Aquatic Sciences

CrossMark

RESEARCH ARTICLE

Effort-based predictors of headwater stream conditions: comparing the proximity of land use pressures and instream stressors on macroinvertebrate assemblages

Gregory J. Pond[1] · Kelly J. G. Krock[1] · Jonathan V. Cruz[2] · Leah F. Ettema[1]

Abstract Environmental agencies are often faced with resource and time constraints in assessing waterbody health. We compared the strengths of varying levels of effort (field measures, laboratory chemistry, land use, and multiple combinations of these) to explain macroinvertebrate assemblage response along a gradient of urban land use intensity among 30 headwater streams in northern West Virginia. Because the spatial arrangement of human disturbance can govern biotic response, land use effects were analyzed at five spatial scales (whole catchment, and 100 m buffer zone at three fixed upstream distances and total stream network upstream of site); instream ecological measures included physical habitat, algal concentrations and water chemistry. Of the five spatial scales, we predicted that riparian land use nearest the site would explain the most variation but that instream measures would be the overall driver of the macroinvertebrate assemblages. Regression analysis evaluated the strength of single and multiple variables in explaining macroinvertebrate multimetric index (MMI) and ordination patterns, and revealed that assemblages were highly responsive to numerous stressors. In contrast to predictions, total upstream network riparian forest cover explained the most variation overall (83%) while specific conductance was the single best instream measure (64%). Stepwise regression models using combinations of field, laboratory, and land use variables all performed reasonably well but we found that a 3-variable model [% forest (catchment), road density, and specific conductance] that minimized colinearity and cost/effort explained 90% of the variation in the MMI. Validation and spatial autocorrelation results suggest that this model could potentially be used to forecast stream condition for prioritizing conservation and remediation efforts in headwaters within the ecoregion, and our general approach would be broadly applicable in other settings.

Keywords Bioassessment · Predictive models · Urbanization · Headwater · Land use alteration · Buffer zone

Electronic supplementary material The online version of this article (doi:10.1007/s00027-017-0534-3) contains supplementary material, which is available to authorized users.

✉ Gregory J. Pond
  Pond.Greg@epa.gov

  Kelly J. G. Krock
  Krock.Kelly@epa.gov

  Jonathan V. Cruz
  Cruz.Jonathan@epa.gov

  Leah F. Ettema
  Ettema.Leah@epa.gov

[1] Office of Monitoring and Assessment, U.S. Environmental Protection Agency, Region III, 1060 Chapline St., Wheeling, WV 26003, USA

[2] Penbay Media Contractor, U.S. Environmental Protection Agency, Region III, 1650 Arch Street, Philadelphia, PA 19103, USA

Introduction

An effective monitoring program demands considerable resources in order to make sound assessments of waterbody condition. Under the US Clean Water Act (CWA), states must evaluate the condition of all of their streams (often thousands of stream segments covering thousands of kilometers). States typically assess streams on a catchment-by-catchment basis defined by hydrologic unit codes (HUCs), rotating sampling efforts among HUCs and years.

Environmental agencies charged with protecting aquatic resources commonly rely on macroinvertebrate assemblages to assess and designate waterbody attainment under various governmental regulations (e.g., aquatic life use of the CWA). Because of the sheer number of streams and associated monitoring costs, spatially-balanced probabilistic approaches are often applied. There are some indications that land use quantification can adequately predict biological condition in streams and rivers at regional scales (e.g., Gergel et al. 2002; Van Sickle et al. 2004; Carlisle and Hawkins 2008; Brown et al. 2012) providing managers with incentive for applying geographic information systems (GIS) tools to assess, explain, and ultimately target waterbodies in need of preservation, restoration or pollution control.

Catchment land alteration (i.e., deviation from natural) drives multiple changes in terrestrial and aquatic ecosystem structure that ultimately shape ecosystem functions (Allan 2004) but these changes are often dependent on both magnitude and proximity to the alteration (Tran et al. 2010). For example, urbanization can affect multiple ecological processes leading to stream degradation. In fact, the 'urban stream syndrome' is a term that captures the consistency by which major hydrological, chemical, and biological changes are observed (Walsh et al. 2005). While stressor pathways are complex, monitoring and assessment strategies often aim to narrow down this complexity using surrogate measures. Land use quantification is commonly used to explain patterns in stream biota and can serve as an effective measure for predicting waterbody health. The spatial proximity of land use pressures (e.g., whole catchment versus riparian corridor) can often dictate the severity of observed aquatic impacts (Roth et al. 1996; Sliva and Williams 2001; Sponseller et al. 2001; King et al. 2005; McBride and Booth 2005; Rios and Bailey 2006; Roy et al. 2007; Tran et al. 2010); however, whole catchment land use has been considered an overly simplistic representation of environmental disturbance (King et al. 2005). Given contradictory findings by some of these investigators on whether catchment or riparian impacts are most important to biological communities (or are concordant), this research area deserves further exploration. Although other analytical methods exist to address proximity effects such as inverse distance-weighting (King et al. 2005; Van Sickle and Johnson 2008), flow-weighting via spatial stream networks (Ver Hoef et al. 2014) and connected impervious surfaces (Roy and Shuster 2009), we chose to emulate other previous studies that examined whole catchment versus riparian land use pressures on stream condition within a local watershed-context.

Land use as a predictor of water quality or biological integrity (e.g., fish or macroinvertebrate assemblages) is enticing because with appropriate models, local or regional evaluations and forecasting can be done using existing GIS data layers (Kristensen et al. 2012; Villeneuve et al. 2015). However, given the complex interplay of land use and its direct and indirect influences on habitat and water quality, localized land use models often suffer from colinearity among the land use classes themselves and with instream (habitat and chemistry) measures (King et al. 2005). This complexity makes it difficult to select appropriate variables without risk of overfitting models that ultimately cannot reliably predict responses in novel locations. Given budgetary constraints in monitoring resources, GIS-based models can provide managers with useful decision tools for assessing and interpreting local or regional waterbody health, but are probably deficient without some level of accompanying instream monitoring data (Gergel et al. 2002).

With an overarching goal to predict regional headwater stream conditions, we aimed to first explain changes in local headwater macroinvertebrate assemblages and component bioassessment endpoints along a gradient of proximate (riparian) and catchment-scale urban land use, and associated instream physical and chemical measures. We focused on headwater streams (1st–2nd order) because of their importance to local and regional biodiversity and their contribution to the overall health of larger rivers and streams (Meyer et al. 2007). Specifically, our objectives were fourfold: (1) we aimed to test which spatial scale of land use pressure could best explain assemblage composition and overall biological condition. We predicted that riparian buffer zone land use affects assemblages more than whole catchment land use. (2) With this acquired knowledge, we compared the best spatial arrangement of land use to instream measures (habitat and chemistry) to test which can better explain biological condition. Here, we predicted that macroinvertebrates respond more consistently to local instream factors compared to the best land use indicator. (3) Furthermore, we sought to develop and compare multivariable composite models to predict future biological condition in these or other unassessed headwater streams within our study ecoregion. Models were based on varying levels of information (land use, field measurements, water chemistry, or combinations of all) that represented differing levels of cost, time and resources needed by environmental agencies to help assess stream health. (4) Based on biological condition endpoints, we inferred stressor thresholds of select land use and instream variables to protect aquatic life in the region.

Methods and materials

Study area

We selected 30 moderate- to high-gradient perennial headwater tributaries (1st–2nd order) of lower Wheeling Creek

(Fig. 1) in the vicinity of Wheeling, West Virginia (WV) that varied in land use pressures (e.g., forested, residential/commercial development, agricultural). Wheeling Creek is a 5th order tributary (~777 km^2) of the Ohio River and drains portions of southwestern Pennsylvania and the northern panhandle area of WV and is within the Upper Ohio South HUC8 watershed. The study area lies within the Permian Hills subecoregion (70a; after Woods et al. 1996) of the Western Allegheny Plateau, a hilly region characterized by Appalachian oak forest in the uplands and mixed mesophytic forest in the valleys, and underlain by sedimentary siltstone, sandstone, shale, limestone, and coal. Surrounding the city of Wheeling in Ohio County, human disturbance intensity is generally dictated by topography, road access, and proximity of parks and basic infrastructure. Human population has varied over the last century in the county, peaking in the mid-1900s (~73,000) versus current day (~44,000). A network of roads spans the entire Wheeling Creek catchment where roads often run parallel to, and cross stream channels in nearly all major tributaries of Wheeling Creek. Our streams ranged from having low to high disturbance intensities (e.g., roadless forested catchments, expansive golf courses, and low- to high-density commercial or residential development). We sampled all sites (50 m reaches) for physical,

chemical, and biological properties in April 2015. Reaches were first selected by stream size (upstream catchment areas ranging from ~0.5 to 1.5 km^2), and secondly to represent positions along streams so as to capture a gradient of catchment and riparian land use pressures. We avoided resource extraction (coal mining and natural gas) land use in this study to focus on primary land uses (commercial, residential, and some agriculture) in the immediate Wheeling vicinity. None of the individual sites were longitudinally connected by flow within the same sub-catchment (non-nested) which allowed for statistical independence. A few developed sites had sections of their upstream or downstream areas piped underground and stream banks of urbanized streams often had extensive non-native vegetation consisting of Japanese Knotweed (*Fallopia japonica*), Multiflora Rose (*Rosa multiflora*), Privet (*Ligustrum vulgare*), Russian Olive (*Elaeagnus umbellata*), and Honeysuckle (*Lonicera* spp.). In order to reduce confounding by some common factors, selection criteria involved choosing sites with full streambed shading (canopy cover, median = 93%; range = 83–98%) and similar stream catchment area (median = 0.87 km^2; range = 0.5–1.3 km^2); this latter criterion typically led to site comparability in mean wetted width (median = 2.0 m; range = 1.3–2.9 m), elevation (median = 290 masl; range = 250–330 masl), general
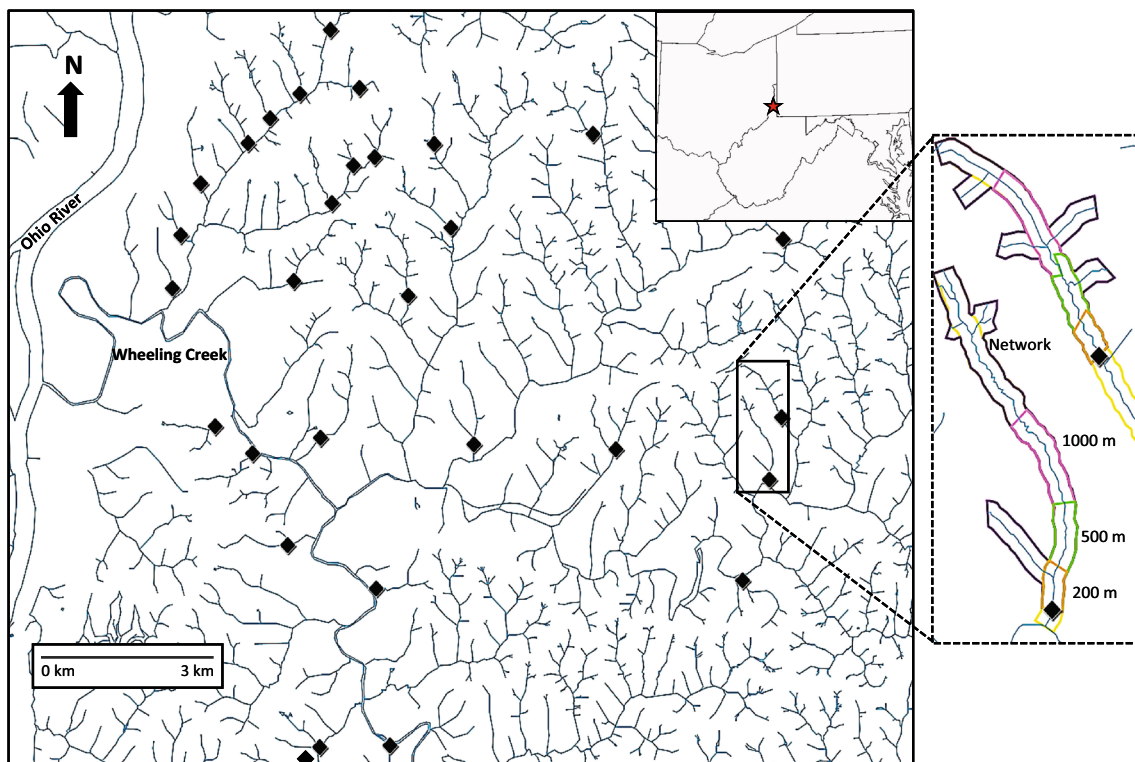

Fig. 1 Map of sample sites along with general study location (*upper right insert*) and magnified view of example buffer zone polygons (*far right insert*)

lithology (Dunkard Group of Permian age), dominant soil type (alfisols; Westmoreland silt-loam), and stream channel morphology (i.e., riffle-pool sequences dominated by cobble riffles). Example site photographs are shown in Online Resource 1.

Macroinvertebrate sampling

We sampled macroinvertebrates in April 2015 using West Virginia Department of Environmental Protection's (WVDEP) riffle kicknet method (WVDEP 2014). In brief, a 0.5-m wide rectangular frame kick-net (595 μm mesh) was used to collect 4–0.25 m^2 samples; the contents were composited to yield a 1 m^2 sample. Samples were fixed with 95% ethanol and returned to the laboratory for processing to a 200 ($\pm$20%)-fixed count random subsample from a gridded pan, and subsequent identification and enumeration. All aquatic insects (including Diptera), crayfish, snails and clams were identified to the genus-level. Following WVDEP protocol, higher classifications were assigned to some taxa: Oligochaeta (family), Nematoda (phylum), Turbellaria (class), and Trombidiformes (order).

Instream measures

Rapid bioassessment protocol (RBP) (Barbour et al. 1999) visual habitat assessments and other site measurements were performed at the time of benthic sampling. The RBP habitat assessment method consisted of rating ten habitat metrics on a scale of 0–20 (total possible score = 200). At each site, we estimated channel slope (hand-held clinometer); mean canopy cover (spherical densitometer) from lower, middle and upper reach of each transect, and mean wetted stream width. Substrate composition was estimated using pebble counts with a gravelometer where approximately 100+ random substrate particles were measured along the reach in a wandering upstream direction. For analytical purposes, pebble count data were reduced to three variables: % fines <1 and <4 mm, and median particle diameter (D_{50}). We used an in situ fluorometer (Bentho-Torch, bbe Moldaenke, Germany) to estimate the dominant type (green, diatoms, blue-green) and mean concentration of benthic algae (benthic Chl a, μg/cm^2) on the surfaces of six stones. Stones (i.e., medium cobble) were chosen in riffle areas of the reach, primarily in the immediate vicinity of benthic macroinvertebrate collections. Two replicate measurements were taken on each stone and averaged across all measurements; diatoms made up >90% of these measurements and we chose to combine all algal types into total benthic Chl a for analytical purposes. A multi-probe sonde (YSI, Yellow Springs, OH, USA) was used in situ to measure specific conductance (μS/cm), dissolved oxygen (both mg/L and % saturation), pH and temperature (°C) at the time of benthic sampling. We also collected water chemistry grab samples (ions, nutrients, metals, TSS, alkalinity) once at each site (top of reach); samples were analyzed at EPA Region III's Environmental Science Center at Fort Meade, MD. In addition, fecal coliform bacteria samples were collected concurrently and analyzed as colony forming units/100 mL (cfu/100 mL) at a local laboratory within 6 h of sampling.

Land use quantification

We used ESRI ArcGIS software and tools (ArcMap 10.3.1, Arc Hydro v2.0) to delineate catchments (3 m digital elevation model) upstream of the 30 sample points and to quantify land use at five spatial scales: (1) whole catchment, and a 100 m wide riparian buffer zone (50 m per side) polygon delineated at upstream (US) segment distances from the lowest reach end of (2) 200 m, (3) 500 m, (4) 1000 m, and (5) for the entire stream network US of site (to the map-derived stream sources). We used the West Virginia Statewide Addressing and Mapping Board (SAMB) for GIS layers of streams, roads, and building points (structures). For the whole catchment, land use classes were derived from the 2011 National Land Cover Database (NLCD; see Homer et al. 2012) at a spatial resolution of 30 m. ArcMap's buffer tool was used in conjunction with the NLCD layer to determine % forest land cover within the 100 m buffer zone (50 m per side) polygons at 200, 500, 1000 m, and entire network US (cumulatively). We combined deciduous-evergreen-mixed forest into one forest classification. Likewise, in the whole catchment, varying intensities of urban development (low, medium, high) and agriculture (hay/pasture, crops) were aggregated. Several catchments had large golf courses which were classified as open-space development (similar to lawns) and we analyzed open-space and urban development both separately, and combined as a total development class. Ultimately, we compared five land use classes: % forest (sub-classes: catchment, buffer 200 m US segment, 500 m US, 1000 m US, and entire network US), % developed urban, % developed open-space, % developed total, and % agriculture. In addition, catchment-scale road density (km/km^2) and structure density (structures/km^2) were calculated using the SAMB data. We also evaluated percent impervious cover (% IC) as an indicator to compare with other studies found in the literature using the 2011 NLCD derived impervious layer (30 m) calculated for each catchment polygon (see Online Resource 2 for site examples).

Statistical analyses

All of the data used in this study were collected in a one-time sample event; thus, our analyses followed a

space-for-time substitution approach where trends or changes in macroinvertebrates due to altered land use or instream measures is inferred from sampling a range of characteristics across space.

Data reduction

With the macroinvertebrate data, we calculated a WV-specific multimetric index (MMI; 100-point scale) called the GLIMPSS (CF) (genus-level index of most probable stream status, with Chironomidae identified to family-level; Pond et al. 2012) specifically calibrated for this region and season (plateau spring, or PL Sp). We used this MMI as our primary response variable that relies on the following component metrics: intolerant (<4 tolerance value) richness, Ephemeroptera richness, Plecoptera richness, clinger richness, modified Hilsenhoff biotic index (mHBI), % Ephemeroptera + Plecoptera + Trichoptera relative abundance (minus *Cheumatopsyche*), and % Chironomidae + Annelida. Taxon tolerance value designations followed WVDEP's assignments. This version of the GLIMPSS is used by WVDEP for monitoring purposes in the study region (WVDEP 2015); we also applied WVDEP's stream condition ratings of "very good" (25th %ile of state reference), "good" (5th %ile of reference), "degraded" (upper bisection of reference 5th %ile), and "severely degraded" (zero to lower bisection of reference 5th %ile) as grouping variables (see "Statistical analyses"). Although not part of the MMI, we documented the range in density (organisms/m^2) (a simple extrapolation estimated from the proportion of grids required to meet the target subsample size from the 1 m^2 kicknet sample; Ligeiro et al. 2013), and ranges of total and EPT richness.

In addition to the MMI, we synthesized overall assemblage structure using multivariate ordination with non-metric multidimensional scaling (NMDS; PC-ORD v. 6, Gleneden Beach, OR, USA). The NMDS was based on Bray–Curtis distances and $\log_{10} (x + 1)$ transformed abundances, omitting rare taxa (observed in <5% of samples) (McCune and Grace 2002). The software rotates (varimax) the final ordination axes so that they are orthogonal (where axis 1 accounts for the most variance); projected axes are thus parametric and site scores (and taxa) can be used in correlation-type analysis as a form of indirect gradient analysis. Thus, we used NMDS 1 as a separate response variable to compare with MMI and to represent fine-scale changes in the assemblage composition (see below). For visualization within the *n*-dimensional ordination plot, we grouped sites according to MMI stream condition (very good, good, degraded, and severely degraded after Pond et al. 2012). This also provided a relative check on how taxonomic composition varied with MMI-based condition assessments.

Instream measures were screened for colinearity and strongest macroinvertebrate response. First, variables were inspected for normality (histogram plots and Shapiro-Wilkes test, Systat v. 13) and transformed accordingly (log, $\sqrt{}$, arc sin $\sqrt{(x)}$) to ultimately improve linearity among predictors and response variables. Several of the chemical variables (e.g., trace metals) were omitted when below detection limits at >90% of sites. While some variables were evaluated individually, we used Spearman correlation to initially inspect for multicolinearity of all variables; in this case, one variable of a correlated pair ($r > 0.75$) was chosen first if it was deemed to show a stronger signal with biotic metrics, and secondarily (in the case of ties) if it had a wider range.

Catchment land use classes were also evaluated for colinearity (i.e., an increase in developed land corresponded to a concomitant reduction in forest). Since natural land cover in this ecoregion is forest (deciduous, conifer, mixed), we assume that non-forest is a proxy for land use alteration. However, to avoid statistical dependency of mutually inclusive land use variables (King et al. 2005), we used a series of Pearson product-moment partial correlations to examine correlation independence of the MMI with % forest catchment, % developed open, % developed urban, % developed total, % agriculture, road density, and structure density. Here, correlations between the MMI and each land use variable were calculated after other land use variables were held constant. From this analysis, we focused on a reduced set of land use indicators for multiple linear regression modeling (see "Constructing and validating composite explanatory/predictive models"), but we compared all land use classes as individual explanatory variables (see below). Although correlated with other land use classes, % IC was evaluated separately with macroinvertebrate indicators.

Testing individual explanatory variables

We used ordinary least-square regression analysis to evaluate magnitude, direction and significance of the relationships between land use and instream measures on macroinvertebrate response indicators (MMI and NMDS 1 scores). For all regressions, model residuals were plotted and inspected for lack of independence (random distributional patterns); we found that transforming the independent variables resulted in normal ± scatter of the residuals along the zero line. For prediction 1 (riparian versus catchment scale), we compared the strength of fit using coefficient of determination (r^2) and Akaike Information Criteria (AIC$_C$; corrected for small sample size) on the relationships between MMI and NMDS 1 scores versus % forest land cover (our assumed proxy) at different proximities (i.e., whole catchment, and 100 m wide riparian buffer within reaches of 200, 500, 1000 m, and total network US). We

also compared MMI and NMDS 1 with catchment-scale road density, structure density, % developed (open-space, urban, and total), % agriculture, and % IC. Prediction 2 was then investigated by comparing the strength and significance of the relationship between macroinvertebrate response (r^2 and AICc) with the best land use variable versus the best instream explanatory factor.

Constructing and validating composite explanatory/ predictive models

For objective 3, we compared the predictive power of combined variables (field measures, water chemistry data, land use, and combinations of these) collected with varying levels of effort (i.e., time and cost) on the MMI (our primary response variable), by running a series of stepwise (backwards selection) multiple linear regressions on the reduced set of variables for each level of effort. Model performance was assessed with adjusted R^2 and AIC$_C$. To avoid multicolinearity in the multiple regressions, we set tolerance $(1-R^2)$ at 0.5, and then inspected variance inflation factors [VIF; $(1/1-R^2)$] for each variable selected using a VIF <4 as an acceptable threshold indicating negligible multicolinearity (Kutner et al. 2004). We compared models based on weight of evidence using ΔAICc and AICc weights (w_i) [where Δ_i is the paired difference between the minimum AICc and model i; w_i is computed as $\exp(-0.5\Delta i)/\sum \exp(-0.5\Delta i)$]. We also considered a relative cost gradient based on typical work hours and expenses for field, laboratory, and GIS measures (see "Results"). The preferred model was selected based on a combination of the best weight of evidence with a lesser resource intensive level of effort. For this preferred model, we evaluated spatial autocorrelation of the predictor variables with Moran's I scatterplots. Prior, we assigned spatial weights using Euclidean distance and a furthest neighbor distance threshold (measured at ~3.1 km based on actual spatial distances). For each predictor variable and MMI (regression residuals) we report Moran's I and pseudo p values based on 999 random permutations.

For comparative purposes, we further evaluated multivariate relationships between macroinvertebrate assemblages, and the candidate composite models while controlling for potential spatial autocorrelation. Here, we used partial Mantel tests (pMantel, PC-ORD v. 6; McCune and Grace 2002) that compared each pair of Bray–Curtis (macroinvertebrate) vs. Euclidean dissimilarity matrices (based on multiple regression parameters from field data, chemical data, land use data, and combined models) while partialing out a spatial proxy (Euclidean matrix based on y-latitude and x-longitude coordinates). The pMantel procedure tests the assumption that the distance among objects (e.g., sites) in one dissimilarity matrix (e.g., macroinvertebrate taxa) is

linearly independent of the distances from the same objects in another matrix (e.g., water chemistry, habitat data). Mantel tests produce a coefficient akin (but lower in magnitude) to correlation analysis (r) with larger values indicating stronger associations; accompanying significance tests were executed with 999 randomized permutations where $p < 0.05$ would reject the hypothesis of no association.

Preferred model validation

To validate the preferred multiple regression model's performance, we compared it to random exclusion cross-validation regressions and an independent dataset. The random exclusion cross-validation technique (Systat v. 13) randomly removed one-fifth of the 30 sites (6 sites), re-ran regressions (999 repetitions), and calculated an average model fit (predicted R^2). The independent validation dataset comprised 13 Western Allegheny Plateau sites sampled by the WVDEP between 2006 and 2009; these sites were screened for seasonality, catchment size and lack of acid coal mine drainage to be comparable to our original data set (these sites all trained toward the Ohio River but belonged to three adjacent HUC8 watersheds; see Online Resource 4). Several of the validation sites were slightly larger than our original sites (~2–4 km^2) but catchment area was not found to be correlated with the MMI from regional reference sites in previous analyses (Pond et al. 2011). Our validation sites provided a gradient of environmental quality (e.g., 22–82% forest; specific conductance = 189–1442 µS/cm) and biological condition (MMI = 23.6–88.5). We also evaluated spatial autocorrelation of the predictor variables and MMI regression residuals with Moran's I scatterplots at the validation sites. Similarly, we assigned spatial weights using Euclidean distance and a furthest neighbor distance threshold that measured ~12 km. For each predictor variable (and MMI residuals), we calculated Moran's I and pseudo p values based on 999 random permutations.

Threshold detection

Existing statistical methods for threshold determination are varied (Dodds et al. 2010). Rather than finding breakpoints where macroinvertebrates significantly change along a stressor gradient, our interest for objective 4 was to simply determine stressor values corresponding with MMI impairment. We used the linear regression fit of the MMI scores with select land use and instream stressor variables to identify central tendency-based thresholds corresponding to a MMI score of 57 (Pond et al. 2012) which demarcates a line between a passing and failing stream condition assessment. Identified threshold values of transformed variables were then back-transformed to report the actual values.

Results

Macroinvertebrates

The 30 benthic riffle samples yielded 105 distinct taxa with total richness ranging from 15 to 32 taxa per site; EPT richness ranged from 4 to 16 genera per site. Estimated macroinvertebrate densities ranged from 420 to 3376 organisms/m^2 (mean = 1530; all samples met the 200-count subsample target for MMI calculations). Among these headwater streams, richness was dominated by aquatic insect genera (e.g., 10 Ephemeroptera, 11 Plecoptera, 10 Trichoptera, and 46 Diptera); however, at a few sites the crustacean amphipod *Gammarus* numerically dominated (~50–75% abundance). MMI scores ranged from 12 (severely degraded) to 89 (very good) across all sites. Based on the GLIMPSS (CF) threshold (score of 57) designated in Pond et al. (2012), 14 sites were non-degraded and 16 sites were degraded.

The NMDS ordination (Fig. 2) plotted sites primarily along axis 1 (77% variance explained after 62 iterations; $p < 0.001$ after 999 permutations); axis 2 explained only 12% variance ($p > 0.05$). Overall stress (i.e., goodness of fit) for the two-dimensional solution was 0.134. Ordination patterns showed that sites were approximately grouped by their corresponding MMI-based condition categories, indicating that MMI represented multivariate taxonomic composition relatively well. The top 10 important taxa [i.e., genera with the strongest correlations (5- and 5+ responders)] with NMDS 1 included: pollution sensitive *Epeorus* ($r = -0.83$), *Amphinemura* ($r = -0.82$), *Ephemerella* ($r = -0.71$), *Diphetor* ($r = -0.65$) and *Leuctra* ($r = -0.64$), and inversely, the pollution tolerant *Cricotopus* ($r = 0.79$), *Cheumatopsyche* ($r = 0.77$), *Diamesa* ($r = 0.74$), *Thienemannimyia* ($r = 0.65$), and *Hydropsyche* ($r = 0.57$). Because NMDS 1 can represent finer fine-scale changes to assemblage composition compared to typical MMIs, it was also used as a response variable in regression analysis (see below).

Environmental data

Preliminary correlation analysis uncovered several instream abiotic variables that were strongly co-related (e.g., specific conductance/chloride (Cl)/sodium (Na)/hardness (total); embeddedness/sediment deposition scores; riparian zone/
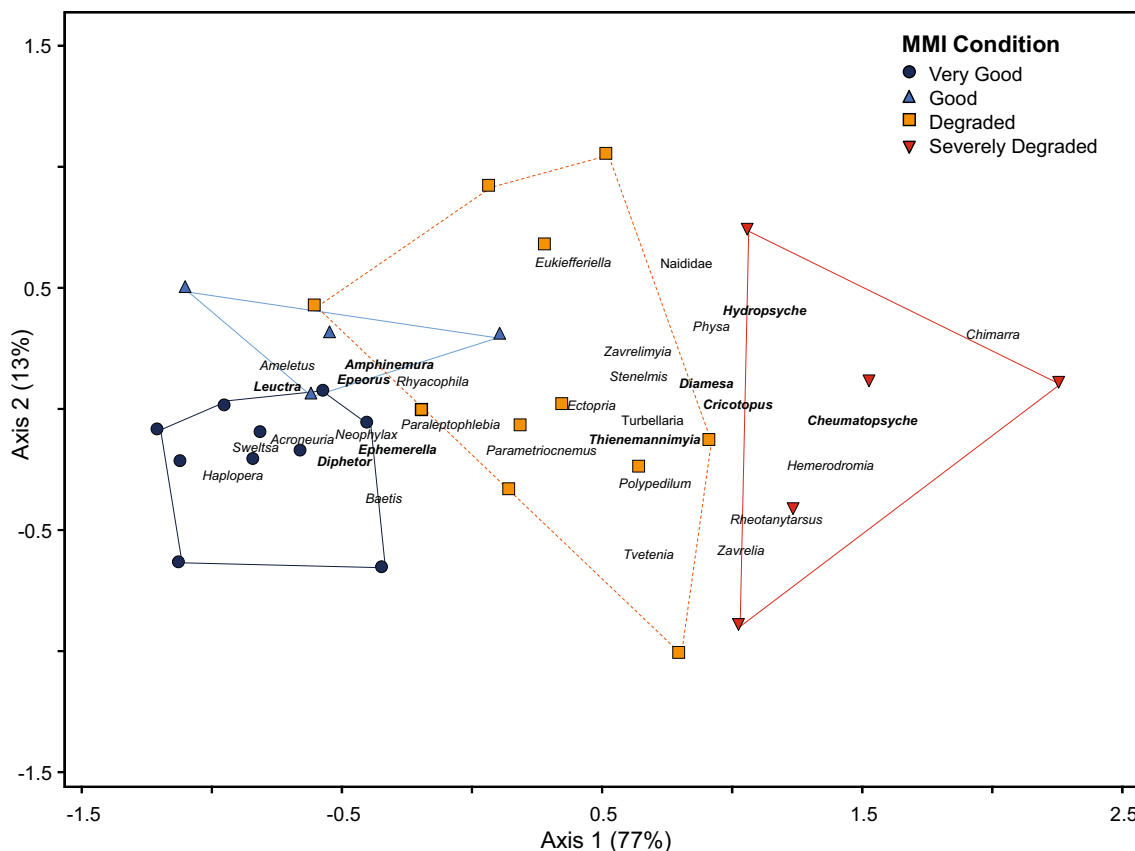

Fig. 2 NMDS ordination with sites grouped by MMI condition category. Only those taxa correlated with dimension 1 or 2 (|r| > 0.4) are plotted for clarity. Genera with the strongest correlations (− and +) with NMDS 1 shown in *bold type*

channel alteration scores; and pairings of substrate particle size fractions). These results reduced the set of instream variables (Table 1) used in regression modeling. Online Resource 3 provides site-specific environmental measures (e.g., catchment area, mean wetted width, channel slope, canopy cover, and stressor variables).

Despite the highly colinear nature of land use classes, partial correlation analysis identified suitable variables for use in the multiple regressions [however, all land use variables were compared with biological measures in the single regression analysis (see below)]. As anticipated, % forest catchment and % developed total were strongly related ($r = -0.93$). Partial correlation analysis of MMI versus % forest catchment (while holding other land use variables constant) often revealed decreased correlations while some maintained strong relationships. For instance, when controlling for road density, the MMI-% forest catchment relationship remained high (partial $r = -0.78$); however, the correlation between MMI and % forest catchment declined in magnitude from $r = -0.89$ to partial $r = -0.36$ when % developed total was controlled. While statistically independent of % forest catchment, % agriculture and structure density were weakly correlated with MMI. The 100 m buffer zone % forest classes (at 200, 500, 1000 m, and entire network US) were also highly colinear. Here, we compared each class with the MMI and found that buffer zone % forest along the entire network US showed the strongest signal with MMI.

Comparison of single explanatory variables

The responses of MMI and NMDS 1 with individual abiotic variables (Table 3) were relatively similar (model r^2) with some difference noted. First, catchment area (a common covariate) was not significantly correlated with any of our biotic or abiotic variables (all Spearman's $|r| < 0.26$, $p > 0.05$, not shown). In contrast to our prediction for objective 1, we found that while several land use classes were significant predictors of biological condition (MMI), our nearest proximity variable (% forest buffer 200 m US) performed worse ($r^2 = 0.47$, AICc $= 262$) than % forest catchment ($r^2 = 0.79$, AICc $= 234$) in explaining MMI or NMDS 1 variation. However, % forest within the buffer zone at longer reach scales (both 1000 m and entire network US) were much better predictors than % forest catchment. Riparian forest along the entire network (% forest network) was the best single predictor (Table 3). Within the buffer zone, an increasing correlation with macroinvertebrate indicators was observed with increasing distance upstream (e.g., 200, 500, 1000 m, r^2 values increased from 0.47 to 0.58 to 0.83, respectively for MMI). Catchment-level % IC ranged from 0 to 44% and was strongly correlated with MMI and NMDS 1 ($r^2 = 0.70$). Structure density and % agriculture

were non-significant predictors of MMI ($p > 0.05$), but MMI actually tended to increase with % agriculture in our data set and it was important in one of the composite models (see below).

Our prediction for objective 2 that the best instream variables would explain macroinvertebrates better than the best land use variables was not supported. Several instream measures (chemistry and habitat) were significantly related with MMI and NMDS 1 (Table 2) but not as strongly as with key land use variables. With MMI, the best predictor was specific conductance ($r^2 = 0.63$, $p < 0.001$) while the weakest was % fines < 1 mm ($r^2 = 0.19$, $p = 0.02$); water hardness (total) and Cl were key single chemical variables ($r^2 = 0.58$ and $r^2 = 0.48$, respectively). With NMDS 1, benthic Chl a ($r^2 = 0.48$) and total habitat score ($r^2 = 0.46$) were the next best explanatory variables after specific conductance ($r^2 = 0.56$). Bank stability score was the best single-habitat metric with MMI ($r^2 = 0.37$), but embeddedness score was more strongly related to NMDS 1. While nitrite + nitrate was not a significant single predictor, it was important in multiple regression modeling with MMI (see below). Although we purposely avoided coal mining land use, elevated sulfate levels were observed at nearly all sites (Table 1), perhaps due to seepage from exposed coal outcrops or old underground mines, including house coal extraction by residents more than a century ago. However, sulfate was not significantly correlated with MMI or NMDS1 scores (not shown). Linear regression plots of the best single-variable predictors (land use and instream measures) of macroinvertebrate assemblages are shown in Fig. 3.

Composite regression models based on level of effort

Our 3rd objective was to develop and compare competing multiple regression models to predict the MMI based on the level of information collected among sites. Despite common intercorrelations between land use classes and the reduced set of instream variables (see Online Resource 4), we were able to minimize multicolinearity in stepwise regressions (all selected variables had VIFs ≤ 2.2). We developed several strong and significant candidate composite models (Table 3) explaining variation in the MMI. The best model combined a mix of field-collected variables, laboratory chemical variables and GIS-derived land use (adj. $R^2 = 0.92$, AIC$_C = 210$): % forest buffer 1000 m US, hardness (total), nitrite + nitrate, and bank stability score. Comparatively, models developed from field-only data (benthic Chl a, specific conductance, and bank stability score; adj. $R^2 = 0.76$, AIC$_C = 241$, $\Delta_i = 30.7$) and lab-only water data (fecal coliform, hardness (total), Cl, and nitrite + nitrate; adj. $R^2 = 0.77$, AIC$_C = 241$, $\Delta_i = 30.7$) performed the worst. But when combined, field + lab

Table 1 Minimum, mean, and maximum values for select environmental variables

Land use	Min	Mean	Max
Catchment area (km^2)	0.48	0.88	1.33
Road density (km/km^2)[a]	0	5.3	12.9
Structure density (no./km^2)[a]	1	71.5	227.3
% Forest catchment[a]	7	54.6	98
% Developed urban	0	10.9	71
% Developed open-space	0	23.7	86
% Developed total	0	35	93
% Agriculture[a]	0	7.5	34
% Forest buffer 200 m US	0	69.4	100
% Forest buffer 500 m US	6	71.2	100
% Forest buffer 1000 m US[a]	2	69.2	100
% Forest buffer network US[a]	1	68.2	100
% Impervious cover	0	6.4	44

Field data	Min	Mean	Max
Stream width (m)	1.3	2.0	2.9
Channel slope (%)	1.1	3.1	7
Benthic Chl a (µg/cm^2)[a]	1.0	5.2	9.3
% Fines <1 (mm)[a]	0	5.6	13.4
% Fines <4 (mm)	1.9	10.2	19.2
Substrate diameter (D_{50}) (mm)	16	34.2	64.0
Canopy cover (%)	82.7	92.4	98.5
Epifaunal substrate score	11	15.0	18
Embeddedness score	12	15.2	19
Channel alteration score	10	15.6	20
Sediment deposition score	9	12.0	15
Bank stability score[a]	10	15.4	19
Bank vegetation scores	8	12.7	16
Riparian zone score	2	14.6	20
Total RBP habitat score	117	143.0	166
pH	7.46	7.92	8.46
Specific conductance (µS/cm)[a]	299	657	1658
Dissolved oxygen (mg/l)	9.5	11.3	13.1
Temperature (°C)	7.6	10.9	15.4

Laboratory data	Min	Mean	Max
Fecal coliform (col/100 ml)[a]	8	109.7	510
Bicarbonate alkalinity (mg/l)	97.2	184.7	268
Hardness (mg/l)[a]	143	271.7	386
Calcium (mg/l)	46	84.9	123
Magnesium (mg/l)	7	14.6	29
Sulfate (mg/l)	40.2	86.5	227
Chloride (mg/l)[a]	2.74	86.7	583
Sodium (mg/l)	4.58	32.8	189
Nitrite+Nitrate (mg/l)[a]	0.05	0.50	1.9
Phosphorus, total (mg/l)	0.01	0.03	0.09
Aluminum, total (mg/l)	0.04	0.25	0.73
Iron, total (mg/l)	0.1	0.391	1.14
Manganese, total (mg/l)	0.003	0.028	0.076

Those marked with an superscript ([a]) were selected in stepwise multiple regression models. Individual habitat assessment metric scores range from 0 to 20, total RBP habitat score ranges from 0 to 200. *US* "upstream"

Table 2 Single variable, ordinary least-squares linear regression results for land use and a subset of instream measures comparing MMI and NMDS 1 as response variables

	MMI					NMDS 1				
	β	r^2	p	Δ_i	w_i	β	r^2	p	Δ_i	w_i
Land use										
% Forest buffer network US	58.9	0.83	<0.001	0	0.58	−2.6	0.86	<0.001	0	0.82
% Forest buffer 1000 m US	49.8	0.83	<0.001	1	0.35	−1.9	0.85	<0.001	3	0.18
% Developed total	−53.4	0.80	<0.001	5	0.05	1.9	0.69	<0.001	24	<0.01
% Forest catchment	66.5	0.80	<0.001	6	0.04	−2.5	0.77	<0.001	16	<0.01
% Impervious cover	−123.9	0.70	<0.001	19	<0.01	4.7	0.70	<0.001	23	<0.01
% Developed urban	−80.3	0.62	<0.001	25	<0.01	2.9	0.59	<0.001	32	<0.01
% Developed open-space	−60.3	0.62	<0.001	25	<0.01	1.9	0.46	<0.001	41	<0.01
Road density	−5.7	0.60	<0.001	27	<0.01	0.2	0.55	<0.001	35	<0.01
% Forest buffer 500 m US	41.6	0.59	<0.001	27	<0.01	−1.6	0.63	<0.001	29	<0.01
% Forest buffer 200 m US	30.1	0.47	<0.001	35	<0.01	−1.2	0.55	<0.001	35	<0.01
Structure density	−0.1	0.09	0.06	50	<0.01	0.03	0.06	0.159	57	<0.01
% Agriculture	37.7	0.09	0.102	51	<0.01	−0.8	0.01	0.379	59	<0.01
Instream measures										
Specific conductance	−111.8	0.63	<0.001	0	0.87	3.9	0.56	<0.001	0	0.78
Hardness, total	−158.6	0.58	<0.001	4	0.12	4.9	0.39	<0.001	9	0.01
Chloride	−12.4	0.48	<0.001	10	0.01	0.5	0.44	<0.001	6	0.04
Total RBP habitat score	1.1	0.46	<0.001	11	<0.01	−0.1	0.46	<0.001	5	0.06
Bank stability score	6.1	0.37	<0.001	17	<0.01	−0.2	0.25	0.005	16	<0.01
Benthic Chl a	−19.9	0.34	0.001	18	<0.01	0.9	0.48	<0.001	4	0.11
Embeddedness score	7.3	0.33	0.001	18	<0.01	−0.3	0.36	<0.001	10	0.01
Fecal coliform	−10.3	0.24	0.006	22	<0.01	0.4	0.29	0.001	13	<0.01
% Fines (<1 mm)	−2.6	0.19	0.018	24	<0.01	0.09	0.16	<0.001	19	<0.01
Nitrite + nitrate	−5.3	0.05	0.242	29	<0.01	0.1	0.02	0.545	24	<0.01

Variables are sorted by decreasing order of MMI model support based on differences in Akaike Information Criterion from minimum (Δ_i), and Akaike weights (w_i). Total RBP habitat score is provided for comparison

efforts yielded a slightly better performing 4-variable model (adj. $R^2 = 0.82$, $AIC_C = 233.6$, $\Delta_i = 23.4$) using benthic Chl a, specific conductance, bank stability score, and nitrite + nitrate. Focusing only on catchment land use, an improved 2-variable model was constructed (adj. $R^2 = 0.84$, $AIC_C = 225.9$, $\Delta_i = 15.7$) using % forest catchment and road density; however, when combined with field data, a better 3-variable model (adj. $R^2 = 0.90$, $AIC_C = 217.6$, $\Delta_i = 7.4$) was constructed that included specific conductance, road density, and % forest catchment.

Considering predictive power with the perceived amount of time and costs (see Online Resource 5), we selected the combined land use–field effort prototype: MMI = 144 − (42.3 × log$_{10}$ specific conductance) − (1.5 × road density) + (42 × arc sin ($\sqrt{}$ forest catchment)) as our preferred predictive model (Fig. 4). An analysis of spatial autocorrelation on this model yielded low univariate Moran's I values indicating random dispersion of these three predictors in geographic space: specific conductance (Moran's $I = −0.06$, $p = 0.39$), % forest catchment (Moran's $I = −0.015$, $p = 0.41$), and road density

(Moran's $I = −0.07$, $p = 0.42$). However, a test on the MMI regression residuals showed some slight spatial autocorrelation (Moran's $I = 0.29$, $p = 0.02$). Here, 4 of the 30 sites having higher or lower lagged residuals were typically located near outer boundaries of the study area with fewer neighbors.

Multivariate assemblage response to stressors approximated results found with the MMI response. As seen in the NMDS ordination (Fig. 2), sites exhibited patterns that corresponded with the biological condition classes as rated by the MMI and taxa with known sensitivity or tolerance to pollution plotted within these classes. Multivariate comparisons of the effort-based regression models using partial Mantel tests (pMantel r and significance) between Bray–Curtis distance (macroinvertebrates) and Euclidean distance (abiotic factors) are shown in Table 4. First, we found no spatial autocorrelation in the biological data matrix (macroinvertebrates versus geographic coordinates; Mantel $r = 0.01$, $p = 0.41$). Despite this result, we conservatively controlled for any underlying geographic influence (e.g., spatial clumping of sites as visualized in

Fig. 3 Linear regression plots of best single variable predictors (% forest network and specific conductance) of MMI and NMDS axis 1. *Dashed lines* demarcate 95% confidence intervals of the fitted regression line
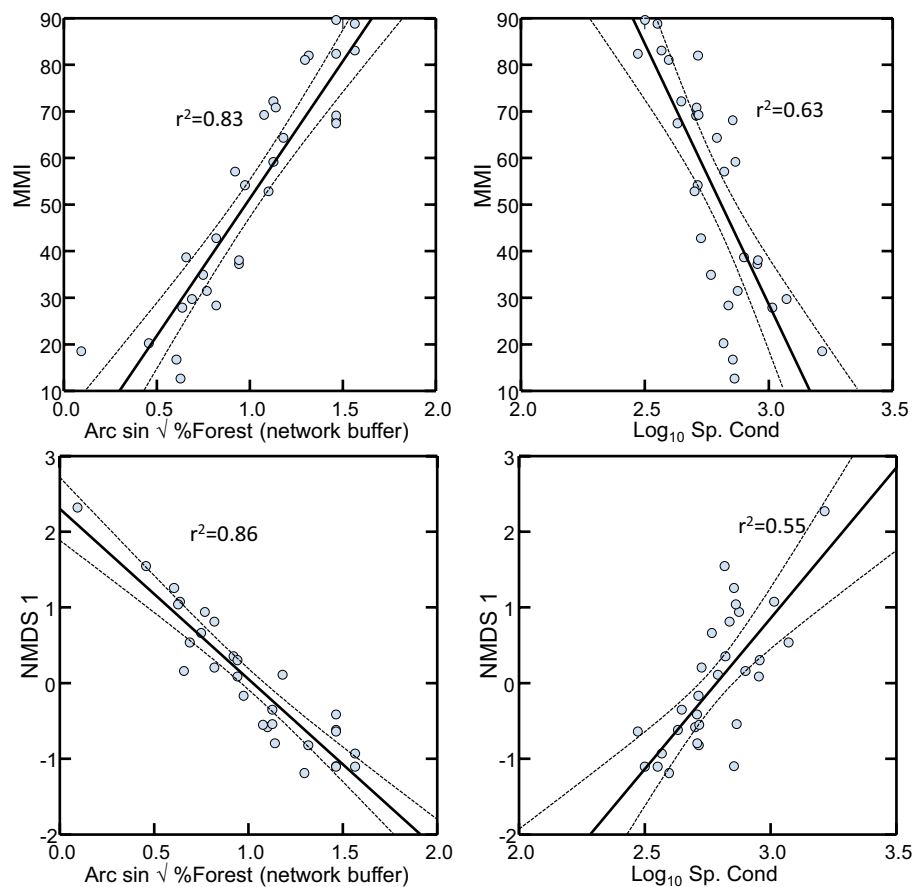

Fig. 1) with pMantel tests. As with MMI-based multiple regression models, all pMantels were significant but we observed better agreement between macroinvertebrates and land use matrices than with field or chemical matrices. However, Euclidean matrices based on combined variables (i.e., select field, chemical, and land use as in Table 3) did not necessarily improve the associations seen in the MMI multiple regressions (i.e., comparing Tables 3, 4); in fact, results of independent field and % forest catchment tests (pMantel $r = 0.42$ and 0.45, respectively) were roughly equivalent to a combined matrix (pMantel $r = 0.44$), while a univariate matrix (% forest buffer network US) had the highest association overall (pMantel $r = 0.70$).

Preferred model validation

Statistical cross-validation of the preferred 3-variable regression model (% forest catchment, road density, and specific conductance) with the MMI revealed an adequate fit (predicted $R^2 = 0.98$). Although this model performed reasonably well in predicting biological condition of the 13 independent sites, it tended to over-predict the more disturbed sites. Here, predicted and observed MMI scores were approximately similar (Fig. 4) but predicted scores

were typically biased high and more variable ($R^2 = 0.86$; $R^2 = 0.83$ with intercept set at 0). These validation sites showed no evidence of spatial autocorrelation (Moran's $I = -0.03$, $p = 0.37$) in the residuals of MMI scores, or the individual predictor variables (see Online Resource 6). Further, we found that 12 of the 13 sites (92%) were correctly classified as impaired or unimpaired using the MMI threshold score of 57 (Fig. 4; Online Resource 6).

Threshold detection

Using an MMI impairment cutoff of <57 (Pond et al. 2012) for our objective 4, we estimated (via regression fitting) that impairment occurred when % forest catchment was <60% (and <70% forest riparian buffer for entire network US), % total developed >30%, % developed urban >15%, and road density >4.5 km/km². However, steep regression slopes (Table 2) suggested that considerable biological changes were occurring at even lower rates of land alteration. For single instream variables, we detected MMI-based impairment when specific conductance was >540 μS/cm. The MMI impairment threshold for Cl corresponded to a mean of 30 mg/l, but this response pattern was more

Table 3 Stepwise multiple regression results for MMI, categorized by differing levels of effort (field measures, laboratory chemical analysis, land use, or combinations)

	Adj R^2	p	VIF	AIC_C	Δ_i	w_i	Variables selected
Field-only measures	0.76	0.005	1.2	241.0	30.7	<0.001	− Benthic Chl a
		<0.001	1.4				− Specific conductance
		0.012	1.3				+ Bank stability score
Laboratory-only chemical analysis	0.77	0.045	1.2	241.0	30.7	<0.001	− Fecal coliform
		0.047	2.2				− Hardness, total
		<0.001	2.2				− Chloride
		0.004	1.6				− Nitrite + nitrate
Combined field + laboratory	0.82	<0.001	1.2	233.6	23.4	<0.001	− Benthic Chl a
		0.042	1.4				+ Bank stability score
		0.003	1.1				− Nitrite + nitrate
		<0.001	1.4				− Specific conductance
Land use (catchment)	0.84	0.002	1.8	225.9	15.9	0.003	− Road density
		<0.001	1.8				+ % Forest catchment
Land use (catchment + buffer)	0.85	0.056	1.1	225.1	15.9	0.003	+ % Agriculture
		<0.001	1.1				− % Forest network US
Combined land use (catchment) + field + laboratory	0.90	0.03	1.9	217.6	7.4	0.03	− Road density
		0.002	2.1				− Specific conductance
		<0.001	2.0				+ % Forest catchment
Combined land use (buffer) + field + laboratory	0.92	<0.001	2.0	210.3	0.0	0.97	+ % Forest buffer 1000 m US
		0.051	1.9				+ Bank stability score
		0.004	1.2				− Hardness, total
		0.014	1.4				− Nitrite + nitrate

Adjusted R^2, p values (for individual variables), Akaike's information criterion (AICc), differences in AICc from minimum (Δ_i), and akaike weights (w_i) of each model are shown. *VIF* variable inflation factors. Slope direction of independent variables used in the models are indicated as positive (+) or negative (−)

variable ($r^2 = 0.48$) and the maximum Cl at an unimpaired site was 72 mg/l.

Discussion

Local watershed or regionally-based assessment tools that use GIS or instream measurements could be cost effective by helping to focus monitoring efforts and prioritizing potential areas for restoration or conservation opportunities. Moreover, an ability to forecast future water quality conditions with appropriate models would benefit local or regional landscape design and management plans (Van Sickle et al. 2004; Merriam et al. 2013). Although the interplay of modified land use and physical–chemical impacts on biotic response is complex, we produced robust empirical models explaining macroinvertebrate response to land use alteration with a relatively small data set (n = 30). We found that GIS-based tools to derive land uses for response variables can be predictive, but recognize that simple bivariate relationships likely miss additive or interactive effects as noted by Bryant and Carlisle (2012).

Regardless of scale or proximity, loss of natural vegetation (particularly in urbanizing areas), can lead to drastic changes in the physical, chemical and biological characteristics of streams (Paul and Meyer 2001; Moore and Palmer 2005; Roy et al. 2003; Cuffney et al. 2010). Although choice of land use indicators is varied in the literature, we chose % forest as the primary indicator for land alteration since it represents the natural state in our study area and had high initial correlation with the MMI. However, % developed total (sum of all open-space, low, moderate, and high urban development) was the near-equivalent inverse of % forest in our study (Spearman $r = -0.94$), and one could infer that either classification would indicate a similar degree of macroinvertebrate response. Recent studies (Wickham et al. 2016) have depicted % IC as a reliable measure and it was a strong predictor in our study, but less effective than % forest. Land use quantification alone might miss some important impacts as urbanized headwater streams (including some of ours) are often buried and piped for considerable lengths (Roy et al. 2009), potentially altering hydrology, chemistry, and invertebrate dispersal. Chemically, increases in pollutants via dissolved salts (e.g., chloride) or other forms of urban runoff can alter
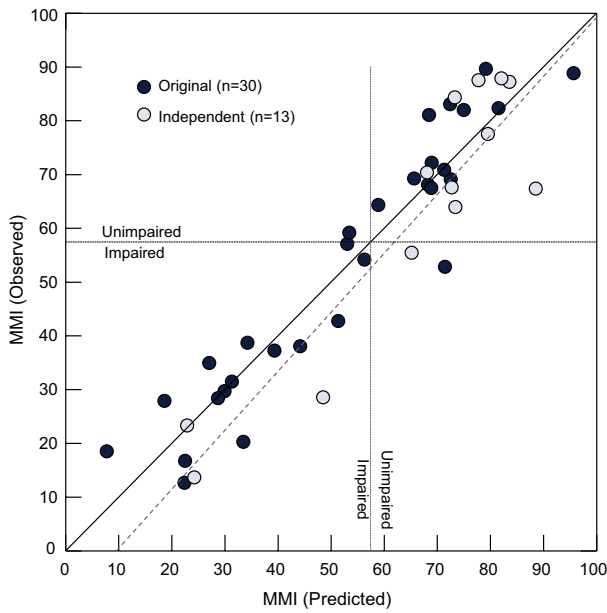
Fig. 4 Observed versus predicted multimetric index (MMI) scores for original and independent validation sites fitted using the recommended multiple regression equation (based on performance and cost/efforts). The MMI was predicted from specific conductance, road density and % forest catchment using the equation: MMI (predicted) $= 144 - (42.3 \times \text{Log}_{10}$ specific conductance) $- (1.5 \times$ road density) $+ (42 \times$ arc sin ($\sqrt{}$ % forest catchment)). *Solid line* is the original fitted regression line ($R^2 = 0.90$); *dashed line* represents regression of independent sites ($R^2 = 0.86$). *Lines* representing MMI impairment thresholds also shown

invertebrate assemblages through direct toxicity (Findlay and Kelly 2011) or alteration of nutrient and microbial processes (Swan and DePalma 2012).

Comparing land use spatial scales

We expected near-field riparian buffer (at 200 m upstream) land use pressure to shape macroinvertebrate assemblages

most, but it was apparent that distant and cumulative disturbances from upstream pressures within the buffer zone were more critical. Riparian buffer zones are frequently restored as a means to improve local stream conditions, but research has confirmed success is unlikely if upland disturbance is a driving factor (Walsh et al. 2007, Wahls et al. 2013). Although the 200 m polygon scale did correlate better with reach factors such as riparian vegetation score, channel alteration score, total RBP habitat score, and fecal coliforms (see Online Resource 4), MMI scores were more variable at this scale, compared with their response to increasing forest cover in further upstream segments, and for the whole catchment. This finding is in contrast to Sponseller et al. (2001) who showed greater forest cover–macroinvertebrate relationships at the local 200 m riparian segment scale. Despite this particular contradiction, we confirmed our expectation that buffer zone disturbance (specifically, remaining % forest cover along the entire stream network) explained MMI variation more than whole catchment forest. The outcome was further supported by the results of the pMantel tests that indicated a stronger relationship between the macroinvertebrate assemblage and % forest buffer network US than % forest catchment. Therefore, from a spatial perspective, our findings complimented Roy et al. (2003) and Tran et al. (2010) but contradicted others (Roth et al. 1996; Sliva and Williams 2001). The differences among these results might be attributed to a host of factors such as study design and stream size, regional setting, scale, temporality, assemblage metrics or water quality indicators. Moreover, we believe shorter (and narrower) headwater stream corridors like ours might be particularly vulnerable to buffer zone alterations due to an apparent link between the low volume of aquatic habitat and the extent of riparian land use alteration. Our finding that land use alteration within the entire upstream riparian corridor drives macroinvertebrates more than near-site pressures has implications for stream restoration needs,

Table 4 Partial Mantel tests (except Space-only based on full Mantel test) between macroinvertebrate assemblages (Bray–Curtis distance) and environmental data (Euclidean distance) based on level of information

Level of information	Mantel r	p	Variables used in distance matrix
Space	0.01	0.411	Geographic coordinates (*x*-longitude, *y*-latitude)
Field|space	0.42	0.002	Benthic Chl *a*, Sp. conductance, Bank stability
Laboratory|space	0.35	0.001	Fecal coliform, hardness, chloride, nitrite + nitrate
Field, laboratory|space	0.34	0.001	Benthic Chl *a*, Sp. conductance, nitrite + nitrate, Bank stability
Land use (catchment)|space	0.45	<0.0001	Road density, % forest catchment
Land use (buffer)|space	0.70	<0.0001	% Forest buffer network US
Combined model Land use (catchment)|space	0.44	0.0002	% Forest catchment, road density, Sp. conductance
Combined model Land use (buffer)|space	0.25	0.008	% Forest buffer 1000 m US, hardness, nitrite + nitrate, bank stability

Significance based on 10,000 permutations. Variables used correspond to regression variable selections in Table 3. *US* upstream

where biological response to instream channel alignment, habitat enhancements, and riparian tree planting is likely scale-dependent (i.e., requiring longer lengths of restored upstream habitat in our case) and with success contingent on the absence of chemical stressors (Bond and Lake 2003; Walsh et al. 2007).

Comparing instream measures versus land use

We found that many of the abiotic variables could significantly explain macroinvertebrate indicators, but with varying power. Here, our results contradicted some previous studies (Roy et al. 2003; Wang et al. 2003; Hale et al. 2016) in that we found land use explained more variation in macroinvertebrates than any of our instream factors. Two of the upstream buffer zone forest classes (1000 m and entire network US) were far stronger single predictors than any single instream chemical or habitat variable. In larger Michigan streams, buffer zone land use and instream variables explained macroinvertebrate indices more precisely than did catchment land use (Lammert and Allen 1999). The influence of instream versus catchment land use factors might be dependent on the degree of catchment disturbance. For example, Midwestern fish assemblages were more responsive to instream variables in unaltered catchments, but as the level of disturbance increased, land use became more important (Wang et al. 2006). We did not explore this angle but believe it is plausible, or simply a statistical artifact that in the absence of a pronounced land use effect, instream variation would be more detectable.

In our headwater reaches, specific conductance was the best single field or laboratory variable describing changes in MMI and NMDS 1. In other studies, specific conductance was a good predictor of macroinvertebrate assemblages in urban (Roy et al. 2003; Morgan et al. 2007; Utz et al. 2016) and coal mining affected catchments (Pond et al. 2008; Griffith et al. 2012; Cormier et al. 2013a, b). In our data set, Na, Cl, and hardness (total) were the strongest correlates ($r > 0.90$) with specific conductance. Although our sampling occurred in April (when road de-icing had ceased), we believed there was considerable road salt deposits (surface residues and groundwater) affecting our streams. We observed the highest Na and Cl concentrations (25–100× above background) at sites draining interstate and county highway corridors and Cl was most correlated with road density (see Online Resource 4). Out of all land use classes, we found that % developed urban had the highest correlation magnitude with specific conductance ($r = 0.86$); consequently, it was necessary to control for this colinearity in multiple regression model selections. In contrast, instream habitat factors (e.g., total RBP habitat score, bank stability score, benthic Chl a, % fines) played significant but lesser roles in determining biological condition

indicating the stronger influence of chemical factors (influenced by land use) on macroinvertebrates.

Combining land use and instream abiotic variables into predictive models

A model that uses various scales of land use information in addition to instream measures might be the most comprehensive (Lammert and Allen 1999; Bailey et al. 2007; Macedo et al. 2014; Villeneuva et al. 2015) and we found this true in our headwater data set. However, one might argue that the addition of instream measures only increased our predictive power slightly and it would be necessary to consider the costs of collecting instream data as opposed to GIS-only methods. This finding was observed by Kristensen et al. (2012) who found instream variables did not appreciably add to the power of predicting fish assemblages in Denmark. They argued that GIS offered cost-effective tools for targeting catchments or reaches for conservation planning or rehabilitation but warned against monitoring streams purely from a desktop. In Brazilian streams, combinations of variables from geophysical, site, and land use data explained the most variation in fish and benthic richness compared to models based on individual measures alone (Macedo et al. 2014). Given that land cover data (such as NLCD or even more refined state and county data sets) are readily available and free to the public, the potential cost-savings can be substantial compared to full water chemistry and habitat monitoring schemes, especially for watershed screening and planning.

Multiple regressions applied to sets of variables (field, laboratory, GIS) based on varying levels of perceived effort (cost and time), were able to develop several strong candidate models. We propose that our reduced cost, 3-variable model that included % forest catchment, road density, and specific conductance could be used by independent investigators to help infer biological condition of small streams in our study region either in targeted reaches or through a spatially balanced probabilistic survey design (e.g., McManus et al. 2016). Although we found no evidence of spatial autocorrelation with these three predictor variables, we noted minor spatial autocorrelation with MMI residuals; however, validation results indicate that this model is capable of detecting degraded or exceptional quality streams in a broader area of the region.

Inferring thresholds of disturbance

Resource agencies would benefit from the ability to forecast stream impairment, and thus threshold identification is a critical undertaking if society is to proactively conserve aquatic resources. Thus, we used our linear regression models to estimate thresholds of various single stressors

and to compare these thresholds with other studies. MMI impairment occurred when % forest catchment was <60% (<70% forest buffer for entire network US), when % total developed was >30%, and when % developed urban was >15%. In Georgia Piedmont streams, Roy et al. (2003) also suggested that 15% urban land cover represented a threshold indicating significant change in macroinvertebrate metrics. Our regression analysis detected impairment of MMI scores when % IC was >5%, a value less than the 10–15% threshold typically found in the literature (Brabec et al. 2002) but was similar to lower threshold values found in Maryland streams (King and Baker 2010; King et al. 2011) and Alaska (Ourso and Frenzel 2003). Utz et al. (2009) and Baker and King (2010) similarly reported low % urbanization and % IC thresholds for many Maryland taxa indicating similar sensitivities (and tolerances) of the top genera that we described from our NMDS ordination analysis.

Our conductivity threshold value (540 µS/cm) was higher than known extirpation thresholds for West Virginia macroinvertebrate genera (~300 µS/cm) identified by Cormier et al. (2013a) and Bernhardt et al. (2012) who used different statistical techniques (and relied on observed genera); however, Cormier et al. (2013b) found a family-level MMI had a conductivity threshold of 180 µS/cm in WV. We note that our streams drained some carbonate lithology that produced elevated background conductivities compared with other parts of the state. Moreover, ions at our impacted sites were dominated by de-icing salts (Cl and Na), compared to SO_4 and HCO_3 from coal mining as in Cormier et al. (2013a). Our regression-based threshold for MMI and Cl (30 mg/l) was lower than the range of that was reported for road-salted Toronto streams, where Cl thresholds of 50–90 mg/l were observed for macroinvertebrates (Wallace and Biastoch 2016).

Conclusions and recommendations

Considering that land alteration influences instream physical and chemical stressors (e.g., % urban developed was strongly correlated with specific conductance; total RBP habitat score was best correlated with riparian buffer zone forest), our multiple regression models provided acceptable subsets of best predictors that captured combined effects without overfitting. Headwater macroinvertebrates were very responsive to a variety of disturbance indicators but network-scale riparian forest cover was the strongest predictor overall and could potentially be used as a single predictor, but we contend that water quality measures strengthen simple land use-focused models.

From a level-of-effort based monitoring perspective (i.e., requiring less money or resources), we recommend that the model: $MMI = 144 - (42.3 \times \log_{10}$ specific conductance$) - (1.5 \times \text{road density}) + (42 \times \arcsin(\sqrt{\text{forest catchment}}))$ can be used to predict headwater stream condition in this region, but we caution its use in streams that are outside of the model's experience (e.g., catchment size, season, ecoregion). This model could be used to quickly characterize a large number of headwater catchments in a watershed to pinpoint areas in need of immediate conservation or remediation for improving receiving streams. Furthermore, our potential stressor thresholds can be evaluated at new sites to help interpret the sources of MMI impairment. Because several other models also performed relatively well (see Table 3), we believe that they too could be explored, depending on the types of information (GIS, instream, or combinations) available to the investigator.

References

Allan JD (2004) Landscapes and riverscapes: the influence of land use on stream ecosystems. Ann Rev Ecol Evolut Syst 35:257–284

Bailey RC, Reynoldson TB, Yates AG, Bailey J, Linke S (2007) Integrating stream bioassessment and landscape ecology as a tool for land use planning. Freshw Biol 52:908–917

Baker ME, King RS (2010) A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods Ecol Evol 1:25–37

Barbour MT, Gerritsen J, Snyder BD, Stribling JB (1999) Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish, 2nd edn. EPA 841-B-99-002. U.S. Environmental Protection Agency, Office of Water, Washington, DC

Bernhardt ES, Lutz BD, King RS, Fay JP, Carter CE, Helton AM, Campagna D, Amos J (2012) How many mountains can we mine? Assessing the regional degradation of central Appalachian rivers by surface coal mining. Environ Sci Technol 46:8115–8122

Bond NR, Lake PS (2003) Local habitat restoration in streams: constraints on the effectiveness of restoration for stream biota. Ecol Manage Restor 4:193–198

Brabec E, Schulte S, Richards PL (2002) Impervious surfaces and water quality: a review of current literature and its implications for watershed planning. J Plan Lit 16:499–514

Brown DG, Walker R, Manson S, Seto K (2012) Modeling land use and land cover change. In Land change science. Springer Netherlands, The Netherlands, pp 395–409

Bryant WL, Carlisle DM (2012) The relative importance of physicochemical factors to stream biological condition in urbanizing basins: evidence from multimodel inference. Freshw Sci 31:154–166

Carlisle DM, Hawkins CP (2008) Land use and the structure of western US stream invertebrate assemblages: predictive models and ecological traits. J N Am Benthol Soc 27:986–999

Cormier SM, Suter GW, Zheng L (2013a) Derivation of a benchmark for freshwater ionic strength. Environ Toxicol Chem 32:263–271

Cormier SM, Suter GW, Zheng L, Pond GJ (2013b) Assessing causation of the extirpation of stream macroinvertebrates by a mixture of ions. Environ Toxicol Chem 32:277–287

Cuffney TF, Brightbill RA, May JT, Waite IR (2010) Responses of benthic macroinvertebrates to environmental changes associated with urbanization in nine metropolitan areas. Ecol Appl 20:1384–1401

Dodds WK, Clements WH, Gido K, Hilderbrand RH, King RS (2010) Thresholds, breakpoints, and nonlinearity in freshwaters as related to management. J N Am Benthol Soc 29:988–997

Findlay S, Kelly VR (2011) Emerging indirect and long-term road salt effects on ecosystems. Ann N Y Acad Sci 1223:58–68

Gergel SE, Turner MG, Miller JR, Melack JM, Stanley EH (2002) Landscape indicators of human impacts to riverine systems. Aquat Sci 64:118–128

Griffith MB, Norton SB, Alexander LC, Pollard AI, LeDuc SD (2012) The effects of mountaintop mines and valley fills on the physicochemical quality of stream ecosystems in the central Appalachians: a review. Sci Total Environ 417:1–12

Hale AN, Noble G, Piper K, Garmire K, Tonsor SJ (2016) Controlling for hydrologic connectivity to assess the importance of catchment- and reach-scale factors on macroinvertebrate community structure. Hydrobiologia 763:285–299

Homer CH, Fry JA, Barnes CA (2012) The national land cover database US Geological Survey Fact Sheet 3020:1–4

King RS, Baker ME (2010) Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. J N Am Benthol Soc 29:998–1008

King RS, Baker ME, Whigham DF, Weller DE, Jordan TE, Kazyak PF, Hurd MK (2005) Spatial considerations for linking watershed land cover to ecological indicators in streams. Ecol Appl 15:137–153

King RS, Baker ME, Kazyak PF, Weller DE (2011) How novel is too novel? Stream community thresholds at exceptionally low levels of catchment urbanization. Ecol Appl 21:1659–1678

Kristensen EA, Baattrup-Pedersen A, Andersen HE (2012) Prediction of stream fish assemblages from land use characteristics: implications for cost-effective design of monitoring programmes. Environ Monit Assess 184:1435–1448

Kutner MH, Nachtsheim CJ, Neter J (2004) Applied Linear Regression Models, 4th edn. McGraw-Hill, Irwin

Lammert M, Allan JD (1999) Assessing biotic integrity of streams: effects of scale in measuring the influence of land use/cover and habitat structure on fish and macroinvertebrates. Environ Manage 23:257–270

Ligeiro R, Ferreira W, Hughes RM, Callisto M (2013) The problem of using fixed-area subsampling methods to estimate macroinvertebrate richness: a case study with Neotropical stream data. Environ Monit Assess 85:4077–4085

Macedo DR, Hughes RM, Ligeiro R, Ferreira WR, Castro MA, Junqueira NT, Oliveira DR, Firmiano KR, Kaufmann PR, Pompeu PS, Callisto M (2014) The relative influence of catchment and site variables on fish and macroinvertebrate richness in cerrado biome streams. Lands Ecol 29:1001–1016

McBride M, Booth DB (2005) Urban impacts on physical stream condition: effects of spatial scale, connectivity, and longitudinal trends. J Am Water Resour Assoc 41:565–580

McCune B, Grace JB (2002) Analysis of ecological communities. MjM Software, Gleneden Beach

McManus MG, Pond GJ, Reynolds L, Griffith MB (2016) Multivariate condition assessment of watersheds with linked micromaps. J Am Water Resour Assoc 52:494–507

Merriam ER, Petty JT, Strager MP, Maxwell AE, Ziemkiewicz PF (2013) Scenario analysis predicts context-dependent stream response to landuse change in a heavily mined central Appalachian watershed. Freshw Sci 32:1246–1259

Meyer JL, Strayer DL, Wallace JB, Eggert SL, Helfman GS, Leonard NE (2007) The contribution of headwater streams to biodiversity in river networks. J Am Water Resour Assoc 43:86–103

Moore AA, Palmer MA (2005) Invertebrate biodiversity in agricultural and urban headwater streams: implications for conservation and management. Ecol Appl 15:1169–1177

Morgan RP II, Kline KM, Cushman SF (2007) Relationships among nutrients, chloride and biological indices in urban Maryland streams. Urban Ecosyst 10:153–166

Ourso RT, Frenzel SA (2003) Identification of linear and threshold responses in streams along a gradient of urbanization in Anchorage, Alaska. Hydrobiologia 501:117–131

Paul MJ, Meyer JL (2001) Streams in the urban landscape. Annu Rev Ecol Syst 2011:333–365

Pond GJ, Passmore ME, Borsuk FA, Reynolds L, Rose CJ (2008) Downstream effects of mountaintop coal mining: comparing biological conditions using family- and genus-level macroinvertebrate bioassessment tools. J N Am Benthol Soc 27:717–737

Pond GJ, Bailey JE, Lowman BM, Whitman MJ (2011) The West Virginia GLIMPSS (genus-level index of most probable stream status): a benthic macroinvertebrate index of biotic integrity for West Virginia's wadeable streams. West Virginia Department of Environmental Protection, Division of Water and Waste Management, Watershed Assessment Branch, Charleston, WV. http://www.dep.wv.gov/WWE/watershed/bio_fish/Documents/20110829GLIMPSSFinalWVDEPDOI.pdf. Accessed 21 Nov 2016

Pond GJ, Bailey JE, Lowman BM, Whitman MJ (2012) Calibration and validation of a regionally and seasonally stratified macroinvertebrate index for West Virginia wadeable streams. Environ Monit Assess 185:1515–1540

Rios SL, Bailey RC (2006) Relationship between riparian vegetation and stream benthic communities at three spatial scales. Hydrobiologia 553:153–160

Roth NE, Allan JD, Erickson DL (1996) Landscape influences on stream biotic integrity assessed at multiple spatial scales. Landscape Ecol 11:141–156

Roy AH, Shuster WD (2009) Assessing impervious surface connectivity and applications for watershed management. JAWRA J Am Water Resour Assoc 45:198–209

Roy AH, Rosemond AD, Paul MJ, Leigh DS, Wallace JB (2003) Stream macroinvertebrate response to catchment urbanisation (Georgia, USA). Freshw Biol 48:329–346

Roy AH, Freeman BJ, Freeman MC (2007) Riparian influences on stream fish assemblage structure in urbanizing streams. Landsc Ecol 22:385–402

Roy AH, Dybas AL, Fritz KM, Lubbers HR (2009) Urbanization affects the extent and hydrologic permanence of headwater streams in a midwestern US metropolitan area. J N Am Benthol Soc 28:911–928

Sliva L, Williams DD (2001) Buffer zone versus whole catchment approaches to studying land use impact on river water quality. Water Res 35:3462–3472

Sponseller RA, Benfield EF, Valett HM (2001) Relationships between land use spatial scale and stream macroinvertebrate communities. Freshw Biol 46:1409–1424

Swan CM, DePalma CA (2012) Elevated chloride and consumer presence independently influence processing of stream detritus. Urban Ecosyst 15:625–635

Tran CP, Bode RW, Smith AJ, Kleppel GS (2010) Land-use proximity as a basis for assessing stream water quality in New York State (USA). Ecol Indic 10:727–733

Utz RM, Hilderbrand RH, Boward DM (2009) Identifying regional differences in threshold responses of aquatic invertebrates to land cover gradients. Ecol Indic 9:556–567

Utz RM, Hopkins KG, Beesley L, Booth DB, Hawley RJ, Baker ME, Freeman MC, L Jones K (2016) Ecological resistance in urban streams: the role of natural and legacy attributes. Freshw Sci 35:380–397

Van Sickle J, Johnson CB (2008) Parametric distance weighting of landscape influence on streams. Landsc Ecol 23:427–438

Van Sickle J, Baker J, Herlihy A, Bayley P, Gregory S, Haggerty P, Ashkenas L, Li, J (2004) Projecting the biological condition of streams under alternative scenarios of human land use. Ecol Appl 14:368–380

Ver Hoef JM, Peterson EE, Clifford D, Shah R (2014) SSN: an R package for spatial statistical modeling on stream networks. J Stat Softw 56:1–43

Villeneuve B, Souchon Y, Usseglio-Polatera P, Ferréol M, Valette L (2015) Can we predict biological condition of stream ecosystems? A multi-stressors approach linking three biological indices to physico-chemistry, hydromorphology and land use. Ecol Indic 48:88–98

Wahl CM, Neils A, Hooper D (2013) Impacts of land use at the catchment scale constrain the habitat benefits of stream riparian buffers. Freshw Biol 58:2310–2324

Wallace AM, Biastoch RG (2016) Detecting changes in the benthic invertebrate community in response to increasing chloride in streams in Toronto, Canada. Freshw Sci 35:353–363

Walsh CJ, Roy AH, Feminella JW, Cottingham PD, Groffman PM, Morgan RP II (2005) The urban stream syndrome: current knowledge and the search for a cure. J N Am Benthol Soc 24:706–723

Walsh CJ, Waller KA, Gehling J, Nally RM (2007) Riverine invertebrate assemblages are degraded more by catchment urbanisation than by riparian deforestation. Freshw Biol 52:574–587

Wang L, Lyons J, Rasmussen P, Seelbach P, Simon T, Wiley M, Kanehl P, Baker E, Niemela S, Stewart PM (2003) Watershed, reach, and riparian influences on stream fish assemblages in the Northern Lakes and Forest Ecoregion, USA. Can J Fish Aquatic Sci 60:491–505

Wang L, Seelbach PW, Lyons J (2006) Effects of levels of human disturbance on the influence of catchment, riparian, and reach-scale factors on fish assemblages. In: RM Hughes, L Wang, PW Seelbach (eds) Landscape influences on stream habitats and biological assemblages. American Fisheries Society Symposium 48, Bethesda, Maryland, pp 641–664

[WVDEP] West Virginia Department of Environmental Protection (2015) Watershed assessment branch 2015 standard operating procedures. Division of Water and Waste Management, Watershed Assessment Branch, Charleston, WV. http://www.dep.wv.gov/WWE/watershed/Pages/WBSOPs.aspx

Wickham J, Neale A, Mehaffey M, Jarnagin T, Norton D (2016) Temporal Trends in the Spatial Distribution of Impervious Cover Relative to Stream Location. J Am Water Resour Assoc 52:409–419

Woods AJ, Omernik JM, Brown DD, Kiilsgaard CW (1996) Level III and IV Ecoregions of Pennsylvania and the Blue Ridge Mountains, the Ridge and Valley, and the Central Appalachians of Virginia, West Virginia, and Maryland EPA/600R-96/077. USEPA, ORD, Corvallis