Development of multi-metric regression models for predicting macroinvertebrate response to changes in stream quality in the Blue River, Kansas and Missouri, USA

Barry C. Poulton
U.S. Geological Survey
Columbia Environmental Research Center
Columbia, Missouri

J.L. Graham, and T. J. Rasmussen
U.S. Geological Survey
Kansas Water Science Center
Lawrence, Kansas
Recent Publication

Effects of Wastewater Effluent Discharge and Treatment Facility Upgrades on Environmental and Biological Conditions of the Upper Blue River, Johnson County, Kansas and Jackson County, Missouri, January 2003 through March 2009

USGS Scientific Investigations Report 2010-5248

Jennifer L. Graham¹, Mandy L. Stone¹, Teresa J. Rasmussen¹, and Barry C. Poulton²

¹U.S. Geological Survey, Kansas Water Science Center, Lawrence, Kansas.


USGS science for a changing world
Overall project scope

- Assess effects of wastewater discharge, and treatment plant upgrades, on water quality, habitat, and aquatic communities (algae, macroinvertebrates) in a 10 km segment of the upper Blue River.

- In 2008, bi-annual community sampling of algae and macroinvertebrates, discrete and continuous water quality monitoring, habitat assessment, and comparisons with previous data collected since 2003.

Objectives of this study component

- Identify macroinvertebrate indicator metrics that best detect changes and discriminate site differences, for each season.

- Determine suite of metrics with lowest redundancy that could be included in models for predicting changes in biological quality for the upper Blue River system.
Sampling Site
WWTF Effluent

Upstream Site (Kenneth Road)
% Urban = 20.7
% Impervious = 5.8

Middle Site (151st Street)
% Urban = 24.9
% Impervious = 7.1

Downstream Site (Blue Ridge Boulevard)
% Urban = 25.8
% Impervious = 7.7
Upstream site

Middle site

Downstream site

- **KDHE Macroinvertebrate Protocol**
- *Triplicate samples twice per year*
Questions…

Can a multiple regression analysis identify meaningful combinations of metrics with low redundancy that can be used for predicting changes in biological quality of specific stream reaches or specific point-source disturbances?

Are there other previously un-examined metrics that also respond to stream changes associated with a discharge, and if so, are they the same metrics currently being used for evaluating biotic condition in all of Johnson County, KS?
Biological Quality – different ways to score sites

A) Use benchmark criteria. Score test sites in relation to metric expectations defined by reference site performance.

B) Use established multi-metric index. Score test sites relative to one another, based on performance of metrics that respond to range (gradient) in human disturbances.

C) Use percent of best available metric value. For each metric, score sites in relation to best attained value among all sites in the watershed (best value = score of 100), then integrate all metric scores to obtain overall site score.

This approach better for predicting stream quality when multi-year data are available and when metrics are co-correlated.
Core Aquatic Life Support Metrics for KS and / or MO

Part of established multi-metric biotic condition score
Mean Score (summed % of best attained metric values)

Sampling Locations

Site 1
Site 2
Site 3

WWTF Discharge

April 2008
August 2008
+ - 1 S.E.
Underlying Assumptions

- Not all indicators included within a multi-metric index will respond to every stressor, or in every season, or with the same magnitude (i.e. “not every community attribute acts the same”)

- Metrics belonging to different metric categories are less likely to be highly correlated than those within the same metric category

- Metric models may have more site discrimination power and/or wider applicability, if the collection of metrics (and resulting index) represents: 1) a greater number of metric categories, and 2) a greater number of taxonomic groups
**Approach** - use results from a specific system or stressor (Blue River with wastewater discharge) to test ability of metrics to discriminate differences in relative biological condition among sites

**Steps – “Metric Filters”**

- Calculate 34 metrics (from RBP, core metrics used by states, etc.)

  - **Metrics w/statistical significance among sites (non-parametric ANOVA)**
  - **STEPDISC procedure** (p ≤ 0.15 entrance criteria) to identify best metric combo
  - **MULTIPLE REGRESSION** with Site Score as dependent variable (R² selection method)

Choose models to predict changes in Score; selection based on:

1) R² and 2) metric redundancy / co-correlation

3 - 5
Upstream values > than one or both downstream sites

E_p
EP_p
★ EPT_p
EPTC_{ratio}
EPT_{rich}
IntKBI_{tol}
P_p
P_{rich}
SC_p
★ SCFC_{ratio}

Upstream values < than one or both downstream sites

Corb_p
DT1_p
DT2_p
★ HydT_p
ODNI_p
★ KBI_{tol}
MBI_{tol}

Middle site > or < than other sites

Filt_p
Olig_p

ANOVA results
15 metrics significant in spring, 12 in late summer, and 5 in both seasons

★ Highly significant (p> 0.001) in one or both seasons
<table>
<thead>
<tr>
<th>Model</th>
<th>w / % Metrics</th>
<th>wo / % Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Metric</td>
<td>(KBI_{tol})</td>
<td>(R^2 = 0.98)</td>
</tr>
<tr>
<td></td>
<td>(P_{rich})</td>
<td>(p &lt; 0.0001)</td>
</tr>
<tr>
<td></td>
<td>(DT2_p)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(KBI_{tol})</td>
<td>(R^2 = 0.83)</td>
</tr>
<tr>
<td></td>
<td>(SCFC_{ratio})</td>
<td>(p = 0.022)</td>
</tr>
<tr>
<td></td>
<td>(Cling_{rich})</td>
<td></td>
</tr>
<tr>
<td>4-Metric</td>
<td>(KBI_{tol})</td>
<td>(R^2 = 0.99)</td>
</tr>
<tr>
<td></td>
<td>(Olig_p)</td>
<td>(p &lt; 0.0001)</td>
</tr>
<tr>
<td></td>
<td>(P_{rich})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(DT1_p)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(KBI_{tol})</td>
<td>(R^2 = 0.92)</td>
</tr>
<tr>
<td></td>
<td>(EPT_{rich})</td>
<td>(p = 0.018)</td>
</tr>
<tr>
<td></td>
<td>(SDI_{div})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SCFC_{ratio})</td>
<td></td>
</tr>
<tr>
<td>5-Metric</td>
<td>(Olig_p)</td>
<td>(R^2 = 0.99)</td>
</tr>
<tr>
<td></td>
<td>(SCFC_{ratio})</td>
<td>(p &lt; 0.0001)</td>
</tr>
<tr>
<td></td>
<td>(EPTC_{ratio})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(P_{rich})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(DT2_p)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(KBI_{tol})</td>
<td>(R^2 = 0.96)</td>
</tr>
<tr>
<td></td>
<td>(EPT_{rich})</td>
<td>(p = 0.024)</td>
</tr>
<tr>
<td></td>
<td>(IntKBI_{tol})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SDI_{div})</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(SCFC_{ratio})</td>
<td></td>
</tr>
</tbody>
</table>

**Bold** = partial sum of squares is also significant \((p > 0.05)\)
<table>
<thead>
<tr>
<th>Model</th>
<th>w / % Metrics</th>
<th>wo / % Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-Metric</td>
<td><strong>SCFC\textsubscript{ratio}</strong> $T_{\text{rich}}$ <strong>ODNI}_p</strong></td>
<td><strong>KBI\textsubscript{tol}</strong> <strong>SCFC\textsubscript{ratio}</strong> $T_{\text{rich}}$</td>
</tr>
<tr>
<td></td>
<td>$R^2 = 0.93$</td>
<td>$R^2 = 0.85$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.002$</td>
<td>$p = 0.016$</td>
</tr>
<tr>
<td>4-Metric</td>
<td><strong>SCFC\textsubscript{ratio}</strong> $T_{\text{rich}}$ <strong>DT5}_p</strong> <strong>ODNI}_p</strong></td>
<td><strong>KBI\textsubscript{tol}</strong> <strong>SCFC\textsubscript{ratio}</strong> $T_{\text{rich}}$ <strong>C}_rich</strong></td>
</tr>
<tr>
<td></td>
<td>$R^2 = 0.98$</td>
<td>$R^2 = 0.93$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.0005$</td>
<td>$p = 0.011$</td>
</tr>
<tr>
<td>5-Metric</td>
<td><strong>IntKBI\textsubscript{tol}</strong> <strong>SCFC\textsubscript{ratio}</strong> $T_{\text{rich}}$ <strong>Cling}_rich</strong> <strong>ODNI}_p</strong></td>
<td><strong>KBI\textsubscript{tol}</strong> <strong>SCFC\textsubscript{ratio}</strong> $EPTC\textsubscript{ratio}$ $T_{\text{rich}}$ <strong>Cling}_rich</strong></td>
</tr>
<tr>
<td></td>
<td>$R^2 = 0.99$</td>
<td>$R^2 = 0.97$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.0001$</td>
<td>$p = 0.011$</td>
</tr>
</tbody>
</table>

**Bold** = partial sum of squares is also significant ($p > 0.05$)
## Metric examples - Johnson County, Kansas

<table>
<thead>
<tr>
<th>Metric</th>
<th>ANOVA Significance</th>
<th>STEPDISC</th>
<th>Best 3-5 M Regressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SC_p$</td>
<td>Y (1)</td>
<td>Y (2)</td>
<td>N</td>
</tr>
<tr>
<td>Tany$_p$</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$MBI_{tol}$</td>
<td>Y (1)</td>
<td>Y (1)</td>
<td>N</td>
</tr>
<tr>
<td>$EP_p$</td>
<td>Y (1)</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>$EPT_p$</td>
<td>Y (2)</td>
<td>Y (1)</td>
<td>N</td>
</tr>
<tr>
<td>$Total_{rich}$</td>
<td>N</td>
<td>Y (1)</td>
<td>N</td>
</tr>
<tr>
<td>$SDI_{div}$</td>
<td>N</td>
<td>N</td>
<td>Y (1)</td>
</tr>
</tbody>
</table>

*KDHE Aquatic Life Support, but not included in 10-metric score*
Other potential macroinvertebrate metrics

$\text{SCFC}_{\text{ratio}}$ $P_{p}$ $T_{\text{rich}}$

High frequency in models, significance in at least one season, chosen by STEPDISC in at least one season

$\text{Corb}_{p}$ $\text{IntKBI}_{\text{tol}}$

Absent or infrequently present in models, but significance in both seasons, and not chosen by STEPDISC
Conclusions - metrics

Models highly significant, valid for Score range of 610-830, etc.

Significant low-redundancy models $R^2$ range of 0.83 to 0.99

Some more important when # of metrics in model greater

$P_p$  $EPT_{rich}$

Some more important when % metrics not included in models

$KBI_{tol}$  $MBI_{tol}$

Some important for ↑ model prediction but not for site discrim.

$ODNI_p$  $Cling_{rich}$  $DT5_p$

Some w/ good site discrim. redundant with currently used ones

$EPT_p$  $P_p$  $ScFc_{ratio}$  $T_{rich}$

* Approach has some potential for E. KS or W. MO streams, and for other stresses