

The "Golden Database" and Use of Statistical and Ecological Methods to Determine Relative Causality of Impact from Emerging and Industrial Chemicals

Scott Dyer^{1*}, Katherine Kapo², Dick De Zwart³, Leslie Faggiano⁴, Christian Mulder³, Leo Posthuma³, Chris Holmes⁵ ¹The Procter & Gamble Company, Cincinnati, Ohio, USA; ²Montani Run LLC, Kenna WV, USA; ³RIVM, The Netherlands; ⁴Univ. Girona, Spain; ⁵Waterborne Environmental Inc., Leesburg, Virginia, USA.

Introduction

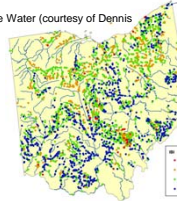
Prospective risk assessments of chemicals may include the use of bioavailability models, SSDs, and mixture toxicity modeling to estimate the proportion of species potentially affected by chemical mixtures. To test such approaches it is imperative that sufficient datasets be compiled of both prospective (modeled data) are matched to documented field-based biological impacts via appropriate geographic information systems (GIS). We compiled to what we believe is the largest database that includes monitored fish and invertebrate communities and compares these to measured and modeled chemical concentrations, risks, instream habitat, landuse, and ecoregions for more than 2000 sites throughout the state of Ohio, USA.

Data Sources

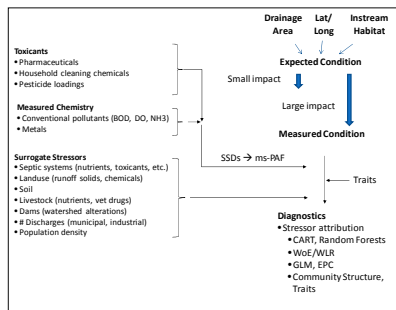
- United States Census of Population and Housing. 1990 & 2000. (<http://factfinder.census.gov/home/saff/main.html>; http://tp.census.gov/census_1990/) (<http://www.census.gov/main/www/cen2000.html>)
- Combined Animal Feeding Operation locations. (http://www.epa.state.oh.us/dsw/cafo/permit_lists.html)
- Crop Land Data Layer and Landcover. (http://www.nass.usda.gov/research/CropLand/metadata/metadata_oh07.htm), (<http://www.epa.gov/mrlc/nlcd-2001.html>)
- Nutrients and Pesticides. (<http://www.agcensus.usda.gov/Publications/2007/index.asp>), (http://water.usgs.gov/nawqa/prisp/usage/maps/compound_listing.php?year=02)
- Soil Characteristics. (<http://soils.usda.gov/survey/geography/ssurgo/>)
- Domestic Wastewater Discharge. (http://www.epa.state.oh.us/dsw/cafo/permit_lists.html), (www.epa.gov/cws/)
- Combined Sewer Overflow Locations. (<http://www.epa.state.oh.us/dsw/gis/cso/index.php>)
- Industrial Discharge Locations. (http://www.epa.state.oh.us/dsw/permits/npdes_info.html)
- Dam Locations. (<http://crunch.tec.army.mil/nidpublic/webpages/nid.cfm>), (<http://www.ohiodnr.com/water/fabid/3361/Default.aspx>)
- Water Chemistry Data (WaterQualityData_2000.mdb, courtesy of Bob Miltner & Dennis Mishne, Ohio EPA)
- River Network and Flows. (<http://horizon-systems.com/NHDPlus/index.php>). Mean and harmonic mean flows calculated by Waterborne Environmental Inc.
- Predicted Riverine Concentrations of Pharmaceuticals and Household Chemicals. Modification of GIS-ROUT model using data from NHDPlus and Domestic Wastewater Data layers D Plus
- Toxicity Extrapolations (Species Sensitivity Distributions). Pharmaceuticals (PhRMA data set and use of ICE (see: www.epa.gov/oeampub/ichain/webice/index.htm), Household chemicals, metals and ammonia (from De Zwart et al., 2006 and P&G), Pesticide SSDs from RIVM
- Monitored Aquatic Community and Habitat. Ohio EPA Division of Surface Water (courtesy of Dennis Mischné)

Biomonitoring

The basis for investigating many diagnostics methods is the robust biomonitoring data set collected by Ohio EPA. These data are used by OEPA to manage receiving water quality. The figure to the right illustrates the Index of Biotic Integrity scores for fish communities. Blue dots = meets OEPA reference quality.



Conceptual Framework



Reference Condition

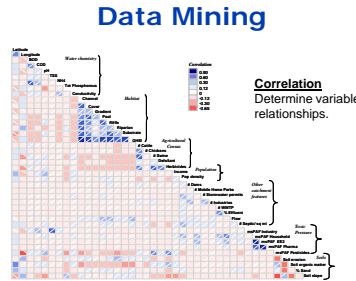
Reference Condition determines "impact" for test sites. We investigated 3 methods:

- Expert Judgment (based on OEPA designations)
- IBI scores = excellent ecological status (pragmatic)
- Expert judgment reference sites with excellent IBI scores

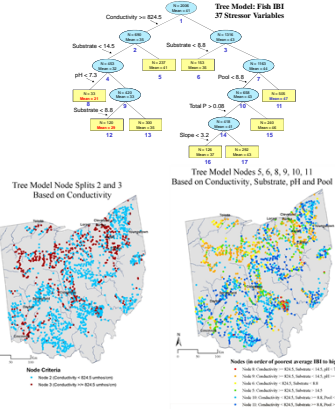
Results:

- Method 1 = due to various hydrological conditions (e.g., dry, high flows) IBI scores did not reflect reference conditions. Result: Great diagnostic model uncertainty.
- Method 2 = nearly half the state's sites met reference site quality. This reduced the ability to distinguish reference from test sites, thus leading to large model uncertainty.
- Method 3 = greatest differences between reference and test sites, thus greater diagnostic power, reduced uncertainty.

Data Mining

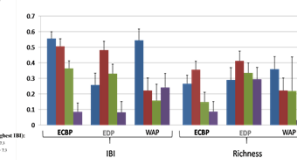


CART-Analysis & Maps
Provides an initial suites of variable relationships to biological dependent variables, such as the IBI.

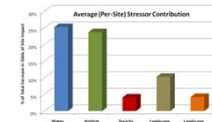
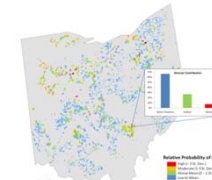


Traits in Diagnostics

Artificial Neural Networks (back-propagation algorithm): Mean and standard deviation of correlation coefficient between observed and predicted values (Test set) for each traits modalities in function of each Ecoregion.



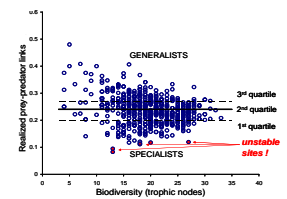
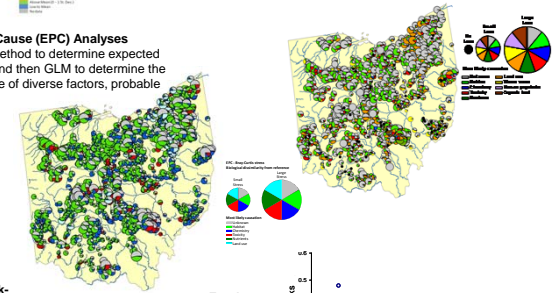
Causal Diagnostics



Weight of Evidence/ Weighted Logistic Regression
Based on locations of impacted sites statewide, the strongest spatial relationships were with water chemistry and instream habitat.

Effect Probable Cause (EPC) Analyses

Uses RIVPACS method to determine expected biological status and then GLM to determine the relative importance of diverse factors, probable cause, per site.



The number of trophic links is close to a quadratic function of fish diversity. As soon the river community becomes dominated by fish specialists (red circles), the food web becomes unstable and the biomass and energy flows across trophic levels collapse.