Abstracts

Thursday, May 1

Session K6: Statistical Analysis Tools

10:00 – 11:30 am | Room 232

Linked Micromaps: Statistical Summaries of Monitoring Data in a Spatial Context

Michael G. McManus[1], Quinn C. Payton[2], Marc H. Weber[3], Thomas M. Kincaid[3] and Anthony R. Olsen[3]
[1]*US Environmental Protection Agency, Cincinnati, Oh.,* [2]*Oregon State University, Corvallis, Oreg.,* [3]*US Environmental Protection Agency, Corvallis, Oreg.*

Abstract
Communicating summaries of spatial data to decision makers and the public are challenging. Linked micromap plots (LMplots) provide a way to simultaneously summarize and display both statistical and geographic distributions of data from surveys such as the National Aquatic Resource Surveys (NARS). The production of LMplots to display statistical summaries associated with areal units or polygons, such as watersheds or ecoregions, can be cumbersome. We identify four steps that streamline the process for making LMplots and incorporate those steps in the R package micromap. Those steps include: smoothing complex boundaries of polygons to just convey regional identity and neighbor relationships, linking the spatial and statistical data, making a draft LMplot, and refining that plot. As an example we make an LMplot summarizing pH data over nine regions from the National Lakes Assessment, with that display showing interquartile ranges and confidences limits. Being able to incorporate such measures of variation is an advantage LMplots have over pin or choropleth maps. For an example of exploratory spatial data analysis, we make an LMplot using stream conductivity and condition from a spatially balanced probabilistic survey of streams in West Virginia to identify local and global outliers. LMplots can also be applied to statistical summaries for administrative units, such as counties, states, or countries, and have been used with public health, demographics, and agricultural data. The ability for viewers to discern both statistical and spatial patterns is a strength of these plots, and we hope that visualizing georeferenced statistics in this manner will become more common. The views in this abstract are those of the authors and do not necessarily represent the views or policies of the U.S. EPA.

STORET Analytical Tools

Amy Wesley-Snider, Megan Tulloch, Danette Boezio, Peter Ilieve and Michele Eddy
RTI International, RTP, N.C.

Abstract
The U.S. Environmental Protection Agency's (EPA's) modernized STOrage and RETrieval (STORET) Data Warehouse provides a central location for the storage of observational data affecting water quality on a national scale. Existing web services can call on the STORET warehouse and provide data access but no analytical tools are currently available to users. RTI completed a pilot project that created an analytical tool to work with STORET Data. Under the STORET Trends Analysis Tool Project RTI obtained a static copy of the STORET database and for 22 characteristic types at all STORET stations executed, in R, the Seasonal Kendall Trend analysis for each characteristic type. Two approaches were used to develop these analytical tools: geospatial and tabular. The geospatial tools are targeted to more advanced users. The tabular solutions provide easier analysis for all levels of users. Both approaches would utilize open source "R" for statistical analysis. A large source of data in the STORET warehouse stems from the National Aquatic Resource Surveys (NARS). These probabilistic survey data from wadeable streams, rivers, lakes, and wetlands provide millions of monitoring results that span the nation. This tool is being expanded to include the capability to do national level screening, in particular against Clean Water Act Section 304(a) national criteria, and/or allow for the entry of state specific water quality criteria. Another tool under development could be used to estimate the extent of stressors and their impact in aquatic biological

conditions essential for determining the general and regional condition of aquatic resources. This additional analytical functionality could be targeted directly to the National Aquatic Resource Surveys (NARS) data. These STORET Analytical Tools will assist states with integrating STORET with other data systems such as ATTAINS. This integration will allow states to easily utilize STORET data in assessment decisions.

Disaggregation of Pollutant Sources through the Implementation of Continuous, Surrogate-Based Regression Equations

James Riddle and Ben Hammond
Woolpert Inc., Columbia, S.C.

Abstract
Watershed pollutant loadings are commonly assessed using computer models that include many assumptions (point discharge data), lumped parameters (*e.g.*, EMCs and CNs), and limited climate data that can be sources of error that lead to misrepresentation of pollutant loading dynamics inherent to nearly every watershed in the country. A more direct and practical approach, which is consistent with research conducted by the USGS office in Kansas, to watershed nutrient loading analysis is to define/establish site-specific relationships between the pollutant of concern (POC) and a continuous water quality parameter that would allow watershed stakeholders to more accurately measure pollutant loads. Woolpert has taken this approach with multiple communities facing major fiscal ramifications of implementing a TMDL that does not accurately disaggregate pollutant loadings. The result of misrepresenting loading rates to a watershed could be wasted resources and little or no improvement of water body conditions.

In 2008 Woolpert, established a continuous monitoring network within a 200 square mile impaired watershed. The six stations, installed along the river, divide the watershed into subwatersheds that have established baseline main stem conditions that allow for the identification of areas with elevated loadings. Woolpert has used site-specific regression equations to estimate these pollutant loads throughout the watershed since July 2010. These equations have been successfully validated each year with as many as 30 additional grab samples. In 2012, Woolpert installed three additional temporary monitoring stations for the purposes of characterizing areas with elevated loadings and specifically urban and rural land uses in smaller sub-watersheds. This allowed Woolpert to approximate specific contributions with a land use based approach to further identify primary pollutant contributors. POC loadings from the three stations will be applied to similar land uses throughout the watershed to approximate loadings for the entire watershed.

The results of Woolpert's continuous, regression-based approach to assessing watershed loads over the past three years indicate that the current regulatory model for total phosphorus is inaccurate. The consistency that Woolpert has observed in regression based loading trends estimated for varying rainfall years supports the use of this approach for approximating pollutant loads and potential sources.

Accounting for Confounders Leads to Clearer Effects- Thresholds for Some Stressors

Claire Buchanan[1], Jacqueline Johnson[2], Carlton Haywood[1] and Adam Griggs[1]
[1]*Interstate Commission on the Potomac River Basin, Rockville, Md.,* [2]*Independent, Crofton, Md.*

Abstract

Multiple stressors generally confound (mix up, conceal) relationships between individual stressors and aquatic communities. We have used a number of analysis techniques to untangle the confounding influences of several anthropogenic stressors in Chesapeake basin fresh and estuarine waters. The result has been clearer effects-thresholds for individual stressors, including nutrients, light attenuation, conductivity, flow alteration, and habitat degradation. The analyses strongly indicate the need to account for natural environmental differences, which are most evident in reference-quality habitats. Once these differences are accounted for, there is convergence in how aquatic communities respond to increasing stress regardless of the stressor. This supports the Biological Condition Gradient concept.