Review of NEIWPCC Report

by NJDEP Office of Science Statistician

August 21, 2012

The treatment of non-detects has not been adequately discussed. On page 232 it is mentioned that phosphorous has a high proportion of non-detects. This indicates that it is an issue and needs to be addressed.

The summary tables present minimal information. Since there are non-detects, skewness and possibly other issues of concern, more thorough summary tables should be presented. They should include the number of non-detects and information on the distribution of the detection limits. Some comments on outliers and skewness should also be included in the table discussion.

On page 67 a mention is made of zero values in the data. Are these non-detects? The treatment of non-detects needs to be discussed for all analyses in this report, including PCA, Regression, Correlation and ANOVA.

The approach taken in using PCA (pg. 18) in this report is not standard and no documentation is presented to justify it. Typically, to develop an index using PCA, the scores of the first few principal components would be examined. If the first eigenvalue (score variance) comprises a large amount of the total variability, then the first principal component might be taken as the index. If weighting the index is desired then the first eigenvalue would be used as a weight. In this report, there do not seem to be any attempts to assess the adequacy of using only the first principal component.

The approach taken in this report is to use the squared component of the eigenvector as a multiplicative weight for that component of the index. The justification is that this weight would be the variance of the component. This claim is not correct. If the variables had been standardized to a variance of 1, then there would be some basis for this, although correlations between variables would also have to be considered. The SAS code in the appendices shows that no variance standardization was done during the PCA analysis and it did not appear to have been done before that. The use of multiplicative weighting should be justified as well as this particular weighting method.

The only justification for combining the weighted and raw indices is given on page 78. This states that it integrates the multiple indicators and their variability. The advantage of this approach is not obvious and requires some justification and documentation. Combining the two indices might serve to blur any useful measure rather than improve it.

The PCA's were performed on one to three variables (pg. 369). A PCA on one variable provides no information and should not be included.

The variables have large coefficients of variation, indicating skewness or some large outliers. Since the eigenvectors in a PCA will be influenced by this, the skewness and outliers should be studied and possibly remedied by transformations.

On page 185 some regressions are discussed. They all have very low R^2. Even though some of these are significant, it would call for a careful examination of the data to determine whether the regression relationship is valid. Residual plots should be presented to look for deviations from a linear fit.