Analyses of Nitrate Groundwater Data in Alachua County and Portions of the Santa Fe River Basin (2013-2014, 2020-2021, 2023-2024)

A report submitted to the Alachua County Environmental Protection Department from AquiferWatch Inc.

by

Rick Copeland, P.G., Ph.D.

September 2024

Analyses of Nitrate Groundwater Data in Alachua County and Portions of the Santa Fe River Basin (2013-2014, 2020-2021, 2023-2024)

SUMMARY

Per an agreement with the Alachua County Environmental Protection Department (ACEPD), AquiferWatch Inc. (AW) agreed to evaluate three sets of groundwater quality data. The data were obtained by ACEPD and AW during three synoptic sampling events within portions of the Santa Fe River basin, including northern and western portions of Alachua County. Data were obtained from wells tapping the Upper Floridan aquifer. The groundwater samples were collected in: (1) September 2014 and May 2015 (2014-2015), (2) November 2020 and May 2021 (2020-2021), and (3) November 2023 and May 2024 (2023-2024). Each data set consisted of between 78 and 1676 samples

Since elevated nitrate concentrations have been observed in the Upper Floridan aquifer for decades within the Santa Fe basin, as an environmental protection department, ACEPD is interested in monitoring changes in nitrate concentrations. For this reason, the objective of this investigation is to determine if the three sets of nitrate data displayed significant changes over time.

Analyses revealed the 2020-2021 and 2023-2024 nitrate data sets were statistically equal. Both data sets had significantly lower concentrations than the 2014-2015 data set. The reason for the decrease is believed to be related to an increase in rainfall across Florida since the late 2000s (Copeland et al., 2023). The increase rainfall resulted in an increase of recharge to the Upper Floridan aquifer, and subsequent nitrate dilution in the aquifer. Another possibility is a reduction in nitrogen loading from land surface to the aquifer for about the past 15 years.

During the 2014-2015 and the 2020-2021 timeframes, the laboratory method detection limit was 0.01 mg/L. However, during the latter period, the limit was raised to 0.20 mg/L. This caused unnecessary obstacles in evaluating nitrate data with low concentrations. It is recommended that in the future, the ACEPD require their analytical laboratory to use the lower detection limit.

METHODS

All analyses procedures and the construction of graphs were conducted in the R statistical platform (R Core Team, 2023) using several statistical packages. Each procedure is described in the text.

DATA ANALYSES

Tests for Normality and Descriptive Statistics

The indicator of concern is nitrate plus nitrite as N (NO3 + NO2 as N) and abbreviated NO3. NO3, as well as many other analytes (indicators) in groundwater, have skewed distributions. That is, they have non-Gausian or non-normal distributions. If data have non-normal distributions, nonparametric statistical methods are recommended. For this reason, tests for normality were conducted for NO3 for each data set. The Shapiro-Wilk test (Conover, 1999) in the "stats" package in R was used. As with other statistical tests used in this investigation, the Shapiro-Wilk test evaluates the probability of the null hypothesis (NH) being true, as compared to the alternate hypothesis (AH). The probability of the NH being true is reported as the p-value. If the p-value is lower than a preset threshold, then the NH is rejected. If the NH is rejected, it is inferred that the AH is true. For this study the threshold was set at 0.05; meaning the NH is rejected if there less than a five percent probability of it being true.

For each set of data, the resulting p-value of each Shapiro-Wilk test was <0.001; inferring the NO3 distributions are non-normal. Of note, since the data are non-normal, the median is a better representation of central tendency than the mean.

Descriptive statistics were produced, using the base package in R. Table 1a displays descriptive statistics for the three data sets **(bold font)**. In addition, the table also displays statistics for the: (a) September/November and the (b) May NO3 data from each data set (not in bold font). The table columns display the data set, the number of observations in each (n), the minimum, Q1, mean, Q2 (Median), Q3, and the maximum values. Note, Q1 is the 25th percentile, the median (Med) is the 50th percentile, and Q3 is the 75th percentile.

Units are in milligrams per liter (mg/L). Note that for the 2014-2015 and the 2020-2021 data sets, the minimum is 0.01. However, the minimum for the 2023-2024 set is 0.20 mg/L. The two values represent the analytical laboratory's method detection limit (MDL). Data labeled at the MDL level are often less than (LT) the laboratory's MDL. Since concentrations can be less than the MDL, to make statistical comparisons, modifications to the data sets were required. Thus, all data reported as LT 0.20 in the 2014-2015 and in the 2020-2021 data were raised to the MDL value used in the 2023-2024 data set. This method is one of several recommended when using censored data (Helsel, 2012).

After adjusting the two earlier data sets, descriptive statistics were recalculated and are presented in Table 1b. The changes using the adjusted, relative to the non-adjusted data are presented in italics and are only found in the Minimum, Q1, Mean columns. Although the modifications are relatively small, they can affect the results of some of the hypothesis tests. One test that was affected will be discussed later.

Table 1a. Descriptive Statistics of Non-adjusted NO3 Data.

Data Set	n	Minimum (mg/L)	Q1 (mg/L)	Mean (mg/L)	Q2 Median (mg/L)	(mg/L) Q3	Maximum (mg/L)
2014-2015	**78**	**0.01**	**0.69**	**1.84**	**1.30**	**2.18**	**10.00**
2020-2021	**167**	**0.01**	**0.10**	**1.35**	**0.57**	**1.23**	**22.00**
2023-2024	**158**	**0.20**	**0.20**	**1.66**	**0.63**	**1.87**	**15.50**
Nov. 2014	40	0.01	0.64	1.76	1.25	2.23	7.70
May 2015	38	0.01	0.72	1.92	1.30	2.13	10.00
Nov. 2020	86	0.01	0.08	1.38	0.56	1.23	22.00
May 2021	81	0.01	0.11	1.33	0.59	1.20	17.00
Nov. 2023	80	0.20	0.22	1.67	0.63	1.94	14.10
May 2024	78	0.20	0.20	1.65	0.63	1.75	15.50

Table 1b. Descriptive Statistics of Adjusted NO3 Data (Used for Data Analyses).

Data Set	N	Minimum (mg/L)	Q1 (mg/L)	Mean (mg/L)	Q2 Median (mg/L)	(mg/L) Q3	Maximum (mg/L)
2014-2015	**78**	**0.20**	**0.69**	**1.84**	**1.30**	**2.18**	**10.00**
2020-2021	**167**	**0.20**	**0.10**	**1.35**	**0.57**	**1.23**	**22.00**
2023-2024	**158**	**0.20**	**0.20**	**1.66**	**0.63**	**1.87**	**15.50**
Nov. 2014	40	*0.20*	0.64	*1.77*	1.25	2.23	7.70
May 2015	38	*0.20*	0.72	*1.94*	1.30	2.13	10.00
Nov. 2020	86	*0.20*	*0.20*	*1.42*	0.56	1.23	22.00
May 2021	81	*0.20*	*0.20*	*1.37*	0.59	1.20	17.00
Nov. 2023	80	0.20	0.22	1.67	0.63	1.94	14.10
May 2024	78	0.20	0.20	1.65	0.63	1.75	15.50

Tests for Differences

Prior to conducting analyses to determine if changes occurred over time, a preliminary question was addressed. Were the September/November and the May data subsets different from each other? If they were different, trend analyses would need to consider variations between them.

Inspection of the three data sets revealed that the same wells were not always sampled during each sampling event. As such, testing for differences between any two sampling event data sets could not be considered paired (dependent) data. For this reason, the Wilcoxon rank-sum (WRS) test (Conover, 1999) was used for the comparisons using the stats package in R. In R, the test is the wilcox.test. Were the monthly distributions (and medians) in each of the data sets statistically equal?

Table 2 displays the results of the monthly comparisons (September/November to May) tests for each data set. All tests used the adjusted data. In Table 2, beginning in Row3, Columns 1, 3, and 5 display the data subsets plus the sample size of each subset (n). Columns 2, 4, and 6 display the corresponding medians. The resulting p-values are 0.936, 0.87, and 0.960, respectively, indicating there are no statistical difference. Further analyses using monthly data were not needed. Consequently, only the data sets: (1) 2014-2015, (2), 2020-2021, and (3) 2023-2024 were considered for further analyses. The sample sizes for the three data sets (Table 1b) are 78, 167, and 158, respectively.

Table 2. Results of Wilcoxon Rank-Sum Test of Monthly Subset Data

Data Set					
2014-2015		**2020-2021**		**2023-2024**	
Subset (n)	Median	Subset (n)	Median	Subset (n)	Median
Sept 2014 (n =40)	1.25	Nov 2014 (n = 86)	0.56	Nov 2014 (n = 80)	0.63
May 2015 (n = 38)	1.30	May 2015 (n = 81)	0.59	May 2015 (n = 78)	0.63
p-value	0.936	p-value	0.870	p-value	0.960

Regarding the three data sets, Figure 1a displays boxplots of the three data sets. Outliers are displayed as circles. The outliers tend to clutter the plots. For this reason, boxplots without the outliers were constructed (Figure 1b). In the figure, note the medians (middle bars in the boxes) of the 2020-2021 and the 2023-2024 data. The median value of the 2023-2024 data appears to slightly greater than the 2020-2021 data. The medians in both data sets are visually less than the median of the 2014-2015 data.

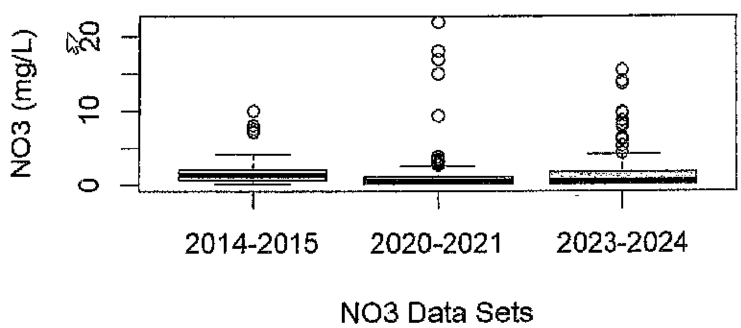

Figure 1a. Boxplots of Three NO3 Data Sets with Outliers

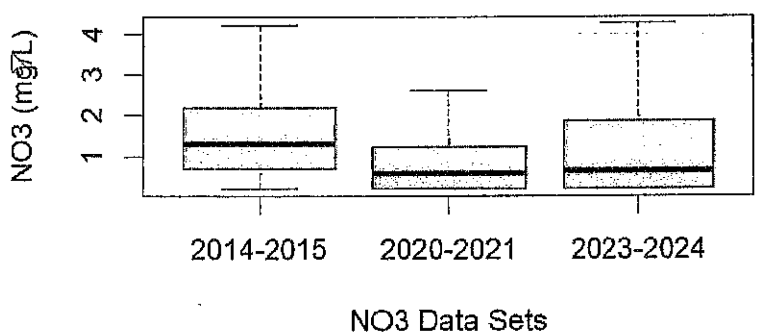

Figure 1b. Boxplots of Three NO3 Data Sets without Outliers.

Are the 2020-2021 and the 2023-2024 data sets statistically different from each other? Two Wilcoxon tests were used to make the comparisons. In the first, data LT 0.20 mg/L in the 2020-2021 data sere were not raised (non-adjusted) to the 0.20 level. In the second, data LT 0.20 were raised (adjusted) to the 0.20 level.

The results of the two tests are found in Table 3. Rows 2 and 3 display the results for the non-adjusted data, whereas Rows 4 and 5 do the same for the adjusted data. Note the medians of the 2020-2021 and the 2023-2024 data sets were not affected by the adjustments. The table displays the Wilcoxon (W) statistics (10846 and 11966, respectively) The resulting p-values using non-adjusted data is 0.006. This suggests the two sets of data are different and the medians in Table 3 suggests that NO3 concentrations in 2023-2024 increased, relative to 2020-2021. If so, this is important information. However, the p-value, using the adjusted data, is 0.113. Although there may an increase, because of the greater MDL value in the 2023-2024 data set, one is not justified in making the claim.

Figure 1b suggests the NO3 concentrations in both the 2020-2021 and 2023-2024 data sets are less than those in 2014-2015. This visual comparison was tested. The latter two data sets were combined and then compared to the early data set (2014-2015), using the Wilcoxon test. Table 4 displays the result. Note, the median of the combined 2020-2024 NO3 data is 0.61 mg/L, whereas the median for the 2024-2015 data is 1.30. Also note, the combined median (0.61) is less than the Q2 concentration of the combined data set Table 1b is 0.69. Table 4 indicates the resulting p-value is <0.001, indicating a decrease in NO3 concentrations since 2014-2015. Visually, Figure 2 displays boxplots of the (2014-2015) and the combined (2020-2024) data sets, with outliers removed.

Table 3. Results of Wilcoxon Rank-Sum Tests, Comparing 2020-20221 and 2023-2024 NO3 Data.

Data Set	Median	W-Statistic	P-value
2020-2021 (Non-Adjusted) (n = 167)	0.57	10846	**0.006**
2023-2024 (Non-Adjusted) (n =158)	0.63		
2020-2021 (Adjusted) (n = 167)	0.57	11866	0.113
2023-2024 (Adjusted) (n = 158)	0.63		

Table 4. Results of Wilcoxon Rank-Sum Test, Comparing 2014 and 2020-2024 NO3 Data.

Data Set	Median	W-Statistic	P-value
2014-2015 (Adjusted) (n = 78)	1.30	893.5	**<0.001**
2020-2024 (Adjusted) (n = 326)	0.61		

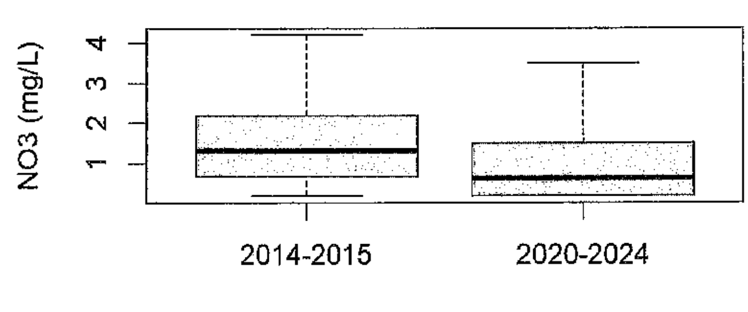

Figure 2. Boxplots of NO3 Concentrations in the Floridan Aquifer System in Alachua County and portion of the Santa Fe River Basin, Florida During Two Periods, 2014-2015 and 2020-2021.

REFERENCES CITED

Conover, W.J., 1999, Practical nonparametric statistics, New York, Joh Wiley and Sons, 367 p.

Copeland. R., Owen, G., and Patarasuk, R., 2023, Trend Analyses of Groundwater Quality int the Upper Floridan Aquifer of Alachua County, Florida, (1987-2021), Tallahassee, FL, AquiferWatch Inc. 31 p.

Helsel, D.R., 2012, Statistics for censored environmental data using Minitab and R, 2nd ed., Nondetects and data analysis: Statistics for censored environmental data, New York, Wiley, 325 p.

R Core Team, 2023, R: A language and environment for statistical computing: R Foundation for Statistical Computing, Vienna, Austria.