Data Compilation and Filtering

The goal of the data compilation effort was to provide a comprehensive inventory of data from active monitoring programs within the CHNEP boundaries. Data were compiled as part of the ongoing WaterAtlas.org support, and will continue to be updated as new data are available.

Data were screened to meet quality objectives developed by the Florida Department of Environmental Protection as described in the FDEP methods (Appendix A). Duplicate values and equipment blanks were not included in the data compilation effort. Values coded as below the detection limit were set to the detection limit. Other quality assurance codes were used to screen the data as described by FDEP methods (Appendix A).

Kendall Tau Trend Test

The core statistical trend used for this project is the seasonal Kendall Tau Test for Trend (Helsel and Hirsch 1982). Implementation of the procedure follows the description provide by Reckhow et al. 1993. The procedure used in the analysis is based upon the Kendall Tau Trend Test developed by Janicki Environmental as described in the Water Quality Data Analysis Report for Charlotte Harbor National Estuary Program (2013), which leverages the Kendall Tau Fortran programs developed by the United States Environmental Protection Agency and available from the USEPA Laboratory in Corvallis, Oregon. The full procedure for data screening and trend analysis was implemented on the Water Atlas using the R statistics package, EnvStats (Millard 2013). The complete R code is available within the data download on the Water Atlas Trend Analysis webpage. Readers are encouraged to view the R code for additional details about the procedure outlined below.

The first step of the analysis averages the time series data to monthly values.

Next, a correlation analysis is performed for each monthly value, the previous month's value, two months prior, etc., until correlation statistics have been calculated for all previous months up to 15 months prior.

In the third step, a determination is made as to whether seasonality exists in the time series of data. An operationally defined and objective test to identify the presence of seasonality is applied. If the data presented by the plot have seasonality, then one would expect the 6 month lag values to be negatively correlated and the 12-month lag values to be positively correlated. The objective test measures the proportional distance between the zero line and the lower 95% confidence limit for the 6-month lag correlation, and the proportional distance between the zero reference line and the upper 95% confidence limit for the 12-month lag correlation. If the sum of the two distances are greater than 1, or if the 12-month lag distance is greater than 1, then seasonality is determined to exist.

If the data are determined to be seasonal, then the data are adjusted for season by subtracting the median monthly value from each data point. The season-adjusted data are then applied to a Kendall Tau algorithm. The Kendall Tau test determines the slope of the time series of data, and p-values for various data conditions.

In the next step of the analysis, the season-adjusted and de-trended data are prepared in the form of a correlogram to test for the presence of autocorrelation in the time series. If the 1-month lag or the 2-month lag is significantly correlated with the present values, then the data are identified as auto-correlated and an adjustment is made to the p-value.

In the final step of each trend analysis the appropriate p-value (corrected for autocorrelation if necessary), significance assessment (based on alpha=0.05), slope, autocorrelation assessment (present/absent), and seasonality assessment (present/absent) of the trend analysis are compiled from each step's output and tabulated in summary tables of trend test results.

References:

Millard SP (2013). EnvStats: An R Package for Environmental Statistics. Springer, New York. ISBN 978-1-4614-8455-4

Reckhow, K. K Hepford and W. Warren Hicks. 1993. Statistical method for the analysis of lake water quality trends. EPA 841-R-93-003.

Appendix A

FDEP Data Exclusion Screens

Removal of results reported in Florida STORET that did not include units, or included units that were inappropriate for the particular analyte: These were excluded as the results could not accurately be quantified, or relied upon for assessment purposes under the Impaired Waters Rule (IWR).

Results reported as negative values: It was concluded that any results reporting a negative value for the substance analyzed represent reporting errors: Credible data could not have any values less than the detection limit (in all cases a positive value) reported. Therefore, results reported as negative values could not be relied upon for assessment purposes under the IWR.

Results reported as any of “888" "8888" "88888" "888888" "8888888" and "999" "9999" "99999" "999999" "9999999": Upon investigation, all data reported using these values were found to be provided by a particular Water Management District. The District intentionally coded the values in this manner to flag the fact that they should not be used, as the values reported from the lab were suspect. The data coded in this manner was generally older.

Removal of J-qualified Results: J-qualified results from this same Water Management District were excluded from the assessments after the District brought to the attention of FDEP that their use of the J-qualifier was not consistent with FDEP’s use of the FDEP J-qualifier.

Removal of extremely old U.S. Geological Survey (USGS) data (beginning of the previous century): This data did not have complete date information available. Accurate date information is required to be able to assess results under the IWR. USGS data using USGS parameter codes of 32230 or 32231 were also excluded from assessments performed under the IWR, based on information in a memo that was sent from USGS.

Removal of results for iron which were confirmed to be entered into the South Florida Water Management District database, DBHYDRO, using the wrong legacy STORET parameter codes. They were excluded from the assessment under the IWR.

Removal of results reported using “K”, "U", "W", and "T" qualifier codes (all of which suggest that the result was below MDLs, i.e., method detection limits), when the reported value of the MDL was greater than the criterion, or the MDL was not provided: In order to be able to compare a non-detect result to a criterion value, it is necessary to know that it was possible to measure as low as the numeric value of the criterion.

Removal of certain results reported using an “I” qualifier code (meaning that the result value was between the MDL and the practical quantification limit, or PQL): These results were excluded from assessments performed under the IWR, where the MDL was not provided, or where the MDLs and PQLs were inconsistent with the rest of the data record,.

Removal of certain results reported for metals using an “I” qualifier code: Where the criterion is expressed as a function of hardness, and the numeric value of the metal criterion corresponding to the reported hardness value was between the MDL and the PQL, these data were excluded from the assessments performed under the IWR.

Removal of results reported using an “L” qualifier code (meaning that the actual value is known to be greater than the reported value) where the reported value for the upper quantification limit was less than the criterion: The reasoning for excluding these data follows a logic somewhat similar to the reasoning to the cases discussed above for results reported as below the method detection limits.

CHNEP Trend Analysis Methods

Removal of results reported with a “Z” qualifier code (which indicates that the results were too numerous to count): These results were excluded because there was no consistency among data providers in how data using this qualifier code were reported: Some data providers entered numeric estimates of bacteria counts, while other data providers entered the dilution factor. As a result, meaningful interpretation of data reported using this qualifier was not uniformly possible.

Removal of results reported with an “F” qualifier code (which indicates female species): Since the IWR does not assess any analytes for which this qualifier code would be appropriate, the intended meaning of the use of this qualifier code is unknown. The reported result is therefore rendered uninterpretable (although there are very few instances of the use of this qualifier code in the IWR dataset, and it is possible that some agencies use this to indicate a field measurement).

Results reported with an “O” qualifier code (which indicates that the sample was collected but that the analysis was lost or not performed): Exclusion of results reported using this qualifier code is self-explanatory.

Removal of results reported with an “N” qualifier code (which indicates presumption of evidence of the presence of the analyte). Comparing concentrations of analytes to criteria from the Florida Standards requires a numeric result value: Presence or absence, for the purposes of assessments performed under the IWR, is not sufficient information upon which to base an impairment decision.

Removal of results reported with a "V" or "Y" qualifier code (which indicate the presence of analyte in both the environmental sample and the blank, or a laboratory analysis that was from an unpreserved or improperly preserved sample): Such data may not be accurate. Use of these codes indicates that the reported result is not sufficiently reliable to be used in IWR assessments.

Removal of certain results reported with a “Q” qualifier code (which indicates that the holding time was exceeded): These data were reviewed to determine if the holding time was exceeded. When appropriate, such data were excluded from the assessments. These reviews were performed manually, not as part of the automated processing of the IWR data.

Removal of result values based on recommendations from FDEP’s Environmental Assessment Section as a result of lab audits performed on behalf of the TMDL program: The data excluded based on lab audits were generally analyte-specific and referred to a specific time frame. While the data issues encountered are variable, lack of acceptable, or verifiable, records is a common issue.