.0055 | 0 |
| oxygen | 1.4984 | 0 |

VIF values between the range $1 \leq VIF \leq 5$, suggest that there is a moderate collinearity, but it is not severe enough to warrant corrective measures.

### *3.8.6 Transformations*

We had tried to log transform our data to meet normality and to achieve homoscedasticity in our dataset. The resulting log transformation increased the adjusted R-squared value to 0.5338 and drastically decreased the residual standard error to 0.05456. After conducting a log transformation of our dependent variable, pH, the BP-test, and KS-tests were the following:

*Breush-Pagan Test:*
`BP = 177162`, `df = 25`, p-value = `2.2e-16`

*Kolmogorov-Smirnov Test:*
`D = 0.062237`, p-value = `2.2e-16`

There was no improvement in normality and no success in trying to achieve homoscedasticity.

### *3.8.7 Outliers*

Our data is highly susceptible to outliers, primarily due to 1) measurement process/tool issues and 2) environmental factors. The data collected over the years (2019-2024) is highly dependent on instrument performance, as the instruments required, pH meters, electrical conductivity meters (EC meters), spectroscopes, dissolved oxygen meters and thermometers are analytical instruments that require daily baselining and reconfiguration even in controlled laboratory settings. It would be to no surprise that we find outliers when analytical instruments are exposed in a volatile environment such as a river. Additionally, environmental contributions such as occasional human pollution can contribute to the presence of outliers.

To analyze outliers, we first plotted a residuals vs. leverage plot to detect outliers or influential points.

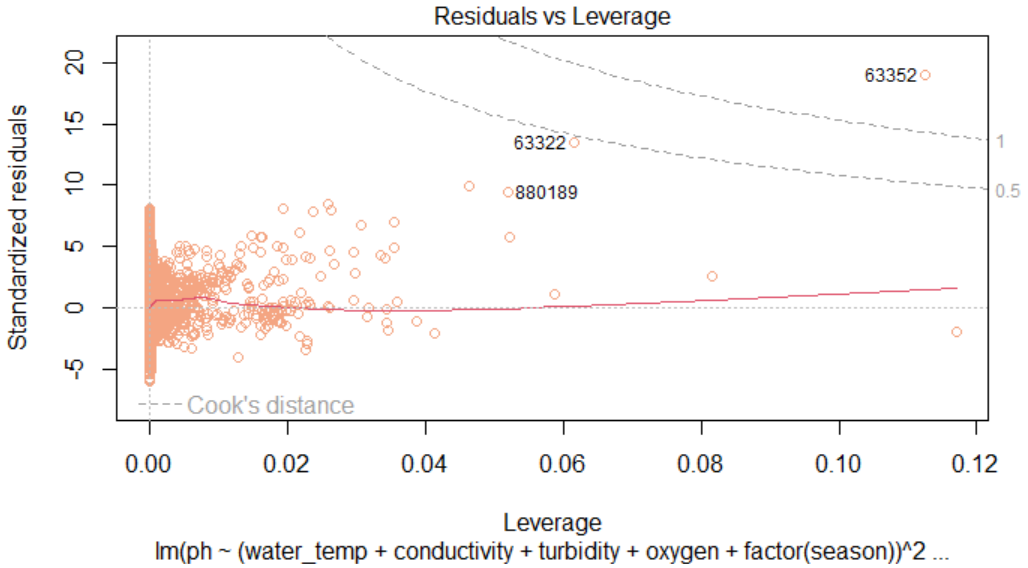


*Figure 10. Residuals vs. Leverage plot for detecting outliers or influential points.*

We find three key outliers identified by the plot are: 880189, 63352, and 63352 from left to right. All other points cluster to the left with minimal Cook's distance.

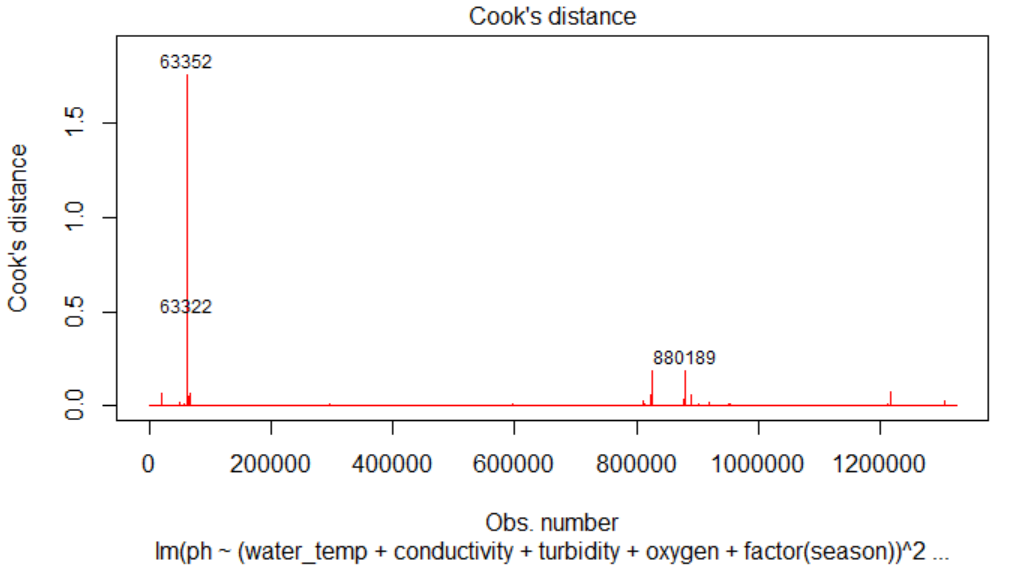


*Figure 11. Cook's Distance plot.*

Figure 11 highlights again the outliers and potential influential points found in Figure 10. From Figure 11, we find the degree of influence each outlier has. We will consider points with >0.5 Cook's Distance.

When we re-run the fitted model, with all the additive and interaction terms mentioned previously, we find that the adjusted R-squared increases slightly (from 0.5369 prev. to 0.5372 after) and that the residual standard error does not decrease (0.2031). Because of these observations, due to the lack of significant improvement of our model (an increase of adjusted R-squared of less than 5%), these points do not pose a heavy influence in our dataset and therefore will remain.

### 3.9 SEASONS & pH

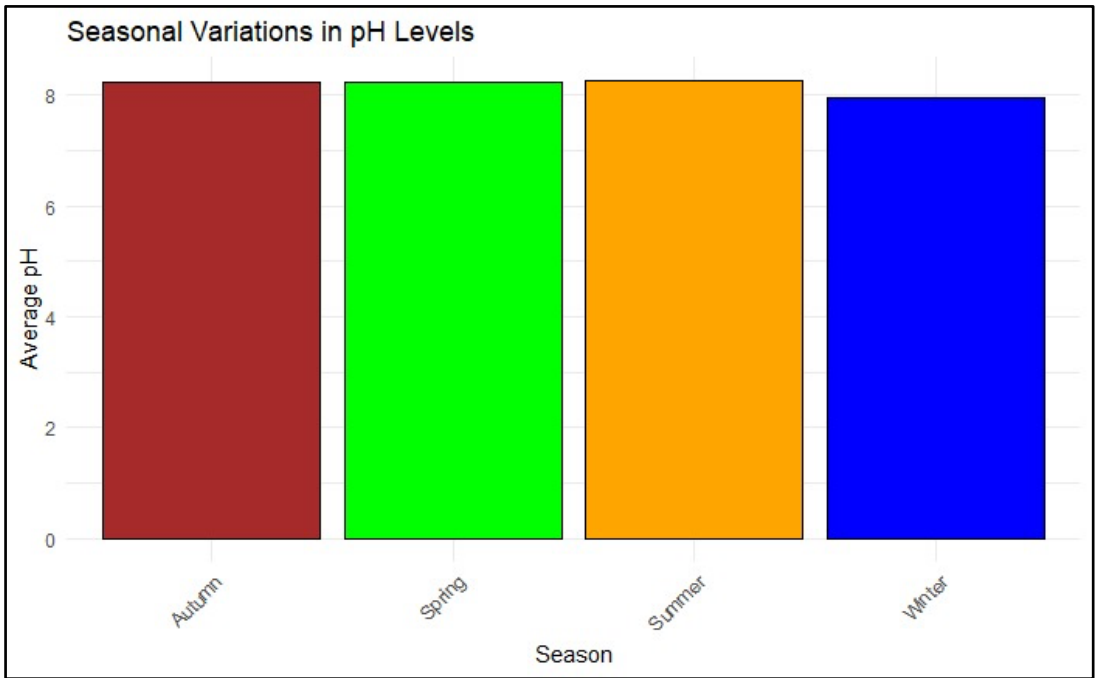


*Figure 12. pH variability in the four different seasons (Autumn, Spring, Summer, Winter) from our dataset.*

We find in Figure 12 that the average pH remains consistently close to 8 across all seasons (autumn, spring, summer, and winter). This indicates that pH levels are stable year-round, with no significant seasonal fluctuations, showing only a slight dip in pH during winter might be due to reduced biological activity, such as less decomposition of organic matter processing, and potentially lower photosynthetic activity, which influences carbon dioxide levels in water and the slight rise in pH in spring could result from increased biological activity, including algae blooms or increased photosynthesis, which consume carbon dioxide and temporarily elevate pH levels.

### 3.10 PREDICTION

The model was trained on data from November 2019 to August 2024, and the model was tested on data from September to November 2024. From the plot it is evident that both the **actual pH values (blue)** and **predicted pH values (red)** follow a similar cyclical pattern. This indicates that the prediction model captures the general trend of the pH variations. There are sections where the predicted pH closely matches the actual pH (e.g., around indices 0 to 10,000 and 25,000 onward), but in other areas, there appear to be larger deviations between the two lines (e.g., between 10,000 to 25,000). This suggests that the model performs moderately well in predicting the general behavior of pH over time but struggles with capturing extremes and variability. To improve the prediction of the model we have to address predictions during peaks and troughs, mostly through additional features or a more powerful model.

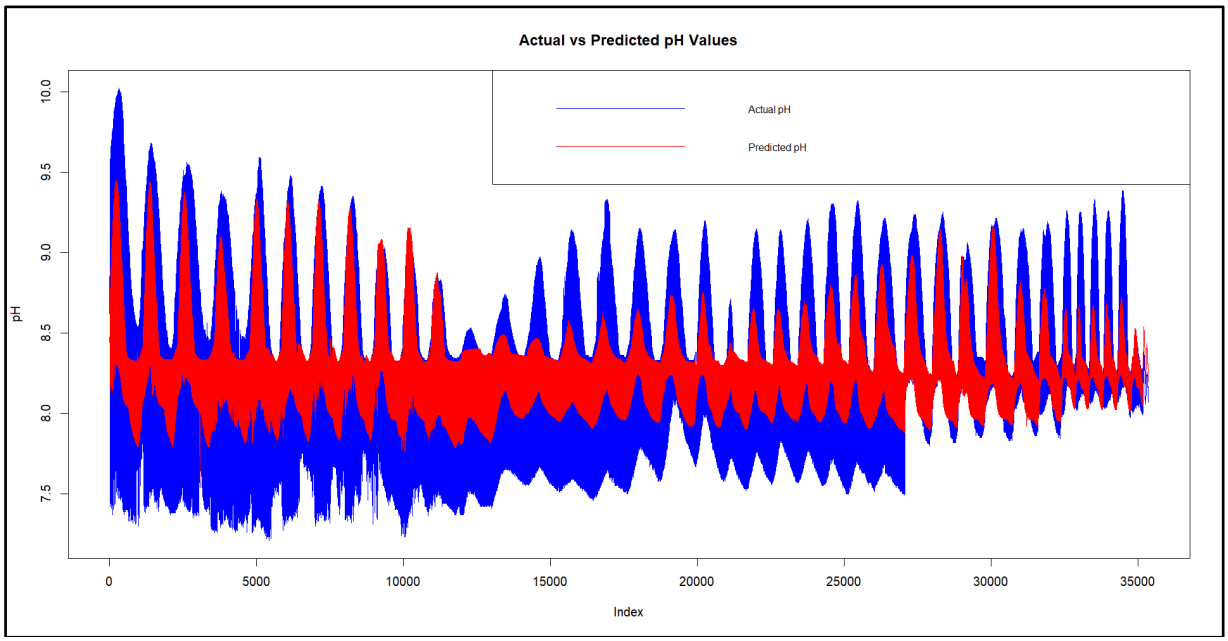


*Figure 13. Actual vs. Predicted pH Values.*

The individual prediction results were fairly good, for instance when we passed the already existing data into the model,

```
water_temp = 14.217, conductivity = 403.9, turbidity = 2.71,
oxygen = 13.05, season = 'Summer'
```

the predicted pH was 8.96, when the original pH was 9.05. This suggests that the model is fairly accurate.

On the day 2021-06-29 at 16:45:00 when water was at the highest temperature,

```
water_temp = 32.205, conductivity = 1687.5, turbidity = 3.02,
oxygen = 19.06, season = 'Summer'
```

the predicted pH was 9.529184, when the original pH was 9.73. This suggests that the model is fairly accurate at the margin value of a parameter.

# 4 CONCLUSION AND DISCUSSION

## 4.1 APPROACH

Overall, the most impactful variables—water temperature, dissolved oxygen, and conductivity—are consistently significant. Seasonal variations alter the relative importance of turbidity and interaction terms, highlighting dynamic relationships.

Water temperature, conductivity, dissolved oxygen, and turbidity significantly influence pH, with their effects varying across seasons.

- Water temperature is a key variable in pH modeling, as it affects dissolved oxygen availability and drives chemical reactions in water.
- Conductivity, particularly important in spring and summer, reflects ionic concentration; higher conductivity values can alter or buffer pH (see Section 3.2.2).
- Dissolved oxygen consistently shows a positive correlation with pH across all seasons, as it is influenced by temperature, biological activity, and water mixing (see Section 3.2.1).
- Turbidity has variable effects, with strong seasonal patterns. It sometimes shows a negative correlation with pH due to organic matter or suspended particles.
- Interaction terms—such as those between water temperature and dissolved oxygen, or between conductivity and turbidity—suggest that these interactions may amplify their influence on pH (see Section 3.2.3).
- Seasonal variations in pH levels are minimal. The average pH remains consistently close to 8 across all seasons (autumn, spring, summer, and winter). This indicates that pH levels are stable year-round, with no significant seasonal fluctuations, showing only a slight dip in winter and a slight rise in spring due to natural processes like decomposition and photosynthesis (see Section 3.7).

## 4.2 FUTURE WORK

While the model is statistically significant, the adjusted $R^2$ value indicates that the dataset is volatile, and pH is influenced by various other environmental parameters that cannot be accurately predicted with the current ordinary least squares models. Ensemble or deep learning models could provide better accuracy and improved predictions. In the future, we plan to implement big data analytics, enabling continuous data streaming via the cloud and more robust model analysis.

## 5 REFERENCES

[1] *Water Quality Monitoring Sonde Data | Open Calgary.* (2024, November 1). Retrieved November 5, 2024, from https://data.calgary.ca/Environment/Water-Quality-Monitoring-Sonde-Data/kc8x-fu3f/

[2] *Open Calgary Terms of Use.* (2024, November 1). Retrieved November 5, 2024, from https://data.calgary.ca/stories/s/Open-Calgary-Terms-of-Use/u45n-7awa

[3] The Conference Board of Canada. (2024). *Water Quality Index.* Retrieved November 6, 2024, from https://www.conferenceboard.ca/hcp/water-quality-index-aspx/

[4] US EPA, O. (2021, June 11). *Factsheets on Water Quality Parameters* [Overviews and Factsheets]. https://www.epa.gov/awma/factsheets-water-quality-parameters

[5] Vitt, D. H., Bayley, S. E., & Jin, T.-L. (1995). Seasonal variation in water chemistry over a bog-rich fen gradient in Continental Western Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, *52*(3), 587–606. https://doi.org/10.1139/f95-059

[6] *Temperature Dependence of the pH of pure Water*. (2013, October 2). Chemistry LibreTexts. https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Acids_and_Bases/Acids_and_Bases_in_Aqueous_Solutions/The_pH_Scale/Temperature_Dependence_of_the_pH_of_pure_Water

[7] *Understanding The Relationship Between pH And Electrical Conductivity*. (2024, September 10). Atlas Scientific. https://atlas-scientific.com/blog/relationship-between-ph-and-conductivity/

[8] Page, C. (n.d.). *Water quality, water hardness and water data*. Https://Www.Calgary.Ca. Retrieved December 1, 2024, from https://www.calgary.ca/content/www/en/home/water/drinking-water/water-quality-water-hardness-water-data.html