

Article

Non-Parametric Analysis of Nitrogen Trends in the form of Nitrate and Nitrite in Rivers and Streams of the Contiguous United States for 1990-2019

Amin Mohebbi ¹ and Simin Akbariyeh ²

¹ Department of Civil Engineering, Construction Management, and Environmental Engineering, Northern Arizona University, 2112 S Huffer Ln, Flagstaff, AZ 86011; amin.mohebbi@nau.edu

² School of Engineering, Brown University, 184 Hope Street, Providence, RI 02912; simin_akbariyeh@brown.edu

* Correspondence: simin_akbariyeh@brown.edu

Abstract: Nitrogen and phosphorous support the ecosystem by supplying nutrients to algae and aquatic plants. Having them in excess results in the eutrophication of waters creating quality problems. In the past, nitrogen has been widely investigated for wells in the context of groundwater flow. However, a national-scale nitrogen assessment in rivers and streams has not received enough attention. In this research, the Wilcoxon rank sum test, as a non-parametric hypothesis testing method, has been applied to nitrogen concentration in the form of nitrate-nitrogen and nitrite-nitrogen in rivers and streams of the Contiguous United States. This approach was particularly selected because of the non-normal and positively skewed nitrogen levels occurring in the surface flow. This method was able to identify the impaired body of waters as well as quantify the confidence, significance, and errors involved. The Northern Appalachians (NAP), Northern Plains (NPL), and Xeric (XER) ecoregions were worsening in the nitrogen-nitrate condition with NAP, and XER needed immediate actions. The nitrite-nitrogen condition did not pose an immediate threat, so mitigation plans should focus more on nitrate-nitrogen remediation. It was shown that the method was superior to the two-sample t-test by yielding lower type II errors.

Keywords: Nitrate-Nitrogen; Nitrite-Nitrogen; Non-Parametric Hypothesis Testing, Type II Error, Rivers and Streams

1. Introduction

Nitrogen and phosphorous in excess are considered contaminants in water quality [1]. These excess nutrients cause the rapid growth of oxygen-utilizing organisms such as algae [2]. Their uncontrolled growth can cover the water surface, reducing the surface reaeration capacity [3]. In addition, their high rate of oxygen consumption does not allow water to reaerate fast enough (even with proper exposure), resulting in hypoxic regions in water bodies [4]. These hypoxic regions cause fish and other aquatic organisms issues, creating an imbalance in the ecosystem [5]. That is why the topic of nutrients as contaminants has been rigorously studied since they were designated as a water quality issue [6].

Nitrogen, in particular, has been the focus of numerous research with different agendas. In particular, in the area of excess nutrients, limiting nutrient [7], sources of excess nitrogen in fertilizers and septic tanks [2], best management practices to control nitrogen from fertilizers and septic tanks [8,9], and impacts of other hydrological parameters such as precipitation on nitrogen concentration [10] have been well documented. These works target mainly groundwater flow as it directly receives nitrogen in different ionic forms such as nitrate-nitrogen ($\text{NO}_3\text{-N}$) and nitrite-nitrogen ($\text{NO}_2\text{-N}$) from infiltration [11,12].

Nitrogen in surface water bodies, including lakes, ponds, oceans, rivers, and streams, has received less attention in the past. The available literature in this area usually does not have sufficient spatial coverage (regional case studies) [13,14] or temporal coverage (only

a year or two years) [15]. While acknowledging these excellent research projects, a national scale assessment with monthly, yearly, and decadal timescales seems justified. Although there are a few national-scale studies, their scope, objective, and approach are different from this project [16–18].

This research aims to assess the quality of the Contiguous United States (CONUS) rivers and streams in terms of nitrogen. The project analyzes nitrogen data trends in the form of nitrate-nitrogen and nitrite-nitrogen for three decades (1990-2019). This analysis is performed by one of the lesser-known hypothesis testing methods called the Wilcoxon rank sum test, which is a non-parametric method. It is expected that this method would provide a better understanding of how the current nitrogen trends are compared to the past and identifying the impaired surface water bodies with a more reliable method.

The study is intended to be concise, applied, and straightforward. With that in mind, first, the methodology, including the study area, data source, and governing equations of the Wilcoxon rank sum test, are covered. After the proposed method is applied, the results are discussed not only on how they address the research goal but also on how they complement the past work. The study concludes by listing contributions that could benefit both researchers and practitioners.

2. Materials and Methods

In this section, the methodology used in the current study is briefly discussed in three distinct sections. First, the study area's geographical bounds and the period chosen for the project are presented. Next, the data processing procedure, a combination of spatial and temporal analysis, is elaborated. For spatial analysis, the codes are developed in ArcMap v10 (<https://www.esri.com/>) with Python v2.7 (<https://www.python.org/>), and for temporal analysis, the codes including the functions and the syntaxes were developed in MATLAB v2019b (<https://www.mathworks.com/>). Finally, a brief introduction to hypothesis testing, particularly the Wilcoxon rank sum test, is discussed.

2.1. Study Spatial and Temporal Extends

The research was conducted on the Contiguous United States (CONUS) rivers and streams over the National Rivers and Streams Assessment (NRSA) reporting regions [15]. NRSA categorizes the CONUS into nine regions based on the ecology and their key stressors. These regions are Northern Appalachians (NAP), Southern Appalachians (SAP), Coastal Plains (CP), Upper Midwest (UMW), Temperate Plains (TPL), Southern Plains (SPL), Northern Plains (NPL), Western Mountains (WMT), and Xeric (XER) (Figure 1). These ecoregions were adopted in the current research, but the data associated with them did not have enough temporal coverage (2008-2009 and 2013-2014) <https://www.epa.gov/national-aquatic-resource-surveys> and therefore was not used. Instead, the data from National Water Quality Monitoring Council (NWQMC) (<https://www.waterqualitydata.us/portal/>) with a longer-term temporal coverage were more suited for this research (1990-2019). This portal offers various water quality measures, including nitrate-nitrogen ($\text{NO}_3\text{-N}$) and nitrite-nitrogen ($\text{NO}_2\text{-N}$) in rivers and streams as part of the United States Environmental Protection Agency's (US-EPA) STORAGE and RETrieval (STORET) campaign. These data went through a rigorous control to assess their quality, spot their anomalies, and understand their distribution.

2.2. Data Quality Control

The datasets acquired from the NWQMC were initially aggregated into monthly data by extracting the month from the date variable and then isolating $\text{NO}_2\text{-N}$ and $\text{NO}_3\text{-N}$ concentration levels based on the extracted month. Note that averaging of any kind was not applied to preserve the data in its original form. Therefore, the term aggregate was used to group the concentration readings into months. This process was performed recursively

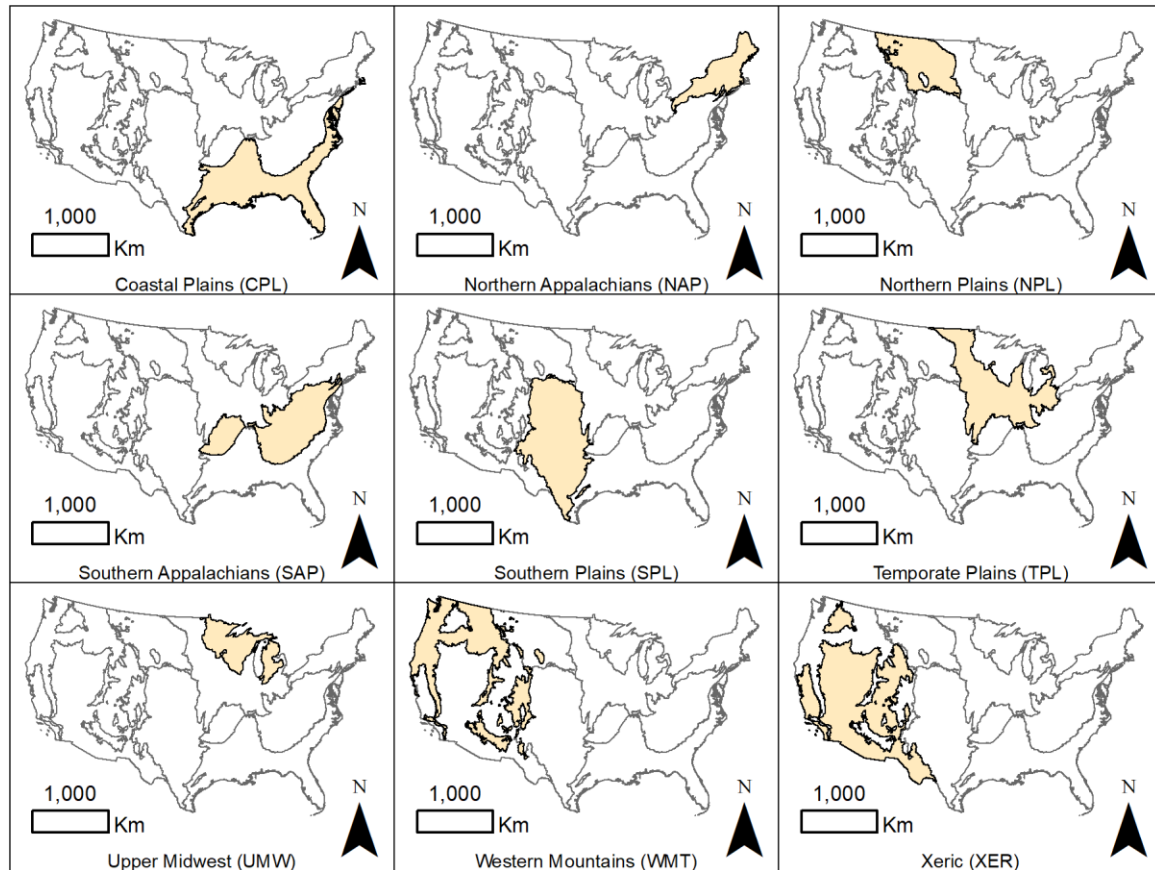


Figure 1. National Rivers and Streams Assessment (NRSA) ecoregions used in the current study

for three 10-year periods (1990-1999, 2000-2009, and 2010-2019), resulting in 12 (months) * 3 (decades) * 2 (NO₂-N and NO₃-N) = 72 datasets. Next, these datasets were superimposed on NRSA regions to be separated based on their location. Nine NRSA reporting ecoregions and 72 datasets resulted in 648 total datasets which will be referred to as month-regions hereafter.

Inspection of the month-regions showed extreme values at both ends of the distribution with negative and/or extremely small values on the left tail and extremely large values on the right tail. These are most likely due to sensor malfunction or sensor operating out of its designed range. Either way, these values were not reliable and must be removed from the datasets prior to further analysis. A threshold value of 0.001 mg/l was defined to remove all the negative, zero, and small values. The large values were treated as outliers if they were more than three scaled Median Absolute Deviations (MAD) away from the median [19]. The scaled MAD is calculated by

$$\text{MAD} = c \times \text{median}|D - \text{median}(D)| \quad (1)$$

with

$$c = -1/[\sqrt{2} \times \text{erfcinv}(3/2)] \quad (2)$$

where *erfcinv* is the Inverse complementary error function. These processes reduced the sample sizes by 10-15% but created more reliable and well-behaved samples that are of interest in hypothesis testing.

2.3. Hypothesis Testing

There are various hypothesis testing approaches in the literature, and each comes with its limitation and scope [20]. The widely used and easy-to-apply statistical software makes it easy to perform hypothesis testing. However, care must be taken when choosing

and applying these classical statistical tools. In particular, all the assumptions must be met, and the approach must be relevant to the problem under study. The former is well understood, but the latter requires a priori knowledge of the problem.

It was briefly discussed in the Data Quality Control section that the project data is observational and therefore not as well behaved as the data generated from mathematical models. This means typical hypothesis testing such as Student's t-test and z-test are not appropriate tools. Due to multiple reasons such as data type, non-normality, outliers, and a priori knowledge about the solute transport in surface water, the Wilcoxon rank sum test was selected as the primary assessment tool, which is discussed in depth later.

The Wilcoxon rank sum test is a non-parametric test that is used to compare two independent populations. The non-parametric tests do not require the sample to be normally distributed or belong to a large dataset to satisfy the central limit theorem assumption¹. Also, this approach indirectly tests the medians rather than the means, which was more preferred in this study because of the extreme data skewness. The test requires mixing the observations from the two populations together while maintaining identifiers about which population they came from. Once mixed, a numeric rank is assigned to each data point from smallest to largest, and these ranks are summed up for each population. If the sum of ranks for two populations is significantly different, perhaps those two populations are not behaving similarly. The following steps are followed to apply this idea:

1. State the parameter of interest.
2. Define a null hypothesis (H_0): $D_1 = D_2$ where D_1 and D_2 are populations 1 and 2 distributions, respectively.
3. Define an alternate hypothesis (H_1): $D_1 \neq D_2$ or in a more specific form $D_1 \geq D_2$ or $D_1 \leq D_2$.
4. Calculate z statistics if the population size is large enough (at least 10) and compare it to the critical z value by

$$z_0 = \frac{T_1 - E(T_1)}{\sqrt{V(T_1)}} \quad (3)$$

with

$$E(T_1) = \frac{n_1 n_2 + n_1(n_1 + 1)}{2} \quad (4)$$

and

$$V(T_1) = n_1 n_2 \left(\frac{n_1 + n_2 + 1}{12} \right) \quad (5)$$

where $E(T_1)$ is the expected value of population 1, $V(T_1)$ is the variance of population 1, n_1 and n_2 are sample sizes for population 1 and 2, respectively and T_1 and T_2 are the sum of ranks for populations 1 and 2, respectively. T_1 and T_2 are related by

$$T_1 + T_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} \quad (6)$$

where an increase in T_1 will cause a decrease in T_2 .

5. Calculate critical z value based on the significance level (α) and type of test, i.e., a one-tailed test or a two-tailed test. Alternatively, the p -value can be calculated and compared with the significance level. The p -value is the smallest level of significance used to reject or fail to reject the null hypothesis.
6. Compare the two values from step 4 and step 5 and reject or fail to reject the null hypothesis based on the outcome. In the z-score method, compare the z_0 to the critical z value and reject the null hypothesis if the $|z_0| < z_{\text{critical}}$ at the designated significance level. In the p -value method, compare the calculated p -value to the significance level. If $|p| < \alpha$ then reject the null hypothesis at the designated significance

¹ Central limit theorem states that for a large enough random samples taken from a population, the distribution of the sample means will be approximately normally distributed.

level. The rejection of the null hypothesis means supporting the alternate hypothesis in both cases.

7. State the outcome in practical terms. This may sound trivial, but rejection or failure to reject the null hypothesis does not mean a failure in the outcome. It all depends on how the null and alternate hypotheses are defined.

3. Results

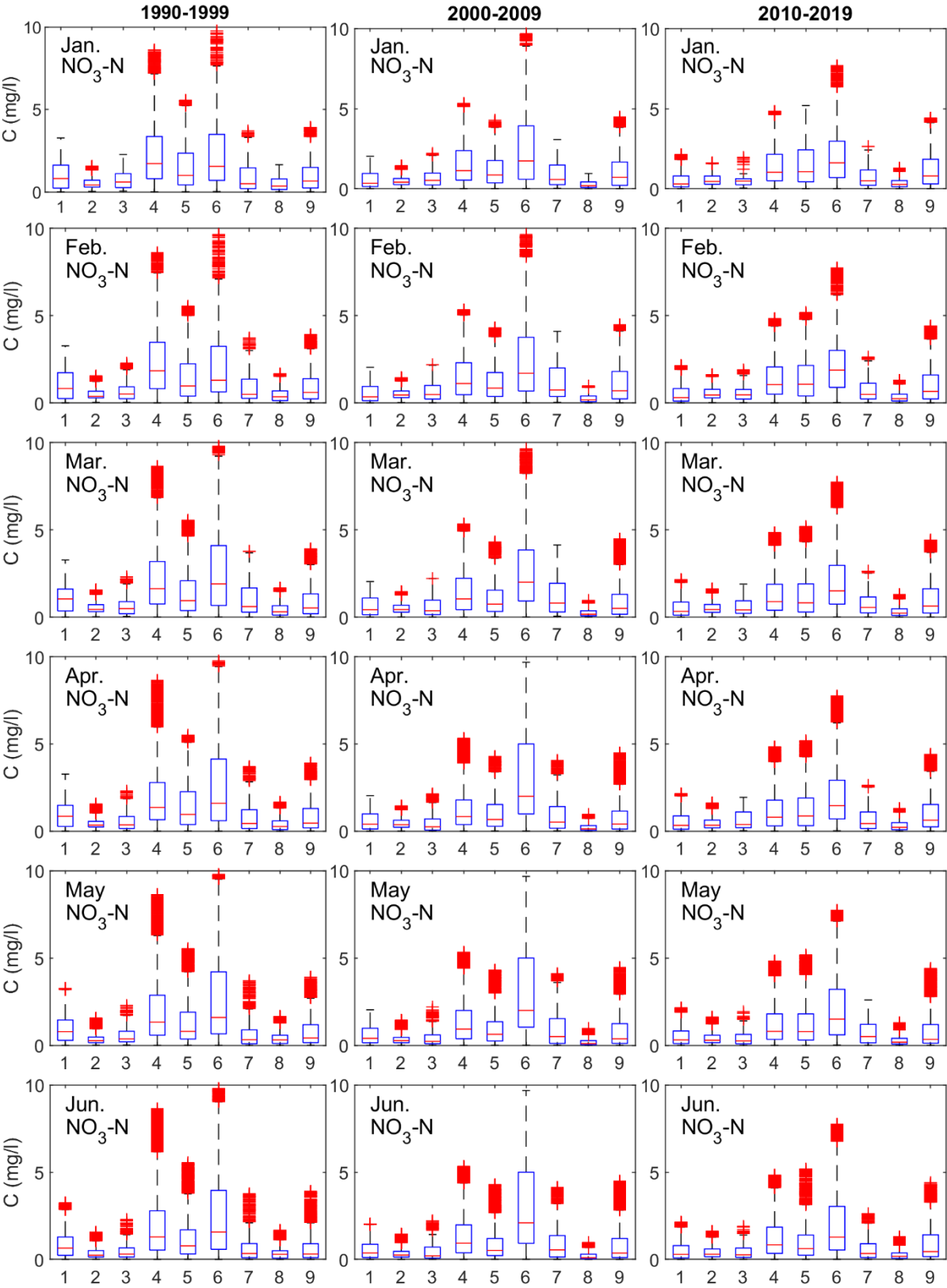
The methodology presented was applied to the raw datasets acquired from the NWQMC. The justification to use a non-parametric test was presented by conducting a feasibility study called the exploratory data analysis phase. This phase covers an in-depth study of data outliers, skewness, and non-normality, as well as descriptive statistics. The descriptive statistics focus on the data median, which is less susceptible to extreme values and is the main parameter of interest in the non-parametric hypothesis testing. Wilcoxon rank sum test was performed after the careful formulation of the null and alternate hypotheses next. Further, the possible errors involved in the process were assessed and compared by the conventional two-sample t-test.

3.1. Data Non-Normality

The box-and-whisker plot of month-regions are presented in Figure 2 for $\text{NO}_3\text{-N}$ and Figure 3 for $\text{NO}_2\text{-N}$, with the concentration (mg/l) in the y-axis, ecoregions (NAP, SAP, CP, UMW, TPL, SPL, NPL, WMT, and XER) paired with numbers 1 to 9 in the x-axis, months in the upper left corner legend, and decades as titles above each figure. These plots were generated by calculating dataset minimum (lower whisker), 25% percentile (box bottom, Q_1), 50% percentile (median, Q_2), 75% percentile (box top, Q_3), and maximum (upper whisker). Inspection of these plots revealed that the median was always closer to the first quartile and the bottom whisker was shorter than the upper whisker. These indicated high positive skewness, or in a technical term, a group of high-frequency and low values clustered around the left tail of the distribution. This was expected because, in water quality measurements, the low concentration levels are more frequently observed than high concentration levels. In addition, the data also failed the normality test of Kolmogorov-Smirnov [21] for all the scenarios, dictating that an assumption of normality for nitrogen concentration measurement in rivers was not valid.

Another observation was the abundance of the outliers denoted as the “+” sign in Figures 2 and 3 located above/below the upper/lower whiskers. These outliers were different from the ones that were already detected and removed by the scaled MAD method. They seemed to be spatially and temporally variable and were flagged based on the $\pm 1.5 * (Q_3 - Q_1)$ rule. These data were not removed from the analysis since they did not have the typical outlier characteristics encountered in the solute concentration measurements. Also, the scaled MAD method was a conservative approach to trim down only 10-15% of the data points as reported before.

According to Figures 2 and 3, it was also observed that there were barely any outliers in the $-1.5 * (Q_3 - Q_1)$ region, which was another indication of the positive skewness. In addition, the number of outliers in $\text{NO}_3\text{-N}$ was more than $\text{NO}_2\text{-N}$, which was attributed to the higher number of collected $\text{NO}_3\text{-N}$ samples. The total number of samples collected between 1990 to 2019 for $\text{NO}_3\text{-N}$ was 7 to 62% more than the number of the samples collected for $\text{NO}_2\text{-N}$. Another reason could be the higher range of the $\text{NO}_3\text{-N}$ samples compared to the $\text{NO}_2\text{-N}$ samples. The range was calculated by subtracting the maximum from the minimum concentration for each month. This range can be seen in Figures 2 and 3, where the $\text{NO}_3\text{-N}$ and $\text{NO}_2\text{-N}$ ranges were roughly between 0 and 10 mg/l and 0 to 0.3 mg/l, respectively.



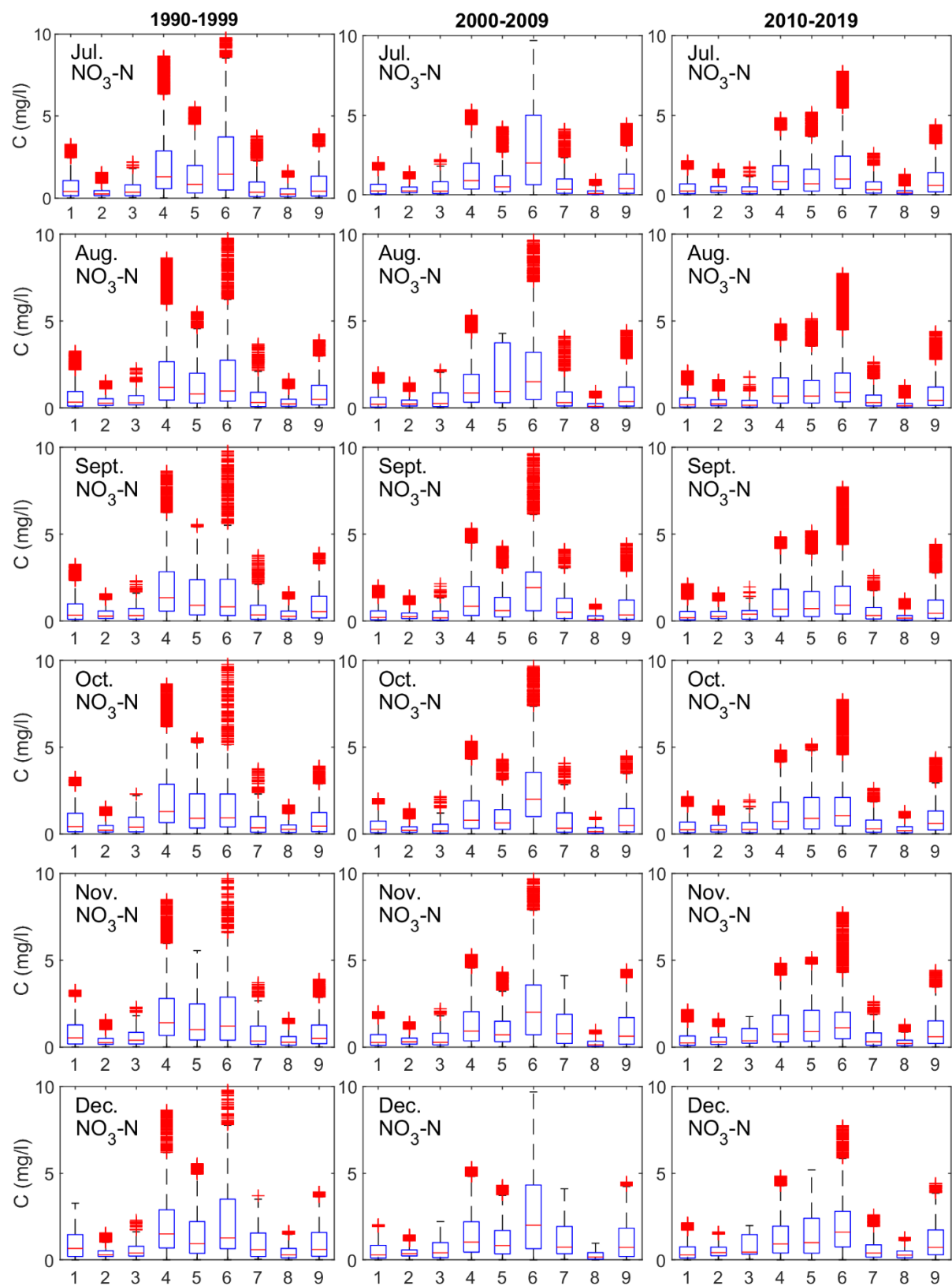
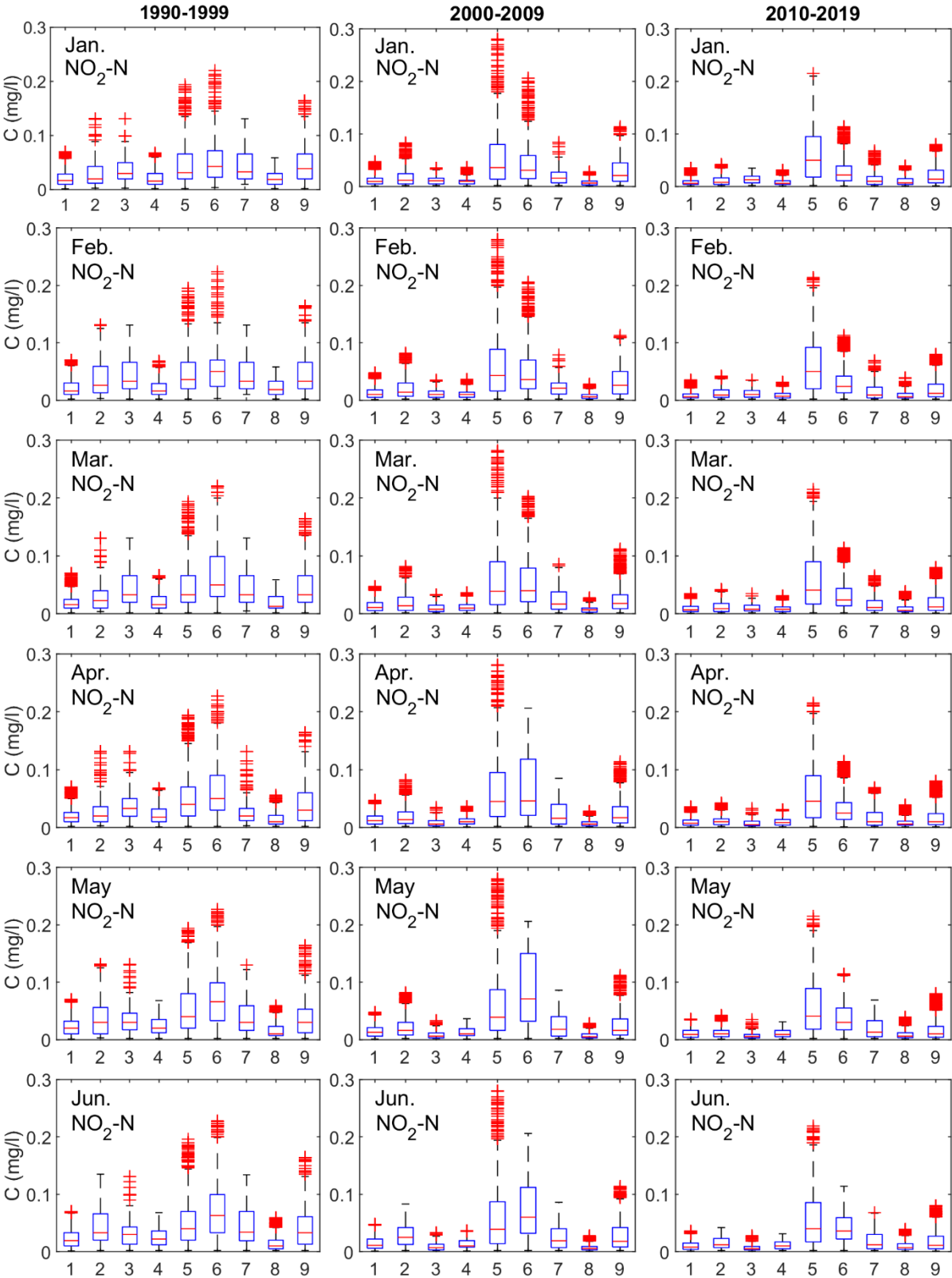


Figure 2. NO₃-N month-region measure of spread by box-and-whisker plot with the x-axis representing National Rivers and Streams Assessment ecoregions and the y-axis representing NO₃-N concentration (C) in mg/l



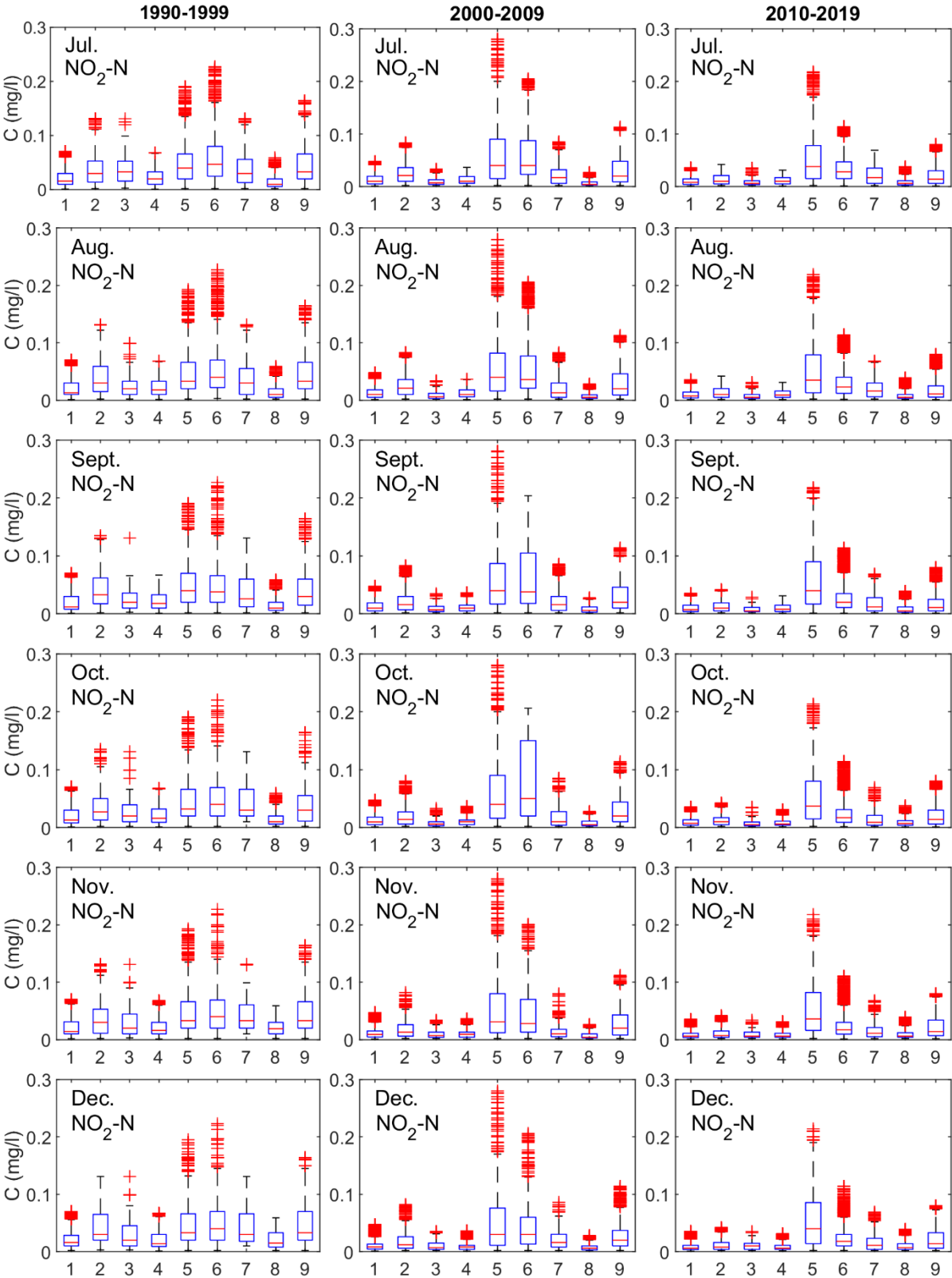


Figure 3. NO₂-N month-region measure of spread by box-and-whisker plot with the x-axis representing National Rivers and Streams Assessment ecoregions and the y-axis representing NO₂-N concentration (C) in mg/l

3.2. Data Seasonality and Trends in NRSA Regions

The monthly, decadal, and regional data trends are summarized in Figures 2 and 3 in the form of quartiles. The obvious observation was that the concentration was mainly a function of space (ecoregion) rather than time (month). This should not be inferred as that the monthly variation did not exist. In fact, later, it is demonstrated that the monthly variation was correlated to water year. However, compared to the regional variations, it can be neglected. Although this was a new finding since both months and regions were studied simultaneously, the main goal of this research was to identify the nutrient-impaired waters and quantify the rate at which they are deteriorating. To that end, 1990-1999 data were designated as the baseline and any increase/decrease in the following decades was relative to these years. The change in NO₂-N and NO₃-N medians for the 2000-2009 period were calculated based on

$$\text{change} = \frac{M_{1990-1999} - M_{2000-2009}}{M_{1990-1999}} \times 100 \tag{7}$$

where $M_{1990-1999}$ is the median of data for the 1990-1999 period and $M_{2000-2009}$ is the median of data for the 2000-2009 period. 2010-2019 period change was calculated in a similar manner. The study objective was to identify the increase in the nutrient levels, so a change with a positive sign was of interest.

The results are presented in Table 1 for NO₃-N and NO₂-N with years in the first column, NRSA ecoregions in the second column, and months in the following columns. The increase in level (positive change) is boxed to locate the impaired month-regions quickly. According to Table 1, the six regions of NAP, NPL, SPL, TPL, UMW, and XER had NO₃-N problems. Out of these regions, SPL and TPL also had NO₂-N problems (Table 1). The increase was between 2% to 137% for NO₃-N and between 3% to 59% for NO₂-N. In addition, Dec. (x10), Jan. (x7), and Feb (x8) had the most impaired waters compared to the other months.

The result presented here was solely based on comparing the data medians. While simple and effective, this method failed to quantify the impact of sample randomness. The identified month-regions must be further investigated, perhaps by more rigorous statistical methods.

3.3. Wilcoxon Rank Sum Test

The Wilcoxon rank sum test was applied to the datasets following the steps established in the Materials and Methods section. The null hypothesis was defined as no *significant* change in NO₃-N or NO₂-N levels compared to the baseline. The alternative hypothesis was defined while having the project goal in mind. As stated before, the goal of this research was to identify the impaired regions with the aggregated monthly data. So, the alternate hypothesis was defined as the increase in the levels significant enough to raise concerns. Therefore, the alternate hypothesis was written as $D_1 \leq D_2$ where D_1 was the 1990-1999 distribution and D_2 was 2000-2009 or 2010-2019 distributions. This required a one-tailed (because of \leq) z-test because of the large sample size. The results were expected to have some overlap with the median method and hopefully narrow down the number of impaired month-regions.

The results of hypothesis testing are listed in Table 2 for NO₃-N and NO₂-N in a similar style as the median method results. Typically, the outcome of hypothesis testing is to reject or fail to reject the null hypothesis. These were abbreviated to R and FtR in Table 2, though in the code written for this purpose, the logical values of 1 and 0 were generated. Like the median method, the Rs were boxed to distinguish the regions with increased NO₃-N and NO₂-N. A comparison was made between Table 1 and Table 2 for NO₃-N and NO₂-N. The results were quite interesting as the hypothesis testing had separated the actual impaired surface water from the ones that were flagged by the median method just

Table 1. NO₃-N and NO₂-N month-region median change from the baseline

Years	Region	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
-------	--------	------	------	------	------	-----	------	------	------	-------	------	------	------

NO ₃ -N Med. Diff. (2000-2009 vs 1990-1999)	CPL	-58%	-58%	-60%	-53%	-49%	-43%	-38%	-37%	-39%	-38%	-51%	-57%
	NAP	-3%	21%	-2%	8%	2%	4%	0%	-17%	-10%	-4%	17%	24%
	NPL	-13%	-6%	-25%	-31%	-38%	-32%	-34%	-13%	-43%	-57%	-31%	3%
	SAP	-34%	-40%	-36%	-39%	-30%	-28%	-31%	-27%	-37%	-39%	-35%	-31%
	SPL	-14%	-13%	-21%	-30%	-21%	-33%	-40%	18%	-34%	-30%	-30%	-11%
	TPL	12%	31%	5%	25%	25%	34%	38%	56%	137%	115%	67%	57%
	UMW	16%	52%	33%	18%	57%	61%	-1%	-3%	46%	-7%	125%	24%
	WMT	-46%	-52%	-47%	-51%	-67%	-66%	-67%	-63%	-66%	-50%	-56%	-42%
	XER	9%	16%	-5%	-11%	-11%	17%	-6%	-27%	-36%	9%	26%	21%
NO ₃ -N Med. Diff. (2010-2019 vs 1990-1999)	CPL	-63%	-63%	-69%	-61%	-61%	-55%	-39%	-45%	-45%	-43%	-56%	-57%
	NAP	8%	21%	-2%	-2%	9%	24%	4%	-15%	-7%	14%	15%	45%
	NPL	-20%	-9%	-15%	5%	-31%	-13%	-36%	-48%	26%	-34%	-9%	13%
	SAP	-40%	-43%	-46%	-41%	-40%	-35%	-37%	-42%	-49%	-44%	-47%	-39%
	SPL	6%	10%	-13%	-9%	-1%	-21%	-16%	-15%	-21%	-1%	-12%	7%
	TPL	5%	45%	-21%	-9%	-6%	-18%	-32%	-8%	11%	13%	-8%	26%
	UMW	0%	-1%	-8%	-1%	57%	0%	-6%	-3%	-9%	-15%	-12%	-32%
	WMT	-24%	-32%	-27%	-17%	-42%	-39%	-46%	-48%	-42%	-31%	-26%	-4%
	XER	21%	10%	21%	37%	-19%	45%	42%	-12%	-17%	35%	20%	21%
NO ₂ -N Med. Diff. (2000-2009 vs 1990-1999)	CPL	-43%	-38%	-31%	-29%	-36%	-43%	-38%	-23%	-17%	-25%	-37%	-48%
	NAP	-40%	-46%	-39%	-32%	-47%	-24%	-30%	-30%	-52%	-48%	-57%	-60%
	NPL	-63%	-70%	-76%	-82%	-80%	-77%	-79%	-70%	-65%	-70%	-65%	-65%
	SAP	-38%	-38%	-38%	-44%	-50%	-55%	-50%	-44%	-44%	-38%	-43%	-42%
	SPL	14%	19%	18%	13%	-3%	-3%	0%	21%	0%	25%	-6%	-9%
	TPL	-28%	-28%	-20%	-8%	8%	-5%	-15%	-10%	0%	25%	-30%	-25%
	UMW	-52%	-36%	-48%	-20%	-40%	-44%	-43%	-57%	-38%	-67%	-70%	-47%
	WMT	-65%	-68%	-46%	-40%	-50%	-45%	-60%	-50%	-40%	-50%	-79%	-60%
	XER	-46%	-21%	-45%	-43%	-47%	-45%	-39%	-39%	-33%	-33%	-39%	-39%
NO ₂ -N Med. Diff. (2010-2019 vs 1990-1999)	CPL	-65%	-63%	-56%	-56%	-55%	-58%	-50%	-42%	-40%	-44%	-58%	-63%
	NAP	-60%	-65%	-61%	-50%	-67%	-64%	-67%	-67%	-70%	-63%	-77%	-73%
	NPL	-57%	-70%	-76%	-85%	-82%	-83%	-82%	-75%	-75%	-75%	-70%	-50%
	SAP	-63%	-56%	-50%	-52%	-55%	-55%	-50%	-50%	-56%	-63%	-63%	-57%
	SPL	59%	39%	24%	14%	3%	0%	-5%	6%	0%	16%	9%	21%
	TPL	-49%	-52%	-52%	-50%	-55%	-43%	-40%	-43%	-47%	-58%	-58%	-55%
	UMW	-70%	-73%	-67%	-50%	-57%	-65%	-43%	-47%	-54%	-70%	-67%	-63%
	WMT	-63%	-68%	-54%	-40%	-40%	-30%	-40%	-50%	-50%	-30%	-68%	-53%
	XER	-64%	-64%	-64%	-67%	-67%	-67%	-58%	-67%	-63%	-53%	-58%	-58%

1990-1999 was designed as the baseline decade. Positive values are boxed.

because their median showed an increase relative to the baseline. The number of the impaired month-regions dropped from 64 to 44 for NO₃-N and 17 to 8 for NO₂-N.

The R or FtR was decided based on the significance level of $\alpha = 0.05$, a standard practice in engineering. The p -values calculated from z-statistics were compared to α ; the values lower than 0.05 fell in the rejection region, rejecting the null hypothesis of no significant change in the median. These p -values are listed in Table 3, with the values smaller

Table 2. NO₃-N and NO₂-N month-region median hypothesis testing using Wilcoxon rank sum test

Years	Region	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
-------	--------	------	------	------	------	-----	------	------	------	-------	------	------	------

NO ₃ -N H ₀ (2000-2009 vs 1990-1999)	CPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	NAP	FtR	R	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	R	R
	NPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	SAP	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	SPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	R	FtR	FtR	FtR	FtR
	TPL	FtR	R	FtR	R	R	R	R	R	R	R	R	R
	UMW	FtR	R	FtR	R	R	R	FtR	FtR	R	FtR	R	R
	WMT	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	XER	FtR	R	FtR	FtR	FtR	R	FtR	FtR	FtR	FtR	R	R
NO ₃ -N H ₀ (2010-2019 vs 1990-1999)	CPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	NAP	FtR	R	FtR	FtR	R	R	R	FtR	FtR	FtR	R	R
	NPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	R
	SAP	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	SPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	TPL	FtR	R	FtR	FtR	FtR	FtR	FtR	FtR	FtR	R	FtR	FtR
	UMW	FtR	FtR	FtR	FtR	R	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	WMT	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	XER	R	R	R	R	FtR	R	R	FtR	FtR	R	R	R
NO ₂ -N H ₀ (2000-2009 vs 1990-1999)	CPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	NAP	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	NPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	SAP	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	SPL	FtR	FtR	R	FtR	FtR	FtR	FtR	R	FtR	R	FtR	FtR
	TPL	FtR	FtR	FtR	FtR	R	FtR	FtR	FtR	FtR	R	FtR	FtR
	UMW	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	WMT	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	XER	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
NO ₂ -N H ₀ (2010-2019 vs 1990-1999)	CPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	NAP	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	NPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	SAP	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	SPL	R	R	R	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	TPL	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	UMW	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	WMT	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR
	XER	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR	FtR

FtR: Fail to Reject, R: Reject @ 0.05 significance. Rejected month-regions are boxed.

than 0.05 boxed for better visualization. The *p*-values in the Wilcoxon rank sum test varied between 0 and 1, where the smaller the *p*-values, the more significant the difference between the two populations was. In the context of nitrogen analysis, smaller *p*-values corresponded to the month-regions associated with a substantial increase in nitrogen levels. This method is more robust than just comparing the medians because it can also quantify the confidence of the result. With a significance of 0.05, there was a 95% confidence to reject or fail to reject the null hypothesis. This has been discussed more thoroughly in the next section.

Table 3. NO₃-N and NO₂-N month-region *p*-values of Wilcoxon rank sum test

Years	Region	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
NO ₃ -N p-value (2000-2009 vs 1990-1999)	CPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NAP	1.00	0.00	1.00	0.12	0.92	0.55	0.67	1.00	1.00	1.00	0.00	0.00
	NPL	0.96	0.64	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	0.69
	SAP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00
	TPL	0.74	0.00	0.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
	UMW	0.32	0.00	0.08	0.05	0.00	0.00	0.28	0.13	0.00	0.79	0.00	0.03
	WMT	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	XER	0.07	0.00	1.00	1.00	0.99	0.01	1.00	1.00	1.00	0.31	0.02	0.01
NO ₃ -N p-value (2010-2019 vs 1990-1999)	CPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NAP	0.22	0.00	1.00	1.00	0.00	0.00	0.03	1.00	1.00	0.08	0.03	0.00
	NPL	0.99	0.91	0.63	0.12	1.00	0.88	1.00	1.00	0.54	1.00	0.22	0.03
	SAP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SPL	0.40	0.12	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.99	1.00	0.28
	TPL	0.93	0.00	1.00	1.00	1.00	1.00	1.00	1.00	0.12	0.01	0.73	0.24
	UMW	0.89	0.94	1.00	0.95	0.00	0.58	0.93	0.62	0.86	1.00	1.00	1.00
	WMT	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98
	XER	0.00	0.04	0.00	0.00	1.00	0.00	0.00	1.00	1.00	0.00	0.01	0.00
NO ₂ -N p-value (2000-2009 vs 1990-1999)	CPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NAP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SAP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SPL	0.41	0.15	0.00	0.05	0.93	0.91	0.67	0.04	0.31	0.03	0.98	0.99
	TPL	1.00	1.00	1.00	0.84	0.00	0.43	0.98	1.00	0.24	0.00	1.00	1.00
	UMW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	WMT	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	XER	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NO ₂ -N p-value (2010-2019 vs 1990-1999)	CPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NAP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	NPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SAP	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	SPL	0.00	0.00	0.00	0.39	0.62	0.89	0.99	0.76	0.19	0.47	0.57	0.14
	TPL	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	UMW	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	WMT	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	XER	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

$p \leq 0.05 \rightarrow$ reject H0 @ 5% significance. Values smaller than or equal to 0.05 are boxed.

3.4. Error Analysis

One of the reasons why hypothesis testing is more superior to other methods is that it is capable of quantifying error. There are two types of errors associated with any hypothesis testing, regardless of the underlying method. These errors are called type I when rejecting a true null hypothesis (false positive) and type II when accepting a false null hypothesis (false negative). These are summarized in Table 4, where there are four possible outcomes based on rejecting or failing to reject a true or false null hypothesis. In this project, a type I error occurred when a flagged region did not have a significant nitrogen

increase. On the contrary, a Type II error happened when a region with a nitrogen problem was designated as a no problem.

Table 4. NO₃-N and NO₂-N type I and type II errors of hypothesis testing

	H_0 is true No significant change in NO ₂ -N and/or NO ₃ -N	H_0 is false significant change in NO ₂ -N and/or NO ₃ -N
Reject H_0	Type I error (α)	Correct ($1 - \beta$)
Fail to Reject H_0	Correct ($1 - \alpha$)	Type II Error β

The seriousness of these errors is dependent on the problem under study. In nitrogen testing, a type I error did not pose a serious issue, whereas a type II error meant a region with a nitrogen problem was ignored. The error of the former was equal to the significance level of α , so there was a 5% chance that the region with nitrogen issues did not have any problems. The latter error is denoted by β and is based on the sample size and population standard deviation. The equation to calculate β for a Wilcoxon rank sum test [22] is not as common as the original z-test and t-test. The calculated β values were different for every month-region (not shown because of their approximate zero values), but they had a range between 0.02% and 0.07%, which in practical terms means it was implausible for the method to have overlooked an impaired month-region.

To further elaborate on the goodness of the Wilcoxon rank sum test, a two-sample t-test [23] type II error was calculated for the datasets. While maximum β values of 20% to 30% are deemed acceptable [24,25], in this research, an error of 1% was chosen as a threshold. The results are presented in Table 5, with values larger than 1% considered errors (boxed for convenience). Based on the data in Table 5, 61 out of 108 month-regions for NO₃-N and 41 out of 108 month-regions for NO₂-N were likely to have been identified incorrectly. The failure of the two-sample was not due to the sample sizes as both large and low sample sizes exhibited high type II errors but was due to the data non-normality. The Kolmogorov-Smirnov normality test conducted at the exploratory data analysis phase showed very small p -values, which here were associated with high type II errors. The small p -values mean the normality test failed with high confidence for these month-regions.

4. Discussion

The national-scale studies in rivers and streams nitrogen assessment are limited in the literature. The lack of national-scale datasets is one of the reasons, while the absence of a unified testing method that all the stakeholders can utilize can be another reason. There were a few promising works in progress in this area, but the challenge was to associate those with the current study as the ecoregions were not comparable. Nevertheless, qualitative comparisons were made to test the performance of the method.

4.1. National Scale

This project adopted the nine ecoregions due to the extensive research conducted by US-EPA as part of NRSA and its predecessor Wadeable Streams Assessment (WSA) [26]

Table 5. NO₃-N and NO₂-N month-region type II error using two-sample t-test

Years	Region	Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
NO ₃ -N beta (2000-2009 vs 2008-2009)	CPL	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	NAP	74%	100%	80%	100%	98%	100%	94%	0%	13%	81%	100%	100%
	NPL	64%	85%	11%	18%	1%	12%	16%	74%	31%	3%	21%	97%
	SAP	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	SPL	12%	17%	0%	0%	0%	0%	0%	100%	0%	0%	0%	34%

	TPL	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
	UMW	99%	100%	100%	100%	100%	100%	94%	92%	100%	88%	100%
	WMT	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	XER	100%	100%	77%	39%	35%	100%	77%	0%	0%	100%	100%
NO ₃ -N beta (2010-2019 vs 1990-1999)	CPL	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	NAP	100%	100%	84%	82%	100%	100%	100%	0%	40%	100%	100%
	NPL	53%	82%	65%	98%	19%	75%	19%	10%	100%	42%	88%
	SAP	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	SPL	100%	100%	13%	36%	92%	0%	6%	13%	1%	94%	25%
	TPL	99%	100%	0%	35%	58%	0%	0%	58%	100%	100%	61%
	UMW	95%	95%	76%	94%	100%	95%	87%	92%	81%	73%	85%
	WMT	0%	0%	0%	7%	0%	0%	0%	0%	0%	0%	83%
	XER	100%	100%	100%	100%	3%	100%	100%	31%	11%	100%	100%
NO ₂ -N beta (2000-2009 vs 1990-1999)	CPL	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	NAP	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	NPL	0%	0%	0%	0%	0%	0%	0%	0%	1%	3%	0%
	SAP	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	SPL	100%	100%	100%	100%	84%	83%	95%	100%	95%	100%	70%
	TPL	0%	0%	0%	11%	100%	24%	0%	5%	95%	100%	0%
	UMW	1%	2%	0%	14%	0%	0%	0%	0%	0%	0%	5%
	WMT	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	XER	0%	1%	0%	0%	0%	0%	0%	0%	0%	0%	0%
NO ₂ -N beta (2010-2019 vs 1990-1999)	CPL	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	NAP	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	NPL	1%	0%	0%	0%	0%	0%	0%	0%	4%	4%	1%
	SAP	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	SPL	100%	100%	100%	100%	99%	95%	66%	100%	95%	100%	100%
	TPL	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	UMW	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
	WMT	0%	0%	0%	0%	0%	0%	0%	0%	0%	1%	0%
	XER	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%

Values larger than 1% are boxed.

in these regions. While acknowledging the excellent work done by the US-EPA, the current project expanded its work by extending the temporal duration of single years to decades, including months, separating NO₂-N and NO₃-N concentrations, and using hypothesis testing as a method of assessment. That is why the study was not directly comparable to the work conducted by US-EPA, but the ecoregions could be qualitatively assessed. US-EPA's assessment was in the form of poor, fair, and good conditions based on the miles of the rivers in each ecoregion. The poor condition, which is of interest in this project, was defined as when the river had nitrogen concentration worse than 95 percent of the distribution of least disturbed sites in an ecoregion.

US-EPA calculated the poor condition for 2008-2009 and 2013-2014 rivers and streams' lengths relative to the NRSA benchmarks (undisturbed sites). The percent changes are reported in the first two rows of Table 6, where an increase in the percent shows worsening. Based on the US-EPA's findings, all the ecoregions were improving in condition except the NPL (22% to 28%) (Table 6). The current study identified the ecoregions worsening as the NAP (x3 to x6), NPL (0 to x1), and XER (x4 to x9) (Table 6, rows 3 to 6). Based on the count of rejected month-regions, the NPL did not seem to be an issue,

but the NAP and XER needed immediate actions. These increases were attributed to urbanization (septic tanks) and extensive agricultural activity (nitrogen-based fertilizers) [15].

Another finding from Table 6 was that most of the nitrogen problems were related to NO₃-N. Since the US-EPA used the total nitrogen, this piece of information was lost. With this information, a targeted mitigation plan can be devised that focuses only on NO₃-N treatment. One of these mitigation plans has been thoroughly investigated for Canning Catchment in Western Australia [27]. Another classic study highlighting the importance of stream restoration in NO₃-N reduction is presented in [28].

4.2. Baseline/Benchmark Selection

The baseline/benchmark is the nitrogen concentration in an undisturbed site used as a point of reference. This choice in the US-EPA’s work was the least disturbed reference site data distribution difference for each NRSA ecoregion. In the current study, the distribution from 1990-1999 was chosen as the least disturbed distribution among the three decadal data available. This choice was made based on the fact that the United States population was the lowest among the study periods, and climate change did not have observable effects in the 90s. Moreover, the goal of this study was more assessing the trends in the stream nitrogen levels rather than emphasizing how the levels are compared to the standard regulated levels. This has been briefly touched in the US-EPA’s work on a national scale but still requires more research given the meteorological, morphological, and geological differences in the ecoregions.

4.3. Regional Scale

Other works from the past are acknowledged here, but due to an increase in the rate of anthropogenic nitrogen generation, some of the results are outdated or not directly relatable to the current study [29–36]. On the other hand, most of the recent literature only focuses on regional scales such as a specific river [37,38], watershed [39,40], or state [13]. These study domains are overlaid on the NRSA ecoregions to understand how they are compared with the current results. The most prominent one is the study conducted over Pennsylvania, located in the Mid-Atlantic region in the northeast of the United States [13]. This state falls on the NAP and SAP ecoregions of the NRSA. According to this study, 53% of the sites exceeded the 25% percentile for total nitrogen. Based on Table 6, the NAP ecoregion was worsening in the NO₃-N condition, which agrees with [13]. The SAP ecoregion does not show a significant change which means most of the nutrient impairment must have been in the northern part of Pennsylvania.

Another noteworthy study was conducted on the NO₃-N contribution of the state of Iowa to the Mississippi River stream network [38]. The state of Iowa is located in the Mid-western United States, with its economy dependent on livestock and crops. More than 95% of its area falls on the TPL ecoregion. The study conducted in [38] was based on the data collected from 1999 to 2016 at 23 sites. The research reported that the NO₃-N loads in the Iowa-inclusive basins were above the 2003 level for ten consecutive years. This agrees with the current results in which the TPL region showed a significant increase in NO₃-N for all the months (Table 2, 3, and 4) and specifically for September (137% increase). In addition to the increase from the baseline, this is also most likely because of the correlation between the seasonal streamflow and NO₃-N concentration, where often the higher NO₃-N concentration was expected in fall and spring [41]. Finally, based on Table 6, the condition in TPL improved (x10 to x2), but this was not addressed in [14] because the data was collected until 2016, which means the condition would have improved if the temporal coverage was longer.

Table 6. NO₃-N and NO₂-N impaired rivers and streams, comparison of US-EPA and current study

	Years	CPL	NAP	NPL	SAP	SPL	TPL	UMW	WMT	XER
Total-N	2008-2009	35%	37%	22%	47%	48%	64%	42%	17%	31%

	2013-2014	31%	35%	28%	36%	46%	59%	39%	15%	18%
NO ₃ -N	2000-2009	-	x3	-	-	x1	x10	x7	-	x4
	2010-2019	-	x6	x1	-	-	x2	x1	-	x9
NO ₂ -N	2000-2009	-	-	-	-	x3	x2	-	-	-
	2010-2019	-	-	-	-	x3	-	-	-	-

x1: one month-region was impaired, x2: two month-regions were impaired, and ...

5. Conclusions

In this study, nitrogen in the form of NO₃-N and NO₂-N was assessed for the rivers and streams of the contiguous United States by applying a non-parametric hypothesis testing method. The following conclusions were the highlights of the research:

- The nitrogen concentration was *more* dependent on space (ecoregions) than time (months).
- The NO₃-N and NO₂-N concentration levels were non-normal, with the lower concentrations having a high frequency.
- Due to the non-normality of data, typical hypothesis testing methods such as z-test and t-test were not applicable and therefore resulted in large type II errors.
- The Wilcoxon rank sum test as a non-parametric method yielded low type II errors when applied to NO₃-N and NO₂-N distributions.
- Most of the nitrogen-impaired waters were a result of excessive NO₃-N, whereas NO₂-N levels did not pose an immediate threat.
- NAP (x3 to x6), NPL (0 to x1), and XER (x4 to x9) ecoregions were worsening in the NO₃-N condition with NAP, and XER needed immediate actions.
- The choice of the baseline when nitrogen levels were relatively assessed could change the outcome.

Author Contributions: Conceptualization, A.M. and S.A.; methodology, A.M.; software, A.M.; validation, A.M.; formal analysis, A.M. and S.A.; investigation, A.M.; resources, A.M.; data curation, A.M.; writing—original draft preparation, A.M. and S.A.; writing—review and editing, S.A.; visualization, A.M.; supervision, N/A; project administration, N/A; funding acquisition, N/A All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: <https://www.waterqualitydata.us/portal/>

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Omernik, J.M. Nonpoint source--stream nutrient level relationships: a nationwide study [phosphorus and nitrogen levels, USA]. *Ecol. Res. Ser. (USA)*. no. 600/3-77-105. **1977**.

2. Withers, P.J.A.; Jarvie, H.P.; Stoate, C. Quantifying the impact of septic tank systems on eutrophication risk in rural headwaters. *Environ. Int.* **2011**, *37*, 644–653.

3. Gelda, R.K.; Auer, M.T. Development and testing of a dissolved oxygen model for a hypereutrophic lake. *Lake Reserv. Manag.* **1996**, *12*, 165–179.

4. Watson, S.B.; Miller, C.; Arhonditsis, G.; Boyer, G.L.; Carmichael, W.; Charlton, M.N.; Confesor, R.; Depew, D.C.; Höök, T.O.; Ludsin, S.A.; et al. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. *Harmful Algae* **2016**, *56*, 44–66.

5. Khangaonkar, T.; Nugraha, A.; Xu, W.; Long, W.; Bianucci, L.; Ahmed, A.; Mohamedali, T.; Pelletier, G. Analysis of hypoxia and sensitivity to nutrient pollution in Salish Sea. *J. Geophys. Res. Ocean.* **2018**, *123*, 4735–4761.

6. Oenema, O.; van Liere, L.; Schoumans, O. Effects of lowering nitrogen and phosphorus surpluses in agriculture on the quality of groundwater and surface water in the Netherlands. *J. Hydrol.* **2005**, *304*, 289–301.

7. Conley, D.J.; Paerl, H.W.; Howarth, R.W.; Boesch, D.F.; Seitzinger, S.P.; Havens, K.E.; Lancelot, C.; Likens, G.E.; others

- Controlling eutrophication: nitrogen and phosphorus. *Science* (80-.). **2009**, 323, 1014–1015.
8. Oenema, O.; Roest, C.W.J. Nitrogen and phosphorus losses from agriculture into surface waters; the effects of policies and measures in the Netherlands. *Water Sci. Technol.* **1998**, 37, 19–30.
 9. Yu, C.; Huang, X.; Chen, H.; Godfray, H.C.J.; Wright, J.S.; Hall, J.W.; Gong, P.; Ni, S.; Qiao, S.; Huang, G.; et al. Managing nitrogen to restore water quality in China. *Nature* **2019**, 567, 516–520.
 10. Baeumler, N.W.; Gupta, S.C. Precipitation as the primary driver of variability in river nitrogen loads in the Midwest United States. *JAWRA J. Am. Water Resour. Assoc.* **2020**, 56, 113–133.
 11. Akbariyeh, S.; Pena, C.A.G.; Wang, T.; Mohebbi, A.; Bartelt-Hunt, S.; Zhang, J.; Li, Y. Prediction of nitrate accumulation and leaching beneath groundwater irrigated corn fields in the Upper Platte basin under a future climate scenario. *Sci. Total Environ.* **2019**.
 12. Mohebbi, A.; Akbariyeh, S.; Li, Y.; Bartelt-Hunt, S. Modeling Fate and Transport of Nitrate in Groundwater Flow: Case Study Gila Bend Basin, Arizona. In Proceedings of the Association of Environmental Engineering and Science Professors; 2017.
 13. Clune, J.W.; Crawford, J.K.; Boyer, E.W. Nitrogen and Phosphorus Concentration Thresholds toward Establishing Water Quality Criteria for Pennsylvania, USA. *Water* **2020**, 12, 3550.
 14. Schilling, K.E.; Jones, C.S.; Wolter, C.F.; Liang, X.; Zhang, Y.-K.; Seeman, A.; Isenhardt, T.; Schnoebelen, D.; Skopec, M. Variability of nitrate-nitrogen load estimation results will make quantifying load reduction strategies difficult in Iowa. *J. Soil Water Conserv.* **2017**, 72, 317–325.
 15. Omernik, J.; Paulsen, S.; Griffith, G.; Weber, M. Regional patterns of total nitrogen concentrations in the National Rivers and Streams Assessment. *J. Soil Water Conserv.* **2016**, 71, 167–181.
 16. Newcomer, M.E.; Bouskill, N.J.; Wainwright, H.; Maavara, T.; Arora, B.; Siirila-Woodburn, E.R.; Dwivedi, D.; Williams, K.H.; Steefel, C.; Hubbard, S.S. Hysteresis patterns of watershed nitrogen retention and loss over the past 50 years in United States hydrological basins. *Global Biogeochem. Cycles* **2021**, 35, e2020GB006777.
 17. Bellmore, R.A.; Compton, J.E.; Brooks, J.R.; Fox, E.W.; Hill, R.A.; Sobota, D.J.; Thornbrugh, D.J.; Weber, M.H. Nitrogen inputs drive nitrogen concentrations in US streams and rivers during summer low flow conditions. *Sci. Total Environ.* **2018**, 639, 1349–1359.
 18. Lu, C.; Tian, H. Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance. *Earth Syst. Sci. Data* **2017**, 9, 181–192.
 19. Leys, C.; Ley, C.; Klein, O.; Bernard, P.; Licata, L. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *J. Exp. Soc. Psychol.* **2013**, 49, 764–766.
 20. Yu, Y.-S.; Zou, S.; Whittemore, D. Non-parametric trend analysis of water quality data of rivers in Kansas. *J. Hydrol.* **1993**, 150, 61–80.
 21. Massey Jr, F.J. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **1951**, 46, 68–78.
 22. Shieh, G.; Jan, S.; Randles, R.H. On power and sample size determinations for the Wilcoxon-Mann-Whitney test. *J. Nonparametr. Stat.* **2006**, 18, 33–43.
 23. Cressie, N.A.C.; Whitford, H.J. How to use the two sample t-test. *Biometrical J.* **1986**, 28, 131–148.
 24. Cohen, J. *Statistical power analysis for the behavioral sciences*; Academic press, 2013;
 25. Mapstone, B.D. Scalable decision rules for environmental impact studies: effect size, type I, and type II errors. *Ecol. Appl.* **1995**, 5, 401–410.
 26. Olsen, A.R.; Peck, D. V Survey design and extent estimates for the Wadeable Streams Assessment. *J. North Am. Benthol. Soc.* **2008**, 27, 822–836.
 27. Polyakov, M.; White, B.; Zhang, F. Cost-effective Strategies to Reduce Nitrogen and Phosphorus Emissions in an Urban River Catchment. *Melbourne, Aust.* **2017**.
 28. Craig, L.S.; Palmer, M.A.; Richardson, D.C.; Filoso, S.; Bernhardt, E.S.; Bledsoe, B.P.; Doyle, M.W.; Groffman, P.M.; Hassett, B.A.; Kaushal, S.S.; et al. Stream restoration strategies for reducing river nitrogen loads. *Front. Ecol. Environ.* **2008**, 6, 529–538.

29. Holland, D.M.; Principe, P.P.; Sickles II, J.E. Trends in atmospheric sulfur and nitrogen species in the eastern United States for 1989--1995. *Atmos. Environ.* **1998**, *33*, 37–49.
30. McMahon, P.B.; Dennehy, K.F. N₂O emissions from a nitrogen-enriched river. *Environ. Sci. & Technol.* **1999**, *33*, 21–25.
31. Smith, R.A.; Alexander, R.B.; Wolman, M.G. Water-quality trends in the nation's rivers. *Science (80-.)*. **1987**, *235*, 1607–1615.
32. Richards, R.P.; Baker, D.B. Trends in water quality in LEASEQ rivers and streams (Northwestern Ohio), 1975--1995. *J. Environ. Qual.* **2002**, *31*, 90–96.
33. Howarth, R.W.; Swaney, D.P.; Boyer, E.W.; Marino, R.; Jaworski, N.; Goodale, C. The influence of climate on average nitrogen export from large watersheds in the Northeastern United States. In *Nitrogen cycling in the Americas: natural and anthropogenic influences and controls*; Springer, 2006; pp. 163–186.
34. Brooks, P.D.; Lemon, M.M. Spatial variability in dissolved organic matter and inorganic nitrogen concentrations in a semiarid stream, San Pedro River, Arizona. *J. Geophys. Res. Biogeosciences* **2007**, *112*.
35. Schaefer, S.C.; Alber, M. Temporal and spatial trends in nitrogen and phosphorus inputs to the watershed of the Altamaha River, Georgia, USA. *Biogeochemistry* **2007**, *86*, 231–249.
36. Sprague, L.A.; Lorenz, D.L. Regional nutrient trends in streams and rivers of the United States, 1993- 2003. *Environ. Sci. & Technol.* **2009**, *43*, 3430–3435.
37. Huizenga, A.; Bailey, R.T.; Gates, T.K. Stream-aquifer and in-stream processes affecting nitrogen along a major river and contributing tributary. *J. Contam. Hydrol.* **2017**, *199*, 24–35.
38. Jones, C.S.; Nielsen, J.K.; Schilling, K.E.; Weber, L.J. Iowa stream nitrate and the Gulf of Mexico. *PLoS One* **2018**, *13*, e0195930.
39. Ator, S.W.; Garcia, A.M.; Schwarz, G.E.; Blomquist, J.D.; Sekellick, A.J. Toward explaining nitrogen and phosphorus trends in Chesapeake Bay Tributaries, 1992--2012. *JAWRA J. Am. Water Resour. Assoc.* **2019**, *55*, 1149–1168.
40. Renwick, W.H.; Vanni, M.J.; Fisher, T.J.; Morris, E.L. Stream Nitrogen, Phosphorus, and Sediment Concentrations Show Contrasting Long-term Trends Associated with Agricultural Change. *J. Environ. Qual.* **2018**, *47*, 1513–1521.
41. Schilling, K.E.; Lutz, D.S. RELATION OF NITRATE CONCENTRATIONS TO BASEFLOW IN THE RACCOON RIVER, IOWA 1. *JAWRA J. Am. Water Resour. Assoc.* **2004**, *40*, 889–900.