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 Mohebbia and Simin Akbariyehb

aDepartment of Civil and Environmental Engineering, University of New Hampshire, Durham; bSchool of Engineering, Brown University, Providence

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 study, the Wilcoxon rank-sum test, 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 chosen 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 identified as the ecoregions worsening in the nitrogen-nitrate condition with NAP, and XER needed immediate action. The nitrite-nitrogen condition did not pose an immediate threat, so mitigation efforts 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.

Methods

In this section, the methodology used in the current study is briefly discussed. 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.

**Study spatial and temporal extents**

The research was conducted on the Contiguous United States (CONUS) rivers and streams over the National Rivers and Streams Assessment (NRSA) reporting regions (Omer- nik et al., 2016). 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 (NO$_3$-N) and nitrite-nitrogen (NO$_2$-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.

**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 NO$_3$-N and NO$_2$- 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 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 (Leys et al., 2013). The scaled MAD is calculated by:

\[
MAD = c \times \text{median}|D - \text{median}(D)|
\]

with

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

where the 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.

**Hypothesis testing**

There are various hypothesis testing approaches in the literature, and each comes with its limitation and scope (Yu et al., 1993). 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/or 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 used to compare two independent populations (Wilcoxon, 1992). The non-parametric tests do not require the sample to be normally distributed (Siegel, 1957) or belong to a large dataset to satisfy the central limit theorem assumption.\(^1\) Also, this approach indirectly tests the medians rather than the means (Harris & Hardin, 2013), 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 (Montgomery & Runger, 2018):

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 and compare it to the critical z value by

\[
z_0 = \frac{T_1 - E(T_1)}{\sqrt{V(T_1)}}
\]
with
\[ E(T_1) = \frac{n_1n_2 + n_1(n_1 + 1)}{2} \] (4)
and
\[ V(T_1) = n_1n_2\left(\frac{n_1 + n_2 + 1}{12}\right) \] (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 populations 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
\[ T_1 + T_2 = \frac{(n_1 + n_2)(n_1 + n_2 + 1)}{2} \] (6)
where an increase in \( T_1 \) will cause a decrease in \( T_2 \).

5. Calculate the critical \( z \) value based on the significance level (\( \alpha \)) 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 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.

**Methodology**

The Wilcoxon rank-sum test was applied to the datasets following the steps established above. The null hypothesis was defined as no significant change in NO\textsubscript{3}-N or NO\textsubscript{2}-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.

**Results**

The methodology presented was applied to the raw datasets acquired from the NWQMC. The justification for using 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 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.

Data non-normality

The box-and-whisker plot of month-regions is presented in Figure 2 for NO3-N and Figure 3 for NO2-N, with the concentration (mg/l) in the y-axis, ecoregions (NAP, SAP, CP, UMW, TPL, SPL, NPL, WMT, and XER) paired with numbers 1–9 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, Q1), 50% percentile (median, Q2), 75% percentile (box top, Q3), 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, low concentration levels are more frequently observed than high concentration levels. In addition, the data also failed the normality test of Kolmogorov–Smirnov (Massey Jr, 1951) and Lilliefors (Lilliefors, 1967) 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 $\pm 1.5\times(Q_3 - Q_1)$ 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 $-1.5\times(Q_3 - Q_1)$ region, which was another indication of the positive skewness. In addition, the number of outliers in NO3-N was more than NO2-N which was attributed to the higher number of collected NO3-N samples. The total number of samples collected between 1990–2019 for NO3-N was 7–62% more than the number of samples collected for NO2-N. Another reason could be the higher range of the NO3-N samples compared to the NO2-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 NO3-N and NO2-N ranges were roughly between 0 and 10 mg/l and 0–0.3 mg/l, respectively.

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 is that the concentration was mainly a function of space (ecoregion) rather than time (month). For example, compare the first ecoregion of Jan. 1990–1999 with the sixth ecoregion of the same month (Figure 2, upper left panel). As it can be seen, there was a considerable difference between the NO3-N levels, with one barely reaching 3 mg/l and the other reaching 10 mg/l. Now, compare the first ecoregion for all twelve months of 1990–1999 (Figure 2, panels on the left from top to bottom). The data is maintaining the same range with only a slight change in quartile distribution.

This should not be inferred as that the monthly variation did not exist. In fact, later, it was 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. 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.

$$\text{change} = \frac{M_{1990–1999} - M_{2000–2009}}{M_{1990–1999}} \times 100 \quad (7)$$

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 NO2-N problems. Out of these regions, SPL and TPL also had NO3-N issues (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 practical and straightforward, this method failed to quantify the impact of sample randomness. The identified month-regions must be further investigated, perhaps by more rigorous statistical methods.

Wilcoxon rank-Sum test

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 Tables 1 and 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 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 17–8 for NO₂-N.

The R or FrR 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 $\alpha$; 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 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.

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.
hypothesis. This has been discussed more thoroughly in the next section.

**Error analysis**

One of the reasons why hypothesis testing is 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. 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 \( \alpha \), so there was a 5% chance that the region with nitrogen issues did not have any problems. The latter error is denoted by \( \beta \) and is based on the sample size and population standard deviation. The equation to calculate \( \beta \) for a Wilcoxon rank-sum test (Shieh et al., 2006) is not as common as the original z-test and t-test. The calculated \( \beta \) 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 (Cressie & Whitford, 1986) type II error was calculated for the datasets. While maximum \( \beta \) values of 20% to 30% are deemed acceptable (Cohen, 2013; Mapstone, 1995), 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 NO\(_3\)-N and 41 out of 108 month-regions for NO\(_2\)-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

![Figure 3. NO\(_2\)-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 NO\(_2\)-N concentration (C) in mg/l](image_url)
high type II errors but was due to the data non-normality. The Kolmogorov–Smirnov and Lilliefors normality tests 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.

**Discussion**

National-scale nitrogen assessment of rivers and streams is 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.

**National scale**

This project adopted the nine ecoregions due to the extensive research conducted by the US-EPA as part of NRSA and its predecessor Wadeable Streams Assessment (WSA) (Olsen &
Table 1. NO$_3$-N and NO$_2$-N month-region median change from the baseline

| Years | Region | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sept. | Oct. | Nov. | Dec. |
|-------|--------|------|------|------|------|-----|-----|-----|-----|------|-----|-----|-----|
| 2000-2009 vs. 1990-1999 | FtR | Fail to Reject | FtR | Fail to Reject | FtR | Fail to Reject | FtR | Fail to Reject | FtR | Fail to Reject | FtR | Fail to Reject | FtR |
| 1990-1999 | CPL | -58% | -59% | -60% | -53% | -49% | -43% | -38% | -37% | -39% | -38% | -51% | -57% |
| 1990-1999 | NAP | -19% | -21% | -2% | 8% | 2% | 4% | 0% | -17% | -10% | -4% | 17% | 24% |
| 1990-1999 | NPL | -13% | -6% | -25% | -31% | -38% | -32% | -34% | -13% | -43% | -57% | -31% | 3% |
| 1990-1999 | SAP | -34% | -40% | -36% | -39% | -30% | -28% | -31% | -27% | -37% | -39% | -35% | -31% |
| 1990-1999 | SPL | -14% | -15% | -21% | -30% | -21% | -33% | -40% | 18% | -34% | -30% | -30% | -11% |
| 1990-1999 | TPL | 12% | 31% | 5% | 25% | 29% | 34% | -36% | 56% | 42% | 11% | 57% | 57% |
| 1990-1999 | UMW | 16% | 52% | 33% | 18% | 57% | 61% | -1% | -3% | 46% | 7% | 125% | 24% |
| 1990-1999 | WMT | -46% | -52% | -47% | -51% | -67% | -66% | -67% | -63% | -66% | 50% | -56% | -42% |
| 1990-1999 | XER | 9% | 16% | -5% | 11% | -11% | 17% | -6% | -27% | 36% | 9% | 26% | 21% |
| 1990-1999 | WMT | -63% | -63% | -69% | -61% | -61% | -55% | -39% | -45% | -45% | -43% | -56% | -57% |
| 1990-1999 | NAP | 8% | 28% | -2% | 5% | 7% | 24% | -9% | -4% | 19% | 26% | 20% | 21% |
| 1990-1999 | NPL | -20% | -9% | 15% | 5% | -31% | -13% | -36% | -46% | 26% | -34% | -9% | 13% |
| 1990-1999 | SAP | -40% | -41% | -46% | -41% | -40% | -35% | -37% | -42% | -49% | -44% | -47% | -39% |
| 1990-1999 | SPL | 6% | 10% | -13% | 9% | -11% | 0% | -21% | 0% | -16% | -15% | -21% | -12% |
| 1990-1999 | TPL | 5% | 45% | 21% | -9% | 5% | -10% | 0% | -18% | -32% | 8% | 13% | 8% |
| 1990-1999 | UMW | 0% | 1% | -8% | -1% | 5% | 0% | 0% | -6% | 3% | -9% | 15% | 12% |
| 1990-1999 | WMT | -24% | -32% | -27% | -17% | -42% | -39% | -46% | -48% | 42% | -31% | -26% | 4% |
| 1990-1999 | XER | 21% | 10% | 21% | 37% | -19% | 45% | 42% | -12% | 17% | 35% | 20% | 21% |

1990–1999 was designated as the baseline decade. Positive values are boxed.

Table 2. NO$_3$-N and NO$_2$-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. |
|-------|--------|------|------|------|------|-----|-----|-----|-----|------|-----|-----|-----|
| 2000-2009 vs. 1990-1999 | FtR | Fail to Reject | FtR | Fail to Reject | FtR | Fail to Reject | FtR | Fail to Reject | FtR | Fail to Reject | FtR | Fail to Reject | FtR |
| 1990-1999 | CPL | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | R | R |
| 1990-1999 | NAP | FR | FR | R | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | NPL | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | SAP | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | SPL | FR | FR | FR | FR | FR | FR | FR | R | FR | FR | FR | FR |
| 1990-1999 | TPL | FR | FR | R | R | R | R | R | R | R | R | R | R |
| 1990-1999 | UMW | FR | FR | FR | FR | FR | FR | FR | FR | FR | R | R | R |
| 1990-1999 | WMT | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | XER | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | WMT | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | CPL | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | NAP | FR | FR | R | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | NPL | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | SAP | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | SPL | FR | FR | R | R | R | R | R | R | R | R | R | R |
| 1990-1999 | TPL | FR | FR | R | R | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | UMW | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | WMT | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | XER | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | WMT | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | CPL | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | NAP | FR | FR | R | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | NPL | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | SAP | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | SPL | FR | FR | R | R | R | R | R | R | R | R | R | R |
| 1990-1999 | TPL | FR | FR | R | R | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | UMW | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | WMT | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | XER | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |
| 1990-1999 | WMT | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR | FR |

FtR: Fail to Reject, R: Reject @ 0.05 significance. Rejected month-regions are boxed.
Tables 3 and 4. NO3-N and NO2-N month-region p-values of Wilcoxon rank-sum test

| Years | NO3 p-value (2000-2009 vs 1990-1999) |
|-------|------------------------------------|
| Region | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sept. | Oct. | Nov. | Dec. |
| CPL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| NAP | 1 | 0.00 | 1 | 0.12 | 0.92 | 0.55 | 0.67 | 1 | 1 | 0.00 | 0.00 | 1 |
| NPL | 0.96 | 0.64 | 0.98 | 1 | 1 | 1 | 1 | 1 | 1 | 0.99 | 0.69 | 1 |
| SAP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SPL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| TPL | 0.74 | 0.00 | 0.41 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1 |
| 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 |
| WM2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XER | 0.07 | 0.00 | 0.01 | 0.99 | 0.01 | 1 | 1 | 1 | 0.31 | 0.02 | 0.01 | 1 |

| NO2 p-value (2010-2019 vs 1990-1999) |
|-------------------------------------|
| Region | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sept. | Oct. | Nov. | Dec. |
| CPL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| NAP | 0.22 | 0.00 | 1 | 0.00 | 0.00 | 0.03 | 1 | 0.08 | 0.03 | 0.00 | 1 |
| NPL | 0.99 | 0.91 | 0.63 | 0.12 | 1 | 0.88 | 1 | 0.54 | 1 | 0.22 | 0.03 | 1 |
| SAP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SPL | 0.40 | 0.12 | 1 | 0.97 | 1 | 1 | 1 | 1 | 0.99 | 1 | 0.28 | 1 |
| TPL | 0.93 | 0.00 | 1 | 1 | 1 | 1 | 1 | 0.12 | 0.01 | 0.73 | 0.24 | 1 |
| UMW | 0.85 | 0.94 | 1 | 0.95 | 0.00 | 0.58 | 0.93 | 0.62 | 0.86 | 1 | 1 | 1 |
| WM2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0.98 | 1 | 1 | 1 |
| XER | 0.00 | 0.04 | 0.00 | 0.00 | 1 | 0.00 | 0.00 | 1 | 0.00 | 0.01 | 0.00 | 1 |

| NO2 p-value (2000-2009 vs 1990-1999) |
|-------------------------------------|
| Region | Jan. | Feb. | Mar. | Apr. | May | Jun. | Jul. | Aug. | Sept. | Oct. | Nov. | Dec. |
| CPL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| NAP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| NPL | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SAP | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 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 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| UMW | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| WM2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XER | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

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

| Reject H0 | Type I error (α) | Type II Error (β) |
|-----------|------------------|-------------------|
| H0 is true No significant change in NO3-N and/or NO2-N | H0 is false significant change in NO3-N and/or NO2-N |
| Reject H0 | Correct (1 − α) | Correct (1 − β) |
| Fail to reject H0 | Type I Error (α) | Type II Error (β) |

Peck, 2008] 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 calculated the poor condition for 2008 (Peck, 2008) in these regions. While acknowledging the excellent work of 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 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 a 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 3–6). Based on the count of rejected month-regions, the NPL did not seem to be an issue, but the NAP and XER needed immediate action. These increases were attributed to urbanization (septic tanks) and extensive agricultural activity (nitrogen-based fertilizers) (Omernik et al., 2016).

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 (Polyakov et al., 2017). Another classic study highlighting the importance of stream restoration in NO3-N reduction is presented by (Craig et al., 2008).

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 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 on 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.

**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 (Brooks & Lemon, 2007; Holland et al., 1998; Howarth et al., 2006; McMahon & Dennehy, 1999; Richards & Baker, 2002; Schaefer & Alber, 2007; Smith et al., 1987; Sprague & Lorenz, 2009). On the other hand, most of the recent literature only focuses on regional scales such as a specific river (Huizenga et al., 2017; Jones et al., 2018), watershed (Ator et al., 2019; Renwick et al., 2018), or state (Clune et al., 2020). 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 in Pennsylvania, located in the Mid-Atlantic region in the northeast of the United States (Clune et al., 2020). 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 NO$_3$-N condition, which agrees with (Clune et al., 2020). 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 important study was conducted on the NO$_3$-N contribution of the state of Iowa to the Mississippi River stream network (Jones et al., 2018). The state of Iowa is located in the Midwestern United States, with its economy dependent on livestock and crops. More than 95% of its area falls on the TPL ecoregion. The study conducted by (Jones et al., 2018) was based on the data collected from 1999 to 2016 at 23 sites. The research reported that the NO$_3$-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 NO$_3$-N for all the months (Tables 2–4) 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 NO$_3$-N concentration, where often the higher NO$_3$-N

| Table 5. NO$_3$-N and NO$_2$-N month-region type II error using the 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. 1990-1999) CPL 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% |
| NAP 74% 100% 80% 98% 100% 94% 0% 13% 81% 100% 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% 100% |
| UMW 99% 100% 100% 100% 100% 94% 100% 100% 100% 100% 100% 100% |
| WMT 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% |
| XER 100% 100% 77% 39% 35% 19% 19% 10% 100% 81% 100% 100% |
| NO$_2$-N beta (2000-2009 vs. 1990-1999) CPL 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% |
| NAP 100% 100% 84% 82% 100% 100% 100% 0% 40% 100% 100% 100% |
| NPL 53% 82% 65% 98% 19% 79% 19% 10% 100% 42% 88% 99% |
| SAP 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% 0% |
| SPL 100% 100% 13% 36% 92% 0% 0% 0% 0% 0% 0% 0% |
| TPL 99% 100% 0% 35% 58% 0% 0% 0% 0% 0% 0% 0% |
| UMW 95% 95% 76% 94% 100% 95% 87% 92% 81% 73% 85% 22% |
| WMT 0% 0% 0% 7% 0% 0% 0% 0% 0% 0% 0% 83% |
| XER 100% 100% 100% 100% 3% 100% 100% 31% 11% 100% 100% 100% |

| Values larger than 1% are boxed. |

| Table 6. NO$_3$-N and NO$_2$-N impaired rivers and streams, comparison of US-EPA and current study |
|-----------------|---|---|---|---|---|---|---|---|---|---|---|---|
| Years CPL NAP NPL SAP 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 - - - |
| 2010–2019 - - - - - x3 - - - |

x1: one month-region was impaired, x2: two month-regions were impaired, and so on
concentration was expected in fall and spring (Schilling & Lutz, 2004). Finally, based on Table 6, the condition in TPL improved (x10 to x2), but this was not addressed in (Schilling et al., 2017) because the data was collected until 2016, which means the condition would have improved if the temporal coverage was longer.

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 resulted from 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 action.
- The choice of the baseline when nitrogen levels were relatively assessed could change the outcome.

Note
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.

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

Disclosure statement
No potential conflict of interest was reported by the author(s).

References
Akbariye, S., et al. (2018). Three-dimensional modeling of nitrate-N transport in vadose zone: Roles of soil heterogeneity and groundwater flux. Journal of Contaminant Hydrology, 211(Febuary), 15–25. https://doi.org/10.1016/j.jconhyd.2018.02.005
Akbariye, S., et al. (2019). Prediction of nitrate accumulation and leaching beneath groundwater irrigated corn fields in the upper platte basin under a future climate scenario. Science of The Total Environment, 514–526. https://doi.org/10.1016/j.scitotenv.2019.05.417.
Ator, S. W., Garcia, A. M., Schwarz, G. E., Blomquist, J. D., & Sekellick, A. J. (2019). Toward explaining nitrogen and phosphorus trends in chesapeake bay tributaries, 1992–2012. JAWRA Journal of the American Water Resources Association, 55(5), 1149–1168. https://doi.org/10.1111/1752-1688.12756
Baeumler, N. W., & Gupta, S. C. (2020). Precipitation as the primary driver of variability in river nitrogen loads in the midwest United States. JAWRA Journal of the American Water Resources Association, 56(1), 113–133. https://doi.org/10.1111/1752-1688.12809
Bellmore, R. A., et al. (2018). Nitrogen inputs drive nitrogen concentrations in U.S. Streams and rivers during summer low flow conditions. Science of the Total Environment, 639, 1349–1359. https://doi.org/10.1016/j.scitotenv.2018.05.008
Brooks, P. D., & Lemon, M. M. (2007). Spatial variability in dissolved organic matter and inorganic nitrogen concentrations in a semi-arid stream, San pedro river, arizona. Journal of Geophysical Research: Biogeosciences, 112(G3). https://doi.org/10.1029/2006JG000262.
Clune, J. W., Crawford, J. K., & Boyer, E. W. (2020). Nitrogen and phosphorus concentration thresholds toward establishing water quality criteria for pennsylvania, USA. Water, 12(12), 3550. https://doi.org/10.3390/w12123550
Cohen, J. (2013). Statistical power analysis for the behavioral sciences. Academic press.
Conley, D. J., Paerl, H. W., Howarth, R. W., Boesch, D. F., Seitzinger, S. P., Havens, K. E., Lancelot, C., Likens, G. E. & Others (2009). Controlling eutrophication: Nitrogen and phosphorus. Science, 323(5917), 1014–1015. https://doi.org/10.1126/science.1167755
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., & Others (2008). Stream restoration strategies for reducing river nitrogen loads. Frontiers in Ecology and the Environment, 6(10), 529–538. https://doi.org/10.1890/070080
Cressie, N. A. C., & Whitford, H. J. (1986). How to Use the Two sample-test. Biometrical Journal, 28(2), 131–148. https://doi.org/10.1002/bimj.4710280202
Gelda, R. K., & Auer, M. T. (1996). Development and testing of a dissolved oxygen model for a hypereutrophic lake. Lake and Reservoir Management, 12(1), 165–179. https://doi.org/10.1080/07438149609354006
Harris, T., & Hardin, J. W. (2013). Exact Wilcoxon signed-rank and Wilcoxon Mann–Whitney ranksum tests. The Stata Journal: Promoting Communications on Statistics and Stata, 13(2), 337–343. https://doi.org/10.1177/1536867X1301300208
Holland, D. M., Principe, P. P., & Sickles II, J. E. (1998). Trends in atmospheric sulfur and nitrogen species in the eastern United States for 1989–1995. Atmospheric Environment, 33(1), 37–49. https://doi.org/10.1016/S1352-2310(98)00123-X
Howarth, R. W., et al. (2006). The influence of climate on average nitrogen export from large watersheds in the northeastern United States. In Luiz A. Martinelli and Robert W. Howarth, eds. Nitrogen cycling in the americas: Natural and anthropogenic influences and controls (pp. 163–186). Springer.
Huizenga, A., Bailey, R. T., & Gates, T. K. (2017). Stream-aquifer and in-stream processes affecting nitrogen along a major river and contributing tributary. Journal of Contaminant Hydrology, 199, 24–35. https://doi.org/10.1016/j.jconhyd.2017.03.003
Jones, C. S., Nielsen, J. K., Schilling, K. E., & Weber, L. J. (2018). Iowa stream nitrate and the Gulf of Mexico. PloS one, 13(4), e0195930. https://doi.org/10.1371/journal.pone.0195930
Khangaonkar, T., Nugraha, A., Xu, W., Long, W., Bianucci, L., Ahmed, A., Mohamedali, T., & Pelletier, G. (2018). Analysis of hypoxia and sensitivity to nutrient pollution in salish Sea. Journal of Geophysical Research: Oceans, 123(7), 4753–4761. https://doi.org/10.1029/2017JC013650
Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology, 49(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013
Lillefors, H. W. (1967). On the kolmogorov-smirnov test for normality with mean and variance unknown. Journal of the American Statistical Association, 62(318), 399–402. https://doi.org/10.1080/01621459.1967.10482916
Lu, C., & Tian, H. (2017). Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: Shifted hot spots...
and nutrient imbalance. Earth System Science Data, 9(1), 181–192. https://doi.org/10.5194/essd-9-181-2017

Mapstone, B. D. (1995). Scalable decision rules for environmental impact studies: Effect size, type I, and type II errors. Ecological Applications, 5(2), 401–410. https://doi.org/10.2307/1942031

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. Journal of the American Statistical Association, 46(253), 68–78. https://doi.org/10.1080/01621459.1951.10500769

McMahon, P. B., & Denney, K. F. (1999). N2O emissions from a nitrogen-enriched river. Environmental Science & Technology, 33(1), 21–25. https://doi.org/10.1021/es980645n

Mohrbebi, A., Akbariyeih, S., Li, Y., & Bartelt-Hunt, S. (2017). Modeling fate and transport of nitrate in groundwater flow: Case study Gila bend basin, arizona. In: Association of Environmental Engineering and Science Professors.

Montgomery, D. C., & Runger, G. C. (2018). Applied statistics and probability for engineers. Wiley Hoboken.

Newcomer, M. E., Bouskill, N. J., Wainwright, H., Maavara, T., Arora, B., Siirila-Woodburn, E. R., Dwiwedi, D., Williams, K. H., Steefel, C., & Hubbard, S. S. (2021). Hysteresis patterns of watershed nitrogen retention and loss over the past 50 years in United States hydrological basins. Global Biogeochemical Cycles, 35(4), e2020GB006777. https://doi.org/10.1029/2020GB006777

Oenema, O., & Roest, C. W. J. (1998). Nitrogen and phosphorus losses from agriculture into surface waters; the effects of policies and measures in the Netherlands. Water Science and Technology, 37(3), 19–30. https://doi.org/10.2166/wst.1998.0167

Oenema, O., van Liere, L., & Schoumans, O. (2005). Effects of lowering nitrogen and phosphorus surpluses in agriculture on the quality of groundwater and surface water in the Netherlands. Journal of Hydrology, 304(1–4), 289–301. https://doi.org/10.1016/j.jhydrol.2004.07.044

Olsen, A. R., & Peck, D. V. (2008). Survey design and extent estimates for the wadeable streams assessment. Journal of the North American Benthological Society, 27(4), 822–836. https://doi.org/10.1899/08-050.1

Omernik, J. M. (1977). Nonpoint source–stream nutrient level relationships: A nationwide study [phosphorus and nitrogen levels, USA]. Ecological Research Series (USA), no. 600/3-77-105, 4.

Omernik, J., Paulsen, S., Griffith, G., & Weber, M. (2016). Regional patterns of total nitrogen concentrations in the national rivers and streams assessment. Journal of Soil and Water Conservation, 71(3), 167–181. https://doi.org/10.2489/jswc.71.3.167

Polyakov, M., White, B., & Zhang, F., 2017. Cost-effective strategies to reduce nitrogen and phosphorous emissions in an urban river catchment. Melbourne, Australia: Cooperative Research Centre for Water Sensitive Cities.

Renwick, W. H., Vanni, M. J., Fisher, T. J., & Morris, E. L. (2018). Stream nitrogen, phosphorus, and sediment concentrations show contrasting long-term trends associated with agricultural change. Journal of Environmental Quality, 47(6), 1513–1521. https://doi.org/10.2134/jeq2018.04.0162

Richards, R. P., & Baker, D. B. (2002). Trends in water quality in LEASEQ rivers and streams (northwestern ohio), 1975–1995. Journal of Environmental Quality, 31(1), 90–96. https://doi.org/10.2134/jeq2002.9000

Schafer, S. C., & Alber, M. (2007). Temporal and spatial trends in nitrogen and phosphorus inputs to the watershed of the altamaha river, Georgia, USA. Biogeochemistry, 86(3), 231–249. https://doi.org/10.1007/s10533-007-9155-6

Schilling, K. E., Jones, C. S., Wolter, C. F., Liang, X., Zhang, Y.-K., Seeman, A., Isenhart, T., Schnoebelen, D., & Sköpec, M. (2017). Variability of nitrate-nitrogen load estimation results will make quantifying load reduction strategies difficult in iowa. Journal of Soil and Water Conservation, 72(4), 317–325. https://doi.org/10.2489/jswc.72.4.317

Schilling, K. E., & Lutz, D. S. (2004). Relation of nitrate concentrations to baseflow in the raccoon river, iowa. Journal of the American Water Resources Association, 40(4), 889–900. https://doi.org/10.1111/j.1752-1688.2004.tb01053.x

Shieh, G., Jan, S., & Randles, R. H. (2006). On power and sample size determinations for the Wilcoxon–Mann–Whitney test. Journal of Nonparametric Statistics, 18(1), 33–43. https://doi.org/10.1080/10485250500743099

Siegel, S. (1957). Nonparametric statistics. The American Statistician, 11(3), 13–19. https://doi.org/10.1080/00031305.1957.10501091

Smith, R. A., Alexander, R. B., & Wolman, M. G. (1987). Water-quality trends in the nation’s rivers. Science, 235(4796), 1607–1615. https://doi.org/10.1126/science.235.4796.1607

Sprague, L. A., & Lorenz, D. I. (2009). Regional nutrient trends in streams and rivers of the United States, 1993–2003. Environmental Science & Technology, 43(10), 3430–3435. https://doi.org/10.1021/es803664x

Watson, S. B., Miller, C., Arhonditis, G., Boyer, G. L., Carmichael, W., Charlton, M. N., Confesor, R., Depew, D. C., Höök, T. O., Ludsín, S. A., & Others (2016). The re-eutrophication of lake erie: Harmful algal blooms and hypoxia. Harmful Algae, 56, 44–66. https://doi.org/10.1016/j.hal.2016.04.010

Wilcoxon, F. (1992). Individual comparisons by ranking methods. In: Breakthroughs in statistics (pp. 196–202). Springer.

Withers, P. J. A., Jarvie, H. P., & Stote, C. (2011). Quantifying the impact of septic tank systems on eutrophication risk in rural headwaters. Environment International, 37(3), 644–653. https://doi.org/10.1016/j.envint.2011.01.002

Yu, C., Huang, X., Chen, H., Godfray, H. C. J., Wright, J. S., Hall, J. W., Gong, P., Ni, S., Qiao, S., Huang, G., & Others. (2019). Managing nitrogen to restore water quality in China. Nature, 567(7749), 516–520. https://doi.org/10.1038/s41586-019-1001-1

Yu, Y.-S., Zou, S., & Whitemore, D. (1993). Non-parametric trend analysis of water quality data of rivers in kansas. Journal of Hydrology, 150(1), 61–80. doi:10.1016/0022-1694(93)90156-4