# Association Between Chemical Characteristics of Stream Water and the Abundance of Fish in Northeast Minnesota

Daniel Ye

Swenson College of Science and Engineering

University of Minnesota Duluth

Mentor: Dr. Igor Melnykov

## Abstract

To study the association between the chemical characteristics of stream water and the abundance of fish, a canonical correlation analysis was carried out between the three major chemicals of streams and the fish counts of the five most common fish species in northeast Minnesota. The chemicals are dissolved oxygen, pH, and total phosphorus. The fish species were Central Mudminnow, Creek Chub, White Sucker, Common Shiner, and Blacknose Dace. The result indicates that Common Shiner has the most extensive adaptability among the five fish species, has nearly no preference for dissolved oxygen. Creek Chub and Blacknose Dace, White Sucker and Common Shiner have similar preferences for water chemistry. In this way, proper regional stocking strategies can be made with a scientific basis to avoid futile activities.


*Keywords*: Canonical correlation analysis · Chemical characteristics · Fish abundance · Minnesota

## Introduction

The chemical characteristics of water have a complex effect on fish, as different fish species require different living conditions. Typically, the chemical characteristics of a body of water remain stable. However, increasing human activities can introduce substantial changes in water chemistry [1]. For example, pesticides, fertilizers, and soil from farm fields drain into lakes and rivers, killing aquatic insects and depleting dissolved oxygen [2]. As the current conditions of the body of water change, it may no longer be suitable for sustaining certain existing fish.

The enhancement of fisheries through stocking or introduction of fish is a practice frequently used by fisheries owners and scientists worldwide [3]. Every year, the Minnesota Department of Natural Resources stocks local lakes with young and adult fish, including walleye, crappie, trout, to increase or maintain healthy fish populations. More than 400,000 rainbow trout, 40,000 brown trout, 40,000 brook trout, 35,000 steelheads, and 2,200 lake trout were stocked in lakes and rivers in Minnesota in 2020 [4].

However, it appears to be hard to control whether the stocking strategy is appropriate, as many stocking programs are carried out without regular evaluation of the potential or actual success of the exercise [5]. If the fish were placed into water with less-than-optimal parameters, their survival rate would significantly decrease, threatening fish abundance and wasting money. Therefore, determining what water conditions are beneficial or detrimental to specific fish species has important implications.

This study investigated the association between chemical characteristics and the abundance of the five most common fish in northeast Minnesota by canonical correlation analysis. The analysis aims to provide a scientific basis for regional fish stocking strategies in northeast Minnesota.

## Materials and Methods

### Overview of the study area

The study area is northeast Minnesota (Figure 1) [6], also known as the Arrowhead Region. It covers an area of more than one-fifth of Minnesota, with 18,221.97 square miles of land area, and includes Aitkin, Itasca, Koochiching, Carlton, Cook, Lake, and St. Louis counties. The region is dotted with hundreds of lakes surrounded by forest and is home to 162 fish species, including Mudminnow, Walleye, Dace, Bass.

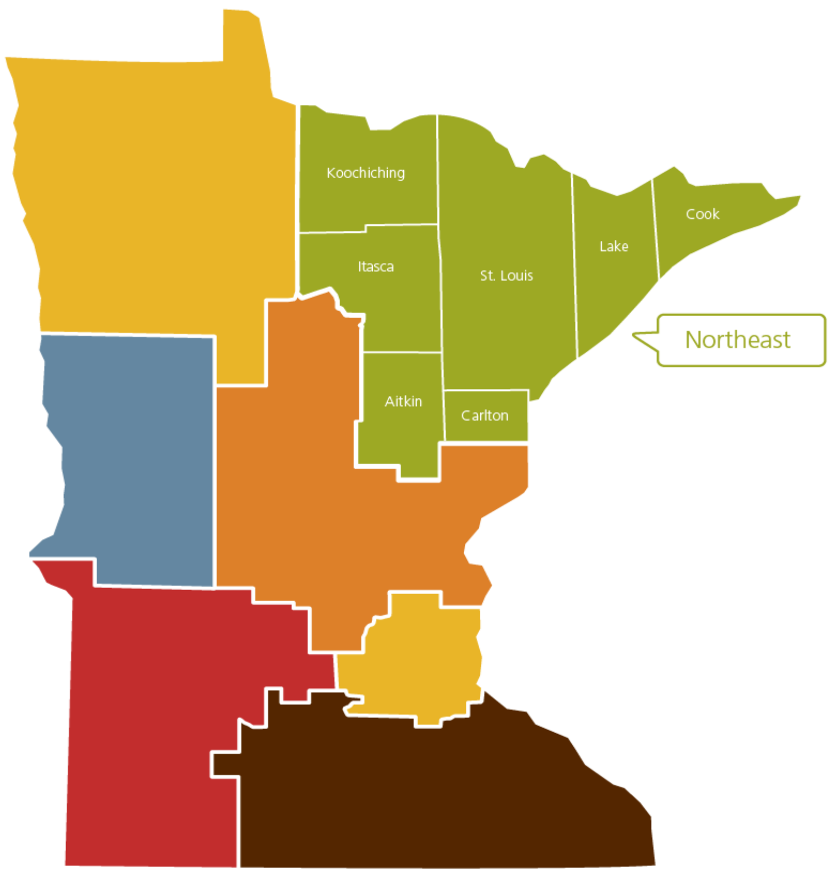


**Figure 1. Northeast Minnesota**

A large portion of the local economy depends on tourism—northeast Minnesota is a popular fishing spot for the Minneapolis–Saint Paul metropolitan area. In 2018, about 222,400 anglers in northeast Minnesota spent $322.6 Million on fishing-related purchases [7]. Therefore, it is of great significance to study the impact of chemical characteristics of water on popular fish species in this area.

**Data acquisition**

The data comes from the records collected by individual monitoring stations on the Minnesota Pollution Control Agency website. The chemical characteristics of water come from the major chemical records, including dissolved oxygen (mg/L), pH, and total phosphorus (mg/L). The fish count records in the corresponding streams show the abundance of fish, which count the estimates of the actual numbers of passing fish. The five most common fish in the local streams were selected as study subjects: Central Mudminnow, Creek Chub, White Sucker, Common Shiner, and Blacknose Dace. In total, 80 streams in the seven-county region were recorded. The time interval of data is from 2010 to 2019.

**Data processing**

Data processing is carried out using canonical correlation analysis (CCA) through R and the package 'CCA.' Canonical correlation analysis is used to reveal the relationship between two multivariate sets of random variables. Compared with other statistical methods of correlation, canonical correlation analysis is particularly suitable for analyzing the standard variable set with a strong correlation between the variables themselves. This relationship can be expressed as a linear combination between more than two $x$ variables in set 1 and more than two $y$ variables in set 2, which correlates to the maximum [8].

The expressions are as follows:

$$\mathrm{U} = a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_kx_k \tag{1}$$

$$\mathrm{W} = b_1y_1 + b_2y_2 + b_3y_3 + \cdots + b_my_m \tag{2}$$

Here, U and W are the canonical variables, $\mathrm{a_i}$ and $b_j$ (i = 1,2,3..., k; j = 1,2,3..., m) are the coefficients of linear combinations. The data processing was performed by the R package 'CCA.'

Canonical dimensions, also known as canonical variates, are latent variables analogous to factors obtained in factor analysis. The number of canonical dimensions is generally equal to the number of variables in the smaller set. However, the number of significant dimensions maybe even smaller [9]. The R package "CCP" was performed to test the significance. The most significant dimensions would be selected to get the canonical correlation expressions.

In the canonical correlation expressions, the variables belong to two groups: water chemicals and fish counts in the corresponding stream. The chemicals in the water are represented by $x_1,\ x_2,\ x_3$, which are defined as dissolved oxygen (mg/L), pH, and total phosphorus (mg/L), respectively. The fish counts are represented by $\mathrm{y}_1,\ \mathrm{y}_2,\ \mathrm{y}_3,\ \mathrm{y}_4,\ \mathrm{y}_5$, which are defined as the fish counts of Central Mudminnow Creek Chub, White Sucker, Common Shiner, and Blacknose Dace, respectively. Ultimately, the coefficients $a_i$ and $b_j$ in expressions (1) and (2) are taken as the main variables to explain the association of different chemicals on different fish populations.

Variables with relatively larger canonical coefficients contribute more to the canonical variables. Similarly, variables whose canonical coefficients have opposite signs show an inverse relationship with each other, and variables with canonical coefficients of the same sign show a direct relationship [10].

Outliers in this study are decided by the standard score, shown as formula (3). $\mu$ and $\sigma$ are the population mean and standard deviation of an overserved value, respectively. The standard score is calculated by subtracting $\mu$ and then dividing the difference by $\sigma$. It is used to quantify the unusualness of the observation when data follows the normal distribution [11].

$$Z = (\frac{x-\mu}{\sigma}) \tag{3}$$

Chemicals and fish counts are assumed to have approximately a normal distribution since normal distribution occurs when many factors influence an outcome simultaneously. The fish count is affected by temperature, chemicals, size of streams, general health of the fish, etc. For any given fish, some of these factors are positive, and some are negative. Most fish counts are close to the average, indicating the fish count approximately satisfies normal distribution**.** Similarly, the amounts of chemicals in the water are result from multiple factors, satisfying normal distribution.

Data with a standard score greater than two or smaller than negative two is recognized as an outlier and excluded from the study. As a result, one stream was excluded because of abnormal recording of total phosphorus, and twelve streams were excluded because of abnormal recording of fish counts.

## Results and Discussion

Canonical correlation analysis was carried on the 67 streams in northeast Minnesota.

For this model, there are three canonical dimensions of which only the first one is statistically significant. The result of significance tests is shown in table 1. It indicates that the first test of the canonical dimensions is significant, with F-value equal to 1.857, P-value equal to 0.031. The second test of significance is insignificant, with F-value equal to 0.511, P-value equal to 0.846. Finally, the last test tests dimension 3 is insignificant, with F-value equal to 0.254, P-value equal to 0.858.

| | Stat | F-value | df1 | df2 | P-value |
|---|---|---|---|---|---|
| 1 to 3 | 0.642853 | 1.857411 | 15 | 160.5139 | 0.031202 |
| 2 to 3 | 0.934146 | 0.511061 | 8 | 118 | 0.846096 |
| 3 to 3 | 0.98746 | 0.253976 | 3 | 60 | 0.858202 |

**Table 1. Test of Significance**

The canonical correlation expressions are shown below as equations 4 and 5.

$$U_1 = 0.283x_1 + 0.687x_2 - 19.407x_3 \tag{4}$$

$$V_1 = -0.047y_1 + 0.009y_2 - 0.015y_3 + 0.003y_4 + 0.008y_5 \quad (5)$$

Among the three observed variables of $x$, $x_3$ has the most significant negative association with the canonical variables $U_1$, $x_2$ has a medium positive association, and $x_1$ has the lowest association. Among the five observed variables of $y$, $y_1$ has the most significant negative association with the canonical variable $V_1$, $y_3$ have a medium negative association, and the other three indicate insignificant association.

The result is visualized by R package 'ggplot2' into biplot (Figure 2). Each dot represents a stream, and the colors of dots represent the year of collection. Each vector represents a variable. Black vectors represent the chemicals, and pink vectors represent the fish species. Vectors in the opposite direction represent negative correlations. A pair of perpendicular vectors show a zero correlation.

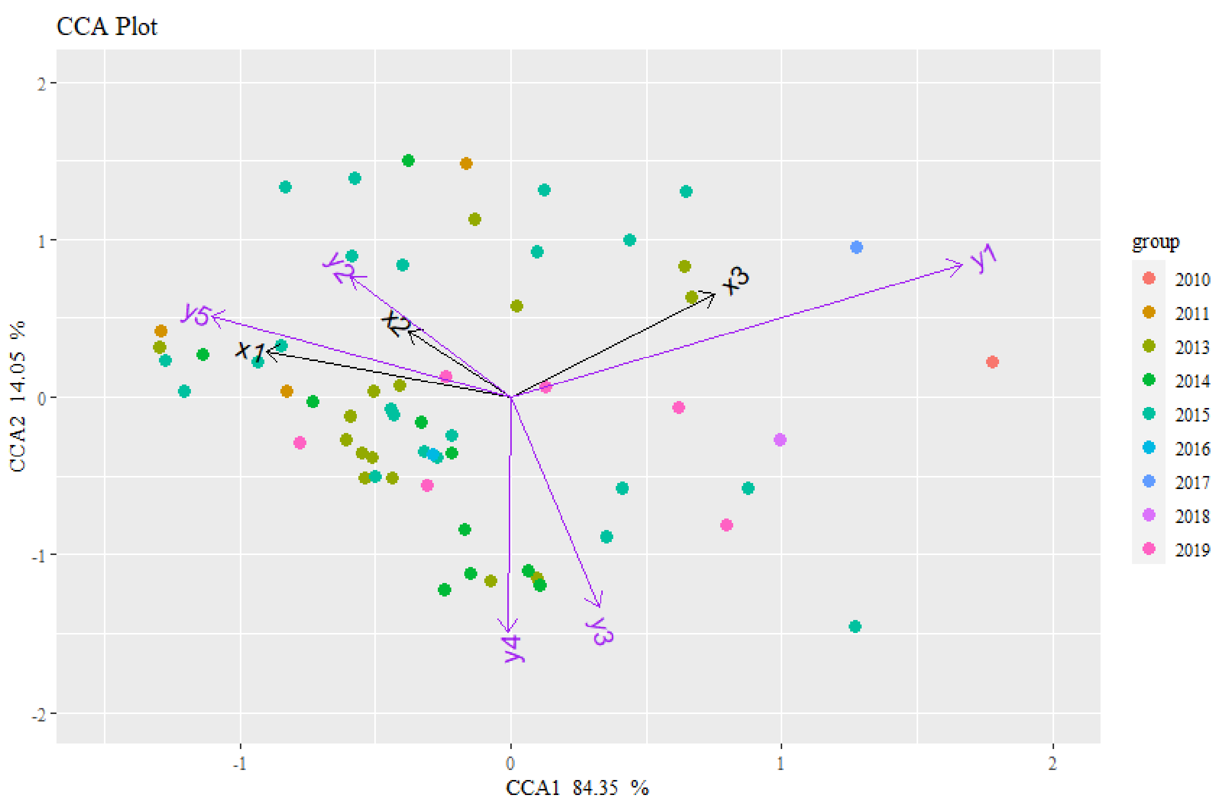


**Figure 2. Biplot of CCA (fish, chemical)**

The more similar the directions of two vectors are, the more positively correlated they are. If a dot has a greater magnitude in a particular vector direction, the stream has a greater value in the corresponding valuable. For instance, the orange dot on the extreme right in Figure 2 represents the Gale Brook stream in Itasca County. Gale Brook's data was collected in 2010, with 1.51 mg/L dissolved oxygen, 6.98 pH, 0.155 mg/L total phosphorus. In the stream, Central Mudminnow has a fish count of 70, Creek Chub has a fish count of 10, White Sucker has a fish count of 1, Common Shiner has a fish count of 1, Blacknose Dace has a fish count of 0.

The result indicates that Central Mudminnow has a strong resilience with high total phosphorus, a moderate resilience with low dissolved oxygen. Creek Chub strongly prefers habitats with higher pH, slightly prefers habitats with higher dissolved oxygen. White Sucker strongly prefers habitats with lower pH, has a moderate resilience with low dissolved oxygen, no preference for total phosphorus. Common Shiner slightly prefers habitats with lower pH, has nearly no preference for dissolved oxygen. Blacknose Dace strongly prefers habitats with higher dissolved oxygen and pH, has a moderate resilience with low total phosphorus.

In general, Common Shiner has the most extensive adaptability among the five fish species. Creek Chub and Blacknose Dace, White Sucker and Common Shiner have similar preferences for water chemistry. In this way, regional stocking strategy can be properly made with a scientific basis to avoid futile stocking activities. It is recommended that all stocking activities can be processed before thorough assessments to maximize the effects and benefits.

## Reference


1. Daskalov, Georgi M.: “Long-term Changes in Fish Abundance and Environmental Indices in the Black Sea”
2. Minnesota Department of Natural Resources: “Improve and Protect Habitat”
3. Cowx, Ian G.: “Stocking strategies. Fisheries Management and Ecology 1: 15–31”
4. Minnesota Department of Natural Resources: “DNR’s 2020 fish stocking successful, despite challenges”
5. Cowx, Ian G.: “An appraisal of stocking strategies in the light of developing country constraints.” Fisheries Management and Ecology 6: 21-34. Science for Environment Policy: Effects of Fish Extinction on Ecosystems
6. Blandin Foundation: Rural Pulse™ Snapshot: Northeast Minnesota, retrieved from https://blandinfoundation.org/learn/rural-pulse-2019/northeast/
7. American Sportfishing Association: “Economic Impacts of Recreational Fishing - Minnesota”
8. Johnson, Richard a. & Wichern, Dean W.: “Applied Multivariate Statistical Analysis Six Edition”
9. UCLA Institute for Digital Research & Education: “Canonical Correlation Analysis | R Data Analysis Examples”
10. Dattalo, Patrick V.: “A Demonstration of Canonical Correlation Analysis with Orthogonal Rotation to Facilitate Interpretation”
11. Frost, Jim: “5 Ways to Find Outliers in Your Data”