A.R.M. Loxahatchee National Wildlife Refuge Total Phosphorus Outlier Analysis and Proposed Alternative Screening Criterion: Distribution Independent Outlier Analysis

For the May 30, 2012 TOC Meeting

Prepared by:
P. Julian[1] and S. Hill[2]

1. Florida Department of Environmental Protection
Office of Ecosystem Projects

2. South Florida Water Management District
Compliance Assessment Unit

Agenda Item 3B

A.R.M. Loxahatchee National Wildlife Refuge Total Phosphorous Outlier Analysis and Proposed Alternative Screening Criterion: Distribution Independent Outlier Analysis.

P. Julian[1]

S. Hill[2]

[1] Florida Department of Environmental Protection, Office of Ecosystem Projects

[2] South Florida Water Management District, Compliance Assessment Unit

BACKGROUND

A proposal for an outlier analysis associated with the Florida Department of Environmental Protection's (FDEP or Department) Data Usability document was initially discussed by the Everglades Technical Oversight Committee (TOC) representatives during the March 2011 TOC meeting. Through discussion among the TOC representatives at subsequent meetings, speculation developed as to whether the use of a conventional outlier analysis, such as a common three standard deviation test (three-sigma test), would classify an inappropriately large volume of data as outliers due to a presumed normal frequency distribution of total phosphorus (TP) concentrations. As part of the February 2012 TOC meeting, Department of Interior (DOI) staff (D. Surratt, Ph. D. and M. Waldon, Ph.D.) presented an alternate proposal to the three-sigma test based on the log-transformed data (Log_{10}). Their analysis is posted on the TOC website.

The DOI analysis concluded that per station, the TP data was not normally distributed and did not follow the assumptions of the three-sigma test. However, the log transformed distribution was approximately log-normal only for data below the 93rd percentile. Overall, DOI's presentation identified that the TP concentration data are neither normally nor log-normally distributed (Figure 1). Assuming the data fit a normal distribution, outlier identification of non-transformed data was conducted, and ultimately back calculated the corresponding number of standard deviations (sigma) necessary for log transformed data, which was 4.1 standard deviations. While the log-normal approach reduces the number of data points identified as outliers, it also does not follow the assumptions of the statistical test due to the distribution of the data for both the transformed and non-transformed data.

In order to avoid complications of parametric statistics and the reliance on data distribution, a non-parametric, distribution-free approach was taken by FDEP and SFWMD (District) staff. As observed by the members of the TOC, rarely are environmental data normally distributed and transformation of the data is typically necessary to facilitate the application of parametric statistics. However, if environmental data, either transformed or non-transformed, do not fit a normal distribution then the application of non-parametric statistics is required for appropriate identification of outliers. Based on the Department and the District's application of a non-parametric statistical test, there is clearly potential for a simpler, more consistent identification of true outliers using a non-parametric approach.

METHODS

Data Source: Total phosphorus data were downloaded from the District's environmental database (DBHYDRO) (South Florida Water Management District 2012). The complete period of record (09/21/1993-01/05/2012) for 14 water quality monitoring stations within the A.R.M. Loxahatchee National Wildlife Refuge (Refuge) (Figure 2) was queried and used for analysis. Consistent with previous analysis (Donatto and Waldon 2012) all data were used for this analysis including flagged/qualified data[1].

Statistical Approach: A non-parametric (distribution-free) approach was applied to detect statistical TP outliers. The frequency of statistical outliers was computed for the upper 95% and 99% confidence limit for both the 95th and 99th percentile of station-specific data and data for all stations combined. Additionally, to determine the minimum detectable difference (effect) of all data collected a power analysis (post-hoc) was conducted as 1-sample T, at power level 0.9999. Moran's I spatial statistical test was also applied to determine the autocorrelation of each station's period of record 99th percentile. All statistical analysis was performed using SAS (Ver. 9.3, Cary, NC, USA), ArcGIS (Ver 10, Redlands, CA, USA) and MiniTab (Ver 15.1, State College, PA, USA).

[1] 3.9% of TP data collected from the 14 water quality monitoring stations within the Refuge contained outliers, Breakdown of outliers can be found in Appendix I, Table 1.

RESULTS

Frequency of statistical outliers and number of outliers are summarized in Tables 1 and 2, respectively. Confidence intervals and percentiles for all stations and each station individually are reported in Table 3 for the period of record between mid-1993 and early 2012. Depending on the degree of scrutiny desired, several levels of discrimination can be applied to data collected. Using the 95% confidence interval (CI) of the 95th percentile resulted in 3.8% of the samples (99 samples) from the combined stations being identified as outliers, and a mean percent frequency for individual stations of 1.89±0.09% (3.57±0.27 samples; mean ± standard error) of samples identified as outliers. This level of scrutiny is very aggressive. Application of 99% CI of the 95th percentile also resulted in 3.8% of the samples (99 samples) from the combined stations being identified as outliers; however, a mean percent frequency for individual stations was calculated to be 0.94±0.07% (1.85 ± 0.19 samples) of samples identified as outliers. The frequency of identification is significantly higher than any outlier detection frequency previously proposed and reflects a high degree of scrutiny. The use of the 95% CI of the 95th percentile could potentially identify samples that simply have elevated yet representative TP concentrations and may not be statistical outliers. On the other hand the application of 99% CI of the 95th percentile identified fewer data points or potential outliers per station, while application of a 95% CI of the 95th percentile identified the same number of data points for the 14 stations when grouped together.

Use of the 99th percentile allows for further flexibility on average when compared to percent outlier detection in other proposed methods. When all stations were grouped together, use of the 99th percentile identified 0.96% of all samples (25 data points) as outliers with a mean of 0.83±0.05% (1.57±0.14) outliers by station. The use of the upper 95% and 99% CI of the 99th percentile resulted in the majority of the data lying within the respective ranges established and not identifying any outliers with the exception being all stations grouped together. This resulted in 0.65% and 0.46% of the data points identified as outliers for the 95% and 99% CI of the 99th percentile respectively.

A large volume of water quality data has been collected at the Refuge sites since early 1993. Based on the amount of available data, FDEP and SFWMD staff investigated whether the statistical power of the available data would yield any detectable difference or effect on the

sample size and standard deviation of the data. Overall (all 14 locations together), the detectable difference for TP data collected within the Refuge is 0.0005 mg/L and the mean between stations is 0.002 ± 0.0002 mg/L; therefore, the current data set is sufficient to draw statistically significant conclusions. Furthermore, the spatial pattern for the 99th percentile for the period of record of each station is spatially auto correlated as indicated by the Moran's I statistical analysis (Appendix I, Figure 1; Moran's Index: 0.497, variance: 0.056, z-score: 2.429, ρ-value: 0.015). Even though this analysis was only conducted using 14 spatial data points, a clear qualitative and quantitative difference is apparent.

DISCUSSION:

FDEP, SFWMD and DOI staff have spent a significant amount of effort and attention to identifying statistical outliers for TP data collected within the Refuge. Based on the assessments by the various agencies, FDEP staff believe there are two potential paths forward for the TOC to consider based on the findings of this analysis and with relation to the specific charge of the TOC: 1) apply the 99th percentile outlier analysis as a statistically valid method for determining TP outliers, or 2) eliminate the need for a statistical outlier test as TOC reserves the right to evaluate all the data, regardless of its status as an outlier, when determining compliance with the Settlement Agreement.

FDEP and SFWMD staff believe that the application of the non-parametric outlier analysis using the 99th percentile (without 95% or 99% confidence interval) is appropriate valid statistical outlier test. The use of this percentile provides a robust and simpler method for identification of statistical outliers independent of distribution and identifies the upper 1% of the data as potential outliers. As observed in DOI's previous analysis (Surratt and Waldon 2012), the rate of outlier detection is comparable. The application of the 99th percentile does not require any specific statistical program other than Microsoft Excel® with the "=percentile (data array,0.99)" function, assuming the data set is structured effectively to determine the percentile for each station or time period. Further evaluation identified a spatial relationship (Appendix I, Figure 1) regarding the 99th percentile of the period of record for all stations throughout the Refuge; therefore, when applying an outlier analysis, such an analysis should be done by stations and not the overall data due to the observed north-south TP gradient.

Despite the findings that a statistically appropriate outlier test is available, further evaluation would be required to determine whether separate statistical outlier tests would need to be applied for Appendix A. However, the question in this particular case becomes whether any statistical outlier test should be applied at all as a tool for identifications of outliers. Regardless of whether data are identified as outliers or not, the TOC reserves the right to take into consideration all data generated to determine compliance with the Settlement Agreement. Therefore, the Department is recommending the application of a statistical outlier test not be adopted at this time. Discussions between Department and DOI technical staff suggest that this recommendation is considered mutually acceptable. The Department proposes to modify the language within the cover memo for the data usability document to appropriately represent that references to the application of statistical tests to determine outliers shall not be construed to apply to compliance data evaluation performed by the TOC. The Department proposes to make these changes and post the final proposed cover memo prior to the next quarterly TOC meeting for discussion and vote.

LITERATURE CITED:

South Florida Water Management District. DBHydro-Hydrometerologic, water quality and hydrogeologic data retrieval system. http://www.sfwmd.gov/dbhydroplsql/show_dbkey_info.main_menu [Accessed February 2012]

Surratt, D., and M. Waldon. 2012. A.R.M. Loxahatchee National Wildlife Refuge Total Phosphorous Outlier Analysis and Proposed Alternative Screening Criterion. February 2012 Everglades Technical Oversight Committee.

United States v. South Florida Water Management District. 1988. Case No. 88-1886-CIV-HOEVELER.

Figures and Tables:

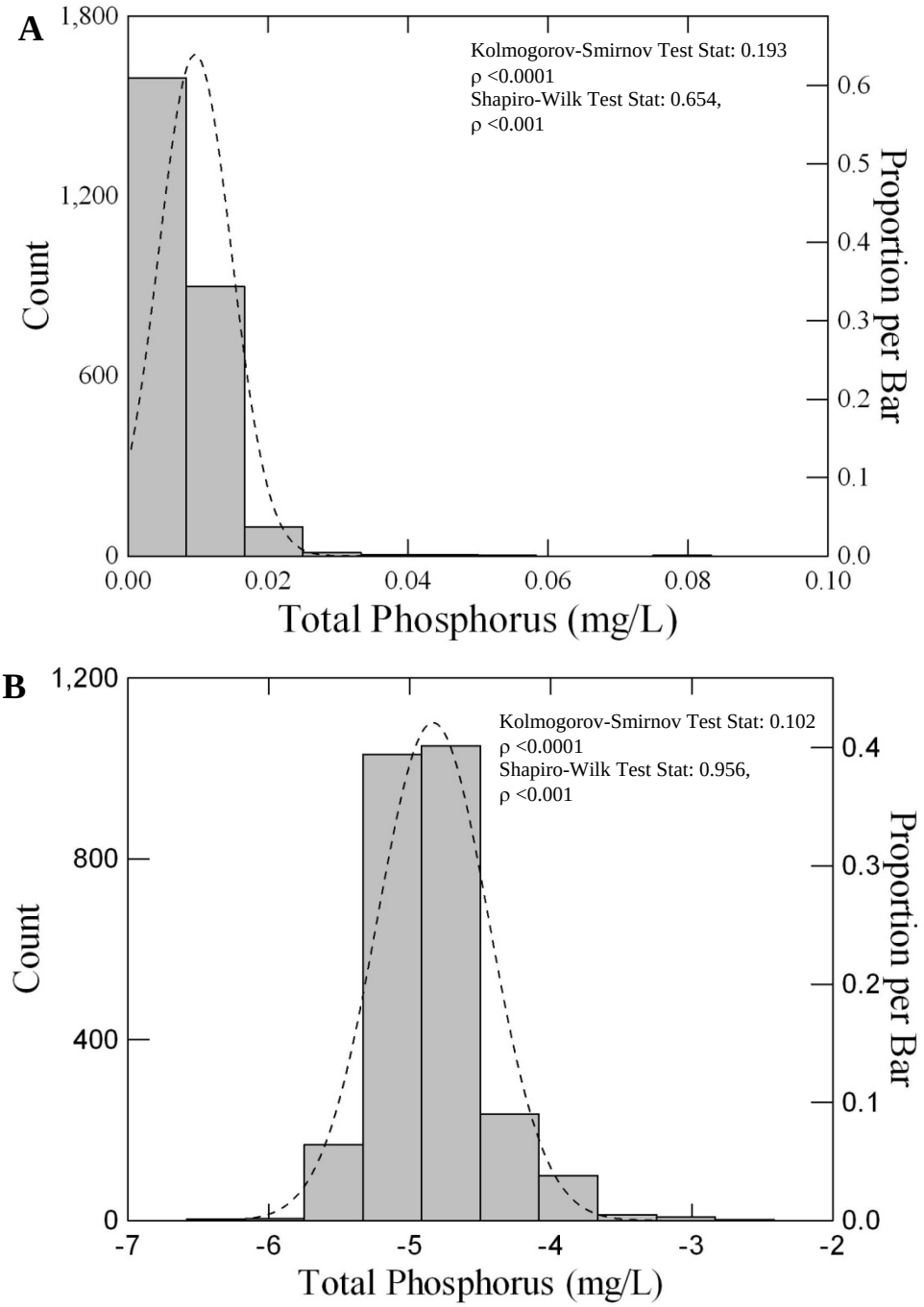


Figure 1. Normal distribution of all TP data for all stations (A) and log-normal distribution of all TP data for all stations (B) for all 14 A.R.M. Loxahatchee NWR water quality stations between mid 1993 (09/21/1993) and early 2012 (01/05/2012).

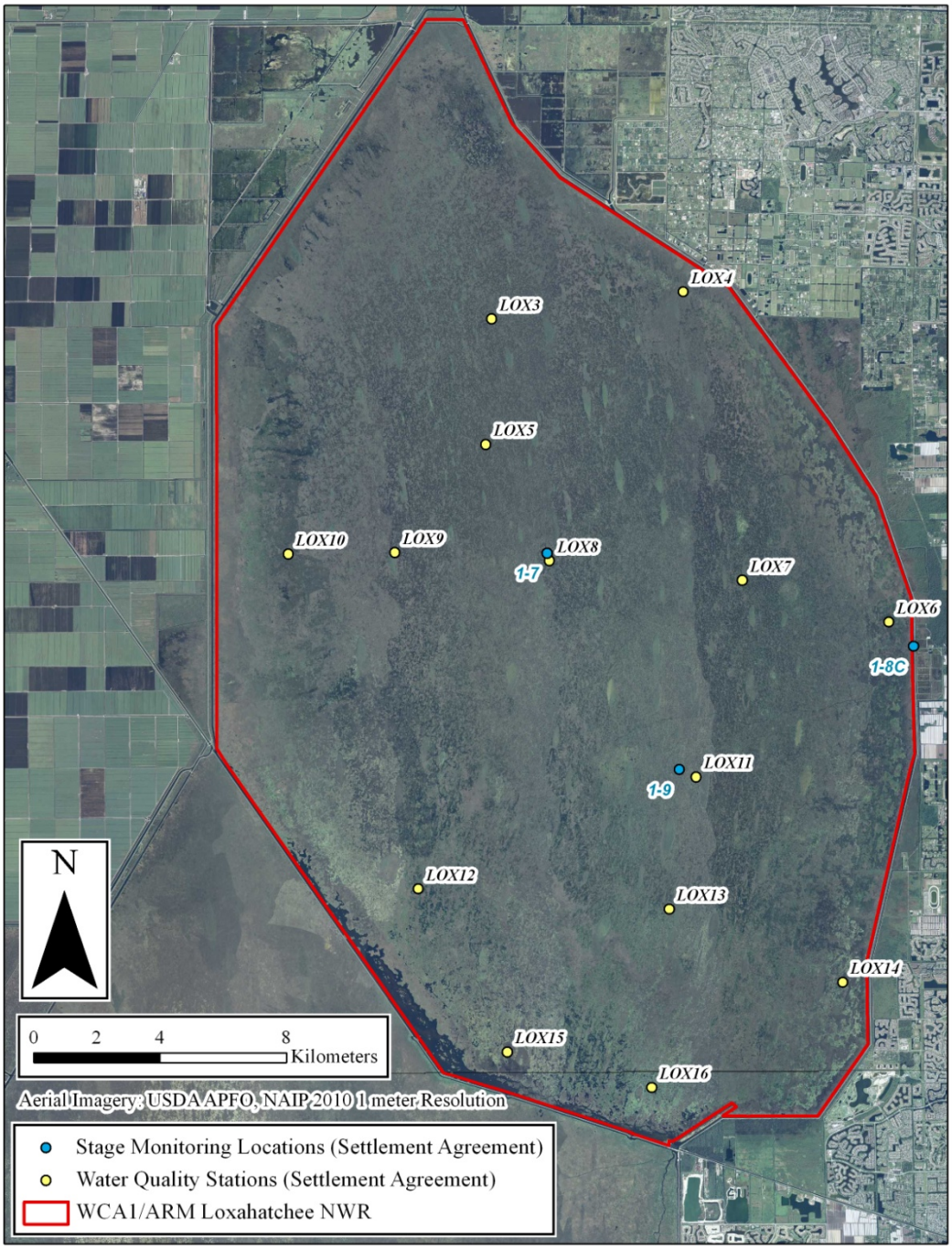


Figure 2. A.R.M. Loxahatchee NWR water quality sampling and stage elevation locations used in Settlement Agreement compliance calculations.

Table 1. Frequency of statistical outliers as percentage of total sample for all 14 stations and each individual station for each proposed limit. Mean and standard error of percentage for all stations are also presented.

		Frequency of Statistical Outliers (as percentage of samples)				
Station ID	**N**	**> Upper 95% CI for 95th Percentile**	**> Upper 99% CI for 95th Percentile**	**> 99th Percentile**	**> Upper 95% CI for 99th Percentile**	**> Upper 99% CI for 99th Percentile**
14 Stations	2614	3.79%	3.79%	0.96%	0.65%	0.46%
LOX10	156	1.28%	0.64%	0.64%	***Out of Range***	***Out of Range***
LOX11	202	1.98%	0.99%	0.99%	***Out of Range***	***Out of Range***
LOX12	220	2.27%	0.91%	0.91%	***Out of Range***	***Out of Range***
LOX13	194	1.55%	1.03%	0.52%	***Out of Range***	***Out of Range***
LOX14	214	1.87%	1.40%	0.93%	***Out of Range***	***Out of Range***
LOX15	216	2.31%	0.93%	0.93%	***Out of Range***	***Out of Range***
LOX16	209	1.91%	0.96%	0.96%	***Out of Range***	***Out of Range***
LOX3	125	1.60%	***Out of Range***	0.80%	***Out of Range***	***Out of Range***
LOX4	159	1.89%	0.63%	0.63%	***Out of Range***	***Out of Range***
LOX5	144	2.08%	0.69%	0.69%	***Out of Range***	***Out of Range***
LOX6	200	2.00%	1.00%	1.00%	***Out of Range***	***Out of Range***
LOX7	203	2.46%	0.99%	0.99%	***Out of Range***	***Out of Range***
LOX8	208	1.44%	1.44%	0.96%	***Out of Range***	***Out of Range***
LOX9	164	1.83%	0.61%	0.61%	***Out of Range***	***Out of Range***
Mean		1.89%	0.94%	0.83%	***Majority out of range***	
SE		0.09%	0.07%	0.05%	***Majority out of range***	

Table 2. Number of statistical outliers for all 14 stations and each individual station for each proposed limit. Mean and standard error of percentage for all stations are also presented.

		Number of Statistical Outliers				
Station ID	**N**	**> Upper 95% CI for 95th Percentile**	**> Upper 99% CI for 95th Percentile**	**> 99th Percentile**	**> Upper 95% CI for 99th Percentile**	**> Upper 99% CI for 99th Percentile**
14 Stations	2614	99	99	25	17	12
LOX10	156	2	1	1	***Out of Range***	***Out of Range***
LOX11	202	4	2	2	***Out of Range***	***Out of Range***
LOX12	220	5	2	2	***Out of Range***	***Out of Range***
LOX13	194	3	2	1	***Out of Range***	***Out of Range***
LOX14	214	4	3	2	***Out of Range***	***Out of Range***
LOX15	216	5	2	2	***Out of Range***	***Out of Range***
LOX16	209	4	2	2	***Out of Range***	***Out of Range***
LOX3	125	2	***Out of Range***	1	***Out of Range***	***Out of Range***
LOX4	159	3	1	1	***Out of Range***	***Out of Range***
LOX5	144	3	1	1	***Out of Range***	***Out of Range***
LOX6	200	4	2	2	***Out of Range***	***Out of Range***
LOX7	203	5	2	2	***Out of Range***	***Out of Range***
LOX8	208	3	3	2	***Out of Range***	***Out of Range***
LOX9	164	3	1	1	***Out of Range***	***Out of Range***
Mean		3.57	1.85	1.57	***Majority out of range***	
SE		0.27	0.19	0.14	***Majority out of range***	

Table 3. Confidence interval and percentile values for all 14 stations together and each individual station.

	Percentile Limits TP Concentration (mg/L)				
Station ID	**Upper 95% CI for 95th Percentile**	**Upper 99% CI for 95th Percentile**	**99th Percentile**	**Upper 95% CI for 99th Percentile**	**Upper 99% CI for 99th Percentile**
14 Stations	0.0170	0.0170	0.0240	0.0297	0.0340
LOX10	0.0200	0.0272	0.0230	***Out of Range***	***Out of Range***
LOX11	0.0190	0.0200	0.0200	***Out of Range***	***Out of Range***
LOX12	0.0193	0.0210	0.0210	***Out of Range***	***Out of Range***
LOX13	0.0200	0.0222	0.0240	***Out of Range***	***Out of Range***
LOX14	0.0150	0.0175	0.0180	***Out of Range***	***Out of Range***
LOX15	0.0174	0.0200	0.0200	***Out of Range***	***Out of Range***
LOX16	0.0170	0.0200	0.0200	***Out of Range***	***Out of Range***
LOX3	0.0391	***Out of Range***	0.0400	***Out of Range***	***Out of Range***
LOX4	0.0277	0.0470	0.0460	***Out of Range***	***Out of Range***
LOX5	0.0229	0.0559	0.0280	***Out of Range***	***Out of Range***
LOX6	0.0181	0.0200	0.0210	***Out of Range***	***Out of Range***
LOX7	0.0189	0.0190	0.0190	***Out of Range***	***Out of Range***
LOX8	0.0200	0.0228	0.0240	***Out of Range***	***Out of Range***
LOX9	0.0205	0.0250	0.0250	***Out of Range***	***Out of Range***

Table 4. Selected descriptive statistics and power analysis for 1-sample T at 0.9999 power level for each station and all stations combined throughout the period of record.

								Power Analysis for 1-Sample T , Power Level=0.9999
	N	**Min**	**Max**	**Arithmetic Mean**	**SE**	**Geometric Mean**	**Standard Deviation**	**Minimum Detectable Difference (effect) at current samlpe size and St Dev**
LOX10	156	0.004	0.043	0.009	0.0003	0.008	0.004	0.002
LOX11	202	0.003	0.026	0.009	0.0003	0.008	0.004	0.002
LOX12	220	0.004	0.047	0.008	0.0003	0.007	0.004	0.002
LOX13	194	0.004	0.035	0.009	0.0003	0.008	0.004	0.002
LOX14	214	0.002	0.020	0.008	0.0002	0.007	0.003	0.001
LOX15	216	0.002	0.034	0.007	0.0002	0.007	0.003	0.001
LOX16	209	0.004	0.078	0.009	0.0004	0.008	0.006	0.002
LOX3	125	0.004	0.05	0.010	0.0005	0.009	0.006	0.003
LOX4	159	0.004	0.054	0.011	0.0005	0.010	0.006	0.003
LOX5	144	0.004	0.08	0.010	0.0006	0.009	0.007	0.003
LOX6	200	0.002	0.044	0.008	0.0003	0.007	0.004	0.002
LOX7	203	0.004	0.027	0.009	0.0002	0.008	0.003	0.001
LOX8	208	0.004	0.048	0.009	0.0003	0.009	0.004	0.002
LOX9	164	0.003	0.034	0.008	0.0003	0.008	0.004	0.002
Overall	2614	0.002	0.08	0.009	0.0001	0.008	0.005	0.0005

APPENDIX I:

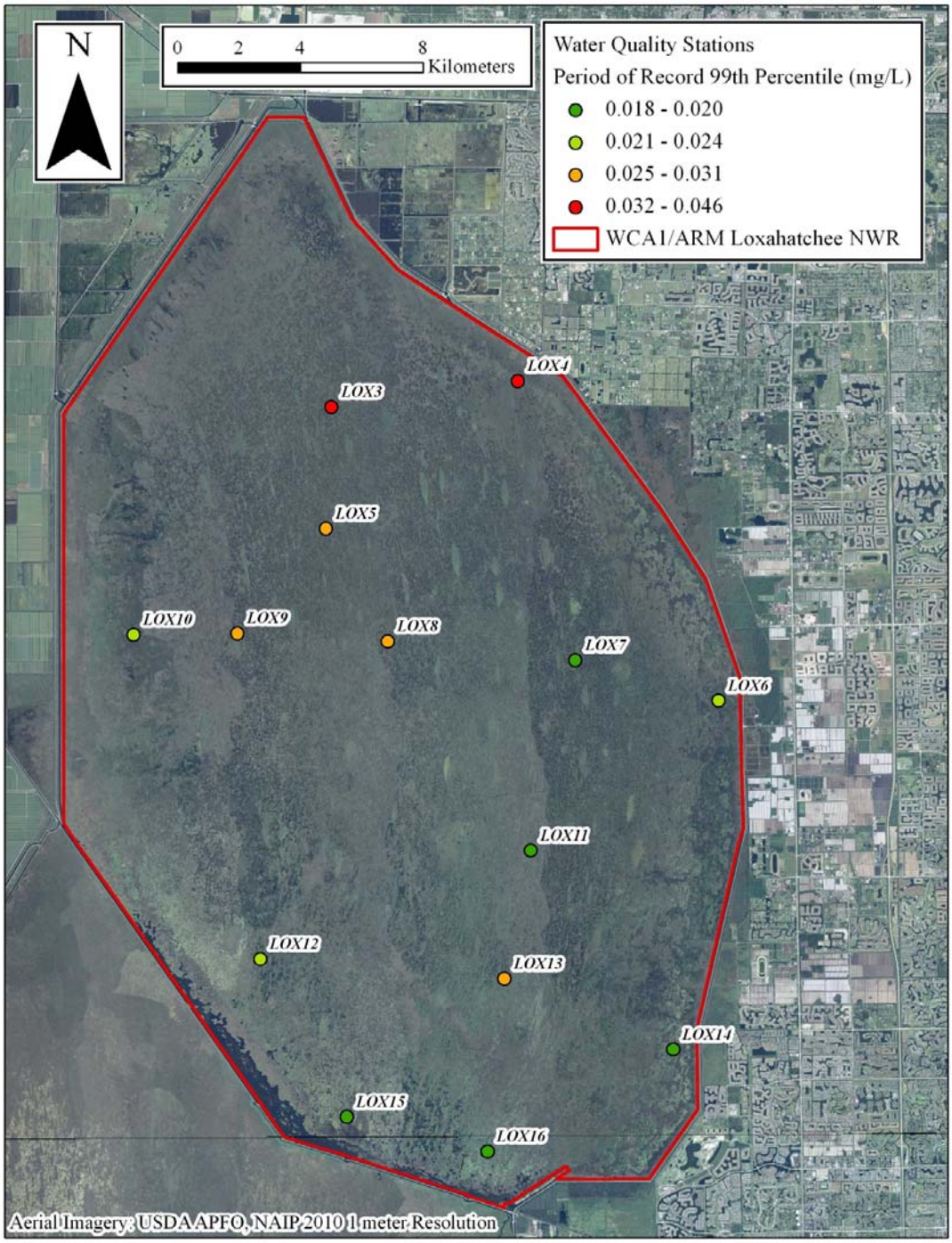
Appendix I, Table 1. data qualifier breakdown for all Total Phosphorous data collected from the 14 water quality stations within A.R.M. Loxahatchee NWR.

Qualifier	Count	Description
?	24	Data is rejected and should not be used.
I,J	1	The reported value is between the lab method detection limit and the lab practical quantitation limit; Estimated value: value not accurate.
J	10	Estimated value: value not accurate.
J3	11	Precision or Accuracy Criteria Not Met
PMF	4	Flag set as Project Managers Request; Invalid Data.
PMR	46	Flag set at Project Managers Request; Invalid Data
Q	4	Out of Holding Time
Y	1	The laboratory analysis was from an unpreserved or improperly preserved sample. The data may not be accurate.

Total Number of Qualifiers: 101

Total Number of Samples: 2614

Percentage of Qualifiers: 3.9%



Appendix I, Figure 1. Spatial representation of the 99th percentile Total Phosphorous concentration (mg/L) throughout the period of record for each of the 14 water quality station within the refuge. Even though there are very few points (<30), the period of record 99th percentile is spatially auto-correlated (Moran's Index: 0.497, variance: 0.056, z-score: 2.429, ρ-value: 0.015).