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

For the May 30, 2012 TOC Meeting

**Florida Department of Environmental Protection &
South Florida Water Management District**

Agenda Item 3B

A proposal for an outlier analysis associated with the Florida Department of Environmental Protection's (FDEP) 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 phosphorous (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. 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.

Applying several non-parametric statistical tests to determine outliers, identified the use of the 99th percentile (Exhibit 1) as a more robust, yet simpler method for identification of statistical outliers independent of distribution and the rate of outlier detection was comparable to the methods presented by DOI staff in their analysis. In addition, the detection efficiency varied spatially when using the 99th percentile throughout the period of record for all stations

throughout the Refuge. This suggests that any outlier analysis may be more appropriately undertaken for individual stationsand not the overall data set due to the north-south gradient of TP.

Despite the findings that a statistically appropriate outlier test is available, the question in this particular case is whether any statistical outlier test should be applied at all as a tool for 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. Based on that charge, the Department is recommending that 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 mutually acceptable. As such, 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.