OpenAccess

# The relative influence of geographic location and reach-scale habitat on benthic invertebrate assemblages in six ecoregions

**Mark D. Munn · Ian R. Waite · David P. Larsen · Alan T. Herlihy**

**Abstract** The objective of this study was to determine the relative influence of reach-specific habitat variables and geographic location on benthic invertebrate assemblages within six ecoregions across the Western USA. This study included 417 sites from six ecoregions. A total of 301 taxa were collected with the highest richness associated with ecoregions dominated by streams with coarse substrate (19–29 taxa per site). Lowest richness (seven to eight taxa per site) was associated with ecoregions dominated by fine-grain substrate. Principle component analysis (PCA) on reach-scale habitat separated the six ecoregions into those in high-gradient mountainous areas (Coast Range, Cascades, and Southern Rockies) and those in lower-gradient ecoregions (Central Great Plains and Central California Valley). Nonmetric multidimensional scaling (NMS) models performed best in ecoregions dominated by coarse-grain substrate and high taxa richness, along with coarse-grain substrates sites combined from multiple ecoregions regardless of location. In contrast, ecoregions or site combinations dominated by fine-grain substrate had poor model performance (high stress). Four NMS models showed that geographic location (i.e. latitude and longitude) was important for: (1) all ecoregions combined, (2) all sites dominated by coarse-grain substrate combined, (3) Cascades Ecoregion, and (4) Columbia Ecoregion. Local factors (i.e. substrate or water temperature) seem to be overriding factors controlling invertebrate composition across the West, regardless of geographic location.

M. D. Munn (✉)
U.S. Geological Survey, 934 Broadway, Suite 300, Tacoma, WA 98402, USA
e-mail: mdmunn@usgs.gov

I. R. Waite
U.S. Geological Survey, 10615 Cherry Blossom Drive, Portland, OR 97216, USA

D. P. Larsen
WED, NHEERL, U.S. Environmental Protection Agency, 200 SW 35th St., Corvallis, OR 97333, USA

A. T. Herlihy
Department of Fisheries and Wildlife, Oregon State University, Corvallis, OR 97331, USA

## Introduction

Benthic invertebrate assemblages are used routinely as indicators of water quality and biological integrity of streams and rivers (Rosenberg and Resh 1993; Karr and Chu 1997). Programs that utilize benthic invertebrate assemblages commonly

rely on some type of a landscape classification system in either the design and/or interpretation of their studies. Landscape classifications, as in ecoregions, are intended to provide resource managers with discrete environmental settings which facilitate the monitoring and management of water resources (Hawkins and Norris 2000). Although numerous studies have addressed spatial patterns in stream biota and similarity among and within ecoregions, studies have shown poor performance for algae (Pan et al. 2000), invertebrates (Waite et al. 2000) and fish (McCormick et al. 2000). What is lacking in the use of geographic classifications schemes is a better understanding of which proximate, or direct, variables among different landscapes are most associated with biological assemblages (Hawkins et al. 2000).

One approach for assessing the influence of direct factors on aquatic assemblages is the use of gradient studies. Gradients are used to evaluate biological responses over a range in a particular stressor (Woodcock and Huryn 2007), with many gradient studies focusing on anthropogenic stressors, as in metals (Carlise and Clements 2003; Woodcock and Huryn 2007), nutrient/organic inputs (Porter et al. 2008), or larger-scale diffuse land use gradients (Cuffney et al. 2000; Hering et al. 2006). However, underlying any anthropogenic gradient are also natural gradients (e.g. temperature, substrate) that can greatly influence our understanding of what is controlling stream assemblages, particularly when we are attempting to understand anthropogenic gradients (Hering et al. 2006). The importance of studying biological assemblages from a gradient perspective has been demonstrated in algae (Biggs 1995; Stevenson 1997; Munn et al. 2002) and invertebrates (Waite et al. 2000; Black et al. 2004).

Although there has been extensive work on watershed to ecoregion assessments, less has been investigated on environmental gradients across large geographic regions. The objectives of this study are to determine (1) what reach-scale habitat variables are important for benthic invertebrate assemblages within and among six ecoregions, (2) whether geographic location (latitude and longitude) is important within and among the six ecoregions, and (3) whether streams classified by a nongeographic system, like substrate, differs in model performance.

## Methods

### Study region

A primary requirement in this study was to obtain datasets that contained a sufficient number of sites within an ecoregion, which relied on similar field and laboratory methods. Two sources of data satisfied this requirement: the U.S. Environmental Protection Agency (EPA), Environmental Monitoring and Assessment Program (EMAP) and the U.S. Geological Survey (USGS) National Water-Quality Assessment (NAWQA) Program. Six EPA Level III ecoregions (Omernik 1987) were included in this study (Table 1, Fig. 1), with site selection based on four criteria: (1) reach-scale habitat data were collected using comparable methods, (2) benthic invertebrate samples were collected using comparable methods, (3) benthic invertebrate taxonomy had similar taxonomic resolution, and (4) data were collected within a 4-year period (1993–1996). These criteria resulted in the selection of data from seven EMAP studies (396 sites) and one NAWQA study (21 sites) for a total of 417 sites in six ecoregions (Table 1). NAWQA data were included to increase the sample size for the Columbia Plateau Ecoregion.

Three of the six ecoregions are in the Pacific Northwest. The Coast Range Ecoregion (Coast) is characterized as having high annual rainfall (58–700 cm), heavily forested lands dominated by coniferous trees, with elevations ranging from 0 to 2,300 m. The Cascades Ecoregion (Cascades) is similar to the Coast Range in that it also has high rainfall (67–546 cm); however, it has much higher elevations (61–4,300 m) with winter snowfall instead of rain and some glacial influence. In contrast, the Columbia Plateau (Plateau) is a semi-arid region with rainfall ranging from 18 to 113 cm/year (Munn et al. 2002), with highly variable landforms. The Plateau is dominated by a combination of natural grasslands and sagebrush, which primarily has been converted to irrigated croplands. The Central Valley (Valley)

**Table 1** Summary of the number of sites by biome and ecoregion

| Biome | Ecoregion level III | Abbreviation | Years | No. sites |
|---|---|---|---|---|
| Coniferous mountains | Southern Rockies (21) | Rockies | 1994–1995 | 84 |
| | Coast Range (1) | Coast | 1994–1995 | 122 |
| | Cascades (4) | Cascades | 1994–1996 | 39 |
| Desert/chaparral | Central California Valley (7) | Valley | 1994–1995 | 83 |
| | Columbia Plateau (10) | Plateau | 1993–1995 | 31 |
| Plains/grasslands | Central Great Plains (27) | Plains | 1994–1995 | 58 |
| Total | | | | 417 |

in California consists of flat plains and a dry climate (15–100 cm/year). Although the historical vegetation consisted of California steppe and tule marshes, the present vegetation is dominated by irrigated crop production and grazing. The Southern Rockies (Rockies) consists of high mountains and tablelands with elevation ranging from 1,500 to 4,300 m and with rainfall ranging from 22 to 152 cm/year. Vegetation includes coniferous forests mixed with alpine meadows. The Central Great Plains Ecoregion (Plains) consists of irregular plains, subarid climate (43–100 cm/year), and intensive dryland and irrigated cropland and grazing (Omernik 1987). Although agriculture is the dominant land use disturbance by area in many western watersheds, other land uses like urban and mining can have a substantial impact. Total land disturbance, defined as the sum of percent agriculture, urban, mining, and transitional, indicated that the three mountainous ecoregions had only 2% to 7% of their land area disturbed, while the other three had 42% to 71% of their land disturbed (James Falcone, U.S. Geological Survey, written commun., 2006).

Development of datasets

EMAP sites were selected using a randomized sampling design with a systematic spatial component using the stream network (Herlihy et al. 2000), whereas NAWQA sites were selected in the study area based on the relative areas of the land use and to represent a complete gradient of environmental conditions (Munn et al. 2002). Only wadeable streams (1–4th order) were sampled, with sampling occurring in the summer or

**Fig. 1** Map showing location of six ecoregions
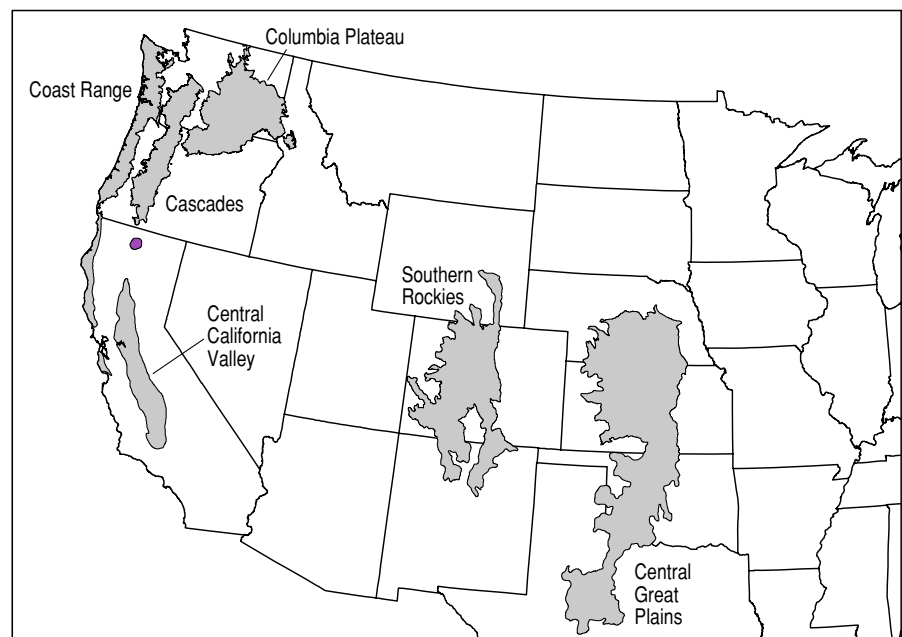
fall period; sampling included reach-scale habitat and benthic invertebrates.

Reach-scale habitat was assessed using standardized protocols (Kaufmann et al. 1999; Meador et al. 1993). Specific site reach lengths ranged from 20 to 40 times the wetted stream width, with habitat measured at 6 (NAWQA) or 11 (EMAP) transects. Reach lengths were always a minimum of 150 m. The habitat variables used in this study included only those that were comparable between the two programs. This resulted in a total of 18 reach-scale habitat variables (Table 2). Longitudinal profiles were used to determine relative areas of major geomorphic units (riffle, pool, and run) along with establishing transects for measurement of channel cross-sectional morphology. Reach slope (SLOPE) was determined using a handheld clinometer, with shear-strength index (SSI) calculated using the formula SSI = (avg depth/2) * slope. The majority of the remaining reach-scale habitat measurements were

**Table 2** Location and environmental variables used to characterize streams in six ecoregions

| Abbreviations | Variables |
| --- | --- |
| Location | |
| LAT | Latitude |
| LONG | Longitude |
| | |
| Reach-level habitat | |
| AQM | Percent aquatic macrophytes |
| BANK | Bank angle (°) |
| BRS | Percent overhanging brush |
| CANOPY | Percent canopy cover |
| CG | Percent coarse substrate (gravel or larger) |
| COND | Specific conductance ($\mu$S/cm) |
| DEPTH | Depth (m) |
| ELEV | Elevation (m) |
| FG | Percent fine substrate (sand or finer) |
| LWD | Percent large woody debris |
| NAT | Percent natural instream cover |
| OHV | Percent overhanging vegetation |
| RIFF | Percent linear riffle habitat |
| SLOPE | Percent stream slope |
| SSI | Shear strength index |
| TEMP | Water temperature (°C) |
| WIDTH | Stream width (m) |
| WXD | Water width × depth (m$^2$) |

made at each transect. Instream-habitat features included percent cover of aquatic macrophytes (AQM); overhanging brush (BRS), which refers to woody debris; overhanging vegetation (OHV), which includes all nonwoody vegetation as in grasses; and natural habitat (NAT), which is the sum of any type of natural-habitat cover. All of these habitat features were estimated at each transect within a 4–10 m band along each transect. Stream width (WIDTH) was determined at each transect as wetted width, with depth (DEPTH) and dominant substrate collected at five points along each transect (three for NAWQA). Percent dominant substrate was determined at each transect point based on the following particle size classifications: bedrock (>4,000 mm); hardpan (solid claylike substrate); boulders (>250–4,000 mm); cobbles (>64–250 mm); gravel (>2–64 mm); sand (>0.06–2 mm); silt, clay, and muck (<0.06 mm); and wood/detritus. Coarse-grained substrate (CG) was defined as gravel or larger (>2 mm) and included bedrock and hardpan, with fine-grained substrate (FG) defined as sand plus silt (<2 mm). Bank angles (BANK) were determined at each transect by placing a rod 90° from the waters edge and measuring the bank angle with a clinometer. Large woody debris (LWD) was determined by measuring all woody material with small-end diameter of at least 10 cm and length of at least 1.5 m, with all LWD in a reach assigned to a size-class category. Water temperature and conductivity were measured using standard field meters. Elevation, latitude, and longitude were determined using 1:24,000 USGS maps.

For EMAP sites, macroinvertebrate samples were collected from riffle or run habitats at each transects. Samples were collected using a semi-quantitative kick sample (500 $\mu$m mesh net). NAWQA samples also were collected using a semiquantitative kick sample (425 $\mu$m mesh net), at three to five riffle or run habitats along the length of the reach. Kick samplers were placed on the stream bottom with the substrate within a specific area disturbed by kicking and then hand cleaning to dislodge organisms. All samples collected within a reach were composited, and preserved using either 95% ethanol or 10% formalin. Invertebrates were sorted using standard 500 count procedures and identified to the

lowest practical taxonomic level (Klemm and Lazorchak 1994; Moulton et al. 2000). Because benthic-invertebrate data used in this study originated from different laboratories, we went through a series of quality-control steps to ensure that all samples had similar taxonomic resolution. First, we merged all data and standardized taxonomic names. Second, we examined the frequency of taxa occurrence and total abundance by individual study. This permitted us to evaluate whether or not all studies had identified invertebrates to the same taxonomic level. If studies used different taxonomic levels, then we aggregated all data to the higher taxonomic level. For example, some studies identified chironomids to the family level, while others identified them to genus. Therefore, we aggregated all chironomid data to the family level. Last, there were cases where specific taxa were identified to both the genus and family level. In this situation we aggregated the data up to family level when the total abundances at the family level exceeded 20%. This strategy resulted in equalizing the percent of taxa within a specific ecoregion with the number of genera (G) in an ecoregion ranging from 62% to 72% and the number of families (F) ranging from 23% to 36%. Within ecoregions this resulted in the following: Coast (27% F, 70% G), Cascades (26% F, 71% G), Plateau (32% F, 65% G), Valley (23% F, 72% G), Rockies (28% F, 70% G), and Plains (36% F, 62% G).

Analyses

Principle-component analysis (PCA) was used to determine whether sites from the six ecoregions combined would separate based on reach-scale habitat variables. Habitat data were square root transformed, normalized, and analyzed using PRIMER© (Clarke and Gorley 2006).

Nonmetric multidimensional scaling (NMS) (McCune and Mefford 1999) was used to summarize patterns in macroinvertebrate assemblage structure and then to determine the association of reach-scale habitat variables and geographic location with the dominant NMS axis. Benthic macroinvertebrate assemblage data were analyzed on the basis of relative abundance (0–100%). Environmental variables used to determine highest correlation (Kendall's tau) with each of the NMS axes primarily included reach-scale habitat variables, conductivity, and the use of latitude and longitude as the geographic location variables. NMS models were divided into three groups for a total of nine models. The first group consisted of all sites combined (West, $n = 417$). The second group consisted of six NMS model runs based on ecoregions (1) Coast ($n = 119$), (2) Cascades ($n = 41$), (3) Valley ($n = 83$), (4) Plateau ($n = 32$), (5) Rockies ($n = 84$), and (6) Plains ($n = 58$). Because substrate greatly varied among the sites, we also test two additional NMS models with site groupings based on substrate type. The two additional NMS runs were made with sites classified as coarse-grain sites (coarse, $n = 209$, average percent fines <33% FG) and fine-grain sites (fine, $n = 208$, average percent fines >33% FG). Each of the two substrate-based models still contained a range of substrates, but a more narrow range.

## Results

### Reach-scale habitat variables

While the PCA results indicated a great deal of variation in site distribution, there was a major separation along the first two axes (Fig. 2). PCA axis I had an eigenvalue of 6.1 accounting for 32% of the variation, whereas PCA axis II had an eigenvalue of 3.0 accounting for 16% of the variation. Sites on the left side of PCA axis I were dominated predominantly by fine-grain substrate, warmer water temperatures, and lower gradient (Plains and Valley), with sites to the right of PCA axis 1 dominated by coarse substrate, higher gradient, and cooler water temperatures (Coast, Cascades, and Rockies). Axis 2 had far fewer explanatory variables; however, the axis was related primarily to stream size. The middle of axis I primarily was dominated by sites that contained a range in these conditions (Plateau, Coast, and Rockies). Table 3 lists a summary of environmental variables by ecoregion. Substrate and water temperature were two of the more common variables found to be correlated with PCA (Fig. 3). Substrate had an increase in percent FG going

**Fig. 2** Principle-component-analysis diagram of 417 sites in six ecoregions and their relative location based on reach-scale habitat features. See Table 1 for ecoregion type
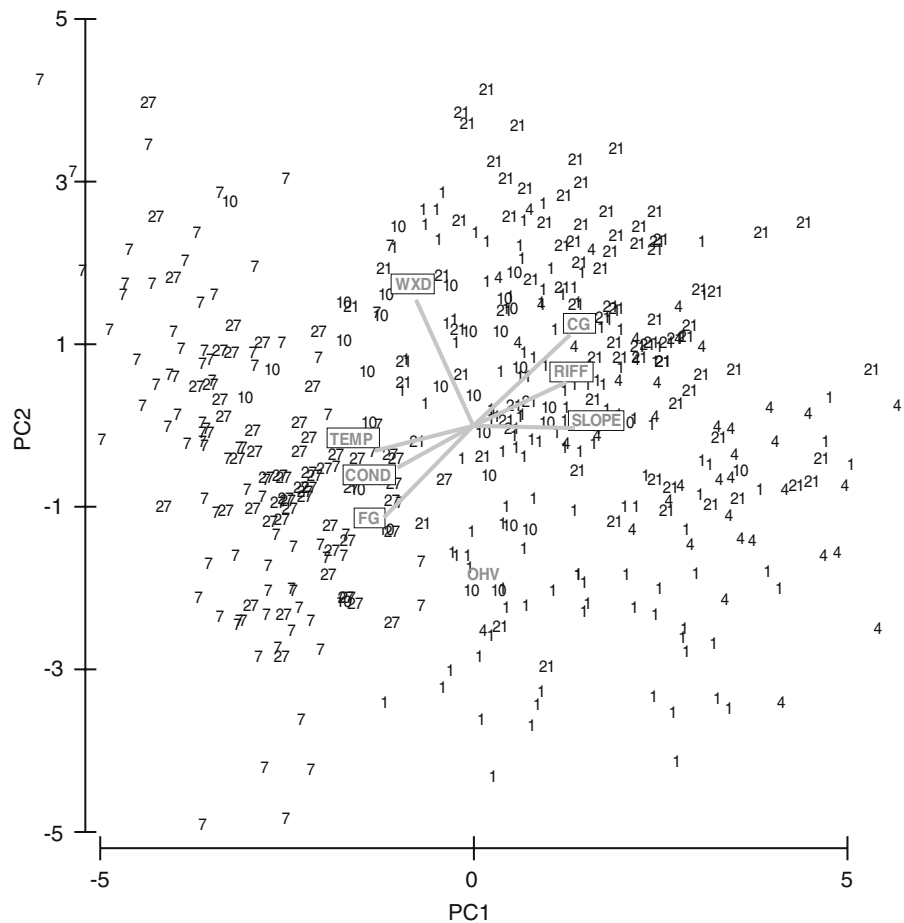


from the Rockies and Cascades to the Plains and Valley (Fig. 3a). The pattern in water temperature was somewhat similar in that the Rockies and Cascades had the lowest temperatures, and the Plains and Valley the highest temperatures (Fig. 3b).

Invertebrate assemblage composition

The composition of benthic invertebrate assemblages varied substantially among the six ecoregions (Table 4). Overall, there were 301 taxa among the 417 sites, with an average of 19 taxa per site. The four ecoregions with more varied stream habitat and a higher percentage of coarse substrate (Rockies, Coast, Cascades, and Plateau) had higher total taxa richness (131–180), a higher average taxa richness per site (19–31), and taxa in these samples tended to be more evenly distributed. In contrast, the two fine-grain dominated

ecoregions (Valley and Plains) contained few total taxa (91–92), but more important a substantially reduced average taxa richness per site (seven to eight). The lower taxa richness in the Valley was due to the dominance by chironomids, snails, amphipods, and oligochaetes; whereas, the low taxa richness in the Plains was due to a complete dominance of chironomids.

Ordination of community assemblages and environmental factors

NMS model performance is determined using model stress and instability. Stress is defined as a measure that expresses whether the distances in the ordination diagram correspond to the dissimilarity values. Stress values of between 10 and 20 commonly are reported in ecological studies; with values above 20 being cause for concern in

**Table 3** Summary of key environmental variables for the six ecoregions combined and individually

| Variables | All N=417 | Coarse-grained substrate | | | | Fine-grained substrate | |
|---|---|---|---|---|---|---|---|
| | | Coast N=119 | Cascades N=41 | Plateau N=32 | Rockies N=84 | Valley N=83 | Plains N=58 |
| AQM | 8 (1) | 33 (0.6) | 2.6 (0.8) | 18 (3.1) | 4.1 (1.4) | 20 (2.4) | 6.6 (1.7) |
| BANK | 42.7 (0.81) | 42.0 (1.6) | 38.6 (1.9) | 33.2 (2) | 51.0 (1.8) | 43.0 (1.5) | 39.4 (2.1) |
| BRS | 8 (0.5) | 10.4 (0.9) | 16.5 (2.6) | 1.3 (.5) | 6.0 (0.9) | 2.4 (0.7) | 9.3 (0.9) |
| CANOPY | 41.6 (1.8) | 73.2 (2.2) | 73.6 (4.0) | 30.5 (4.8) | 32.4 (3.0) | 22.7 (2.9) | 0.52 (0.05) |
| CG | 39.3 (1.8) | 45.2 (2.5) | 59.4 (3.5) | 56.6 (5.7) | 74.8 (2.7) | 3.6 (1.4) | 2.8 (1.3) |
| COND | 365.5 (32) | 82.7 (4.5) | 62.9 (7.9) | 369.4 (28.4) | 162.9 (21.9) | 610.4 (103.4) | 1100.5 (119.6) |
| DEPTH | 49.2 (2) | 38.1 (2.2) | 32.6 (3.6) | 38.5 (4.1) | 41.9 (2.0) | 89.1 (7.0) | 43.1 (3.1) |
| ELEV | 792 (48.9) | 175.7 (13.3) | 851.1 (58.9) | 459.6 (43.6) | 2668.3 (37.6) | 46.3 (4.7) | 533.2 (16.7) |
| FINES | 46.4 (1.9) | 30.8 (2.2) | 20.9 (2.6) | 34.8 (5.6) | 16.0 (1.9) | 85.2 (3.2) | 91.2 (2.13) |
| LWD | 0.05 (.005) | 0.1 (0.01) | 0.13 (0.02) | 0.006 (0.005) | 0.024 (0.005) | 0.008 (0.003) | 0.025 (0.004) |
| NAT | 45 (2) | 54.4 (2.9) | 77.2 (7.5) | 43 (4.2) | 54.9 (3.6) | 25.4 (3.2) | 23.4 (2.1) |
| OHV | 15 (1) | 14.3 (1.4) | 20.2 (2.6) | 16.7 (2.6) | 12.0 (1.7) | 20.9 (2.7) | 10.3 (1.4) |
| RIFF | 24.9 (1.3) | 29.4 (1.7) | 10.6 (2.7) | 33.1 (4.6) | 45.4 (2.9) | 0.78 (0.4) | 4.72 (1.2) |
| SLOPE | 2.2 (0.18) | 2.8 (0.38) | 5.8 (0.90) | 1.2 (0.27) | 2.7 (0.40) | 0.15 (.02) | 0.66 (0.11) |
| SSI | 31.2 (2.1) | 35.6 (3.2) | 64.6 (7.8) | 18.2 (5.4) | 51.1 (6.6) | 5.9 (0.97) | 13.3 (2.42) |
| TEMP | 15.7 (0.25) | 13.5 (0.2) | 10.1 (0.5) | 16 (0.68) | 12.1 (0.4) | 21.0 (0.3) | 21.4 (0.48) |
| WIDTH | 7.7 (0.38) | 7.3 (0.54) | 5.5 (0.92) | 8.4 (1.59) | 7.0 (0.46) | 8.4 (0.91) | 9.4 (1.74) |
| WXD | 5.0 (0.4) | 3.8 (0.4) | 2.6 (2.6) | 4.1 (1.2) | 3.4 (0.3) | 10.1 (1.6) | 4.8 (1.1) |

Values are mean and SE of mean

**Fig. 3** Distribution of **a** percent fine data and **b** water temperature for each of the six ecoregions. See Table 1 for ecoregion type
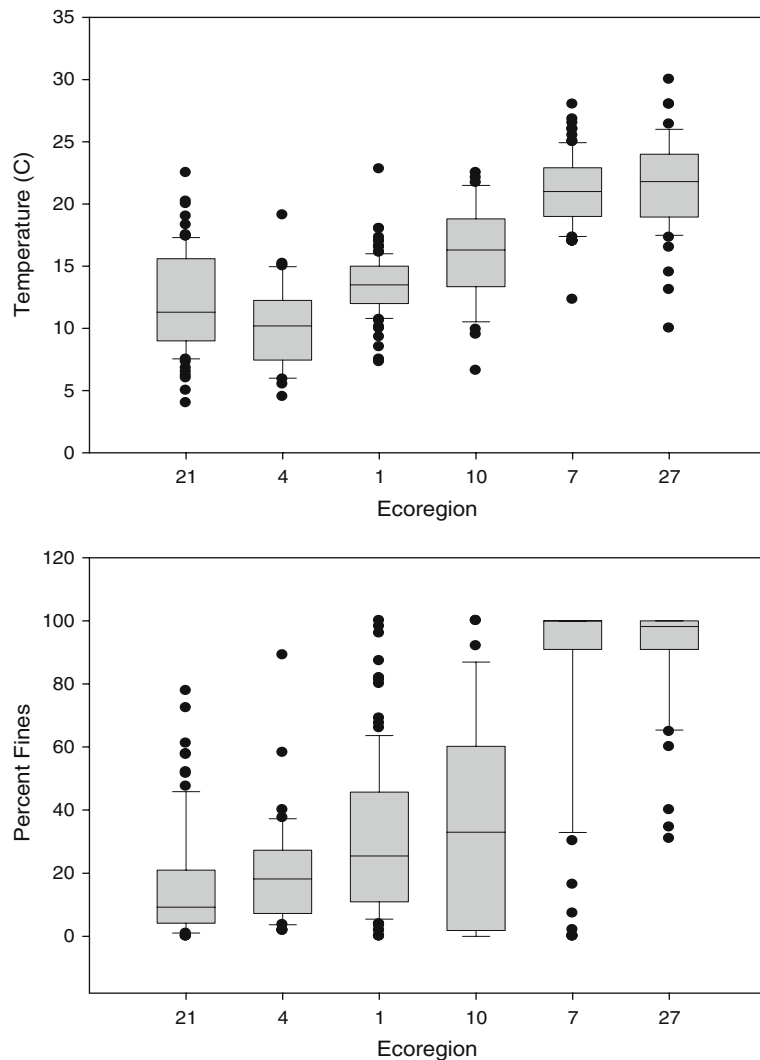


regards to interpretation (McCune et al. 2002). The second statistic used for model assessment is instability, which is defined as the standard deviation in stress over the previous ten iterations, with the number of iterations increasing with increased instability. An instability value of $<10^{-4}$ is sometimes used as a common acceptability criterion (McCune et al. 2002). These model criteria are meant to serve as general guidelines for the model performance.

The West model ($n = 417$) had poor performance, with high stress (24.5), high instability (0.004) and moderate $r^2$ (0.55) (Table 5). The most important environmental variables that were most correlated with the NMD axes were features of stream morphology; i.e. substrate (CG)

and LWD. Other important variables included TEMP, COND, ELEV, and LAT, a measure of geographic location.

The best NMS model performance was observed among some of the ecoregion models (Table 5). The Coast, Cascades, Plateau, and Rockies, all of which tended to share some common features, had acceptable model stress levels (12.3–15.8) and high $r^2$ values (0.74–0.84). The most consistently important environmental variables for these five groups were water temperature, and variables associated with channel morphology (i.e. substrate) or hydraulics (i.e. slope), with temperature and substrate the only variables important in all four models. These were the same habitat variables that were important

**Table 4** Summary of taxa richness and evenness for the six ecoregions

| Ecoregion level III | Total sites | Total taxa | Avg. site richness | Avg. site evenness | Avg. site percent (range) chironomids |
|---|---|---|---|---|---|
| Rockies | 84 | 136 | 19 | 0.67 | 22 (0–93) |
| Coast | 122 | 180 | 29 | 0.65 | 29 (0–87) |
| Cascades | 39 | 142 | 31 | 0.73 | 18 (2–58) |
| Plateau | 31 | 131 | 21 | 0.61 | 32 (1–85) |
| Valley | 83 | 91 | 7 | 0.60 | 29 (0–100) |
| Plains | 58 | 92 | 8 | 0.63 | 44 (0–100) |
| Total | 417 | 301 | 19 | 0.65 | |

in the PCA model. Additional environmental variables correlated with the NMS axes included longitude, elevation, and cover (canopy and overhanging vegetation). Elevation was important only in the Southern Rockies dataset, whereas longitude was important in the Cascades, and Plateau ecoregions. Riparian cover was important in the Coast, Cascades, and the Rockies, the only systems with extensive mountain terrain. The Valley and Plains models, both of which were dominated by fine-grain sediment, had poor model performance. The Valley model had a stress of 24.6, which is outside of acceptable model performance, and a $r^2$ of only 0.59. The environmental variables that were most correlated with the NMD axes were LWD, RIFF, DEPTH, and COND. Substrate was less important in the Valley due to most sites being dominated by fine-grained sediment. The Plains dataset contained insufficient information to develop an appropriate statistical solution using NMS.

The third group of models included the sites grouped by substrate regardless of location. The Coarse model was a bit unique in that the model had an acceptable stress level, high instability, and a moderate $r^2$ value (Table 5). This dataset consisted of all sites with <33% fine-grained sediment and therefore contained sites from most all of the ecoregions. The dominant reach-scale variables that were correlated with the NMS axes included temperature, slope, and substrate. This is the only dataset where both latitude and longitude were correlated with the NMS axes. In contrast, the Fine model had poor performance with a high stress value (27.3) and a low $r^2$ (0.58). The environmental variables associated with the first two axes included RIFF, LWD, and LAT.

## Discussion

### Reach-scale habitat variables

NMS models that included highly dispersed sites were West, Coarse, and Fine. The habitat variables at these sites were most correlated with the NMS axis, which included those associated with water temperature and stream morphology features, as in substrate (FG and CG) and LWD. Substrate (i.e. FG or CG) was a key variable in all ecoregion based NMS models with the exception of the fine-grain dominated Valley (Table 5). Once sites were partitioned into Fine and Coarse, substrate was reduced to a less important variable for the coarse-grained group (CG, $r = -0.28$), but dropped out as a variable for the Fine model. Substrate has long been recognized as a major controlling factor for benthic invertebrates in that it commonly explains much of the invertebrate assemblage composition and distribution of populations (Rempel et al. 2000; Richards et al. 1993). The importance of substrate in controlling benthic invertebrates is related to organic matter retention (Culp et al. 1983), biotic interactions of predation and competition (Lancaster et al. 1990), and providing in-stream flow refugia from hydraulic stress (Lancaster and Hildrew 1993). The role of substrate in our study was shown by the importance of percent CG in the ecoregion specific analysis for four of the ecoregions. For example, in the Rockies we found that substrate was important for two of the three NMS axes. This also was the conclusion by Griffith et al. (2001), which examined the same dataset using other statistical methods. They reported that percent boulder and/or percent cobble were some

**Table 5** NMS: proportion of variance represented by each axis, based on the $r^2$ between distance in the ordination space and distance in the original space. Environmental variables presented are the variables with the highest correlation value (Kendall's tau) that are not correlated with each other (less than 0.5)

| Model | Ecoregion level | Proportion of variance for each axis $r^2$ (correlation coefficients for environmental variables by axis) | | | Total $r^2$ | Stress 10–20 | Instability <0.0001 |
|---|---|---|---|---|---|---|---|
| | | I | II | III | | | |
| All sites model | West ($n = 417$) | 0.37<br>TEMP (0.44)<br>LWD (−0.33)<br>LAT (−0.33) | 0.18<br>CG (0.39)<br>COND (−0.26)<br>ELEV (0.26) | | 0.55 | 24.5 | 0.004 |
| Ecoregion models | Plateau ($n = 32$) | 0.50<br>TEMP (0.37)<br>LONG (−0.30) | 0.22<br>LWD (0.26)<br>FG (−0.26) | 0.12<br>LWD (0.31)<br>SLOPE (0.24) | 0.84 | 12.3 | 0.0003 |
| | Rockies ($n = 84$) | 0.27<br>TEMP (−0.41)<br>ELEV (0.37) | 0.25<br>CG (−0.31)<br>SSI (−0.23) | 0.22<br>FG (0.18)<br>CAN (−0.20) | 0.74 | 14.6 | 0.0005 |
| | Coast ($n = 119$) | 0.35<br>FG (0.30)<br>SLOPE (−0.18) | 0.29<br>CG (0.29)<br>BRS (−0.21) | 0.16<br>DEPTH (−0.32)<br>TEMP (−0.27) | 0.81 | 15.2 | 0.00008 |
| | Cascades ($n = 41$) | 0.27<br>TEMP (−0.43)<br>WIDTH (−0.26) | 0.25<br>CG (−0.35)<br>SSI (−0.26) | 0.24<br>CAN (0.43)<br>LONG (0.42)<br>SLOPE (0.42) | 0.76 | 15.8 | 0.0005 |
| | Valley ($n = 83$) | 0.32<br>LWD (0.28)<br>COND (0.22) | 0.27<br>RIFF (0.24)<br>DEPTH (−0.21) | | 0.59 | 24.6 | 0.0056 |
| Substrate models | Coarse ($n = 209$)<br>(<33% SAFN) | 0.27<br>TEMP (−0.45)<br>SLOPE (0.37) | 0.27<br>LAT (−0.13)<br>FG (0.11) | 0.21<br>LONG (0.32)<br>CG (−0.28) | 0.76 | 19.0 | 0.005 |
| | Fine ($n = 208$)<br>(>33% SAFN) | 0.37<br>RIFF (0.48)<br>LWD (0.37) | 0.23<br>LWD (−0.35)<br>LAT (−0.26) | | 0.58 | 27.3 | 0.008 |

of the dominant variables explaining invertebrate assemblage distributions. In contrast, we found that substrate played no important role in the Valley; however, Griffith et al. (2003) reported that substrate was an important variable. Our study did not find this same conclusion even though similar datasets were used. A possible explanation for this discrepancy is that our study grouped substrate into larger size fraction categories along with having to aggregate chironomids to the family level in order to compare results with other ecoregions.

Water temperature also consistently played an important role in distinguishing benthic invertebrate distributions at both the large, multi-ecoregion scale, and within most individual ecoregions (Table 4). In regions with high relief, water temperature and elevation were often times correlated, as was the case with the Coast ($r = -0.35$), Rockies ($r = -0.45$), and Cascades ($r = -0.56$) ecoregions. However, in ecoregions like the Plateau, water temperature and elevation were not highly correlated ($r = -0.12$) due to the influence of ground water and surface-irrigation water used for agriculture (Munn et al. 2002). Water temperature commonly is reported to be an important variable controlling benthic invertebrate assemblages (Tate and Heiny 1995; Griffith et al. 2001; Friberg et al. 2001), due to growth and life cycle requirements (Sweeney and Vannote 1984).

Conductivity was an important factor in several of the NMS models. Conductivity tends to be an integrator variable across multiple scales and is believed to be an indicator of water quality conditions, and nutrient enrichment specifically (Munn et al. 2002). Conductivity is, therefore, a variable that indicates several high-level watershed processes and, as such, can explain considerably more variability than flow, water temperature, and elevation (Biggs 1995).

Geographic location

Although biogeography helps explain the distribution of aquatic organisms, geographic location (i.e. latitude and longitude) rarely is used as an explanatory variable. This is likely because benthic studies tend to focus on catchment to ecoregion scale, and only more recently have researchers begun to focus on regional to national scales. In our study we found that latitude and/or longitude were important explanatory variables in five of the eight NMS models. Geographic location consistently was more important in models that incorporated sites from all six ecoregions. For example, latitude and/or longitude were correlated with at least one of the NMS axes for the West, Coarse, and Fine models. Furthermore, latitude and longitude were not correlated with any of the reach-scale habitat variables, inferring that geographic location is independent of reach-scale habitat and therefore biogeography is a partial explanation for differences among benthic invertebrate assemblages across large geographic areas. In contrast, the role of geographic location was not important in the Coast, Rockies, and Valley; whereas, longitude was important in the Cascades and Plateau. Streams that drain the western side of the Cascades, which runs north–south, are fed by a combination of glaciers, snowfall, and substantial rainfall along with a significant change in elevation (61–4,300 m). The Plateau is east of the Cascades and is bisected by the Columbia River. Streams within this ecoregion range from coldwater streams that drain the Cascades to an arid, highly fragmented stream network that consists of natural spring-fed streams to highly altered agricultural streams (Munn et al. 2002). The fact that the Cascades and Plateau ecoregions had the greatest variation in canopy cover, elevation, and water temperature (Table 3) indicates that the importance of longitude in the NMS models is a reflection of more widely varied stream conditions going from west to east in both ecoregions regardless of distance.

Potapova and Charles (2002) examined diatoms from across the USA and determined that about one third of the total variance was due to geographical factors that were not correlated with environmental factors. Geographic factors influence species distribution due to distance, transportation, and physical barriers (Potapova and Charles 2002; Kristiansen 1996). The importance of geographic location also has been noted in stream-fish assemblages. Van Sickle and Hughes (2000) and McCormick et al. (2000) reported that fish assemblages at sites were more similar the closer the sites were, and less similar the greater

the distance between. Van Sickle and Hughes (2000) went on to add that geographic location was in fact more important than the larger-scale landscape classifications, as in ecoregions. The use of biogeographic location in the analyses of invertebrates is somewhat limited. Corkum (1989) reported that latitude and distance from the Pacific Ocean were important factors in Pacific Northwest invertebrate assemblages. Latitudinal gradients are more well established as a controlling factor for benthic invertebrate assemblages in coastal-marine waters, with Engle and Summers (1999) reporting that local distributions were determined by natural habitat variables, larger geographic distributions were influenced more by water temperature, which is related to latitude.

NMS models and taxonomic resolution

While our study did not directly assess taxonomic resolution, the merger of datasets from multiple sources required that taxonomic levels be normalized between datasets. The importance of taxonomic resolution has been an important topic of discussion in bioassessment studies due to its potential influence on data interpretation (Resh and McElravy 1993). While our study was able to maintain between 62% to 72% at the genus level and 23% to 36% at the family level, we had to aggregate chironomids to the family level. Waite et al. (2004) reported that while family-level taxonomy is sufficient for many bioassessment studies, the use of genus-level taxonomy is preferred when investigating finer level responses as in natural history and indicator species. They also reported that aggregating chironomids to the family level greatly reduced their ability to distinguish among many potential impacts (Waite et al. 2004). Therefore, our inability to utilize genus level chironomid data likely influenced our NMS models, particularly in the fine-grain dominated ecoregions like the Valley and Plains.

Implications for bioassessment and monitoring

Findings from this study illustrate several important considerations that must be addressed in order to design bioassessment studies. These include biogeography, taxonomic resolution, and the role of environmental gradients. As studies become larger in spatial scale, it may be important to consider whether biogeography will play a role in influencing results, particularly if taxonomic based information is used in analyses. It also is critical to consider the taxonomic resolution. Many benthic ecologists agree that lower level taxonomic resolution provides more information; however, it may be that this is most critical in systems that are dominated by fine-grain substrate and therefore chironomid assemblages. Last, if the purpose of a study is to address a specific environmental condition, i.e. nutrients, then it is critical that the study design address the implications of other environmental factors that can mask the specific variables of interest. For example, in our study we found that substrate was clearly a major factor in controlling benthic invertebrate assemblages and may in fact have masked the influence of other variables. Therefore if substrate is not the variable of interest it is important to sample invertebrates from as similar habitat as is possible to minimize its effect on results.

## References

Biggs, G. J. F. (1995). The contribution of disturbance, catchment geology and landuse to the habitat template of periphyton in stream ecosystems. *Freshwater Biology, 33*, 419–438.

Black, R. W., Munn, M. D., & Plotnikoff, R. W. (2004). Using macroinvertebrates to identify land-cover optima at multiple scales in the Pacific Northwest, USA. *Journal of the North American Benthological Society, 23*, 340–362.

Carlise, D. M., & Clements, W. H. (2003). Growth and secondary production of aquatic insects along a gradient of Zn contamination in Rocky Mountain streams. *Journal of the North American Benthological Society, 22*, 582–597.

Clarke, K. L., & Gorley, R. N. (2006). *PRIMER v6: user manual/tutorial. PRIMER-E Ltd*. United Kingdom: Plymouth.

Corkum, L. D. (1989). Patterns of benthic invertebrate assemblages in rivers of northwestern North America. *Freshwater Biology, 21*, 191–205.

Cuffney, T. F., Meador, M. R., Porter, S. D., & Gurtz, M. E. (2000). Responses of physical, chemical, and biological indicators of water quality to a gradient of agricultural land use in the Yakima River Basin, Washington. *Environmental Monitoring and Assessment, 64*, 259–270.

Culp, J. M., Walde, S. J., & Davis, R. W. (1983). Relative importance of substrate particle size and detritus to stream benthic macroinvertebrate microdistribution (Carnation Creek, British Columbia). *Canadian Journal of Fisheries and Aquatic Sciences, 40*, 1568–1574.

Engle, V. D., & Summers, J. K. (1999). Latitudinal gradients in benthic community composition in Western Atlantic estuaries. *Journal of Biogeography, 26*, 1007–1023.

Friberg, N., Milner, A. M., Svendsen, L. M., Lindegaard, C., & Larsen, S. E. (2001). Macroinvertebrate stream communities along regional and physico-chemical gradients in Western Greenland. *Freshwater Biology, 46*, 1753–1764.

Griffith, M. B., Husby, P., Hall, R. K., Kaufmann, P. R., & Hill, B. H. (2003). Analysis of macroinvertebrate assemblages in relation to environmental gradients among lotic habitats of California's Central Valley. *Environmental Monitoring and Assessment, 82*, 281–309.

Griffith, M. B., Kaufmann, P. R., Herlihy, A. T., & Hill, B. H. (2001). Analysis of macroinvertebrate assemblages in relation to environmental gradients in Rocky Mountain streams. *Ecological Applications, 11*, 489–505.

Hawkins, C. P., & Norris, R. H. (Eds.) (2000). Landscape classifications: Aquatic biota and bioassessments. *Journal of the North American Benthological Society, 19*, 367–556.

Hawkins, C. P., Norris, R. H., Gerritsen, J., Hughes, R. M., Jackson, S. K., Johnson, R. K., et al. (2000). Evaluation of the use of landscape classifications for the prediction of freshwater biota: Synthesis and recommendations. *Journal of the North American Benthological Society, 19*, 541–556.

Hering, D., Johnson, R. K., Kramm, S., Schmutz, S., Szoszkiewick, K., & Verdonschot, P. F. M. (2006). Assessment of European streams with diatoms, macrophytes, macroinvertebrates and fish: A comparative metric-based analysis of organism response to stress. *Freshwater Biology, 51*, 1757–1785.

Herlihy, A. T., Larsen, D. P., Paulsen, S. G., Urquhart, N. S., & Rosenbaum, B. J. (2000). Designing a spatially balanced, randomized site selection process for regional stream surveys: The EMAP Mid-Atlantic pilot study. *Environmental Monitoring and Assessment, 63*, 95–113.

Karr, J. R., & Chu, E. W. (1997). *Biological monitoring and assessment: Using multimetric indexes effectively. EPA/235-R97-001, U.S.* University of Washington, Seattle: Environmental Protection Agency.

Kaufmann, P. R., Levine, P., Robison, E. G., Seeliger, C. & Peck, D. V. (1999). Quantifying physical habitat in wadeable streams. EPA/620/R-99/003. Environmental Monitoring and Assessment Program, U.S. Environmental Protection Agency, Research Triangle Park, N.C. 61 p.

Klemm, D. J., & Lazorchak, J. M. (Eds.) (1994). *Environmental monitoring and assessment program— surface waters and Region 3 regional environmental monitoring and assessment program. 1994. Pilot field laboratory methods for streams. EPA/620/R-94/004. U.S.* Washington DC: Environmental Protection Agency.

Kristiansen, J. (1996). Dispersal of freshwater algae: A review. *Hydrobiologia, 336*, 151–157.

Lancaster, J., & Hildrew, A. G. (1993). Flow refugia and the microdistribution of lotic macroinvertebrates. *Journal of the North American Benthological Society, 12*, 385–393.

Lancaster, J., Hildrew, A. G., & Townsend, C. R. (1990). Stream flow and predation effects on the spatial dynamics of benthic invertebrates. *Hydrobiologia, 203*, 177–190.

McCormick, F. H., Peck, D. V., & Larsen, D. P. (2000). Comparison of geographic classification schemes for Mid-Atlantic stream fish assemblages. *Journal of the North American Benthological Society, 19*, 385–404.

McCune, B., Grace, J. B., & Urban, D. L. (2002). *Analysis of ecological communities*. Gleneden Beach, Oregon, USA: MjM Software Design.

McCune, B., & Mefford, M. J. (1999). *PC-ORD. Multivariate analysis of ecological data. Version 4.0.* Gleneden Beach, Oregon, USA: MjM Software Design.

Meador, M. R., Hupp, C. R., Cuffney, T. F., & Gurtz, M. E. (1993). Methods for characterizing stream habitat as part of the National Water-Quality Assessment Program. U.S. Geological Survey Open-File Report 93-408.

Moulton, S. R. II, Carter, J. L., Grotheer, S. A., Cuffney, T. F., & Short, T. M. (2000). Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Processing, taxonomy, and quality control of benthic macroinvertebrate samples: U.S. Geological Survey Open-File Report 00-212, 49 p.

Munn, M. D., Black, R. W., & Gruber, S. J. (2002). Response of benthic algae to environmental gradients in an agriculturally dominated landscape. *Journal of North American Benthological Society, 21*, 221–237.

Omernik, J. M. (1987). Map supplement: Ecoregions of the conterminous United States. Scale 1:750,000. *Annals of the Association of American Geographers, 77*, 118–125.

Pan, Y., Stevensen, R. J., Hill, B. H., & Herlihy, A. T. (2000). Ecoregions and benthic diatom assemblages in Mid-Atlantic Highlands streams, USA. *Journal of North American Benthological Society, 19*, 518–540.

Porter, S. D., Mueller, D. K., Spahr, N. E., Munn, M. D., & Dubrovsky, N. M. (2008). Efficacy of algal metrics for assessing nutrient and organic enrichment in flowing waters. *Freshwater Biology, 53*, 1036–1054.

Potapova, M. G., & Charles, D. F. (2002). Benthic diatoms in USA rivers: Distributions along spatial and environmental gradient. *Journal of Biogeography, 29*, 167–187.

Rempel, L. L., Richardson, J. S., & Healey, M. C. (2000). Macroinvertebrate community structure along gradients of hydraulic and sedimentary conditions in a large gravel-bed river. *Freshwater Biology, 45*, 57–73.

Resh, V. C., & McElravy, E. P. (1993). Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In D. M. Rosenberg, & V. H. Resh (Eds.), *Freshwater biomonitoring and benthic macroinvertebrates* (pp. 159–194). New York: Chapman and Hall.

Richards, C., Host, G. E., & Author, J. W. (1993). Identification of predominant environmental factors structuring stream macroinvertebrate communities within a large agricultural catchment. *Freshwater Biology, 29*, 285–294.

Rosenberg, D. M., & Resh, V. H. (Eds.) (1993). *Freshwater biomonitoring and benthic macroinvertebrates*. New York: Chapman and Hall.

Stevenson, R. J. (1997). Scale-dependent determinants and consequences of benthic algal heterogeneity. *Journal of the North American Benthological Society, 16*, 248–262.

Sweeney, B. W., & Vannote, R. L. (1984). Influence of food quality and temperature on life history characteristics of the parthenogenetic mayfly, Cloeon triangulifer. *Freshwater Biology, 14*, 621–630.

Tate, C. M., & Heiny, J. S. (1995). The ordination of benthic invertebrate communities in the South Platte River Basin in relation to environmental factors. *Freshwater Biology, 33*, 439–454.

Van Sickle, J., & Hughes, R. M. (2000). Classification strengths of ecoregions, catchments, and geographic clusters for aquatic vertebrates in Oregon. *Journal of the North American Benthological Society, 19*, 370–384.

Waite, I. R., Herlihy, A. T., Larsen, D. P., & Klemm, D. L. (2000). Comparing strengths of geographic and nongeographic classifications of stream benthic macroinvertebrates in the Mid-Atlantic Highlands, USA. *Journal of the North American Benthological Society, 19*, 429–441.

Waite, I. R., Herlihy, A. T., Larsen, D. P., Urquhart, N. S., & Klemm, D. J. (2004). The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: An example from the Mid-Atlantic Highlands, U.S.A. *Freshwater Biology, 49*, 474–489.

Woodcock, T. S., & Huryn, A. D. (2007). The response of macroinvertebrate production to a pollution gradient in a headwater stream. *Freshwater Biology, 52*, 177–196.