Proposed methods to Calculate Nutrient Standards for Reservoirs

Outliers:

- The Chlorophyll *a* data sets from reservoirs were plotted using normal probability plots, and it was concluded that the raw data in all the sets did not meet the assumptions of a normal distribution.
- Tukey's Boxplot method was used (Tukey, 1977) for exploratory data analysis and possible outlier identification because the method does not have a distributional requirement.
- Previous analyses by the USGS that were shared with stakeholder groups utilized a simple visual identification of outliers. The 1.5 IQR (InterQuartileRange) excluded approximately the same numbers of points, therefore the 1.5IQR was used to identify outliers.

Distributions:

- Distribution issues were complicated by the multiple detection limits and reporting limits across and within many of the datasets.
- A maximum likelihood method of estimation possibly could have produced more accurate estimations of data points below the detection limit and their effect on the central tendency (Singh, 2004). However, because no distributional model seemed to represent the raw data well, the consideration of many of the methods became essentially unviable.
- The simple substitution method of one half the detection limit for the detection method was ultimately chosen in spite of the documented lower performance of simple substitution methods (Helsel & Hirsch, 1992). This method was chosen due to the issues of multiple limits, and the lack of a distributional model that described the data.

Very Low Reported Values:

- There were some concerns about data which was reported as having very low concentrations.
- Local labs, including the Houston TCEQ lab, do not use extraction techniques capable of very low limits, as they are expensive and require very specialized equipment and personnel experienced in the procedures.
- Practical detection limits are an outcome of simply setting the volume to be delivered to the lab. Today labs are currently acknowledging a reporting limit somewhere in the range of 3 to 5 ug/L Chl *a*, depending on the lab and the detection technique used.
- Very low reported values were raised to be equal to a 1 µg/L value. The value of 1 µg/L was originally chosen due to the behavior of log transformations when the parameter of interest is below one. This value was retained as this would also place these values close to the one half detection limit of the current reporting range specified by both protocols.
- On average most datasets had few of these data points, and while staff felt it was appropriate; the overall effect of the adjustment should be very small.

Prediction Interval:

Prediction intervals were calculated with the 99% confidence level. A calculated method using the following equation as outlined in Hahn and Meeker, 1991 is:

$$\hat{\tilde{y}}_{10} = \bar{x} + t_{(1-\alpha, n-1)} (1/m + 1/n)^{1/2} s$$

where: Y is the interval, x bar is the mean, *t* is the appropriate students *t* for the desired alpha, m is the future number of samples which is considered to be 10, and n is the number of samples in the original dataset.