Correlates of water quality factors and amphibian species richness in counties along Lake Ontario, Canada

Alina Chen, Vedika Jha, Sol Park, Fangming Teng

Introduction

Amphibians are a class of cold-blooded vertebrates that depend heavily on aquatic ecosystems for survival and reproduction (Wake & Koo, 2018). Therefore, these animals are considered good indicators of environmental quality, and most importantly, water quality (Odum & Zippel, 2008). Many amphibian species have been going extinct within the past three to four decades predominantly due to water quality decline and habitat destruction from urbanization and the resulting pollution of waterways (Boyer & Grue, 1995; Walsh et al., 2005).

Low water quality including an increase in chemical contaminants from urbanization seems to be the leading cause of amphibian diversity and abundance decline (Calderon et al., 2019). This effect is largely due to amphibians' skin permeability which makes them highly susceptible to alterations in water quality (Boyer & Grue, 1995). The pattern of urbanization and runoff of chemical pollutants correlating with decreased amphibian species richness has been identified in many locations across the globe from tourist localities in Argentina and Mexico, as well as areas in Ontario, Canada including the Algoma, Muskoka, Sudbury area, and Holland Marsh (Calderon et al., 2019; Loder et al., 2019). However, it is yet to be explored in the Lake Ontario localities.

Of the various physical and chemical characteristics of water quality, there appears to be consensus across multiple studies that an increase in nitrate concentration in water results in a decrease in amphibian species richness and abundance (Rouse et al., 1999; Calderon et al., 2019). High nitrate concentrations can result in low egg mass density, low hatching success, and physical and morphological abnormalities among amphibians (Hatch & Blaustein, 2000; Camargo et al., 2005). The sources of these nitrates likely come from the runoff of nitrogen fertilizers and human and industrial waste (Calderon et al., 2019).

Some previous research that has found low-pH being a strong predictor of amphibian diversity and abundance, whereas a recent study by Loder et al. found otherwise when they detected amphibians in low-pH water bodies and generally found that there was no relationship between water chemistry and amphibian diversity or presence in central Ontario (Loder et al., 2019). Although some research found a negative correlation between phosphate concentrations and amphibian species richness and abundance, overall there seems to be no consensus on the effect of phosphate on amphibian communities (Calderon et al., 2019). It is inferred that the relationship with phosphate is more complex and indirect as it is linked to eutrophication which can cause anoxic conditions (Calderon et al., 2019).

The water quality of aquatic environments in eastern Canada should be given special focus as the soils are thin, poorly buffered, and have a past history of high acidic deposition that can result in the increased runaway of nitrate and other chemicals into lakes and wetlands (Loder et al., 2019). Despite previous findings, improvements in water quality including acid status for major water bodies in eastern Canada reveal little chemical recovery (Loder et al., 2019). To address this problem, an updated and thorough analysis of the relationship between water quality and species richness in the Lake Ontario area is needed. Species richness, the number of species inhabiting a community is the simplest but an important measure of species diversity because the greater the biodiversity in an ecosystem, the greater the stability when encountering disturbances such as urbanization (Albrecht et al., 2021).

Given this, the objectives of this study were to determine the variations in water quality factors across eight counties surrounding the Lake Ontario region in Canada and to examine the effects of water quality factors on amphibian species richness in these eight counties. We will test two hypotheses: 1) The values of different indicators of water quality vary between counties along Lake Ontario in Canada 2) Lake Ontario water quality measured within 5km of the county's shore is correlated with the amphibian species richness of that county. If water quality is correlated with amphibian species richness, we predict poor water quality will be correlated with lower species richness of amphibians in a given county. Initially, we did not know which water quality factors would be included in our hypothesis test, thus we ran a principal component analysis (PCA) in order to determine which factors to keep for downstream analysis. However, based on the findings of previous studies, we predicted that important measures of water quality will include pH and nitrate concentration.

Methods

Data description and manipulation

<u>Amphibian data:</u>

The amphibian dataset that we used for this project was obtained from the Great Lakes Marsh Monitoring programme in NatureCounts, which contains volunteer-collected data concerning amphibians sighted in marshes throughout the Great Lakes Basin, the United States, and Canada since 1994, with 91 700 observations as of November 20, 2022. This dataset is useful in determining the long-term changes in species richness of the amphibian communities in the wetlands.

Our hypothesis centres around Lake Ontario within Canada, thus we removed data from the United States of America (USA). We then removed columns such as catalog number and institution code that did not provide any pertinent information for our analyses. The columns we kept were YearCollected, County, ScientificName, Genus, SpecificEpithet, Locality, DecimalLongitude, DecimalLatitude, and Family. After removing rows that lacked species information, we had 43,577 observations left.

Next, we filtered for the counties surrounding Lake Ontario: Durham, Frontenac, Lennox-Addington, Prince Edward, Northumberland, Toronto, Peel, Halton, Hamilton-Wentworth, and Niagara, which were identified using Google Maps. Since species richness can change over time and we are interested in a recent analysis of species richness, we only included observations from 2015 to 2019, which left us with 10 species in the counties on Lake Ontario. In addition, there were the most observations of amphibians in the years 2015-2019 with the exception of 2008. We were then left with 4420 observations for analyzing the amphibian species richness of each county. We grouped the data by county to determine the number of observations of different species in each county using the *dplyr* package in R (Figure 1). Finally, we recorded the number of species per county to combine it with the water quality dataset (Table S1 in Supplementary Material).

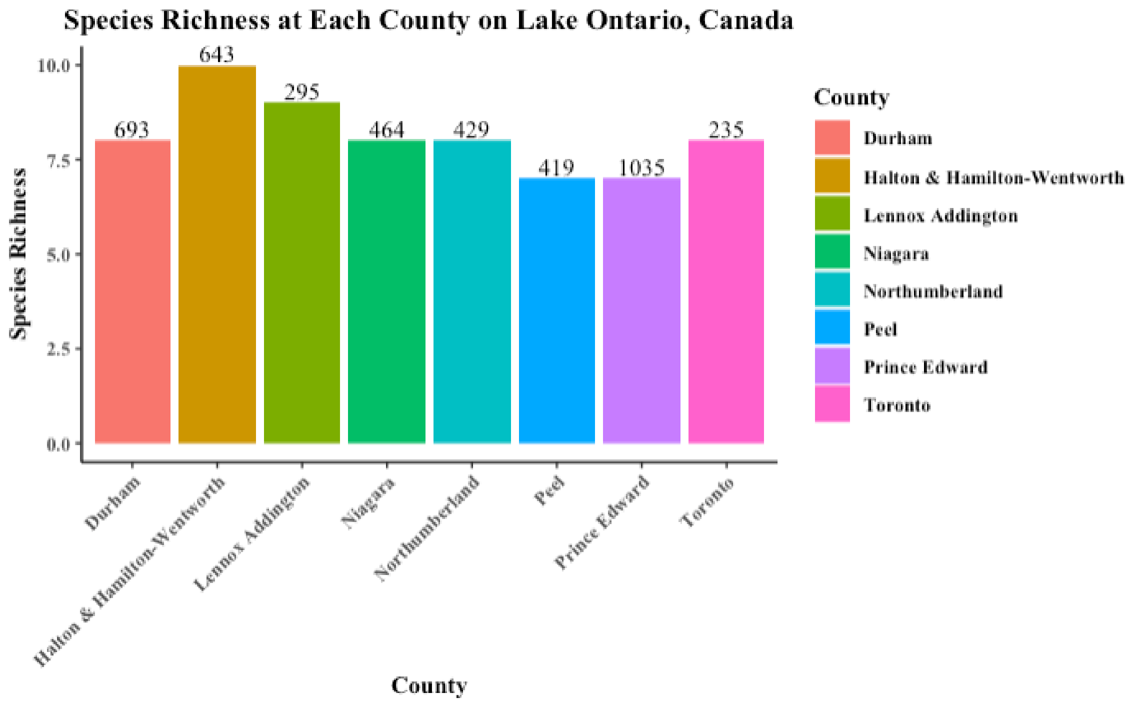


Figure 1. Species richness at each county on Lake Ontario, Canada. The numbers above each bar represent the number of observations per county between 2015-2019. Frontenac county is not present in this figure as this county was removed in the following step. Halton and Hamilton-Wentworth were combined into one county due to shared water quality data.

Water quality data:

The Government of Canada provides water quality and surveillance monitoring data from Lake Ontario from 2000 to 2021. Each site measures several different factors indicative of water quality, including concentrations of chemicals such as phosphorus, iron, and dissolved oxygen, or scalar measures such as conductance, temperature, transparency, and pH. In total, about 1000 different factors were measured, though many were only measured at a few sites.

To match with the manipulation done in the amphibian dataset and because water quality changes over time, we only kept observations from 2015-2019 for the water quality data. We disregarded water quality measurement factors with fewer than 100 observations, as we would not be able to perform proper statistical analyses with factors with too few observations. These two filtering steps reduced the number of water quality factors from over 1000 to 38. We then separated each factor into its own column for analysis. Values were averaged for factors that had multiple replicates measured.

We filtered for data close to our selected counties by importing the dataset's longitude and latitude values into Google Maps and discarding any points in the USA or further than 5km from the shoreline. The remaining data points were mapped to the closest county and exported. Figure 2 shows the original water quality data (from 2015 to 2019) and the filtered water quality data. At this stage, we had to combine two neighbouring counties (Halton and Hamilton-Wentworth) due to shared water quality data, and the county Frontenac was removed due to lack of water quality data, resulting in 8 total counties in our analysis. The county data was combined with the water quality data.

After this stage of filtering, we removed more water quality factors, either due to high numbers of missing values, a lack of biological relevance, or because another factor that was retained measured the same value with a different method. The factors excluded at this stage are summarized in Table S2. We also had to remove all data from the year 2019 due to several missing factor measurements, including biologically important measures such as nitrate and ammonia. This left us with 17 factors that each represented different indicators of water quality and 83 total observations. We visually assessed for outliers in each water quality metric by generating a boxplot (Figure S1). However, no outliers were removed for downstream analysis as this would remove too many observations in the following PCA step and would lead to an even smaller sample size. We also calculated the means of each factor per county to use in our regression.

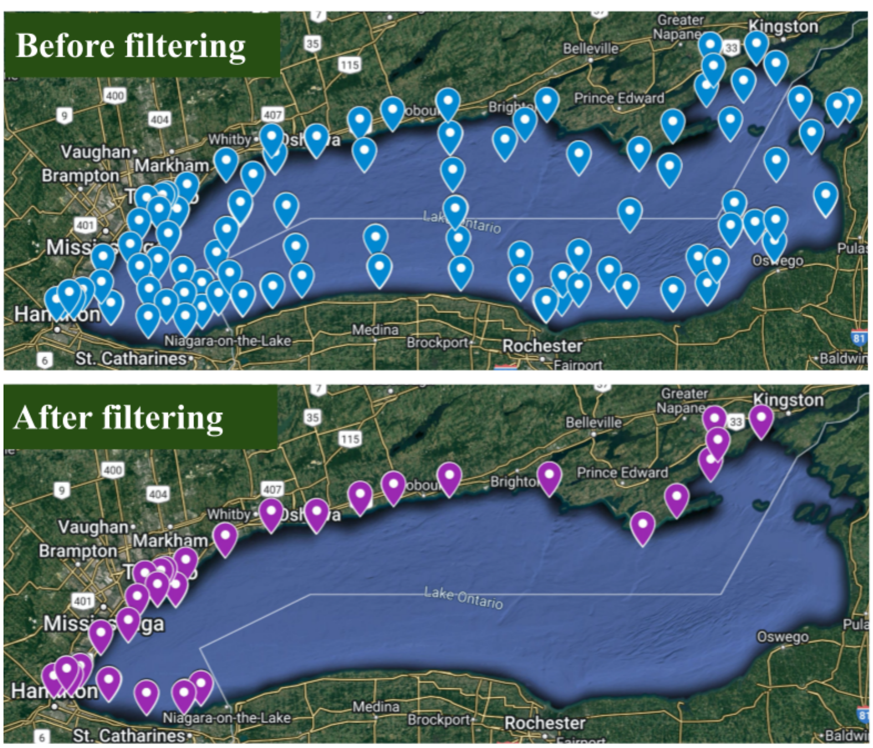


Figure 2. Map of points for water quality data. The top figure displays all locations of water quality metrics taken from 2015-2019 from the Great Lakes Water Quality Monitoring and Surveillance Data. The bottom figure shows the water quality measurement locations kept for analysis after filtering for locations that were within 5 km of each county.

Data analysis

Principal Component Analysis (PCA) was performed to examine the correlation between various water quality metrics. We conducted one PCA using individual data points and one using the means of water quality metrics as this was needed for the linear regression to compare species richness by county. The first principal component (PC1) that explained most of the variation was extracted, converted into a data frame, and combined with the water quality data to use as a predictor of amphibian species richness for downstream analysis. We used the *ggfortify* package in R to visualize the PCA and the general correlation between water quality factors to determine which factors may be used for further analysis.

To test the first hypothesis, a one-way analysis of variance (ANOVA) was performed to examine if water quality significantly differs between the eight counties, and a Tukey's honestly significant difference (Tukey's HSD) test was performed to test pairwise differences of counties if results of the ANOVA were significant. Before performing all the ANOVA and Tukey's HSD tests, Shapiro's test was performed to test the assumption of normality and the Levene's test was performed to test the assumption of homogeneity of variance.

To test the second hypothesis, whether water quality factors are correlated with species richness, we performed a multiple linear regression. We also plotted a residual vs. fitted plot and a Q-Q plot to test the assumption of homogeneity of variance and the assumption of normality of a linear regression respectively. We also performed log, squared, cubed, square root, and reciprocal transformations to test if these would better fit the assumptions of a linear regression. We used the ggpredict function from the *ggiraphExtra* package in R to visualize our final multiple linear regression model.

Results

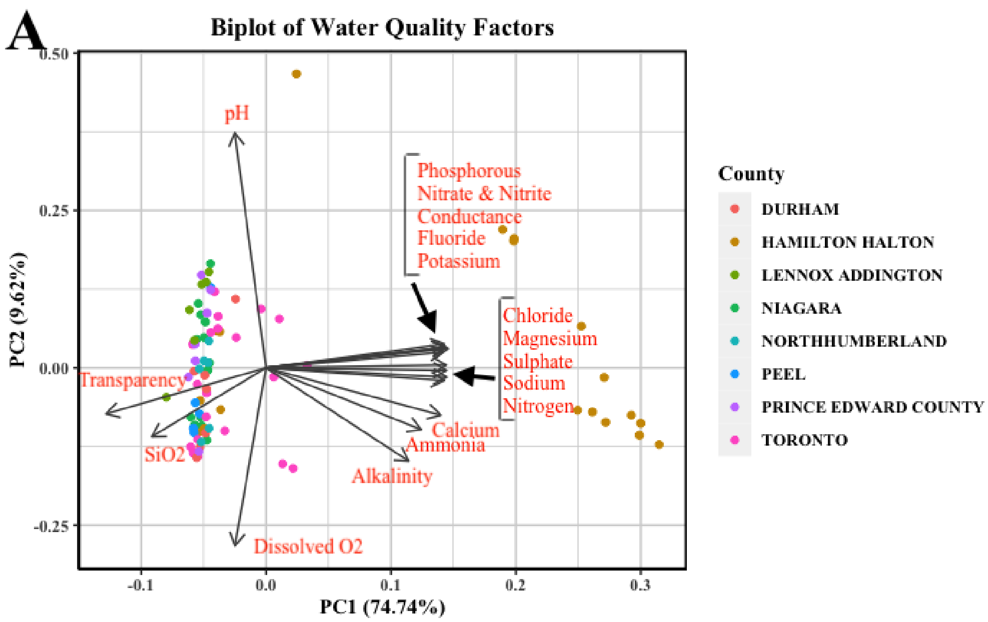
Overview of results

In order to test our two hypotheses, we conducted three main types of analyses: PCA, ANOVA, and linear regression. Our PCA yielded two variables for downstream analysis: the PC1 axis, which represented "water chemistry," and pH. Our ANOVA testing if water quality differs significantly between the counties showed a significant difference in water chemistry between counties but no significant difference in pH between counties. Our linear regression testing if water quality is correlated with species richness showed a significant positive relationship between water chemistry and species richness and no relationship between pH and species richness.

Principal Component Analysis (PCA)

As mentioned, we conducted two PCAs, one using individual data points (n = 83) and a second using mean values of water quality factors of each county (n = 8) in order to fit into our multiple linear regression model. We found that the two PCA show similar patterns (Figure 3). In both PCAs, most of the water quality factors load strongly on PC1 with the exception of pH and concentration of dissolved oxygen that load strongly on PC2. Most chemicals in water (nitrogen, nitrate & nitrite, alkalinity, filtered fluoride, chloride, sulphate, phosphorus, calcium, potassium, magnesium, and sodium) were correlated with each other in PC1, which we will refer to as “water chemistry”. Transparency and silica negatively correlated with these chemicals and dissolved oxygen concentration negatively correlated with pH. The only difference between the PCA of all samples vs. the PCA of means was that PC1 of the mean data explained more variation (85.98%) than PC1 of all the samples (74.74%). The grouped county, Halton and Hamilton-Wentworth had higher water chemistry compared to other counties as seen in the PCA plots.

Based on both PCAs, we chose to retain PC1 (representing water chemistry) for downstream analysis. In order to capture the variation among PC2, we chose to retain pH in the downstream analysis as an additional measure of water quality, as it contributed the most to PC2 variation. Additionally, previous studies have indicated that pH is an important measure of water quality (Freda et al., 1991).



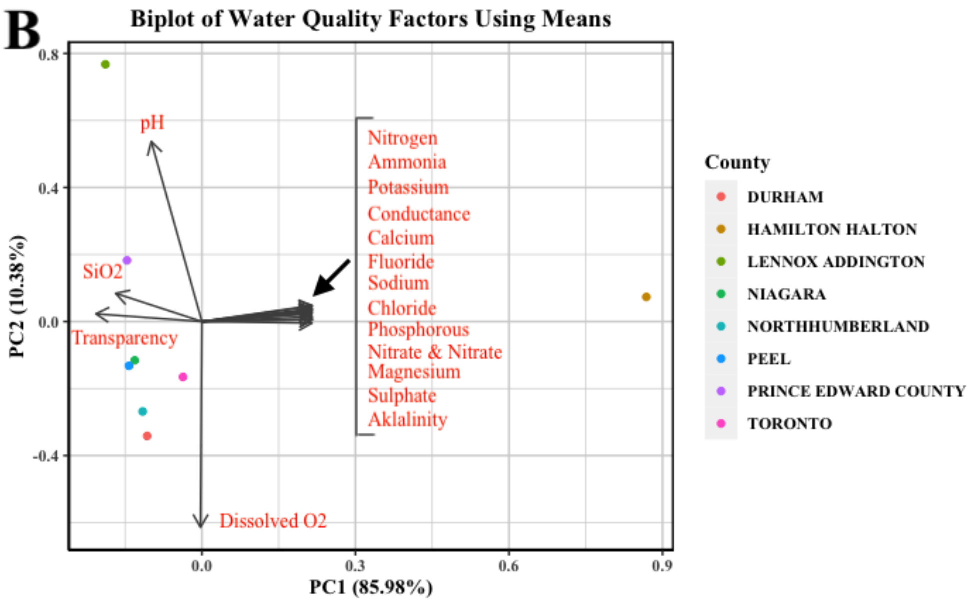


Figure 3. Biplot of water quality factors. Grouped labels are shown in descending order of PC2 axis values. SiO2 is shorthand for Silica, and dissolved O2 is shorthand for dissolved oxygen. A) Shows results when all values were used. PC1 explains 74.74% of the data's variation, and the majority of the factors vary mostly along the PC1 axis (compared to PC2). B) shows results when mean values for each county were used, and shows similar patterns to A, with PC1 explaining 85.98% of the variation. PC1 was retained for downstream analysis as well as pH. County colouring shows that Hamilton Halton county contains more extreme values on the PC1 axis compared to all other counties.

ANOVA

We performed a one-way ANOVA to test if water quality factors differ significantly by county using the PC1 axis (water chemistry) and pH. The ANOVA result shows that the PC1 axis differs significantly between counties ($F(7, 75) = [14.25]$, $p < 0.001$) while pH shows no significant difference between counties ($F(7, 75) = [1.354]$, $p = 0.238$). We then conducted a Tukey's HSD test on the water chemistry ANOVA and found that seven pairs of counties had means in water chemistry that were significantly different at $p < 0.05$.

We then investigated whether these tests met the assumptions of an ANOVA. For the PC1 axis, the Shapiro-Wilk test showed that the water chemistry data is non-normally distributed ($W = 0.56465$, $p < 0.001$) (Table S3). Levene's test indicated that our PC1 axis data does not have equal variances ($F = 12.766$, $p < 0.001$). For pH, the Shapiro-Wilk test demonstrated that pH is non-normally distributed ($W = 0.94936$, $p = 0.00254$) (Table S4). Levene's Test indicated equal variance in pH ($F = 0.7932$, $p = 0.5953$). We assumed independence of observations as each water quality data was taken independently of one another. Our dataset likely violated these assumptions due to the small sample size.

Table 1. The one-way ANOVA results for water chemistry between counties.

	df	Sums of Squares	Mean Square	F	p
County	7	594.7	84.95	14.25	1.26e-11***
Residuals	75	447.2	5.96		

* p < 0.05, ** p < 0.01, ***p < 0.001

Table 2. The one-way ANOVA results for pH between counties.

	df	Sums of Squares	Mean Square	F	p
County	7	0.1459	0.02084	1.354	0.238
Residuals	75	1.1548	0.01540		

Multiple linear regression

Multiple linear regression was used to examine the relationship between water chemistry and pH on amphibian species richness The fitted regression model was: $Species\ Richness\ =\ -\ 93.47013\ +\ 0.24484(Water\ Chemistry)\ +\ 12.02407(pH)$. The overall regression was statistically significant ($R^2 = 0.5878$, $F(2, 5) = 5.991$, $p = 0.04704$). The results of the regression show a significant correlation between water chemistry and amphibian species richness ($\beta = 0.24484$, $p = 0.018$) (Table 3) showing that the greater the water chemistry values (i.e. higher concentration of chemicals in water), the higher the species richness (Figure 3). However, pH did not have a significant correlation with species richness ($\beta = 12.02407$, $p = 0.178$).

When testing the assumptions of a linear regression, the residual plot showed that the assumption of homogeneity of variances is violated (Figure S2). This is likely due to our small sample size. Performing transformations did not remedy this violation, thus we chose to use our un-transformed data in analysis. However, the normal Q-Q plot showed that the data is normally distributed.

Table 3. Results of multiple linear regression using water chemistry and pH as a predictor of amphibian species richness.

	Estimate	Standard Error	t	p
Intercept	-93.47013	64.86655	-1.441	0.209
Water Chemistry	0.24484	0.07073	3.461	0.018*
pH	12.02407	7.67709	1.566	0.178

Residual standard error	0.6363 on 5 degrees of freedom
Multiple R^2	0.7056
Adjusted R^2	0.5878
F-statistic	5.991 on 2 and 5 DF, p-value: 0.04704*

* p < 0.05, ** p < 0.01, ***p < 0.001

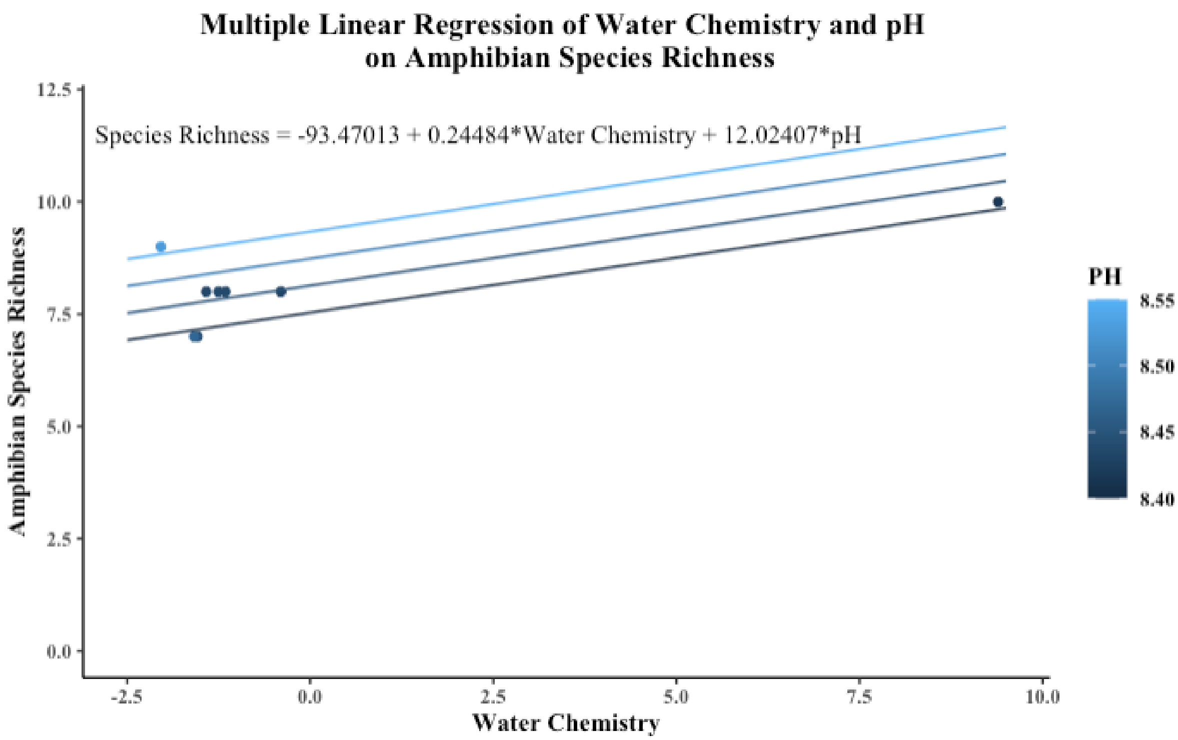


Figure 4. A multiple linear regression model of water chemistry and pH on amphibian species richness based on counties near Lake Ontario, Canada. The overall regression was statistically significant ($p < 0.05$). An increase in water chemistry is significantly correlated with an increase in amphibian species richness ($p < 0.05$). Despite the positive relationship between pH and amphibian species richness, this was not significant according to the model ($p = 0.178$).

Discussion

Overview

In this study, we explored the relationship between Lake Ontario water quality and amphibian species richness across Ontario counties located on Lake Ontario. We hypothesized that water quality would differ significantly between counties, and we hypothesized that decreased water quality will decrease species richness. To test these hypotheses, we first conducted a PCA on our complete water quality data (n=83) as well as the county means of each water quality factor (n = 8) to assess the relationship and importance of 17 different measures of water quality, and select important measures for downstream analysis. We selected the axis explaining most of the variation (PC1) to represent "water chemistry" for downstream analysis, as well as pH. Our ANOVA testing whether water quality differed significantly between counties found significant differences in water chemistry but no significant differences in pH. Our linear regression assessing the correlation between water quality and amphibian species richness found a significant positive correlation between water chemistry and species richness, but no significant relationship between pH and species richness. The correlation between water chemistry and species richness showed that increased water chemistry, representing an increased concentration of several chemicals such as nitrate, potassium, and phosphorus, was related to increased species richness.

These results meant that we could not accept either of our hypotheses unless we disregard pH as a measure of water quality. Even when assessing the significant correlation between water chemistry and species relationship, the direction of correlation went against our prediction. We predicted that as water quality increases, species richness would also increase. However, based on previous literature, increasing values of our water chemistry variable should represent decreasing water quality, and thus we see an inverse correlation in which decreasing water quality increases species richness. There are many possible explanations for these results, which may be due to differential effects of various components of water chemistry, complex environmental dynamics, small sample size, or limitations on water quality measures.

Interpretation of our results is complicated due to the makeup of the water chemistry variable. We have grouped different measures of water quality together because they all vary together, but different components may have different impacts on the environment and the amphibian ecosystem. Previous literature has investigated the effects of various chemicals on water quality, as well as the effects of water quality factors on species richness, and for most of the chemicals in our water chemistry variable, there is consensus on their detrimental effects on organisms when present in water. For example, there is consensus within the literature that chemicals such as nitrate decrease water quality and decrease species richness, including amphibian species richness specifically (Rouse et al., 1999). Similar trends can be seen for chloride, potassium, and other chemicals (Sadowski, 2010). In this way, our data disagrees with the literature. However, some chemicals may have more complex impacts on environments that are less well understood. For example, phosphorus is generally perceived as being detrimental to an ecosystem as it can cause eutrophication and algal bloom, but its effects on amphibian communities are poorly understood (Egea-Serrano et al., 2012).

One main limitation of this study is the small sample size, which may be an alternative explanation for the results. Despite our initial large numbers of samples for both water quality and amphibian data, we were constrained by missing values in the water quality data as well as only examining species richness for 8 counties bordering Lake Ontario, within which there is little variation in the diversity of species present. The small sample size limits the power of our statistical tests to assess significant differences and means extreme values have a larger impact on the results of a given test. Particularly, our test results were largely influenced by Hamilton-Wentworth county, which showed much larger values for water chemistry compared to every other county (Figure 2; Figure 3), and was the only county that showed significant differences in water chemistry with other counties (according to the Tukey test). Hamilton-Wentworth also had the largest species richness, and thus largely contributed to the result of our linear regression, which found that as water chemistry increases so does species richness. A larger sample size would allow us to better discern whether Hamilton-Wentworth's water chemistry and species richness values could be considered outliers or if they represent a broader trend.

Another limitation of this study is that our water quality data does not measure the water quality of the amphibians' marsh habitats but the water quality of a site in Lake Ontario closest to each marsh's county. Since we were unable to find data on marsh water quality, we chose to use lake data. The water quality of the lake is related to the water quality of its watersheds (including marshes). It can also be indicative of overarching patterns of pollution, urbanization, and runoff in a given area, factors that can all impact marsh water quality. However, there still may be other factors such as the distance of a marsh from the nearest urban centre, the population density near a marsh, and the types of industries present near a marsh can locally impact the water chemistry and thus water quality of a given marsh. This may be another reason why our results do not agree with our hypothesis or other studies, the majority of which measure water quality directly in the marsh or other amphibian habitats.

The last limitation is that our species richness data comes from citizen science data. Although citizen science projects such as the Great Lakes Marsh Monitoring can provide a vital resource for scientists, it is important to note that these data may be unreliable due to a lack of consistency, scientific method, and an increased potential for bias (Jäckel et al., 2021). For example, our species richness data may be skewed towards marshes that are closer to large cities or otherwise easier to access, and thus volunteers visit more frequently.

Further directions

Future studies aiming to assess the relationship between water quality and amphibian species richness in Ontario could use water quality data from the marshes themselves for a more accurate assessment of water quality. Additionally, by dividing data more locally into specific marshes or localities within counties, future studies could increase their sample sizes and thus get a better sense of overall trends. With a larger sample size, it may be possible to split the water chemistry variable into individual variables and explore each factor's relationship with species richness separately, allowing us to have a more in-depth understanding of how water chemistry, and thus water quality, impact amphibian species richness. A deeper understanding of how water quality relates to species richness is necessary to inform conservation strategies that can protect amphibians around Lake Ontario from the negative effects of urbanization.

References

Albrecht, J., Peters, M. K., Becker, J. N., Behler, C., Classen, A., Ensslin, A., Ferger, S. W., Gebert, F., Gerschlauer, F., Helbig-Bonitz, M., Kindeketa, W. J., Kühnel, A., Mayr, A. V., Njovu, H. K., Pabst, H., Pommer, U., Röder, J., Rutten, G., Schellenberger Costa, D., … Schleuning, M. (2021). Species richness is more important for ecosystem functioning than species turnover along an elevational gradient. *Nature Ecology & Evolution*, 5(12), 1582–1593. 10.1038/s41559-021-01550-9

Boyer, R., & Grue, C. E. (1995). The need for water quality criteria for frogs. *Environmental health perspectives*, 103(4), 352–357. 10.1289/ehp.95103352

Calderon, M., Almeida, C., González, P., & Jofré, M. (2019). Influence of water quality and habitat conditions on amphibian community metrics in rivers affected by urban activity. *Urban Ecosystems*, 22. 10.1007/s11252-019-00862-w

Camargo, J., Alonso, A., & Salamanca, A. (2005). Nitrate Toxicity to Aquatic Animals: A Review With New Data for Freshwater Invertebrates. *Chemosphere*, 58, 1255-67. 10.1016/j.chemosphere.2004.10.044.

Freda, J., Sadinski, W.J. & Dunson, W.A. (1991). Long term monitoring of amphibian populations with respect to the effects of acidic deposition. *Water Air Soil Pollution*, 55, 445–462. 10.1007/BF00211205

Hatch, A. & Blaustein, A. (2000). Combined Effects of UV-B, Nitrate, and Low pH Reduce the Survival and Activity Level of Larval Cascades Frogs (Rana cascadae). *Archives of Environmental Contamination and Toxicology*, 39, 494–499. 10.1007/s002440010132

Jäckel, D., Mortega, K. G., Sturm, U., Brockmeyer, U., Khorramshahi, O., & Voigt-Heucke, S. L. (2021). Opportunities and limitations: A comparative analysis of citizen science and expert recordings for bioacoustic research. *PloS one*, 16(6). 10.1371/journal.pone.0253763

Loder, A. L., Weeber, R., Wong, S. N. P., Spooner, I. S., & Mallory, M. L. (2019). Correlates of Waterbody Characteristics and the Occurrence or Diversity of Larval Amphibians in Central Ontario, Canada. *Bulletin of environmental contamination and toxicology*, 103(4), 571–578. 10.1007/s00128-019-02698-8

Odum, R. A. & Zippel, K. C. (2008). Amphibian water quality: approaches to an essential environmental parameter. *International Zoo Yearbook*, 42, 40-52. 10.1111/j.1748-1090.2008.00053.x

Rouse, J. D., Bishop, C. A., & Struger, J. (1999). Nitrogen pollution: an assessment of its threat to amphibian survival. *Environmental health perspectives*, 107(10), 799–803. 10.1289/ehp.99107799

Sadowski, E.K. (2010). The impacts of chloride concentrations on wetlands and amphibian distribution in the Toronto region. *Prairie Perspectives*, 144.

Wake, D. B., & Koo, M. S. (2018). Amphibians. *Current biology: Cell Biology*, 28(21), R1237–R1241. 10.1016/j.cub.2018.09.028

Walsh, C. J., Roy, A., Feminella, J., Cottingham, P., Groffman, P., & Morgan II, R. (2005). The Urban Stream Syndrome: Current Knowledge and the Search For A Cure. *North American Benthological Society*, 24, 706-723, 10.1899/0887-3593(2005)024\[0706:TUSSCK\]2.0.CO;2.