Exploring Turbidity and Environmental Factors: A Multi-Linear Regression Analysis of Beach Observation Data in Toronto (2008-2023)*

Bolin Shen

April 13, 2024

This study investigates the dynamics of water quality in Toronto's beaches, focusing on turbidity. Utilizing the extensive 'Toronto Beach Observation Dataset,' collected by Bylaw Enforcement Officers, the analysis explores intricate interactions between environmental factors and turbidity levels. By comparative analyses between Marie Curtis East Beach and Sunny Side Beach reveal, differences in turbidity levels are discovered, indicating variations in water quality. Through the application of multi-linear regression modeling, these relationships are quantified, providing valuable insights to guide informed decision-making. However, there are weaknesses and challenges in this study, such as data limitations and model uncertainties, highlighting the need for continuous research. Overall, this study contributes to understanding water quality dynamics, facilitating interventions for preserving beach health.

Table of contents

*Code and data supporting this analysis is available at: https://github.com/Brian031205/Exploring-Beach-Observation

1 Introduction

Observing beaches is important for urban planners, policymakers, and citizens to ensure public health and safety. Regular monitoring beach conditions allows local governments to identify and address potential risks or dangers, such as issuing warnings or closures based on water quality or other hazards. Data collected over time provides insights into beach conditions, informing decisions on beach conditions, informing decisions on management, infrastructure enhancement, and conservation efforts (OpenDataToronto 2024b).

This study focused on Toronto's beaches, utilizing the Toronto Beach Observation Dataset, accessible through the Toronto Open Data portal (OpenDataToronto 2024a). Through observation and analysis, researchers in Toronto have collected a wide range of data on key environmental factors across ten beaches. This dataset provides valuable insights into the complex interaction between environmental variables and beach conditions, with a specific focus on turbidity. The estimand of this study is the impact of climate and environmental factors — such as wind speed, rainfall amount, air and water temperature, wave action, and waterfowl—on beach turbidity, examining the inherent influence of these variables on the likelihood of turbidity occurrences. The objective is to deepen our understanding of beach dynamics and environmental phenomena.

The findings reveal significant fluctuations in turbidity level across Toronto's diverse beaches, reflecting different responses to environmental factors over time. Through an exploration of the nature of beach observations in Toronto, the analysis identifies consistent trends in climatic factors and their effects on beach turbidity, as well as differences in waterfowl populations reflecting environmental variations. Moreover, it highlights changes over time of beach ecosystems. This study contributes to our understanding and effective management of urban beaches, essential for public health and safety. Utilizing the Toronto Beach Observation Dataset, it explores the complex connections between environmental factors and beach conditions, particularly focusing on turbidity. These insights emphasize the need for effective management strategies to address diverse challenges and differences. Furthermore, the study highlights the importance of continuous monitoring to facilitate evidence-based decision-making by urban planners, policymakers, enabling proactive measures for sustainable beach management and public well-being.

The paper begins by introducing the broader context and motivation behind our study. Section 2 elaborates on data sources and the important variables. Section 3 introduces the specifics of the linear regression model. This is followed by the presents of the data resulted from the model in Section 4. The final section includes discussions on Investigating the Dynamics of Water Quality in Toronto's Water Bodies; Comparative Analysis of Turbidity Trends and Environmental Factors; Model Training, Prediction, and Evaluation; Insights from Linear Regression Analysis and Predictive Modeling; as well as Weakness and Implications for Future Research.

2 Data

The data I used in this paper are obtained from the Open Data Toronto Portal, accessed through the library `opendatatoronto` (Gelfand 2022). The dataset, which covers polls conducted by the city from 2015, was first cleaned before analyzing by the open-source statistical programming language R (R Core Team 2022), using functionalities from `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `dplyr` (Wickham et al. 2022), `readr` (Wickham, Hester, and Bryan 2022), `tibble` (Müller and Wickham 2022), `modelsummary` (Arel-Bundock 2022), `here` (Kirill Müller 2020), `janitor` (Firke 2021) and `knitr` (Xie 2014).

2.1 Data Survey

In environmental research, understanding the interaction among various factors is important. This study explores the dynamics influencing water quality in Toronto's water bodies, with a particular focus on turbidity. Turbidity, representing the clarity of water, stands as the dependent variable under examination, offering a perspective to analyze the effects of several independent variables.

Seven key variables have been identified, each contributing unique insights into Toronto's aquatic environments. While turbidity taking precedence in the analysis, the study extends beyond its individual influence. Beach name, data collection date, waterfowl presence, wave action, rainfall amount, wind speed, air temperature, and water temperature emerge as important independent variables, each capable of influencing turbidity levels over the thirteen-year period from 2008 to 2023. Among these variables, beach name and wave action are categorical, while waterfowl presence, rainfall amount, wind speed, air temperature, and water temperature are numerical. This study aims to discover the intricate relationships among these variables, examining the intricate dynamics governing water quality in Toronto's aquatic ecosystems.

Table 1: Hanlan's Point Beach Environmental Observation Data Table

[!h]

beachName	windSpeed	dataCollectionDate	airTemp	waterTemp	waterFowl	rainAmount	waveAction	turbidity
Hanlan's Point Beach	10	2010	22	17.5	30	2	LOW	1.7
Hanlan's Point Beach	20	2010	12	13.0	18	20	MOD	4.0
Hanlan's Point Beach	26	2010	16	17.5	36	15	LOW	1.0
Hanlan's Point Beach	15	2010	22	17.1	17	19	MOD	0.7
Hanlan's Point Beach	7	2010	20	19.3	12	4	HIGH	4.5
Hanlan's Point Beach	4	2010	24	18.2	30	7	MOD	3.8
Hanlan's Point Beach	5	2010	22	16.1	3	45	LOW	1.7
Hanlan's Point Beach	9	2010	21	17.9	60	0	MOD	0.8
Hanlan's Point Beach	27	2010	27	23.6	140	1	MOD	0.9
Hanlan's Point Beach	10	2010	27	19.0	5	1	MOD	1.4

2.2 Measurement

The beach observations dataset, collected by City Officers following strict protocols for accuracy and reliability, forms a large-scale dataset at regular intervals. This dataset allows for exploration into the potential impact of environmental factors on water turbidity. Toronto's beaches are regularly observed by city staff, with instrumentation such as thermometers or turbidity meters used where necessary to ensure precision. Trained personnel estimate other observations, such as fowl numbers and wave action. These observations are taken from mid-May to mid-September, providing an extensive overview of beach conditions during the peak seasons (OpenDataToronto 2024a). This systematic data collection approach ensures the documentation of a broad range of environmental factors, facilitating detailed analyses and informed decision-making in environmental management and policy development.

2.3 Data summary

The "Toronto Beach Observation Dataset" includes a thorough collection of data gathered from 10 diverse beaches in Toronto, such as Hanlan's Point Beach. Table 1 illustrates the observation findings of seven natural phenomena at Hanlan's Point Beach across various time intervals in 2010. These phenomena include wind speed, air temperature, water temperature, waterfowl presence, rainfall amount, wave action, and turbidity. Each entry in the table represents a distinct observation instance, enabling researchers and stakeholders to examine trends, patterns, and fluctuations in environmental conditions at Hanlan's Point Beach over time.

Due to limited observation data available for many beaches, I chose to contrast the datasets of two beaches with complete records: Marie Curtis East Beach and Sunny Side Beach. As the data for Sunnyside Beach spans from 2008 to 2023 and for Marie Curtis Park East Beach from 2010 to 2023, the analysis focuses on the data from 2010 to 2023. The following line charts demonstrate the comparison of data from these two beaches.

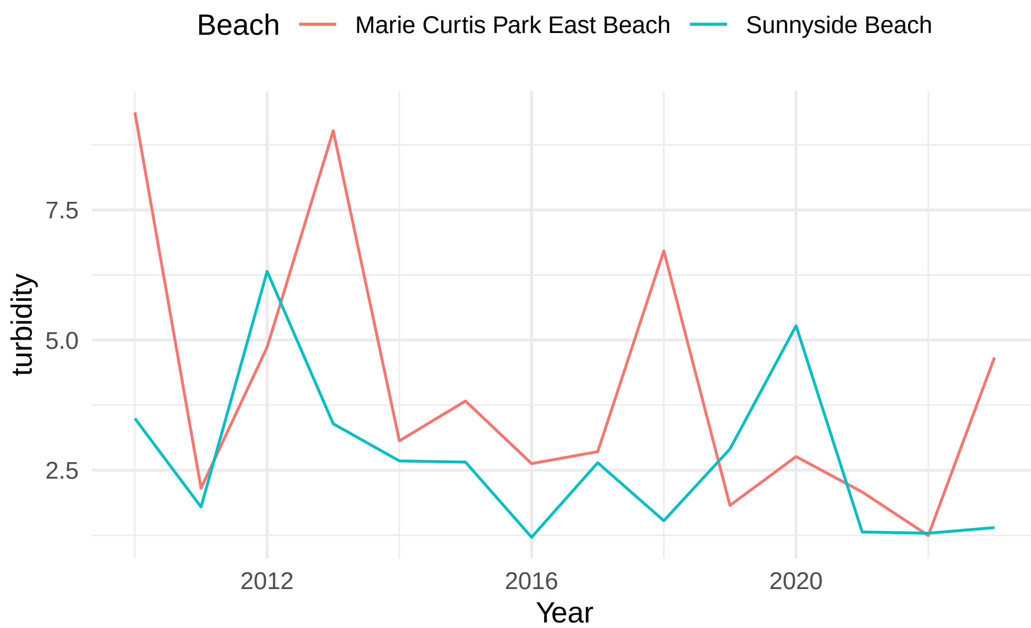

Figure 1: Comparison of Turbidity Levels at Marie Curtis Park East Beach and Sunny Side Beach (2010-2023)

Figure 1 depicts the turbidity trends at Marie Curtis Park East Beach and Sunny Side Beach from 2010 to 2023. Turbidity, an indicator of water clarity, is shown on the y-axis, while the x-axis indicates the years. The red line represents the turbidity at Marie Curtis Park East Beach, and the blue line shows Sunny Side Beach. Marie Curtis Park East Beach exhibits more significant fluctuations in turbidity, indicating greater variability. Both beaches reach their lowest turbidity levels in 2022. Overall, turbidity levels are consistently higher at Marie Curtis Park East Beach compared to Sunny Side Beach, indicating potential differences in water quality between the two locations and implying poorer environmental conditions at Marie Curtis Part East Beach.

Figure 2 depicts the windspeed variations at Marie Curtis Park East Beach and Sunny Side Beach from 2010 to 2023. The x-axis represents the years, from 2010 to 2023, while the y-axis indicates windspeed. Both lines show similar fluctuations, reaching their peak in 2012 at about 18, indicating similar and consistent windspeed patterns between the two beaches over the observed period. Interestingly, in 2022, when the turbidity levels were at their lowest, windspeed were relatively high. This implies a potential deviation from the typical correlation between windspeed and turbidity, highlighting the influence of specific environmental conditions and geographical factors on this relationship.

Figure 3 compares the air temperature data between Marie Curtis Park East Beach and Sunny Side Beach from 2010 to 2023. The x-axis represents the years, while the y-axis indicates the

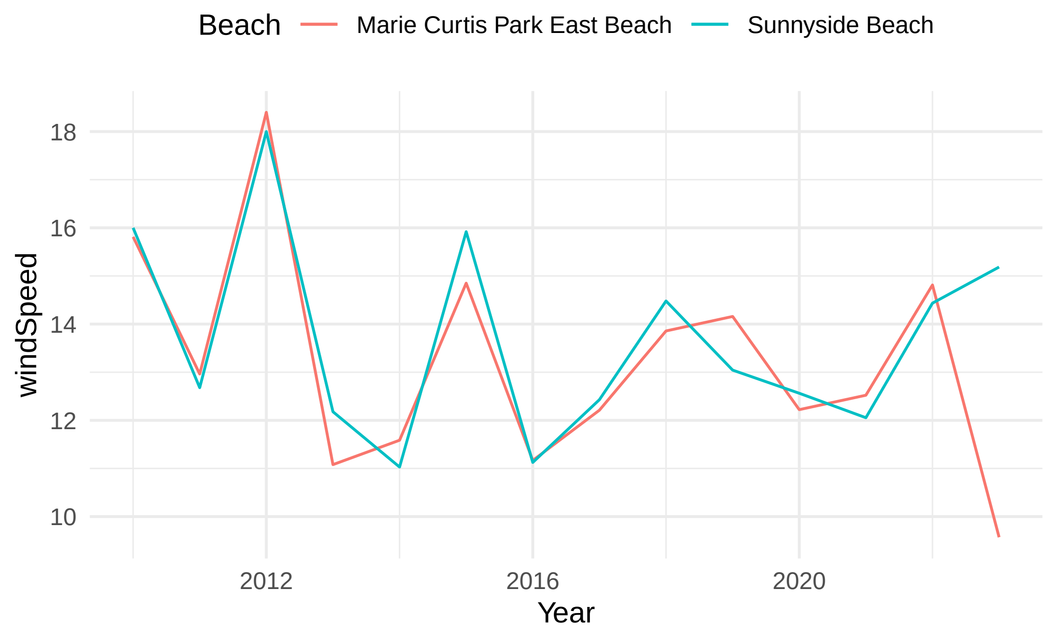

Figure 2: Comparison of Windspeed between Marie Curtis Park East Beach and Sunny Side Beach (2010-2023)

air temperature. The graph reveals remarkably similar fluctuations in air temperature between the two beaches, as evidenced by the closely aligned red and blue lines. Furthermore, there are similar peaks and troughs in air temperature over the observed period. For instance, in 2012, both beaches experienced peak temperatures of about 24.5 degrees Centigrade, followed by a sharp decline to around 16 degrees in 2017. This indicates a consistent pattern of weather conditions between the two beach locations.

Figure 4 compares the waterfowl populations between Marie Curtis Park East Beach and Sunny Side Beach from 2010 to 2023. The x-axis represents the years, from 2010 to 2023, while the y-axis indicates the waterfowl population. The red line peaked at 80 in 2010 for Marie Curtis Park East Beach, followed by a sharp decline and mild fluctuations afterward. In contrast, the blue line, depicting Sunny Side Beach, demonstrates regular fluctuations and maintains a consistently higher waterfowl population from 2011 to 2023. This suggests the environmental factors, such as water quality and vegetation cover, may favor the presence of waterfowl at Sunny Side Beach. This observation aligns with the findings from Figure 1, indicating lower turbidity and potentially better environmental conditions at Sunny Side Beach compared to Marie Curtis Part East Beach.

Figure 5 compares the rainfall amounts between Marie Curtis Park East Beach and Sunny Side Beach from 2010 to 2023. The x-axis represents the years spanning from 2010 to 2023, while the y-axis indicates the amount of rainfall. Remarkably, both lines exhibit similar fluctuations

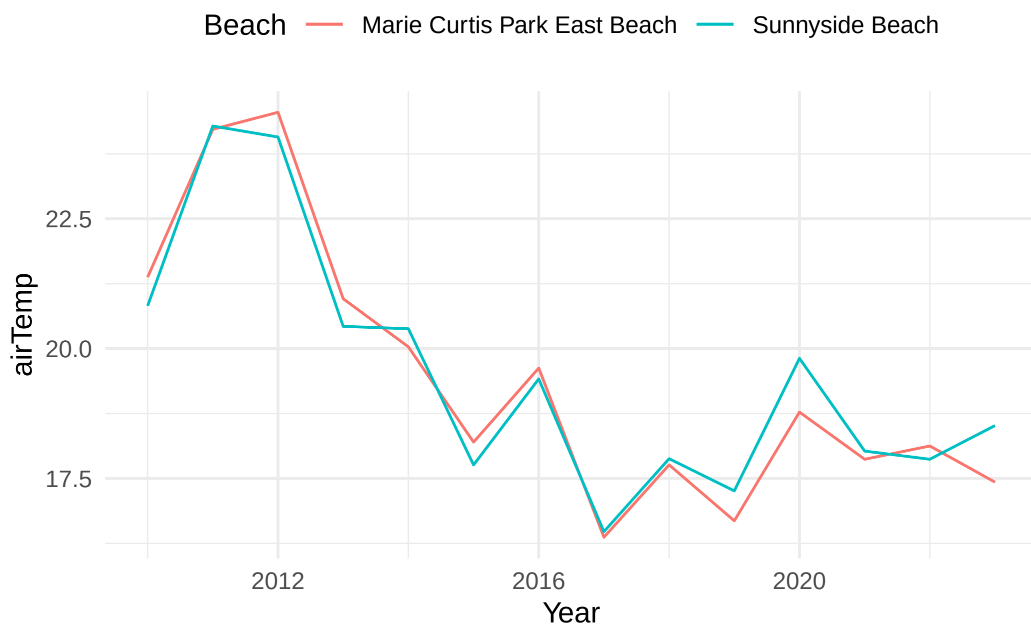

Figure 3: Comparison of Air Temperature between Marie Curtis Park East Beach and Sunny Side Beach (2010-2023)

over the years, with notable peaks occurring in 2010, 2015, 2018, and 2020, where both beaches experienced the high rainfall amounts. Comparing these results with the findings of Figure 1, there appears to be a generally positive correlation between rainfall amount and turbidity for the two beaches. However, it is interesting to note that despite the high rainfall amount in 2018 for Sunnyside Beach, there is a low turbidity level, which may due to the clarity of rainwater or other specific environmental factors.

Figure 6 displays the turbidity levels across ten different beaches, representing the average turbidity values from 2008 to 2023. Each bar represents the turbidity measurement at a specific beach. Bluffers Beach Park exhibits the highest turbidity level at 18, followed by Kew Balmy Beach at 14, and Woodbine Beaches at 7.5. Conversely, beaches such as Centre Island Beach, Gilbralter Point Beach, Hanlan's Point Beach and Sunnyside Beach exhibit lower turbidity levels. This visualization provides a clear comparison of turbidity levels among the selected beaches, highlighting the variations in water clarity across different locations.

Figure 7 presents the population of waterfowl across ten different beaches, displaying the average counts from 2008 to 2023. Each bar represents the number of waterfowl observed at a specific beach. Centre Island Beach leads the highest count at 67, followed closely by Ward's Island Beaches at 66 and Bluffers Beach Park at 60. Hanlan Point Beach, Kew Balmy Beach, and Bluffers Beach Park all have approximately 47 waterfowl. Interestingly, beaches with high turbidity levels, as depicted in Figure 6, also demonstrate a larger presence of waterfowl. This

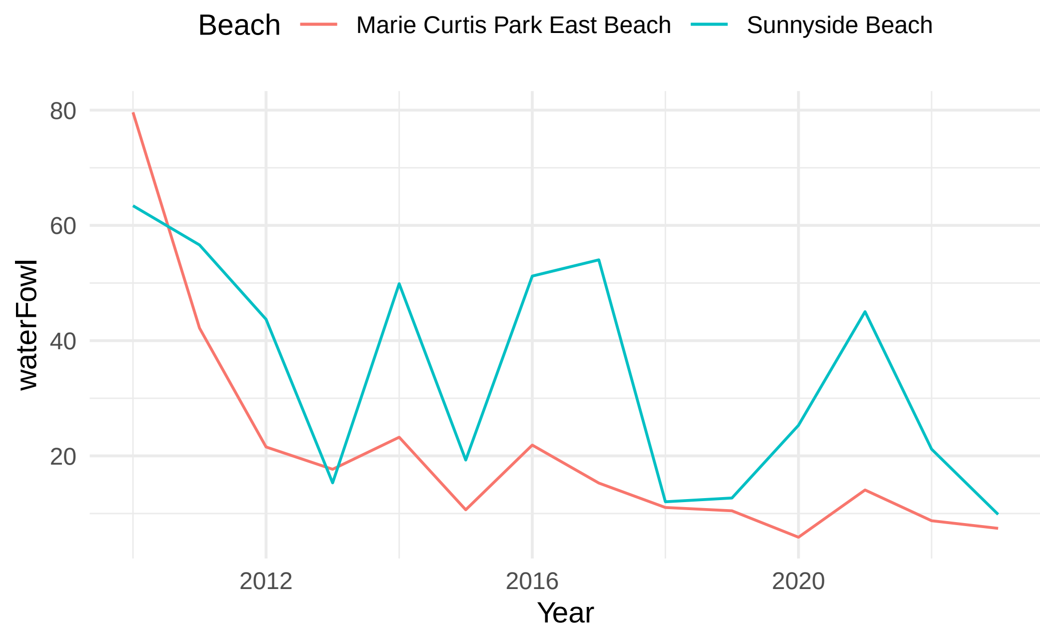

Figure 4: Comparison of Waterfowl Population between Marie Curtis Park East Beach and Sunny Side Beach (2010-2023)

suggests a correlation between waterfowl activities and turbidity levels, with foraging, nesting, and feces contributing to localized increases in turbidity through sediment disturbance and nutrient introduction.

3 Model

The objective of the model is to analyze the relationship between environmental factors (such as, wind speed, rainfall amount, air and water temperature, waterfowl, wave action) and water turbidity, quantifying its strength and direction. The study utilizes a multiple linear regression model to analyze the dataset sourced from the Toronto Open Data portal.

A model is trained and constructed by utilizing 80 percent of the whole dataset, with the remailing 20 percent used for prediction. Formulated based on survey data, the multi-linear regression model is suited to the study's objectives (Equation 1). Subsequently, the model is used to predict climate and environmental characteristics across various beaches in Toronto.

$$\begin{aligned} turbidity = &\beta_0 + \beta_1 x_{beachName} + \beta_2 x_{windSpeed} + \beta_3 x_{dataCollectionDate} + \beta_4 x_{airTemp} + \\ &\beta_5 x_{waterTemp} + \beta_6 x_{waterFowl} + \beta_7 x_{rainAmount} + \beta_8 x_{waveAction} \end{aligned} \tag{1}$$

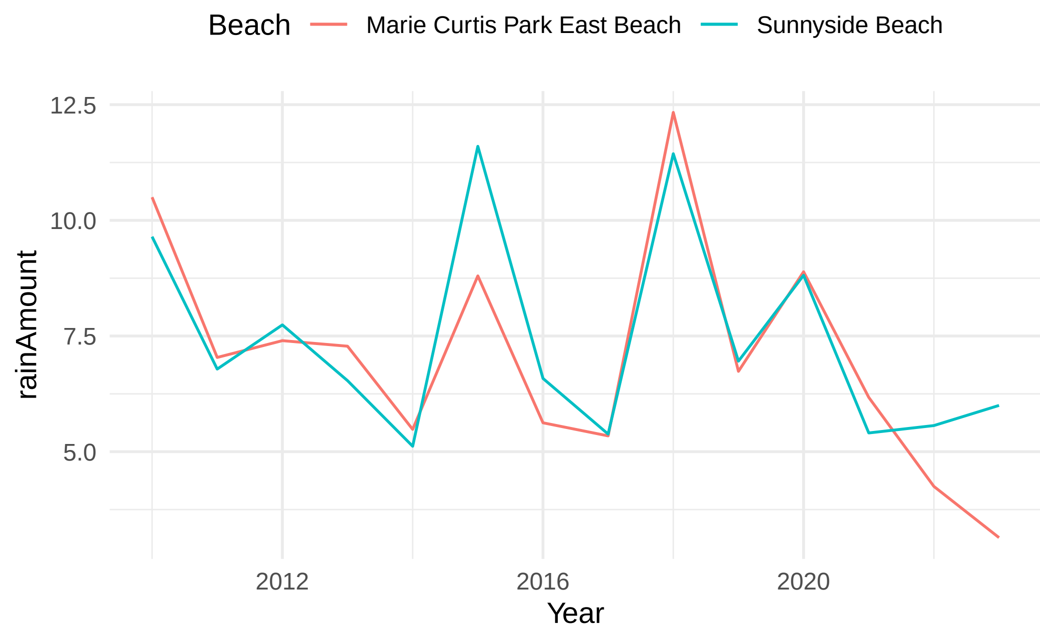

Figure 5: Rain Amount Comparison between Marie Curtis Park East Beach and Sunny Side Beach (2010-2023)

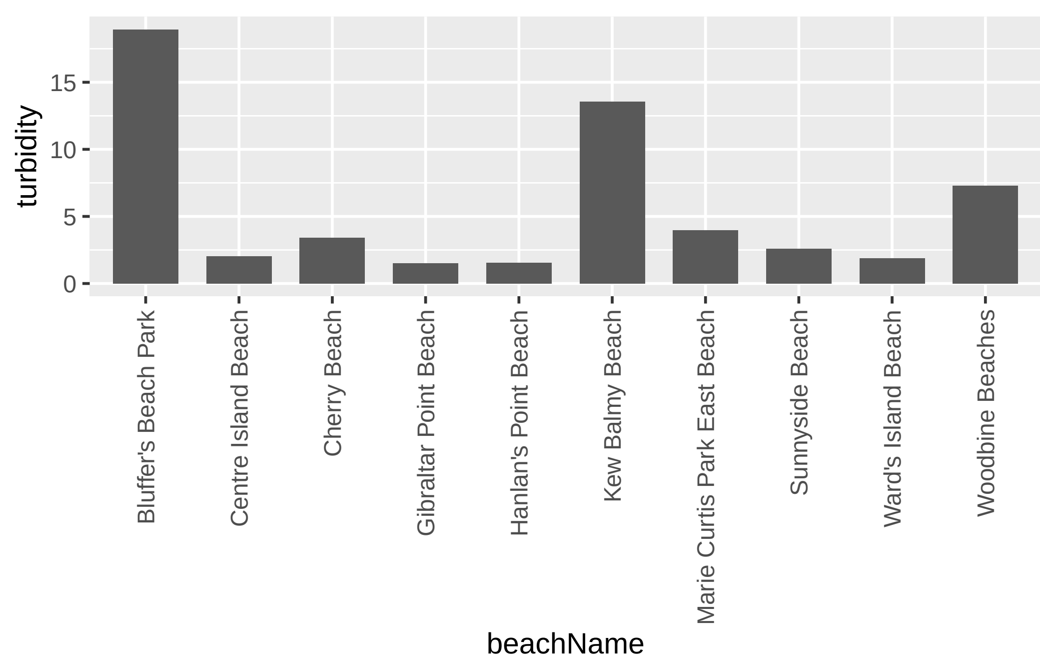

Figure 6: Turbidity Comparison Across Ten Beaches

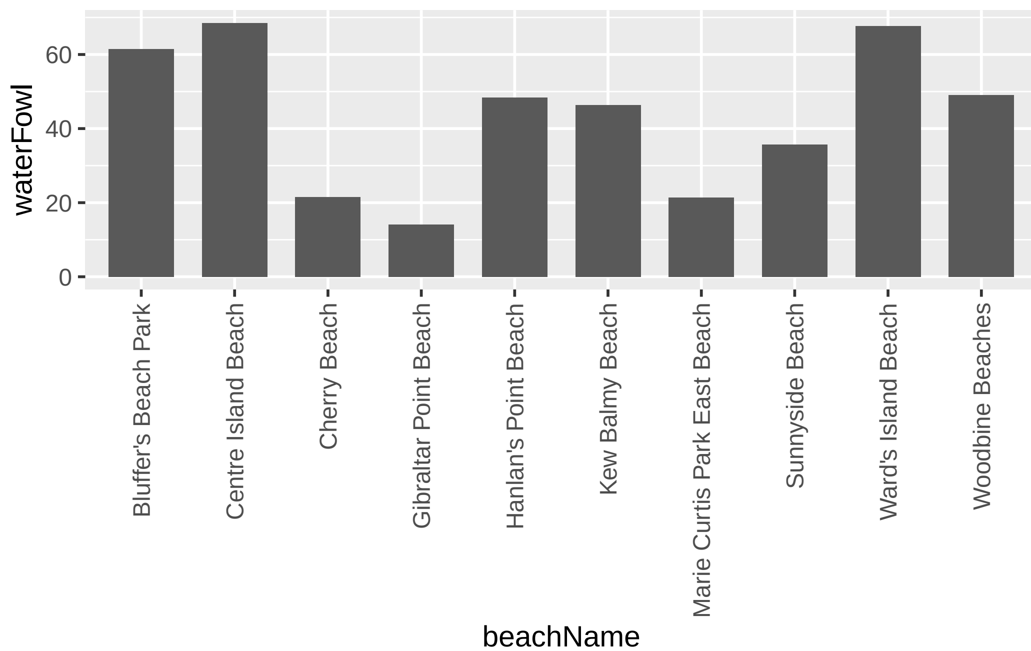

Figure 7: Waterfowl Population Across Ten Beaches

In equation 1, *turbidity* represents the water turbidity for each beach, while all other variables are denoted by x_i.

In this equation, each β represents a coefficient determined through regression analysis. The variables selected for this project include beach name, wind speed, data collection date, air and water temperature, waterfowl presence, rainfall amount and wave action. These environmental factors were chosen as they are established indicators of beach turbidity and are useful for understanding its variations.

After developing the multi-linear regression model, I utilize the predict() function in R(R Core Team 2022) to apply the model and generate predictions. This enables estimating the environmental characteristics across different beaches in Toronto and explore potential relationships between these independent variables with beach turbidity. The stan_glm() function is used for this regression analysis. The linear-shaped distribution of this dataset makes linear regression a suitable approach.

After completing the modeling process, I utilize the remaining 20% of the data to make predictions. Analyzing these prediction results allows me to assess the reliability of the model. Some of the predicted data agrees well with the given data, which indicates that the model is somewhat reliable (See Table 3). However, I have also observed instances of negative numbers in the predicted results, which requires consideration of the limitations of the model. The presence of negative values suggests potential inaccuracies in predicting the target variable in

Table 2: The Modeling Result

term	estimate	std.error	conf.low	conf.high
(Intercept)	422.6160435	253.3552404	1.8696642	825.9630944
beachNameCentre Island Beach	-11.3265426	2.9277144	-16.1757654	-6.6740004
beachNameCherry Beach	-7.6287345	2.8701575	-12.3134043	-2.9510326
beachNameGibraltar Point Beach	-12.5650889	2.8466662	-17.2997577	-7.9909286
beachNameHanlan's Point Beach	-11.7384015	3.0739821	-16.8824517	-6.6089789
beachNameKew Balmy Beach	1.8643513	2.4684229	-2.3382578	5.9812444
beachNameMarie Curtis Park East Beach	-7.5902389	1.7855362	-10.6591866	-4.6497624
beachNameSunnyside Beach	-8.3331170	1.7961554	-11.3784990	-5.3723886
beachNameWard's Island Beach	-11.3479847	2.9262797	-16.1597962	-6.3831071
beachNameWoodbine Beaches	-5.4290089	2.2397919	-9.2590332	-1.6599008
rainAmount	0.1209161	0.0327554	0.0665126	0.1753800
airTemp	-0.0960857	0.0859423	-0.2370016	0.0495648
waterTemp	-0.0216406	0.0301918	-0.0700988	0.0292286
windSpeed	-0.0184829	0.0469334	-0.0968800	0.0586400
dataCollectionDate	-0.1984462	0.1251172	-0.3996480	0.0106574
waterFowl	-0.0018160	0.0081912	-0.0150174	0.0112739
waveActionLOW	-9.5985464	1.4086096	-11.9688067	-7.3181980
waveActionMOD	-7.6717431	1.5688717	-10.1929213	-5.0832081
waveActionNONE	-8.7440385	1.6892144	-11.7043215	-6.0695418

certain cases, or there are other unaccounted factors affecting the prediction outcomes. Therefore, I need to consider both the positive and negative results of the model predictions to fully understand their validity and applicability.

4 Result

Table 2 shows the estimated coefficients for the multi-linear regression equation derived from the Toronto Beach Observation dataset. These coefficients, fitting into Equation 1, are calculated based on the dataset's data. The table is made using kable function from `knitr` (Xie 2014) and formatted using `kableExtra` (Zhu 2020).

In the data modeling process utilizing the Toronto Beach Observation dataset, I divided 80% of the data for model training and allocated the remaining 20% for prediction. Figure 8 illustrates the modeling results, where the x-axis represents the estimate, and the y-axis depicts the coefficients corresponding to each other. This visualization provides insights into the coefficients variations across different beaches concerning the estimated values. Analyzing these coefficients provides a better understanding of the factors influencing the model's predictive capability and individual contributions of each other to the overall estimation.

Figure 8 displays the coefficients derived from linear regression analysis conducted on the Toronto beach observation dataset. Each coefficient estimate is accompanied by error bars representing the confidence interval, providing insights into the uncertainty associated with the estimates. When interpreting these coefficients, it is important to note that positive values suggest a strong positive linear regression relationship between the independent variables and turbidity, while negative values suggest a lack of linear regression relationship. These coefficients offer valuable information about the strength and direction of the relationships between environmental factors and turbidity levels in beach waters, which is of use in understanding the dynamics influencing water quality.

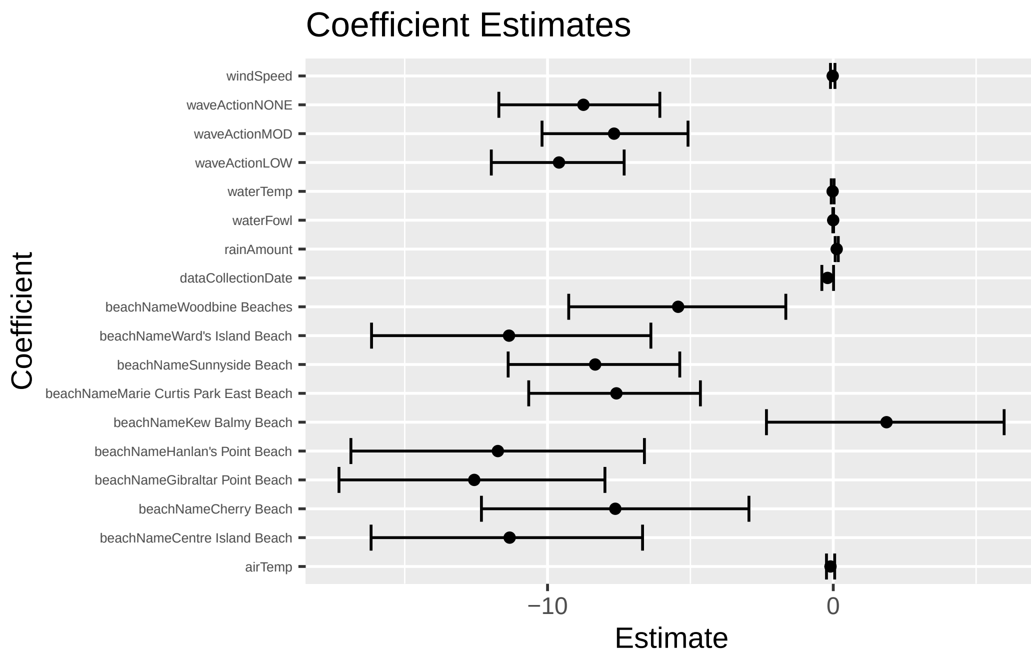

Figure 8: The Predicting Result

Utilizing the results of my linear regression model, I can formulate an equation that adheres to the structure outlined in equation 1, incorporating specific β coefficients for each variable. Due to the numerous variables involved, detailing the equation fully is challenging. Essentially, the equation integrates the β value corresponding to a variable if an individual's characteristic matches that variable. Table 3 offers examples illustrating how the probability varies based on different variables.

Based on the model I established, I randomly selected 10 prediction results, detailed in Table 3. The analysis of these results reveals some findings on beach conditions and environmental factors. Firstly, the model demonstrates its capability to predict various environmental variables across different beaches, as seen in Table 3 where the values of Hanlan's Point Beach agree well with the original data given in Table 1. This indicates its potential in providing valuable insights for beach monitoring and management. Secondly, the variability observed in predicted values across beaches and time intervals suggests distinct patterns and trends in environmental

Table 3: Example of Prediction model

[!h]

beachName	windSpeed	dataCollectionDate	airTemp	waterTemp	waterFowl	rainAmount	waveAction	predict
Hanlan's Point Beach	26	2010	16	17.5	36	15	LOW	0.7362578
Hanlan's Point Beach	9	2010	21	17.9	60	0	MOD	0.6499090
Hanlan's Point Beach	17	2010	21	20.2	100	2	MOD	0.6206824
Hanlan's Point Beach	8	2010	18	18.7	15	0	MOD	1.0170560
Ward's Island Beach	4	2010	24	18.9	75	7	MOD	1.6782565
Ward's Island Beach	27	2010	27	23.6	36	1	LOW	-1.7130361
Centre Island Beach	4	2010	24	19.6	40	7	MOD	1.6714430
Centre Island Beach	9	2010	21	18.4	80	0	HIGH	8.6438525
Centre Island Beach	8	2010	18	17.9	50	0	MOD	1.3330632
Sunnyside Beach	7	2010	20	19.2	90	4	LOW	2.6277983

data at each location. This highlights the importance of considering site-specific factors when assessing beach conditions. Additionally, the inclusion of multiple variables such as wind speed, air temperature, water temperature, waterfowl presence, rainfall amount, and wave action emphasize the complex interaction of environmental factors influencing beach conditions, which is important for effective beach management and environmental protection. Lastly, comparing predicted values with actual observations enables the evaluation of the model's accuracy and reliability, facilitating further refinement and improvement of predictive capabilities, which contributes to more informed decision-making and improved beach monitoring practices.

5 Discussion

In this section, it is important to explore the implications of the analysis conducted on the Toronto beach observation dataset. The examination of various numerical variables alongside macro variables such as beach names and years, offers valuable insights into beach conditions and turbidity trends over time. The focus on two specific beaches with extensive data coverage allows for a better understanding of the beach environments.

5.1 Investigating the Dynamics of Water Quality in Toronto's Water Bodies

Environmental research depends on understanding the complex interaction among various factors influencing water quality. This study explores the intricate dynamics shaping water quality in Toronto's water bodies, with a specific focus on turbidity as the dependent variable. Turbidity, representing water clarity, serves as a key indicator of overall water quality and ecosystem health. However, this study recognizes the multilevel nature of water quality and aims to explore the influence of several independent variables on turbidity levels. Seven key variables were identified for analysis, each offering unique insights into Toronto's aquatic ecosystems. While turbidity is central to the analysis, other variables such as beach name, data collection date, waterfowl presence, wave action, rainfall amount, wind speed, air temperature, and water temperature were also considered due to their potential to influence turbidity levels. These variables were selected based on their relevance and significance in understanding the complex dynamics of water quality. Over a thirteen-year period from 2008 to 2023, this study seeks to reveal the intricate relationships among these variables, providing valuable insights into the factors shaping water quality in Toronto's water bodies. Through rigorous analysis and exploration, this research aims to contribute to a deeper understanding of the environmental processes driving water quality dynamics and inform effective management and conservation strategies for Toronto's aquatic ecosystems.

5.2 Comparative Analysis of Turbidity Trends and Environmental Factors

The "Toronto Beach Observation Dataset" offers a repository of data from ten diverse beaches across Toronto. However, due to limited observation data for many beaches, a focused comparative analysis was conducted using complete records from Marie Curtis East Beach and Sunny Side Beach. This analysis primarily examined data from 2010 to 2023, enabling a detailed exploration of turbidity trends and environmental factors influencing water quality. The findings revealed significant differences in turbidity levels between Marie Curtis Park East Beach and Sunny Side Beach, with the former consistently showing higher levels through the observation period. This variation suggests potential differences in water quality and environmental conditions between the two beach locations. Additionally, the analysis investigated the relationships between turbidity and various environmental factors, such as wind speed, air temperature, waterfowl presence, rainfall amount, and wave action. By illustrating these relationships, the study aimed to examine the intricate dynamics influencing turbidity levels in Toronto's beaches. Insights derived from this analysis provide valuable information for understanding the factors influencing water quality variations and can guide management and conservation efforts to improve beach environments and safeguard public health and safety.

5.3 Model Training, Prediction, and Evaluation

The modeling process began with dividing the dataset into 80% for training and 20% for prediction, enabling the development of a multi-linear regression model to examine the relationship between environmental factors and water turbidity in Toronto's beaches. Utilizing the large-scale Toronto Beach Observation Dataset, the model was trained to identify patterns and associations among various independent variables and turbidity levels. Subsequently, the model's predictive capabilities were evaluated using the predict() function in R, providing valuable insights into its reliability in estimating environmental characteristics across diverse beaches in Toronto. Comparing predicted values with observed data confirmed the model's accuracy, showing its potential as a valuable tool for beach monitoring and management. However, the analysis also found negative values in some prediction outcomes, indicating potential inaccuracies or limitations in the model's predictive abilities. These inconsistencies suggest further investigation is needed to understand underlying causes, such as unconsidered variables, measurement errors, or complexities in the relationship between environmental factors and water turbidity.

5.4 Insights from Linear Regression Analysis and Predictive Modeling

The estimated coefficients derived from linear regression analysis provide insights into the relationships between environmental factors and beach water turbidity. These coefficients offer information on the strength and direction of these relationships, for better understanding the complex forces governing water quality in coastal ecosystems. By quantifying the influence of each variable on turbidity levels, stakeholders can prioritize and address key environmental factors affecting beach health. Moreover, predictive modeling utilizes these coefficients to formulate an equation depicting the probabilistic relationships between environmental variables and turbidity. Despite the complexity of multiple variables, this equation can be used as a foundational framework for understanding the interaction between environmental factors and turbidity levels. However, it's important to acknowledge uncertainties and challenges associated with predictive modeling, including data quality and external variables not considered. Continuous validation and refinement of the model are necessary to enhance predictive capabilities and ensure applicability across diverse beach environments. Overall, integrating linear regression analysis and predictive modeling offers an effective approach to understanding and managing water quality in beach ecosystems, enabling informed decision-making and proactive management strategies for coastal environments.

5.5 Weakness and Implications for Future Research

While this study offers valuable insights into the relationships between environmental factors and water turbidity in Toronto's beaches, several limitations suggest methods for future research. Firstly, reliance on observational data collected by city staff may bring about biases and limitations. Variations in data collection protocols, equipment calibration, and observer expertise could impact data strength. Lessening these biases by standardized protocols and increased training for data collectors could improve data quality. Secondly, focusing on turbidity may overlook other parameters influencing beach health and ecosystem situation, such as nutrient concentrations and microbial contamination. Future research could adopt a multidimensional approach to make thorough assessments. Thirdly, although linear regression provides a useful framework, it may oversimplify complex relationships inherent in natural systems. Exploring advanced modeling techniques could examine these complexities more accurately. Extending the time scope beyond the study's thirteen-year period and broadening the geographical focus beyond Toronto could increase the generalizability and applicability of future research findings.

References

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Kirill Müller, Jennifer Bryan. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2022. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2022. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.