Forecasting Water Quality (Newsletter)*

June 2, 2015

Predicting concentrations of chemicals in surface waters is a major component of permitting decisions, from NEPA impact assessments and NPDES point source discharge to mine closure and Superfund liability bond releases. Decision delays are costly for operators, and regulators are too often sued by those claiming that decisions were based on inadequate data.

Usual approaches to forecasting chemical concentrations are to build complex numeric ecosystem models or predict concentrations of single chemicals rather than the entire set of chemicals of interest. While the usual approaches can produce results, too often that answer is incorrect. The proper approach applies statistical time series models to the set of chemicals as a complete composition. This analytical approach is called compositional data analysis (CoDA) and is slowly gaining recognition for its usefulness in addressing complex environmental data analyses. CoDA has been extensively used for about 30 years to analyze social, economic, political, and geochemical data.

To understand CoDA, and why it provides valid answers to permitting decisions, picture a pie chart: a legislative district's proportions of voters registered as members of the Democratic, Green, Libertarian, and Republican parties. Unaffiliated voters, and those registered with other parties, are not included in the data set. This pie chart contains only the parties of interest. The sum of all party proportions (compositional components) is 100%. The proportions reflect the influence of all unmeasured factors affecting each component. We want to forecast how the proportions change with time. The same approach applies to forecasting market share, plant species on a reclaimed area, production costs, or water chemicals suitable for a designated beneficial use.

The information basis for decisions is contained in the ratios of the compositional components, not in the individual proportions or percentages. Using the components' individual proportions or percentages in statistical models changes their values so they no longer sum to 1.00 (or 100%) and they no longer are a closed composition. Transforming raw proportions to ratios and calculating natural logarithms retains the composition's structure, allowing the transformed values to be validly used in statistical models. Model output is transformed back to original units for ease of interpretation. These forecasts

are technically sound, legally defensible and support management and regulatory decisions.

Statistical models do not yield results with insufficient or incomplete data. For example, there must be more samples than components. The pie chart example above has four components (political party affiliations). Therefore, there must be more than four sets of data (polling results) to analyze changes and yield meaningful results. This requirement can be used to refute claims that insufficient data were used in the decision-making process. Descriptive summary statistics can be calculated even for small data sets, and those results can be used to support decisions until additional data are acquired.

When natural resource projects apply for environmental permits there can be tens of millions of dollars depending on the outcome. Water chemistry concerns are more difficult to properly address than are other environmental components such as flora and fauna because chemical concentrations usually are compared to maximum concentration limits (MCL) based on toxicity. Unfortunately, those toxicity levels are not necessarily determined locally and for local taxa. The assumption that these MCLs can be so used has not been empirically validated. Using statistical models on compositional data allows new results to be quickly generated each time results of new samples are added to the database. If the results are significantly different from those before, the knowledge is immediately available so appropriate determination of cause and effect can be made.

Twenty-first century environmental analysis tools should be used to address twenty-first century concerns. Applying regression and time series models to CoDA data is only one example of how to benefit all stakeholders with robust and objective support for decisions.

—

All newsletters, white papers, and other technical resources can be freely downloaded from `http://www.appl-ecosys.com/publications/`.