Technical Comments on the Use of a Drought Severity Classification System in Surface Water Quality Monitoring

On June 6, 2013, Bill Harrison from the TCEQ SWQM team presented his findings on developing a statistical model to evaluate effects of drought on water quality. To that end, the model utilized the Drought Severity Classification System developed by the National Drought Mitigation Center (NDMC). The NDCM Classification System aggregated various indicators of drought in soils, hydrology, and climatology into five categories ranging from D0, which is abnormally dry, to D4, exceptional drought. Compiling weekly information on the Drought Severity Classification System, the SWQM team conducted a single factor Analysis of Variance (ANOVA) to determine whether the drought categories impacted water quality. That is, the team attempted to correlate an increase in drought severity with an increase in contaminant concentration[1]. The results were mixed. Some water quality parameters showed good correlation while for others, the correlation was ambiguous. In fact, based on the presentation, the highest percentage of sampled sites for any one parameter that showed significant impacts from drought was only about 50%. If the end result of the analysis is to account for the effects of drought on the water quality, then higher rates of significance are needed. Otherwise, basing listings/delisting on this analysis are at best similar to a coin toss.

To ameliorate this problem, the City of Austin Watershed Protection Department is offering the following suggestions. First, the statistical analysis should consider impact on water quality from factors other than flow. During the presentation, Mr. Harrison acknowledged that other factors, such as spring flow, point sources, and reservoir releases, have impacts on the water quality that were not assessed in the statistical model. The City would also suggest "Eco-Region" has an additional factor since geographic differences often contribute to flow regimes.

Second, water quality parameters are often correlated to one another. For instance, conductivity can be thought of as a function of chloride concentration. If the analyses on the parameters are performed separately, then rates of significance may be overinflated. This will obscure the true impact from drought and create difficulties in evaluating the data. The City recommends performing a multi-factor Multivariate Analysis of Variance (MANOVA). Additional problems that should be considered with the data include autocorrelations (also called serial correlations) in the Drought Severity Classification System. That is, drought classifications that are high one week will tend to be high the next week. This will result in a biased data set that needs to be corrected.

Third, diagnostics of the model should be provided. Doing so, verifies that the assumptions used in the model are appropriate. For example, ANOVA assumes that the residuals are normally distributed with a constant variance. Diagnostic plots of the residuals could validate the model and make the process more transparent. Furthermore, if the plots do not validate the assumptions of the model, the plots could point to some underlying structure in the data that can be transformed to make the data fit the assumptions. However, this cannot be determined until these diagnostic plots are shown.

[1] One may think of single factor ANOVA as a specific form of linear regression.

Finally, the overall goal of the analysis seems to be headed in a direction where a decision on whether to list or delist a water body is made based on the severity of the drought. Like all decisions based on statistics, errors will be made just due to random chance. Thus, it seems important to construct an operating characteristic curve to examine the errors involved in these decisions. (One possible curve can be constructed by implementing the heuristic given in the addendum on the next page.) Given an operating characteristic curve for the model, stakeholders and interested parties can look at the error rates for both decisions, and a balance between the competing interests of falsely listing and falsely delisting impairments based on drought can be reached.

Statistical analysis is a powerful tool into looking for effects in water quality due to flow. However, the relationship developed by the analysis is typically non-linear with many correlations between the water quality parameters. More insight into this problem can be gained by using all of statistical tools available to the user. This includes multivariate and multifactor techniques to avoid other confounding and/or nuisance factors, and time series analysis to look for time lags. Diagnostic plots can be useful in examining the structure of the data and help in developing and validating an appropriate statistical model. Finally, an operating characteristic curve is a way to assess the uncertainty in the statistical model. Applying these suggestions may provide a path forward towards such a model.

Addendum: Constructing an Operating Characteristic Curve for a Statistical Model Examining the Relationship between Drought and Water Quality

Suppose that the water quality of a water body is not impacted by drought. Data is collected at that site, and a statistical model is developed. Then, there are two possible outcomes. Based on the data collected, the model would either show that indeed there is no impact in the water quality from the drought or it would show that there is an impact from drought. In the case of the model showing no impact, then a correct decision was made since in reality, that water body is not impacted by drought. In the latter case, an error was made since the model is saying there is an impact, when, in fact, there is no impact. This error can come about due to measurement or sampling errors. If one formulates this in terms of hypothesis testing, then the null hypothesis is that there is no effect in water quality from drought, and the alternative hypothesis is that there is an effect in water quality from drought. Equivalently, the null hypothesis can be stated as the slope of the line through the treatments is zero and the alternative is that the slope of the line through the treatments is greater than zero. If this set of hypotheses is chosen, then the error committed is a Type I error.[2]

It is important to remember that the slope obtained from the data is just one realization of all possible slopes that could have been obtained. Because sampling produces randomness in the data, a different sampling campaign could have resulted in a different slope.

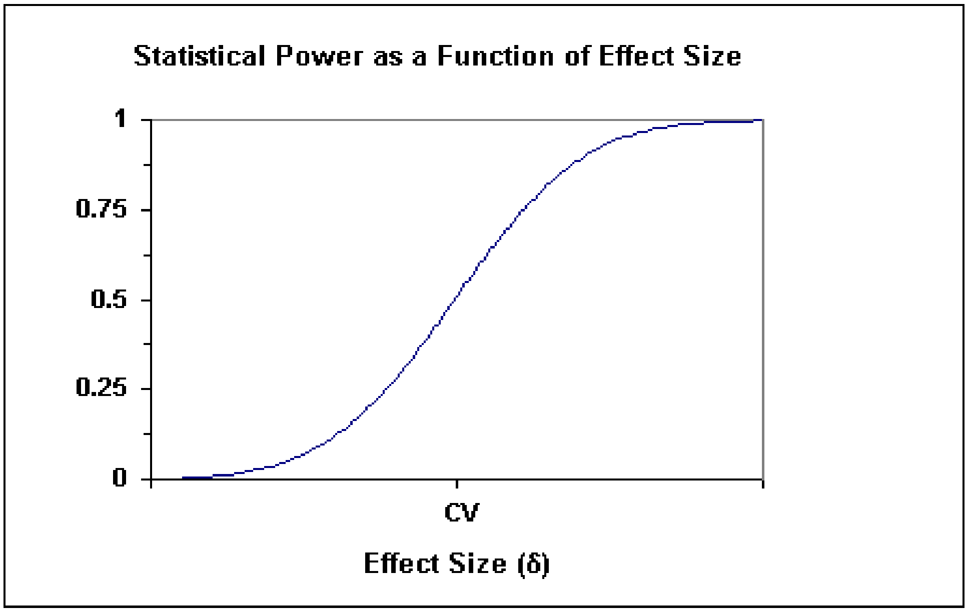
Now to begin this heuristic, assume that the observations are normally distributed about a slope of zero. That is, assume that the data showed that there is no impact from drought. (This assumption can be checked through diagnostic plots during the model selection phase.) Since the slope of the line is a linear combination of the observations, then the slope of the line will also follow a normal distribution with mean of zero and a variance of σ^2/S_{XX}. (Any text on linear regression can be consulted to calculate these variables.) A simulation can now be developed to construct the operating characteristic curve.

One realization is drawn from this normal distribution of slopes and compared to the critical value, $k \cdot \sigma^2/S_{XX}$, where k is some x-value of the t-distribution. If this realization is above the critical value, then an error has been made within this simulated test. If it is below the critical value, then the correct decision was made. Perform this simulation at least 1,000 times and count the number of times an error was made. The percentage of errors made can be quantified as the Type I error rate.

Now, to simulate whether a model would show an impact from drought when there actually is a drought, shift the mean of the normal distribution of slopes towards higher values so that the slope of the line is positive. Denote the difference between the shift and the zero slope as the *effect size*. For every effect size, take 1,000 realizations from the distribution and compare to the critical value. If the realization is

[2] Another set of hypotheses could be that the null hypothesis states that the slope of the line is greater than zero, and the alternative hypothesis is equal to zero. In this set of hypotheses, the burden of proof is on the analyst to show that there *is no* impact on water quality from drought. In the set of hypotheses used in the main body of the text, the burden of proof is to show that there *is* an impact due to drought. The following guidance documents are excellent sources of information providing cases when one set of hypotheses should be used versus the other set of hypotheses: the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) and the Multi-Agency Radiological Laboratory Analytic Protocols Manual (MARLAP).

above the number, then the correct decision was made (since there really is an impact). However, if the realization is below the number, then a Type II error was made (the model indicated no impact, when, in fact, there was an impact.) Run this simulation at least 1,000 times for at least each of the 8 effect sizes and count the number of Type II errors. That is the Type II error rate for that effect size. A plot of the Type I and Type II error rates should look like the plot below with the Type I error rate plotted at an effect size of zero.



Once all interested parties involved agree on the operating characteristic curve of the statistical model, using the model to make a decision to either list or delist a water body (based on whether the model shows an impact from drought) represents a powerful tool.

Comment received from the City of Austin 6/10/2013