

Quantifying Variability in Four U.S. Streams Using a Long-Term Dataset: Patterns in Biotic Endpoints

Camille A. Flinders¹ · Douglas B. McLaughlin² · Renee L. Ragsdale¹

Received: 20 October 2014 / Accepted: 15 April 2015 / Published online: 1 May 2015
© Springer Science+Business Media New York 2015

Abstract Effective water resources assessment and management requires quantitative information on the variability of ambient and biological conditions in aquatic communities. Although it is understood that natural systems are variable, robust estimates of long-term variation in community-based structure and function metrics are rare in U.S. waters. We used a multi-year, seasonally sampled dataset from multiple sites ($n = 5\text{--}6$) in four streams (Codorus Creek, PA; Leaf River, MS; McKenzie and Willamette Rivers, OR) to examine spatial and temporal variation in periphyton chlorophyll *a*, and fish and macroinvertebrate metrics commonly used in bioassessment programs. Within-site variation of macroinvertebrate metrics and benthic chlorophyll *a* concentration showed coefficient of variation ranging from 16 to 136 %. Scale-specific variability patterns (stream-wide, season, site, and site-season patterns) in standardized biotic endpoints showed that within-site variability patterns extended across sites with variability greatest in chlorophyll *a* and lowest in Hilsenhoff's Biotic Index. Across streams, variance components models showed that variance attributed to the interaction of space and time and sample variance accounted for the majority of variation in macroinvertebrate metrics and chlorophyll *a*, while most variation in fish metrics was attributed to sample variance. Clear temporal patterns in measured endpoints were rare and not specific to any one

stream or assemblage, while apparent shifts in metric variability related to point source discharges were seen only in McKenzie River macroinvertebrate metrics in the fall. Results from this study demonstrate the need to consider and understand spatial, seasonal, and longer term variability in the development of bioassessment programs and subsequent decisions.

Keywords Variability · Long term · Fish · Macroinvertebrates · Chlorophyll *a* · Bioassessment · Metrics

Introduction

Aquatic ecologists accept that the abundance and distribution of stream biota can be highly variable, and that this variability occurs both spatially and temporally. This heterogeneity is mediated in large part by inherent variability in physical habitats, which can vary considerably from fine (millimeters to meters) to broad scales (tens of meters to kilometers). Variability is also influenced by temporal variation in ecosystem drivers such as flow, temperature, and water quality characteristics that can occur seasonally (i.e., within-year variation) and over many years (i.e., annual variation). Seasonal changes in stream environments that occur to varying degrees depending on region and local climate have been studied (Linke et al. 1999; Johnson et al. 2012). However, longer term variability that can encompass extreme events such as droughts and flooding is less understood.

Structure and function metrics as descriptors of species–environment relationships are the cornerstone to many programs assessing stream condition (Barbour et al. 1999). One of the key objectives of bioassessment programs developed

✉ Camille A. Flinders
cflinders@ncasi.org

¹ National Council for Air and Stream Improvement,
P.O. Box 1259, Anacortes, WA 98221, USA

² National Council for Air and Stream Improvement, Western
Michigan University, A-114 Parkview Campus,
Mail Stop 5436, Kalamazoo, MI 49008, USA

in the United States, Europe, and Australia is to evaluate the presence and extent of anthropogenic impacts (Simpson and Norris 2000; United States Environmental Protection Agency 2002). Fish and benthic macroinvertebrate community metrics are more widely used than periphyton-based assessment (Association of Clean Water Administrators 2012), although assessment of periphyton biomass in terms of chlorophyll *a* (chl *a*) and algal-based metrics have been developed (Cao et al. 2007; Stevenson et al. 2008). Site condition is typically evaluated using individual metrics, Indices of Biotic Integrity (IBI), and evaluation of the observed versus expected (O/E) taxa in the context of biological response along a gradient of environmental conditions ranging from minimally impacted to severely degraded. Although assessments are often based on a single sample collected from a stream segment (Carter and Resh 2001), the underlying assumption of these programs is that samples generated are representative of site condition, and that natural variability can be distinguished from the effect being assessed or accounted for by sampling protocols. The performance of ecological assessments is strongly linked to the accuracy with which freshwater environments are characterized, with site-specific modeling approaches suggested as a way to improve both accuracy and precision of community predictions (Hawkins et al. 2010).

Although regulatory bioassessment, criteria, and guidelines in many regions in the U.S. and elsewhere rely on community-based endpoints, the long-term variability of these endpoints in the context of variation at different spatial scales is relatively understudied. This may be due to the paucity of long-term datasets with which to make these assessments (Jackson and Füreder 2006). However, understanding variability of assessment endpoints over time and how these patterns might change across spatial scales can result in more precise assessment of stream condition, establish regulatory criteria with more meaningful targets, and increase the effectiveness of management decisions (e.g., setting total maximum daily load (TMDL) limits or developing remediation activities).

Long-term datasets provide a baseline framework with which to evaluate change, and are essential to understanding and interpreting rare events, ecological processes operating on long time frames, and highly variable systems (Franklin 1988; Lindenmayer and Likens 2009). Findings from long-term studies have shown evidence of delayed responses and revealed recurring temporal patterns and surprise outcomes resulting in shifts in understanding of lotic systems and new avenues for research (Lindenmayer et al. 2010; Dodds et al. 2012). With this in mind, a long-term study was initiated in 1998 in four U.S. streams to evaluate abiotic and biotic changes related to inputs of pulp and paper mill effluent (Hall et al. 2009a). The Long-Term Receiving Water Study (LTRWS) uses an integrative approach that examines both

environmental characteristics (i.e., water quality measurements and physical habitat assessments) and community-level assessments of fish, benthic macroinvertebrates, and periphyton. Although the study is designed specifically to address questions related to pulp and paper effluent exposure, expansive spatial and temporal scales were chosen to differentiate point source stressor responses from variation that occurs naturally or due to other anthropogenic factors within sites and over a stream continuum, and to evaluate patterns in the context of seasonal and long-term annual variability.

Although efforts have been made to address questions related to spatial variability in bioassessment endpoints (e.g., Rabeni et al. 1999; Gregg and Stednick 2000), relatively few studies address spatial and seasonal variability in bioassessment metrics using long-term data (e.g., Mazor et al. 2009). The dataset resulting from our ongoing study is uncommon in that it is comprised of fine-scale replicate measurements, multiple assemblage types, and seasonal collections from multiple sites in each of four streams. Additionally, the study spans a longer time period than is typical for many studies, and includes streams with different physical characteristics (e.g., size, temperature, dominant substrates) and drainage land uses. This provides a unique opportunity to evaluate variability in biological response across different spatial and temporal scales and across different ecoregions. Using data collected between 1998 and 2012, we sought to characterize spatial, temporal, and spatio-temporal variability (stream-wide, season, site, and site and season) in commonly used fish and macroinvertebrate assessment metrics and periphyton chl *a*; evaluate whether variability patterns are assemblage or endpoint dependent; and estimate the relative contributions of spatial (within and among sites) and temporal components (seasons and years) to endpoint variability. Finally, we determined the presence of temporal patterns in measured endpoints using polynomial regression and whether these were endpoint, assemblage, or stream dependent. Individuals collecting fish, macroinvertebrate, and periphyton samples, and performing taxonomic identifications and chl *a* measurements were consistent throughout the study. As such, variance estimates attributable to spatial and temporal differences are more robust than might otherwise be found in long-term datasets. Results from this and other similar research can provide important information on variability that can help guide management decisions.

Methods

Study Sites

The study was initiated in Codorus Creek (York County, PA), the McKenzie River (Lane County, OR), and the

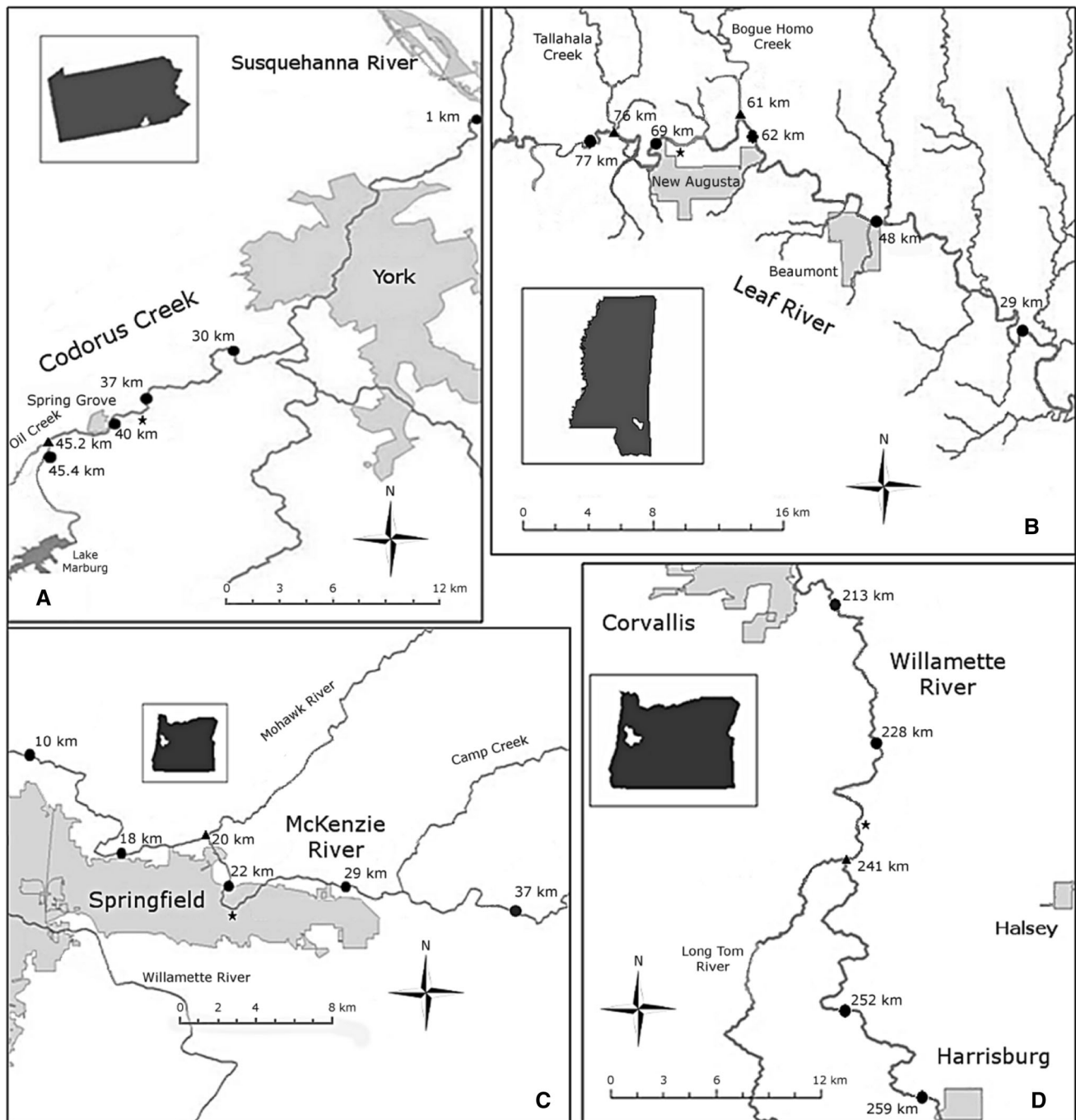


Fig. 1 Biota sampling locations **a** Codorus Creek, **b** the Leaf River, **c** the McKenzie River, and **d** the Willamette River. Site numbers are USGS stream kilometer distance to the stream confluence

Willamette River (Benton and Lane Counties, OR) in 1998, and in the Leaf River (Forrest and Perry Counties, MS) in 1999 (Fig. 1). The Leaf River is a warm-water stream dominated by sand substrates (>90 % sand at all sites), whereas the remaining three streams are cold water and consist predominantly of gravel-pebble (2–64 mm diameter) and cobble (64–256 mm diameter) substrates. Temperature patterns in Codorus Creek are influenced by

cold-water discharge from an upstream dam and by warm-water discharged from pulp and paper mill non-contact cooling water systems and treated effluent. In addition to differences in substrate composition and temperature regimes, streams differed in size and are located in three ecoregions with differing climate, precipitation patterns, and underlying geology, making study results broadly applicable to other systems. In all streams, biological

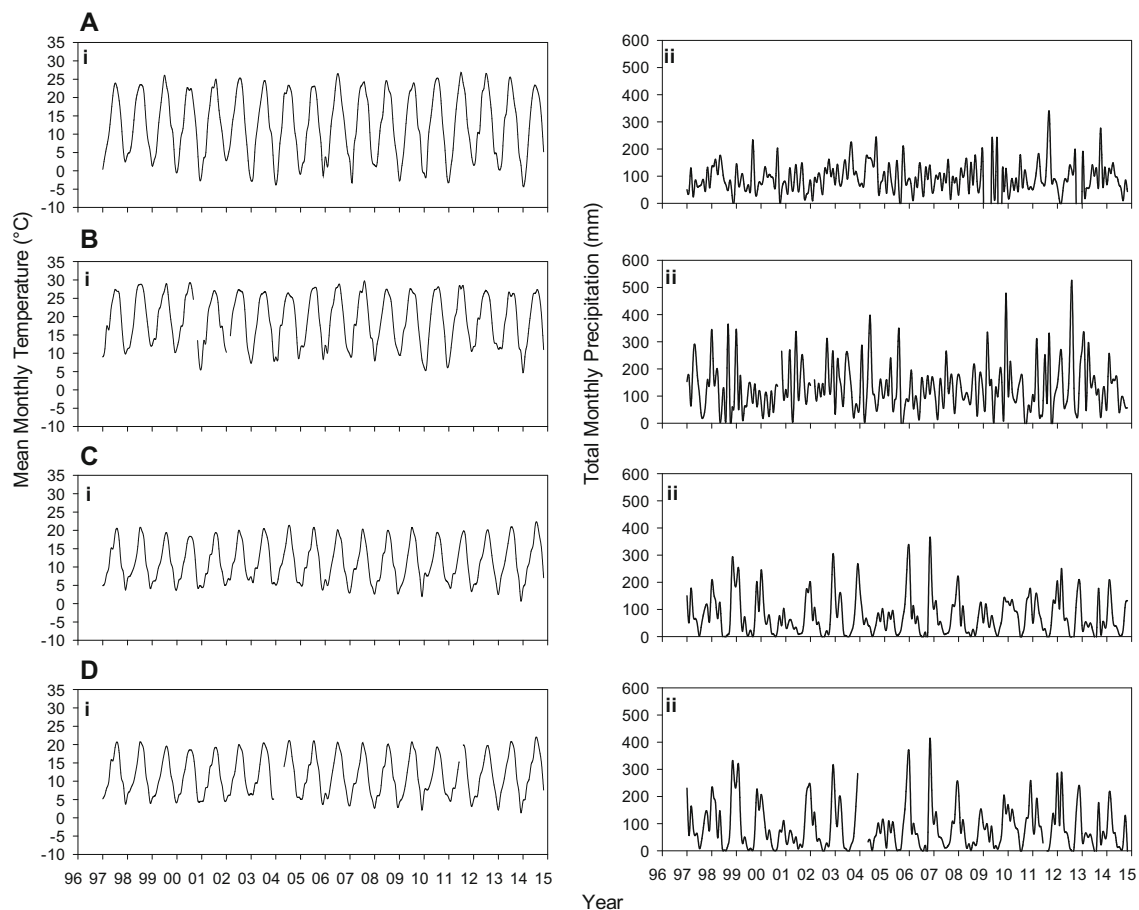


Fig. 2 Patterns in (i) mean monthly temperature and (ii) total monthly precipitation from the National Atmospheric and Oceanic Administration's National Climatic Data Center (<http://www.ncdc.noaa.gov/cdo-web/search?datasetid=GHCNDMS>) for **a** York, PA (GHCND: USC00369933),

b Hattiesburg, MS (GHCND: USC00223887), **c** Corvallis, OR (GHCND: USC00351862), and **d** Eugene, OR (GHCND: USW00024221) representing climate conditions between 1997 and 2015 at Codorus Creek, the Leaf River, the McKenzie River, and Willamette River, respectively

sampling sites were selected to represent similar conditions in terms of substrate, current velocity, and depth. In the larger Leaf, McKenzie, and Willamette rivers, all sites had an open canopy cover with minimal to no shading. Codorus Creek had a forested canopy composed largely of deciduous trees. Canopy cover was mostly consistent across sites, but changes in riparian habitat and stream width were unavoidable with two sites differing from others in terms of channelization and riparian cover (15 km) and greater stream width and substrate size (1 km). With the exception of Codorus Creek, accessibility dictated that water quality sampling locations differ slightly from biological sampling locations. Mean monthly temperature in the vicinity of each study stream followed predictable seasonal trends, and total monthly precipitation patterns showed few significant events that affected flows during the study period (Fig. 2). Drought conditions between 1997 and 2002 affected Leaf River stream flows (NOAA 2002) early in the study, and a storm surge associated with Hurricane Katrina

in August 2005 resulted in low dissolved oxygen that likely affected biotic assemblages more so than elevated storm-related streamflow (Schaefer et al. 2006). Snowmelt from the Cascade Range that contributes significantly to flows in the McKenzie and Willamette Rivers were greater in 1999 and 2008 than in other years (Fig. 3). Stream characteristics and study site locations are described in Table 1 and, in more detail, elsewhere (Hall et al. 2009b).

Field Sampling

Fish, macroinvertebrate, and periphyton samples were collected from sites in each stream during the spring and/or fall (Table 2). Collections during both seasons were not always feasible due to rainfall-related high flows and logistical challenges. In Codorus Creek, most spring samples were collected during March and April, while most fall samples were collected in late September and October. In the McKenzie and Willamette Rivers, spring and fall

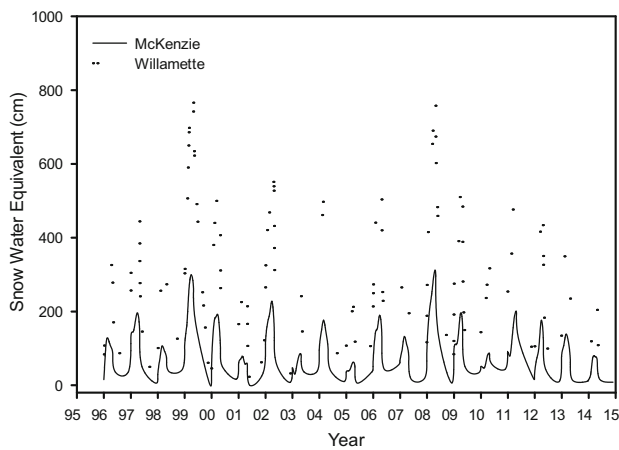


Fig. 3 Total monthly snow water equivalent for the McKenzie and Willamette Rivers between 1996 and 2015 obtained from the US Department of Agriculture Natural Resources Conservation Service data (<http://www3.wcc.nrcs.usda.gov>)

samples were typically collected in May and September, respectively. In the Leaf River, macroinvertebrate and chl *a* samples were most often collected in May and October.

Biotic communities were sampled using standard protocols (Barbour et al. 1999). The fish community was sampled from both banks in Codorus Creek, and small-bodied, near-shore fish from a single bank in the McKenzie and Willamette Rivers using a backpack electrofisher (battery-powered Smith-Root LR-24). Lack of habitat complexity made backpack electrofishing sampling for small-bodied fish ineffective in the Leaf River. A sampling effort of 30 min was targeted for each site with typical reach lengths of ~100 m in Codorus Creek and ~400 m in the McKenzie and Willamette Rivers. Sampling was

always conducted at depths <80 cm, and electrofisher settings (volts, amps, watts, frequency, and duty cycle) were dependent on water conditions (conductivity and temperature). Two netters followed the electrofisher upstream along a transect such that as much of the wadeable available habitat was sampled as was practical. Collected fish were placed into aerated buckets containing river water and monitored. For all streams and sites, fish were sorted in the field, identified to species according to Nelson et al. (2004), and their length and weight recorded before being returned unharmed to stream.

Replicate macroinvertebrate community samples ($n = 5$) were collected from riffle habitat in Codorus Creek and the McKenzie and Willamette Rivers using a Hess sampler (0.086 m², 243- μ m mesh size). Efforts were made to sample in the same riffles during each sampling event, but flow-related shifts in riffle location were unavoidable during some years. Sampling locations within riffles were randomly selected by blindly tossing a weighted marker into the stream and setting the sampler immediately upstream of the marker. A sampling effort in which sediment was disturbed to a depth of 10 cm for 2 min was consistent across streams. Individual Hess samples were kept separate and preserved with 10 % buffered neutral formalin. All samples were collected within approximately 50 m² at a site, and samples from sites within each stream were typically collected within a 2-day period. Because unstable sand substrates dominate the Leaf River, Hester–Dendy multiplate samplers (HD samplers) (Hester and Dendy 1962) were used to assess macroinvertebrate communities. Although we did not expect variability between Hess and HD samplers to be readily comparable (Barton and Metcalfe-Smith 1992), data from HD samplers are useful for

Table 1 Characteristics of study streams and drainages

Characteristic	Codorus Cr.	Leaf R.	McKenzie R.	Willamette R.
Biological sampling sites (<i>n</i>)	6	5	5	6
Reach length (km)	45	50	28	48
Mean reach width \pm SD (m)	18.4 \pm 8.6	64.1 \pm 23.4	64.0 \pm 16.7	93.4 \pm 20.1
Mean discharge (m ³ /s) ^a	2.6	87.4	117.0	304.0
Drainage area (km ²)	712	2802	3367	29,728
Land use (%) ^b				
Forest	27.0	64.0	87.4	76.5
Agriculture	66.6	22.3	2.4	12.8
Urban + industrial	0.9	0.5	0.1	0.6
Water + wetland	2.2	9.3	0.7	1.7
Pulp and paper mill discharge location (km)	39.7	68.7	23.0	237
Major tributaries location (km)	Oil Cr (45)	Tallahala Cr (76) Bogue Homo Cr (62)	Mohawk R. (21)	Long Tom R (240)

^a Discharge from 1998 to 2013 measured at USGS gauging stations 01574500, 02474560, 14162500/14165500, 14166000

^b Wicczorek and LaMotte (2010)

Table 2 Number of collections with respect to stream, endpoint, and season

Stream	Site	Number of collections					
		Fish		Invertebrates		Chl <i>a</i>	
		Spring	Fall	Spring	Fall	Spring	Fall
Codorus Creek	45	8	13	6	12	6	12
	40 ^a	7	12	6	12	6	12
	38	7	13	6	12	6	12
	30	7	13	6	12	6	11
	15	7	11	6	11	6	11
	1	7	10	5	11	5	11
Leaf River	78			9	13	5	9
	72			9	13	5	9
	67			8	12	7	9
	51			9	13	5	9
	46			9	12	5	9
McKenzie River	36	10	11	12	8	12	8
	30	11	11	12	10	12	11
	21			12	10	12	11
	15 ^b	10	11	12	10	12	11
	10	10	11	12	9	11	10
Willamette River	257	11	10	10	10	10	11
	251	9	10	10	10	10	11
	238	9	11	10	10	10	11
	235			9	10	10	10
	232 ^c	11	10	10	10	10	11
	217	9	8	10	10	10	11

At each collection event, $n = 5$ replicate macroinvertebrate and chl *a* samples collected from Codorus Creek and the McKenzie and Willamette Rivers, and $n = 3$ samples collected from the Leaf River

^a Corresponding fish site = 41

^b Corresponding fish site = 18

^c Corresponding fish site = 219

examining within-stream spatial and temporal variability. Three HD samplers (total area 0.089 m² per sampler) were deployed per site and attached by tethers to a float and anchoring block approximately 50 cm below the water surface. Samplers were allowed to colonize for 5–6 weeks prior to removal. Upon retrieval, algae were first removed from the surface plate for periphyton analysis (description below) before the entire HD sampler was preserved with 10 % neutral buffered formalin.

Periphyton was removed from natural cobble substrates ($n = 5$) in riffles in Codorus Creek and the McKenzie and Willamette Rivers using toothbrushes and scalpels (Biggs and Kilroy 1996). In the Leaf River, algae were removed from the top surface of HD samplers using blades and toothbrushes. The algal slurry was washed into separate graduated cylinders and the total volume recorded, homogenized, and a known volume filtered through a 47 mm Gelman A/E glass fiber filter (Gelman Sciences, Ann

Arbor, MI, USA). Quality assurance of the field subsampling method was ensured by filtering duplicate sub-samples from the same rock, with the relative percent difference between field duplicates always <10 %. Filters were wrapped in aluminum foil, kept on dry ice until returned to the laboratory, and stored frozen until analysis for chl *a*. Before 2006, substrate sample area of each rock was estimated by measuring the maximum length and width of each rock sampled to the nearest 1 mm using digital calipers (Dudley et al. 2001). In 2006 and later, substrate sampling area was determined by wrapping the sampled area in aluminum foil and measuring the area of foil scans using ImageJ (Rasband, W.S., ImageJ, U.S. National Institutes of Health, Bethesda, Maryland, USA, <http://imagej.nih.gov/ij/>, 1997–2014). The resulting area estimate (length–width) is positively related to rock surface area measurements derived from estimates using aluminum foil methods (e.g., Aloï 1990).

Laboratory and Data Analysis

Macroinvertebrates were removed from HD samplers in the laboratory. Macroinvertebrate taxa from all streams were identified to the lowest practical taxon (typically species) and enumerated, with full counts conducted for all samples by Benthic Aquatic Research Services (Lawrence, KS). Chironomidae individuals were identified to species using slide mounts of mouth parts. Taxonomy personnel were consistent over the course of the study, and the database reconciled with taxonomic updates when they occurred. Voucher samples of all taxa were retained for quality assurance and control purposes. Ambiguous taxa (e.g., early instars) were infrequent and identified to the lowest taxon possible (typically Genus) and adjusted for in metric calculations. Total chl *a* for most periphyton samples was determined using a Spectronic Genesys 8 Spectrophotometer (Thermo Scientific, Waltham, MA, USA). In some cases, when chl *a* was below the spectrophotometer detection limits, concentrations were determined using a Turner TD-700 Fluorometer (Turner BioSystems, Sunnyvale, CA, USA). Both analyses were conducted according to standard methods (American Public Health Association 2000). The chl *a*/m² on each rock was calculated by dividing the total sample chl *a* (mg) by the sample area of the rock (m²).

We translated fish and macroinvertebrate species abundance data into ecologically relevant community-based metrics representing measures of stream structure and function. Fish community data were standardized for sampling effort based on sampling time. Fish abundance and %intolerant taxa were calculated for each site and collection date. Tolerance information was assigned according to ecoregion and, where possible, stream-specific autecological information obtained from multiple sources (e.g., Simon 1991; Hall et al. 1996; see Flinders et al. 2009 for details). In the rare case in which a discrepancy on species tolerance was encountered, the more sensitive designation was assigned. For example, we followed Simon (1991) and Hall et al. (1996) in classifying banded killifish (*Fundulus diaphanus*) as intermediately tolerant, although others consider this a tolerant species (Barbour et al. 1999). Calculated macroinvertebrate metrics included Taxa richness, %dominant taxon, and %Ephemeroptera, Plecoptera, and Trichoptera abundance (% EPT). Hilsenhoff's Biotic Index (HBI) (Hilsenhoff 1987) was also calculated on the basis of autecological information from multiple sources and, where possible, assigned region-specific tolerance values (Hilsenhoff 1987, 1988; Idaho Department of Health and Welfare 1993; Merritt and Cummins 1996; Barbour et al. 1999; Maxted et al. 2000). These metrics were selected because they are among those most often measured as part of state bioassessment programs (USEPA 2002; ACWA 2012). For example,

invertebrate taxa richness is an endpoint in 42 of the 47 states in which macroinvertebrate structure and function metrics are used for bioassessment, including those where the study streams are located.

Replicate macroinvertebrate and periphyton samples collected from each site during each sampling event were examined in two ways. Endpoint values (i.e., macroinvertebrate metric value and chl *a* concentration) were determined for individual samples and within-site variation on each sampling date evaluated based on the replicate means and standard deviation (coefficient of variation, CV, calculated as (standard deviation/mean) * 100 %). Although CVs are commonly used as a measure of variability (Sandin and Johnson 2000; Mazor et al. 2009), they are sensitive to endpoint means (Sokal and Rohlf 1995). Because CVs are less broadly comparable across metrics, locations, and time periods, we also performed a variance components analysis for each stream to partition the variance associated with site, season, year, interaction terms, and sample variance (residual or error term). Variance component estimates are calculated using assumptions and mathematics associated with ANOVA mean and sums of squares. Thus, percent variability estimates are comparable across metrics and sources of variation, and applicable to other streams or time periods having similar patterns of variability around a metric's grand mean (Larsen et al. 2001). For macroinvertebrate metrics and chl *a*, for which we had within-site replicates during each sampling event, models included site, season, year, and interaction terms. Because fish were sampled from a single transect during each sampling event and thus, not replicated, variance estimates for fish metrics were calculated separately for the main and interaction of space and time with models for site and season and site and year. Variance attributed to inter- and intra-annual variability regardless of site was also estimated for the main (season, year) and interaction (season * year) terms. Because not all sites were sampled in all seasons and years, variance components were calculated using restricted maximum likelihood (REML) which eliminates the occurrence of negative variance component values (Larsen et al. 2001).

We were interested in temporal patterns in endpoint response and evaluated the presence of statistically significant monotonic or polynomial trends in measured bioassessment endpoints using least squares polynomial regression (i.e., linear, quadratic, and cubic models). Because unbalanced seasonal sampling and gaps in yearly collections may bias temporal trends (Legendre and Legendre 1998), we evaluated the presence of temporal patterns in bioassessment endpoints separately for spring and fall, as well for the full dataset at each site.

Because the primary objective of this study was to evaluate variability in typical stream assessment endpoints

Table 3 Stream-wide summary statistics for measured biological endpoints

Stream	Endpoint	Statistic								
		<i>n</i>	Mean	SD	CV (%)	Minimum	25 %ile	Medium	75 %ile	Maximum
Codorus Creek	Fish abundance	115	37.1	30.3	82	1.6	17.5	29.4	47.8	218.0
	%Intolerant fish	115	30.4	20.1	66	0.0	15.0	29.7	42.2	92.6
	Taxa richness	105	61.9	12.0	19	38.0	52.0	60.0	70.0	98.0
	HBI	105	5.3	0.5	10	4.2	4.9	5.4	5.8	6.6
	%EPT	105	19.5	15.0	77	0.5	6.0	17.1	33.5	61.3
	%Dominant taxa	105	31.1	13.8	45	10.9	20.4	27.1	37.6	78.6
Leaf River	Chl <i>a</i> (mg/m ²)	104	138.5	123.7	89	10.1	53.0	94.2	188.5	602.3
	Taxa richness	107	41.5	7.0	17	26.0	37.0	42.0	45.0	61.0
	HBI	107	5.1	0.5	9	4.2	4.7	5.0	5.3	6.5
	%EPT	107	57.0	17.2	30	16.4	45.2	60.3	70.2	86.1
	%Dominant taxa	107	29.4	12.7	43	11.2	20.9	25.4	34.6	72.4
	Chl <i>a</i> (mg/m ²)	72	19.0	18.1	95	0.9	9.9	14.0	22.5	122.6
McKenzie River	Fish abundance	85	118.4	57.8	49	16.8	75.0	107.6	143.3	281.5
	%Intolerant fish	85	64.6	24.6	38	8.8	47.7	64.9	86.5	99.4
	Taxa richness	107	76.6	12.3	16	41.0	6.0	76.0	87.0	103.0
	HBI	107	4.6	0.8	17	2.7	4.1	4.6	5.3	6.3
	%EPT	107	28.1	13.8	49	6.2	16.2	27.5	38.6	63.3
	%Dominant taxa	107	25.5	9.6	38	9.0	17.9	23.8	32.1	52.4
Willamette River	Chl <i>a</i> (mg/m ²)	110	39.3	30.7	78	6.6	16.2	28.6	52.0	157.9
	Fish abundance	98	158.2	86.1	54	0.3	96.4	142.7	201.3	405.4
	%Intolerant fish	98	35.7	21.1	59	3.1	18.6	30.4	50.3	100.0
	Taxa richness	119	63.9	12.4	19	31.0	56.0	63.0	74.0	89.0
	HBI	119	5.4	0.7	12	3.7	5.0	5.5	6.0	6.9
	%EPT	119	18.2	11.8	65	0.9	9.1	15.6	25.9	58.1
Willamette River	%Dominant taxa	119	32.6	14.6	45	13.2	20.3	29.5	43.2	74.8
	Chl <i>a</i> (mg/m ²)	125	139.7	89.1	64	13.4	80.3	118.2	172.5	463.3

(and not the condition reflected by endpoint values) and to compare variability across endpoints relative to river-wide median condition, endpoints were standardized by dividing the endpoint value at each site and sampling date by the river-wide endpoint median. This effectively eliminated scale while maintaining the distribution of the data to highlight variability and more easily compare patterns across different endpoints. It is important to point out that in percentage-based metrics (e.g., %EPT abundance, HBI), the data are constrained and variability reduced relative to endpoints based on raw data (e.g., taxa richness, chl *a* concentrations). Percentiles of the distributions of standardized endpoints were determined to characterize stream-specific variability patterns in biotic endpoints at four spatial and temporal scales: stream-wide patterns across sites and seasons (i.e., all data for each stream), season-specific patterns across sites, site-specific patterns across seasons, and site- and season-specific patterns. Temporal patterns and distribution of macroinvertebrate

endpoints and chl *a* patterns were quantified using values from pooled site replicates collected from each site and sampling event. All statistical analyses conducted using SYSTAT 12.02 (Systat Software, Inc., San Jose California USA) and R 2.10.0 (R Foundation for Statistical Computing).

Results

Stream-wide biological characteristics differed with endpoint, but most were not markedly different across streams (Table 3). Mean HBI and %dominant taxa were most similar across streams with similarity generally extending across other statistical measures (i.e., minimum values, median, etc.). Stream-wide taxa richness was similar in Codorus Creek and the McKenzie and Willamette Rivers but lower in the Leaf River. Measures of %EPT abundance and chl *a* showed the greatest difference across streams.

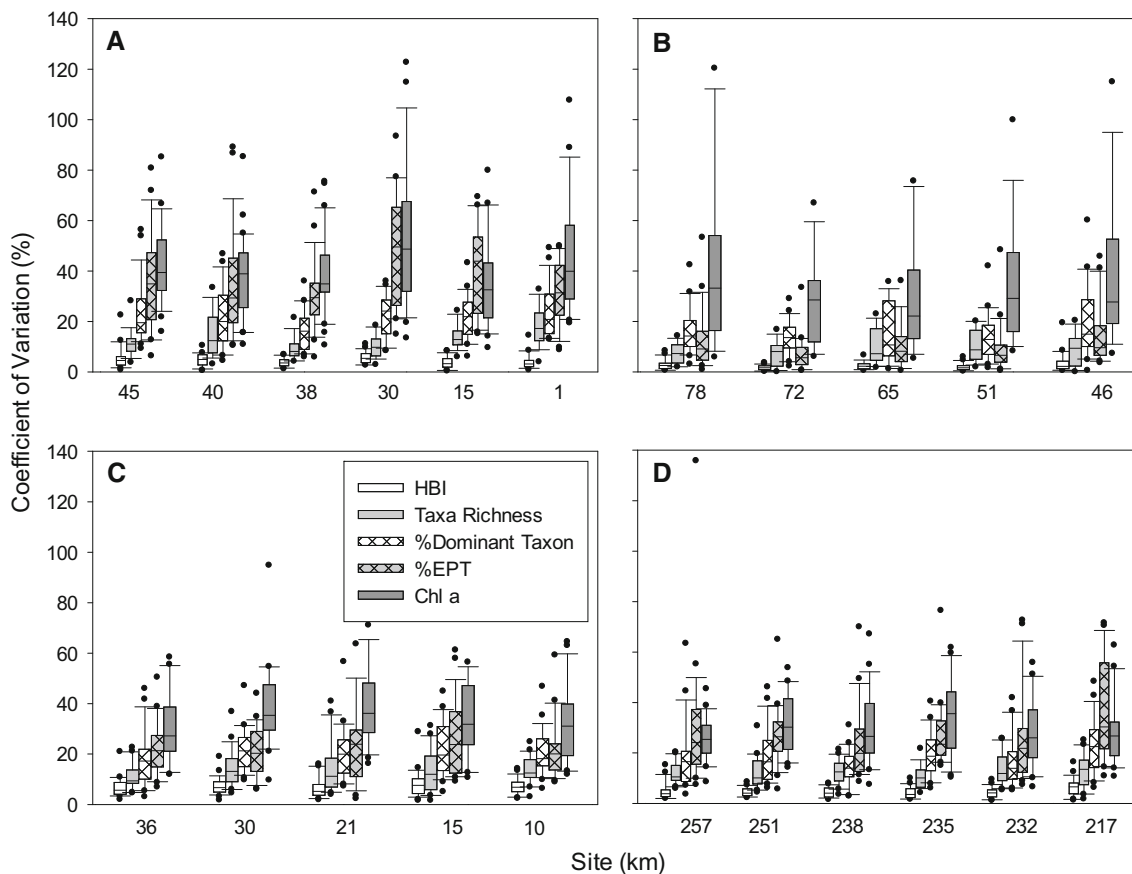


Fig. 4 Patterns of coefficient of variation of macroinvertebrate endpoints and chl *a* concentration calculated from within-site replicate samples collected from **a** Codorus Creek, **b** the Leaf River, **c** the McKenzie River, and **d** the Willamette River from 1998 to 2012

Overall, Codorus Creek and the Willamette River showed the greatest similarity in metric patterns across most summary measures, although consistency in mean values did not necessarily extend to other measures. For example, although mean chl *a* in Codorus Creek (138.5 mg/m²) and the Willamette River (139.7 mg/m²) was similar, variability and range of values differed.

Variability among replicates collected at a single site and sampling event for macroinvertebrate endpoints and chl *a* concentration as determined by CVs was endpoint, stream, and site specific (Fig. 4). At all sites and streams, within-site variability was the lowest in HBI and taxa richness. The median CV of HBI at each site was generally <5 %, with maximum CV almost always <15 %, while taxa richness had a median within-site CV typically between 7 and 12 %. Other macroinvertebrate endpoints showed considerably higher variability within sites, but this differed across streams and sites within streams. Within-site variation in %EPT abundance was the greatest at Codorus Creek sites (median CV = 35–50 %), and greater than that of %dominant taxon in all streams except the Leaf River. In all rivers, chl *a* concentration was the most

variable within a site with median CV values >25 % at most sites during the study.

Dataset distribution patterns in biological metrics differed across endpoints, streams, and the spatial and temporal filter applied to the dataset. Among the three streams for which data were available, the range of fish abundance was the greatest in Codorus Creek with values as high as nearly 8× the river-wide median value, compared to less than 3× the river-wide median in the McKenzie and Willamette Rivers (Fig. 5). Seasonal differences were the greatest in the McKenzie River with a greater number of fish collected in the fall than in the spring. Fish abundance was the most spatially variable in Codorus Creek, with site-specific seasonal differences inconsistent across sites. The %intolerant fish taxa were least variable in the McKenzie River, although all streams showed spatial and spatio-temporal differences in variability.

Among the macroinvertebrate metrics examined, HBI was the most consistent with most data falling between 0.8 and 1.2× the river-wide median values regardless of the stream or spatial/temporal filter applied (Fig. 6). Exceptions were seen when a seasonal filter was applied to

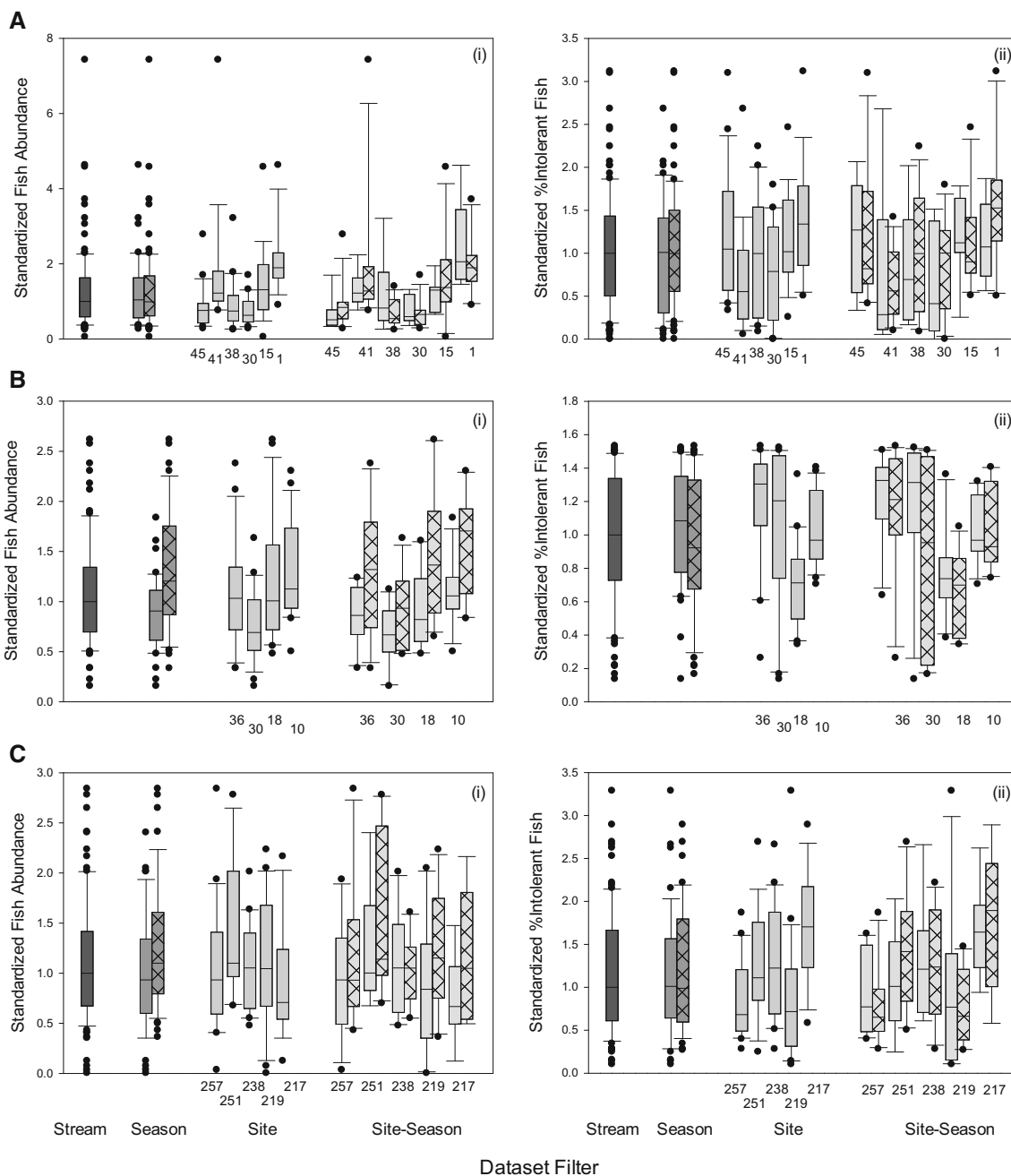


Fig. 5 Variability patterns of standardized (i) fish abundance and (ii) %intolerant fish taxa with respect to spatial and temporal dataset filters collected from **a** Codorus Creek, **b** the McKenzie River, and **c** the Willamette River from 1998 to 2012. Endpoints were standardized by dividing the endpoint value at each site and sampling

date by the river-wide endpoint median. In assessments of seasonal components, *unhatched bars* represent spring samples and *hatched bars* represent fall samples. Numbers associated with the site and site-season filters correspond to site numbers in Table 1

Codus Creek data, with HBI consistently lower in the fall than in the spring. There were greater seasonal differences in taxa richness in Codorus Creek and the Leaf River than in the McKenzie or Willamette rivers, although the range of values was generally consistent across streams (Fig. 7). Similarly, values for %dominant taxa were more variable in Codorus Creek and the Leaf River than in the McKenzie

and Willamette Rivers (Fig. 8). Percent dominant taxa were consistently lower in the fall in the Willamette River and at some sites in Codorus Creek. Among macroinvertebrate metrics examined, patterns in %EPT abundance were the least consistent across streams (Fig. 9). In Codorus Creek, %EPT ranged from 0.03 to 3.8× the river-wide median value and varied both spatially and

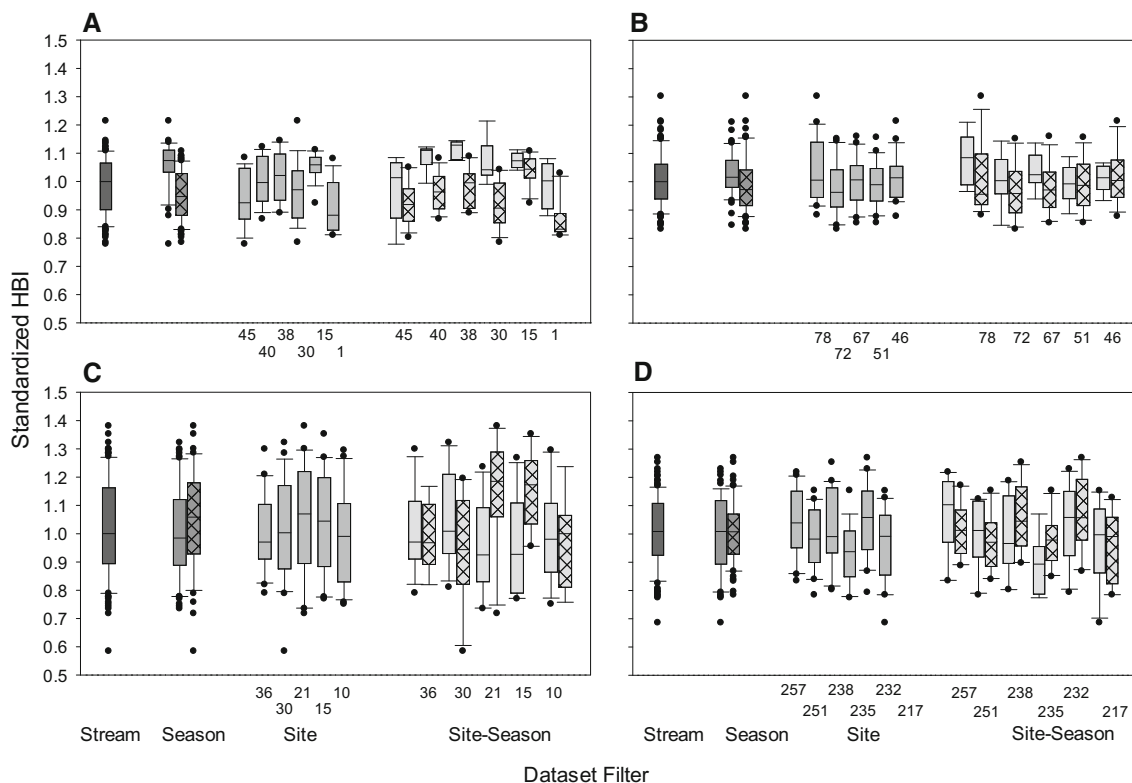


Fig. 6 Variability patterns of standardized Hilsenhoff’s Biotic Index (HBI) with respect to spatial and temporal dataset filters collected from **a** Codorus Creek, **b** the Leaf River, **c** the McKenzie River, and **d** the Willamette River from 1998 to 2012. Site condition calculation is based on pooled replicate samples collected on each sampling date. Endpoints were standardized by dividing the endpoint value at each

site and sampling date by the river-wide endpoint median. In assessments of seasonal components, *unhatched bars* represent spring samples, and *hatched bars* represent fall samples. *Numbers* associated with the site and site-season filters correspond to site numbers in Table 1

temporally with consistently greater values occurring in the fall. Patterns of %EPT in the Willamette River were similarly variable, although spatial and temporal differences were less pronounced.

Periphyton biomass in terms of chl *a* concentration was the most variable of the biological endpoints examined (Fig. 10). In Codorus Creek, chl *a* concentrations were lower and more consistent in the fall with this seasonal pattern seen at most sites when both a spatial and temporal filter was applied to the dataset. McKenzie and Willamette River chl *a* patterns generally showed greater consistency regardless of the spatial or temporal dataset filter, although chl *a* in the Willamette River was typically lower and less variable in the fall than in the spring. Excluding data above the 90th percentile, chl *a* concentrations in the Leaf River were relatively consistent regardless of the spatial and temporal filters applied.

Variance components analysis showed that spatial and temporal variability in macroinvertebrate metrics and chl *a* concentrations was stream and endpoint specific (Table 4). Variance attributed to spatial and seasonal components alone was largely limited to Codorus Creek

where variance across sites accounted for 9.5–20.3 % of variation in macroinvertebrate metrics and chl *a*, with seasonal variation accounting for 16.9–21.4 % of the variation for most endpoints. An exception was seen in %dominant taxa which showed little seasonal variation (<1 %). With the exception of taxa richness, variance attributed to annual variation was high in the Leaf River ranging from 21 to 43 %. In contrast, in the McKenzie and Willamette Rivers, variation attributed to site, season, and year was typically negligible or less than 10 %. In all streams, and the McKenzie and Willamette Rivers, in particular, the interaction of space and time accounted for the greatest amount of variability in macroinvertebrate metrics and chl *a*. That is, patterns in these endpoints varied differently at a given site, season, and/or year than at other sites and times. Variation associated with samples, represented by the error term, was always the lowest in the Leaf River and accounted for 19–44 % in the other streams.

In all streams and for both fish abundance and %intolerant fish, sample variance typically ranged from 60 to 100 % and almost always accounted for the majority of

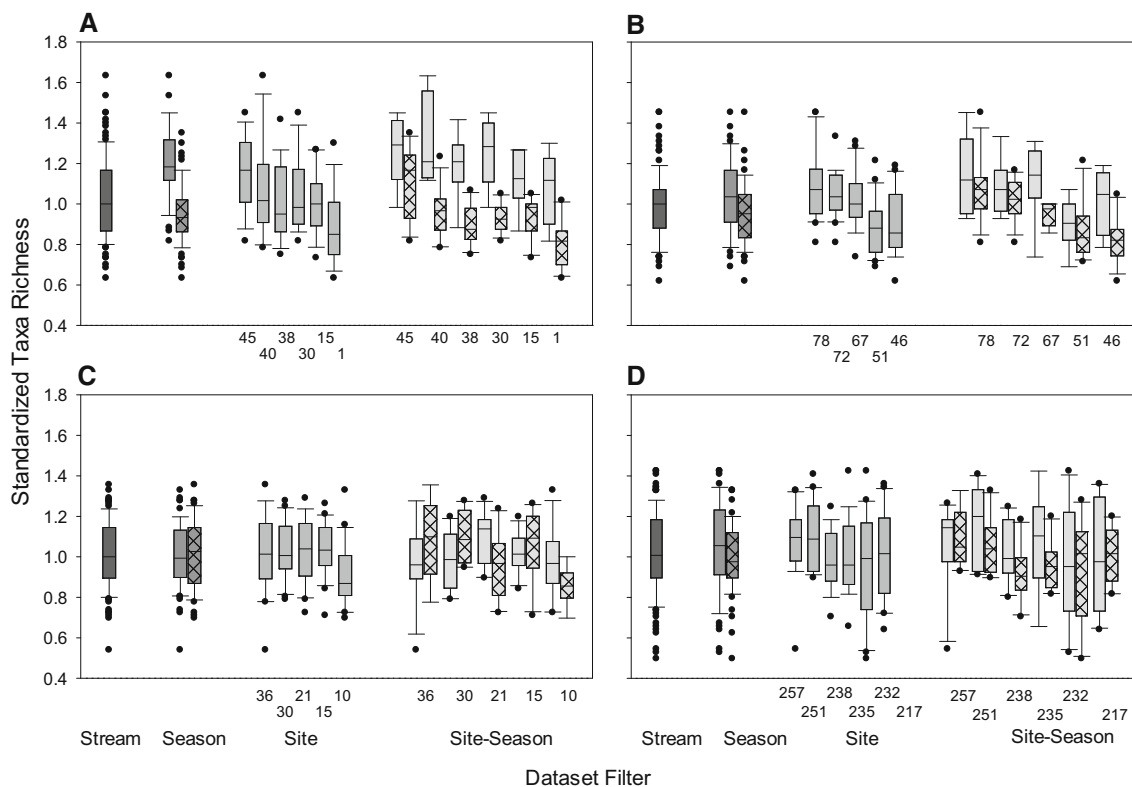


Fig. 7 Variability patterns of standardized invertebrate taxa richness with respect to spatial and temporal dataset filters collected from **a** Codorus Creek, **b** the Leaf River, **c** the McKenzie River, and **d** the Willamette River from 1998 to 2012. Site condition calculation is based on pooled replicate samples collected on each sampling date. Endpoints were standardized by dividing the endpoint value at each

site and sampling date by the river-wide endpoint median. In assessments of seasonal components, *unhatched bars* represent spring samples and *hatched bars* represent fall samples. Numbers associated with the site and site-season filters correspond to site numbers in Table 1

variance in analyses to assess spatial and temporal variance (in terms of season and year) and intra- and inter-annual variance (Table 5). An exception was seen in %intolerant taxa in the McKenzie River where the interaction of site and year accounted for 43.3 % of variability. Seasonal variation in both fish metrics was negligible in Codorus Creek and the Willamette River but accounted for nearly 26 % of variation in McKenzie River fish abundance. Intra-versus inter-annual variability was stream and metric dependent, with sample variance accounting for almost all of the variance in Codorus Creek fish metrics, and 60–87 % in the McKenzie and Willamette Rivers.

Temporal patterns of fish and macroinvertebrate metrics, and chl *a* assessed for each site using polynomial regression were inconsistent and generally weak regardless of the model (i.e., linear, quadratic, or cubic) or whether the model assessed full or seasonal datasets. Fewer than 7 % of models showed significant relationships between biotic patterns and sampling date, with some significant relationships based on datasets with low sample sizes. For example, a significant cubic relationship between chl *a* concentration and time was seen at Leaf River site 46 km in the spring, but this was based

on only five samples. Restricting our focus to endpoints and sites where sampling frequency was balanced between spring and fall in nearly all years (endpoints at most McKenzie and Willamette River sites) still showed few significant temporal relationships which were site and metric specific. For example, at Willamette River site 217 km, there was a significant negative linear relationship between macroinvertebrate richness in the fall, but not with the overall dataset, significant cubic relationships between sampling date and HBI (full dataset) and %EPT abundance (spring and full dataset), and no relationship between time and %dominant taxa (Fig. 11).

Discussion

It has long been recognized that lotic systems are spatially and temporally variable, and researchers have made considerable strides to define (Li and Reynolds 1995; Cooper et al. 1997) and explain patterns in biota (e.g., Cardinale et al. 2002; Brown 2003). Despite this, efforts to quantify spatial and seasonal variation of biological endpoints using

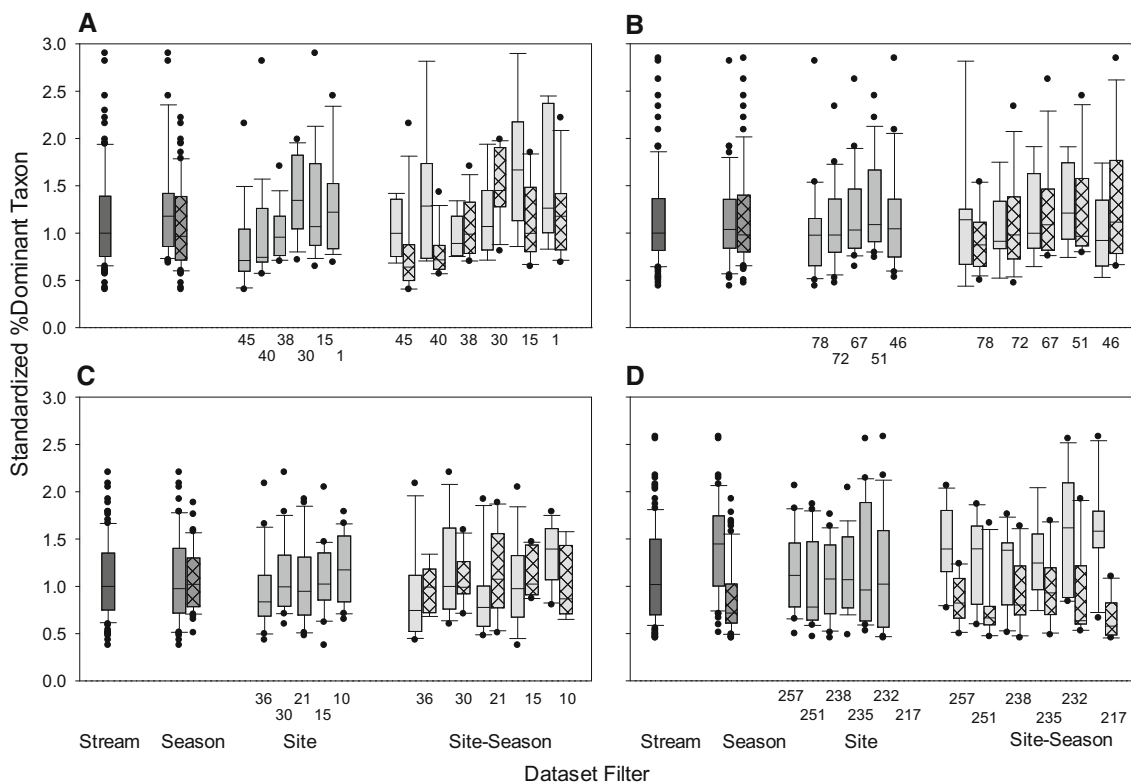


Fig. 8 Variability patterns of standardized %dominant taxa with respect to spatial and temporal dataset filters collected from **a** Codorus Creek, **b** the Leaf River, **c** the McKenzie River, and **d** the Willamette River from 1998 to 2012. Site condition calculation is based on pooled replicate samples collected on each sampling date. Endpoints

were standardized by dividing the endpoint value at each site and sampling date by the river-wide endpoint median. In assessments of seasonal components, *unhatched bars* represent spring samples, and *hatched bars* represent fall samples. *Numbers* associated with the site and site-season filters correspond to site numbers in Table 1

long-term datasets are limited. An understanding of how variable an endpoint is over time, and how this variability may fluctuate spatially within- or across sites and with seasonality, is valuable from a management perspective because it captures rare or long-term events that better enables naturally occurring responses to be separated from those due to anthropogenic influences. This knowledge then informs decision-making regarding remediation and criteria development.

Using a 13-year dataset from four streams, we found that there was considerable spatial and temporal variation in most fish, macroinvertebrate, and periphyton endpoints evaluated, and that this variation was endpoint and stream dependent. We found high within-site variation across replicate macroinvertebrate and chl *a* samples (CV = 16–136 %). This is consistent with those of other researchers who, in hierarchical scale studies, found that the greatest variation to occur at the sample level (e.g., Li et al. 2001; Robson et al. 2005). Patterns of small-scale variability in macroinvertebrate endpoints and periphyton biomass were consistent across streams and sites within streams. However, HBI was more consistent within a site (median CV = 1.7–7.6 %) compared

to other metrics (taxa richness, median CV = 7.3–17.3 %; %dominant taxon, median CV = 10.7–24.2 %; %EPT abundance, median CV = 5.6–49.6 %; and chl *a*, median CV = 22.2–39.9 %). Similar site- and endpoint-specific levels of variation over extended timeframes have been seen in other studies, with coefficient of variation at a site ranging from 21 to 302 % (Mazor et al. 2009). This highlights the need to for sampling protocols that minimize within-site sampling error, and some bioassessment program methods attempt to account for this through the collection of replicate or multi-habitat composite samples. For example, samples collected as part of Florida’s Stream Condition Index (SCI) consist of a composite of 20-D frame dip net sweeps of the most productive habitats in a 100 m reach of stream (Florida Department of Environmental Protection 2011), which is thought to effectively capture site condition. Although it has been emphasized elsewhere (e.g., Gebler 2004), accurate bioassessment necessitates sampling protocols designed to understand and address variability to achieve true estimates of site condition.

Variability attributable to spatial and temporal (seasonal differences within years, and across years) components was

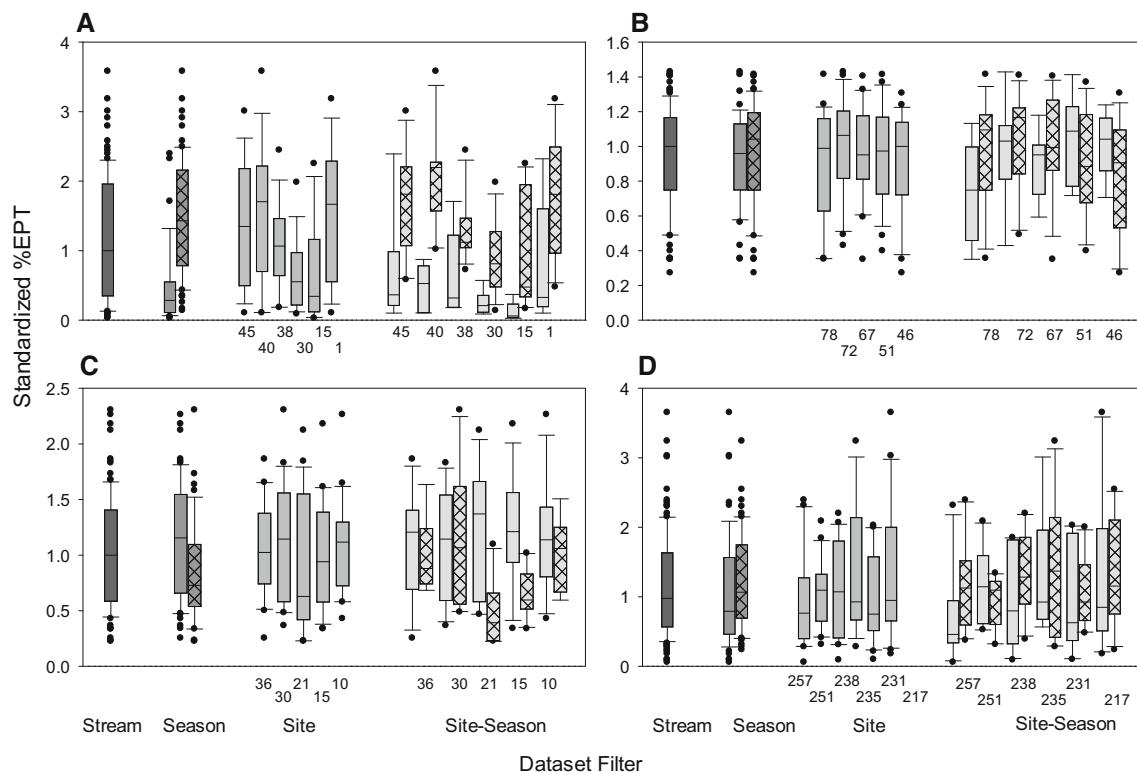


Fig. 9 Variability patterns of standardized %EPT abundance with respect to spatial and temporal dataset filters collected from **a** Codorus Creek, **b** the Leaf River, **c** the McKenzie River, and **d** the Willamette River from 1998 to 2012. Site condition calculation is based on pooled replicate samples collected on each sampling date. Endpoints

were standardized by dividing the endpoint value at each site and sampling date by the river-wide endpoint median. In assessments of seasonal components, *unhatched bars* represent spring samples, and *hatched bars* represent fall samples. Numbers associated with the site and site-season filters correspond to site numbers in Table 1

also stream and endpoint specific. In Codorus Creek, most macroinvertebrate metrics and chl *a* varied due to spatial and seasonal differences that often accounted for ~30 % of variation in the overall dataset. In the larger rivers, these endpoints were minimally influenced by independent spatial and seasonal components and better characterized by differences across years and the interaction of space and time. Differences in variance partitioning patterns across streams are likely due to several factors including stream- and region-specific environmental conditions, stream size, and sampling method. Biological communities are influenced by environmental conditions at multiple spatial scales (Mykrä et al. 2004) ranging from ecoregions (Omernik 1995) to microhabitats (Hart and Finelli 1999; Downes et al. 2000), with sometimes subtle shifts in environmental characteristics such as flow, substrate, temperature, and other water quality variables (e.g., conductivity and dissolved oxygen) resulting in patchiness within and across sites. In our study streams, spatial differences in environmental characteristics are more pronounced in Codorus Creek than in other streams. For example, the temperature regime across sites differs due to hypolimnetic flow from an upstream dam, and inputs of

warmer water from a tributary stream, non-contact industrial cooling water, and effluent discharge, and there are some differences in substrate and shading across sites. In contrast, sites in the remaining three streams are much more spatially consistent in terms of temperature and physical habitat characteristics. Additionally, because these streams are much larger than Codorus Creek, changes in water quality and habitat conditions due to tributary, industrial, and non-point source inputs are tempered by the higher volume of water in the channel.

Regional climate with a greater range of annual temperatures, in conjunction with stream size, is also a likely driver of higher seasonal variance in benthic invertebrates and chl *a* in Codorus Creek than in other streams. Biotic assemblages and associated measures (e.g., Taxa Richness) can be influenced by seasonally related patterns in temperature, flows, and life cycle stages (e.g., Johnson et al. 2012; Lunde et al. 2013). Although many bioassessment programs account for this by limiting sampling to certain times of the year, some programs have sampling windows that extend across seasons which has the potential to influence the characterization of resident assemblages. For example, macroinvertebrate samples collected by West

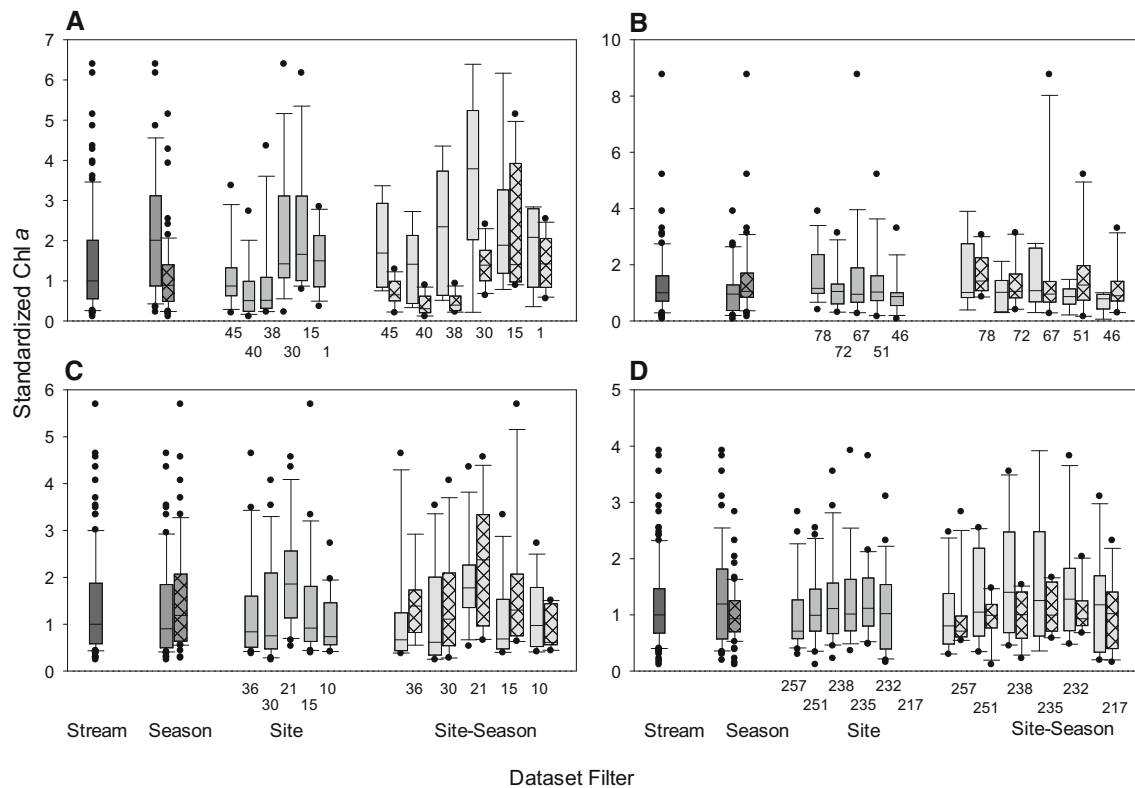


Fig. 10 Variability patterns of standardized chl *a* with respect to spatial and temporal dataset filters collected from **a** Codorus Creek, **b** the Leaf River, **c** the McKenzie River, and **d** the Willamette River from 1998 to 2012. Site condition calculation is based on pooled replicate samples collected on each sampling date. Endpoints were

standardized by dividing the endpoint value at each site and sampling date by the river-wide endpoint median. In assessments of seasonal components, *unhatched bars* represent spring samples, and *hatched bars* represent fall samples. *Numbers* associated with the site and site-season filters correspond to site numbers in Table 1

Virginia Department of Environmental Protection are acceptable for use in stream bioassessment if they are collected between April 15 and October 15 (WV DEP 2009). Results from our long-term dataset and a dataset of even longer duration demonstrate that metric response can differ markedly between seasons, and that the extent to which seasonality affects these patterns differs with the metric examined, stream, and even sites within a stream. Using a 20-year macroinvertebrate dataset, Mazor et al. (2009) found that seasonal variation for some assessment measures was small (e.g., Coleoptera richness) but large for others (e.g., EPT richness), which they suggested could be related, in part, to seasonal shifts in biological interactions. In our study, seasonal response patterns of macroinvertebrate metrics and chl *a* were not consistent with stream or metric, and variability attributed to seasonal patterns may reduce the precision of some endpoints. While the response of individual metrics can be muted when aggregated and scored in a multi-metric index (Schoolmaster et al. 2012), managers seeking to evaluate changes in site condition over time or with respect to a reference/least impacted site should have an understanding of and account for stream/

site- and metric-specific seasonal patterns to ensure management decisions are based on the best available data.

In fish metrics, variation among samples (error/residual variance) accounted for the majority of variation at most streams compared to that driven by site, season, year, or their interaction terms. Although temporal patterns in fish assemblage structure have been examined (e.g., Bêche et al. 2009; Resh et al. 2013), the precision of fish metrics at difference spatial and temporal scales has not been as well studied as macroinvertebrate-based metrics. Fish bioassessment endpoints may be biased due to fish mobility. When conditions become unsuitable due to water quality or insufficient resources, fish will disperse to other more favorable locations (Schlosser 1990) which may result in fish assemblages that are poorly linked to site condition. Indeed, in examining the relationship between fish bioassessment metrics and environmental condition, Hitt and Angermeier (2008) demonstrated that fish in headwater tributaries with limited dispersal opportunities were more strongly linked with environmental measures than fish in main stem tributaries. This supports the findings of Fayram et al. (2005) who showed that IBIs of

Table 4 REML variance components estimates of main and interactions of sites, seasons, years for macroinvertebrate metrics and periphyton chl *a* measured in the four study streams

Endpoint	Source	Stream			
		Codorus	Leaf	McKenzie	Willamette
Taxa richness	Site	20.3	17.5	0.0	1.8
	Season	21.0	9.2	0.0	0.0
	Year	0.0	4.0	0.0	0.4
	Site * season	0.0	1.8	4.9	0.7
	Site * year	0.0	3.5	15.4	0.0
	Season * year	6.4	12.6	11.5	37.0
	Site * season * year	21.4	23.2	24.2	24.3
	Error	30.9	28.3	44.1	35.7
HBI	Site	16.6	1.8	0.0	8.5
	Season	19.8	7.2	0.0	0.0
	Year	2.4	41.4	0.0	1.1
	Site * season	3.4	3.0	13.6	0.1
	Site * year	2.8	0.2	2.1	2.0
	Season * year	8.6	3.9	10.8	31.7
	Site * season * year	20.1	31.1	49.9	37.0
	Error	26.3	11.4	23.5	19.7
%EPT	Site	12.7	0.0	0.0	1.6
	Season	16.9	0.0	6.4	0.0
	Year	0.0	37.8	0.0	17.7
	Site * season	1.0	8.0	11.1	0.0
	Site * year	4.0	6.9	1.3	0.0
	Season * year	19.3	8.8	13.5	17.7
	Site * season * year	21.5	25.3	43.9	39.1
	Error	24.6	13.2	23.8	23.7
%dominant taxa	Site	9.5	0.0	0.0	0.0
	Season	0.3	0.0	0.0	25.7
	Year	0.0	21.0	8.3	0.0
	Site * season	8.9	0.3	2.5	1.8
	Site * year	4.3	1.6	16.1	0.0
	Season * year	13.8	2.7	0.0	20.1
	Site * season * year	30.4	55.5	29.3	23.9
	Error	32.7	19.0	43.9	28.4
Chl <i>a</i> (mg/m ²)	Site	9.8	0.0	6.9	4.2
	Season	21.4	0.0	0.0	0.0
	Year	0.0	42.9	0.0	7.3
	Site * season	7.6	0.0	0.7	2.3
	Site * year	10.2	30.3	0.0	3.1
	Season * year	4.5	1.5	32.7	41.6
	Site * Season * year	15.9	12.1	33.5	17.5
	Error	30.6	13.2	26.1	24.0

headwater stream fish assemblages were more variable than those in larger drainages. Additionally, capture efficiency can vary greatly due to abiotic conditions such as stream flow, depth, and water clarity even among experienced collectors (Peterson and Paukert 2009). In our streams, fish metric variability was the greatest in Codorus Creek, although this was largely driven by

patterns at a single site during the fall. This stream has not only the smallest drainage area (~710 vs. 2800–30,000 km²) but also a greater diversity of fish ($n = 52$ taxa vs. 24 and 28 taxa collected from the McKenzie and Willamette Rivers using backpack electrofishing), which may account for some stream differences in variability patterns of fish.

Table 5 REML variance components estimates of the main and interactions of site and season, site and year, and season and year for fish metrics in Codorus Creek, and the McKenzie and Willamette Rivers

Model	Source	Codorus		McKenzie		Willamette	
		Abundance	%intolerant	Abundance	%intolerant	Abundance	%intolerant
Site, Season	Site	25.2	10.0	13.2	22.2	2.3	21.0
	Season	0.0	0.0	25.9	3.3	5.6	0.0
	Site * season	1.2	0.0	0.0	0.0	0.0	0.0
	Error	73.6	89.9	60.9	74.4	92.1	79.0
Site, Year	Site	25.7	9.5	14.6	17.8	4.3	20.4
	Year	4.5	0.0	11.5	18.8	21.2	7.2
	Site * year	0.0	13.7	0.0	43.3	10.1	5.3
	Error	69.8	76.8	73.9	20.2	64.5	67.1
Season, Year	Season	0.0	0.0	29.5	0.6	9.6	0.0
	Year	0.0	0.0	11.2	14.6	20.4	0.0
	Season * year	5.3	0.0	0.3	0.0	4.9	13.4
	Error	94.7	100.0	59.0	84.8	65.1	86.6

Variability of biota can increase in response to natural and anthropogenic stressors, and researchers have emphasized the value of examining variability (e.g., range, coefficient of variability) as a dependent variable to evaluate stressor response (Palmer et al. 1997). The data evaluated in this study were collected as part of an effort to evaluate the physical, chemical, and biological response to pulp and paper mill effluent and other watershed stressors. As such, Codorus Creek, and the Leaf, McKenzie, and Willamette Rivers have effluent discharges located at sites 39.7, 68.7, 23.0, and 237 km, respectively. Apparent effluent-related shifts in metric response (relative to upstream sites) or variability were seen only in the McKenzie River and limited to the fall. Values and variability of fall macroinvertebrate taxa richness and %EPT abundance decreased downstream of the discharge, while the values and variability of %dominant taxa and HBI increased. Effluent responses in other endpoints and streams were either absent or masked by naturally occurring or other sources of variability. Fall changes in McKenzie River macroinvertebrates may result from greater effluent concentration due to low stream flows, or individual or additive response to other seasonally related changes in water quality such as temperature and dissolved oxygen. It is important to reiterate that large and statistically significant differences in individual endpoints may be less important in the context of management decision tools (i.e., IBI's and regulatory thresholds) because aggregating individual bioassessment endpoints into a multi-metric index can temper the response of individual metrics (Schoolmaster et al. 2012; Zuellig et al. 2012; Mazor et al. 2014).

Clear and consistent temporal patterns in fish and macroinvertebrate metrics and chl *a* were not seen, with most endpoints showing different temporal patterns at different

sites. This is not altogether surprising as the study was designed to capture seasonal variability and year-to-year variability associated with variable climate conditions (e.g., drought and high precipitation/flow years), with drivers of these patterns expected to be region specific. For example, drought conditions that affected much of the southeastern U.S. in the early years of the study (1997–2002) (NOAA 2002; Wang et al. 2010) that contributed to reduced flows, and short-term but significant water quality changes occurring in the days following Hurricane Katrina (Schaefer et al. 2006) only had the potential to influence biota patterns in the Leaf River. Many researchers assessing long-term biota patterns have shown high temporal- and site-specific variability. In most cases, long-term variability was associated with climate patterns (e.g., Bêche and Resh 2007a, b) and anthropogenic influences (Pyron et al. 2006), although biological factors such as species invasions (e.g., Hill and Lodge 1999) and disease (e.g., Kohler and Hoiland 2001) can also influence temporal variability. High temporal variability can pose challenges for precisely assessing streams. Many bioassessment programs evaluate site condition based on samples collected during a single site visit (Carter and Resh 2001), which may (Pyron et al. 2008) or may not be (Gebler 2004) sufficiently robust for precise bioassessment. However, conclusions based on short-term datasets may be very different from those based on longer term studies (Dodds et al. 2012). As such, resource managers should determine the environmental risk associated with making decisions and predictions based on limited temporal datasets and develop approaches for evaluating management outcomes and increasing dataset confidence through continued monitoring.

The results from this study and those cited within highlight the challenges and uncertainties facing lotic system managers. First and foremost, natural variability is

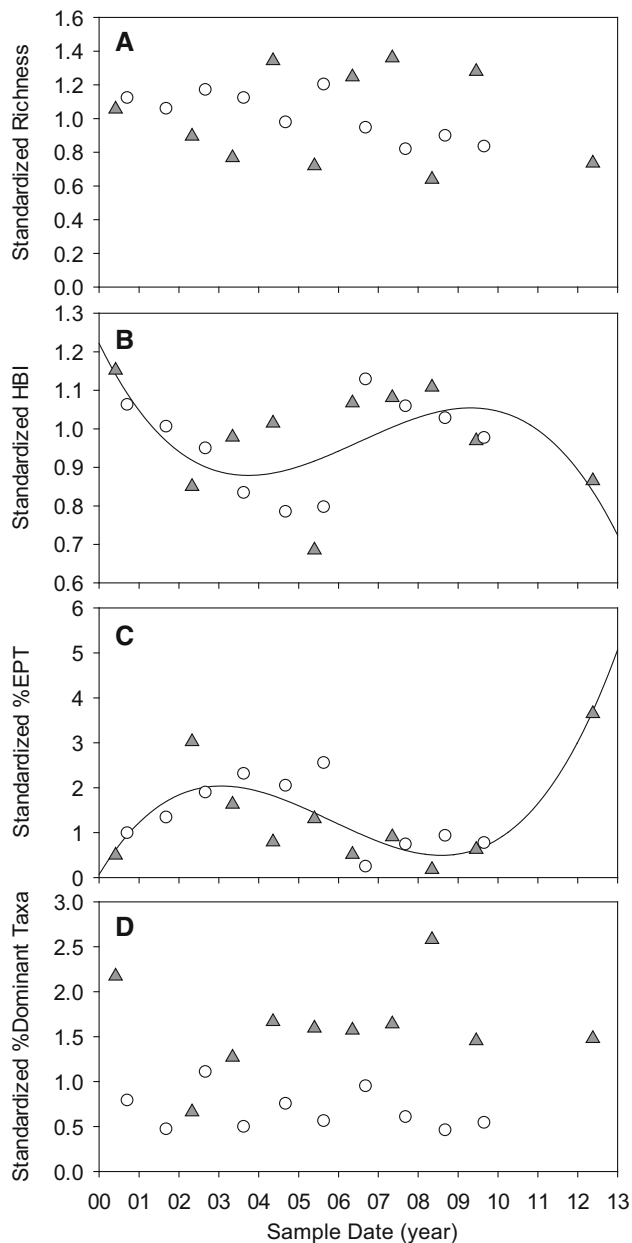


Fig. 11 Temporal patterns in standardized macroinvertebrate **a** richness, **b** HBI, **c** %EPT abundance, and **d** %dominant taxa at Willamette River site 217 km. *Gray triangle symbols* represent samples collected in the spring, and *open circles* represent samples collected in the fall. *Solid line* shows significant polynomial relationship between endpoint and sample date based on analysis of the full dataset

inherent in stream systems, and sample sizes and frequency must be sufficient to best estimate true site condition with a reasonable degree of confidence. Understanding seasonality and scheduling sampling accordingly may aid in reducing variability and improve the ability to detect change (Leung and Dudgeon 2011). Because variability is endpoint specific, the development of regulatory criteria must

consider variability and the influences of natural and anthropogenic factors in the selection of regulatory endpoints. Sufficient understanding of variability in selected regulatory endpoints is essential for setting criteria to ensure that stressor response is detectable and not masked by natural variability alone. The high degree of site-specific variability observed within individual surface waters suggests that the development and implementation of watershed- or region-wide management criteria and benchmarks should be approached carefully so that their value to managers and stakeholders is maximized. Further, high variability may lead to more errors in management decisions, and regulators should be aware of and cautious toward under- or over-protection of streams (Gebler 2004). Finally, because assessment of condition of a stream reach can differ substantially from that of larger aggregations (i.e., sub-catchments), development of site-specific criteria may offer more accurate and precise ecosystem protection than regionally developed criteria. Although costly, taken together, these precautions offer greater confidence that assessed condition is sufficiently accurate and subsequent management decisions are based on information that reflects an appropriate level of confidence.

Acknowledgments B. Arthurs and J. Ikoma played a significant role in field sampling, fish identification, and data entry, with field assistance during the course of the study from R. Ragsdale, J. Thomas, D. McGarvey, B. Streblow, R. Philbeck, D. Brodhecker, A. Helfrich, T. Pearce-Smith, F. Howell, and J. Napack. Discussions with M. Dubé, W. Landis, W. Minshall, J. Rodgers, and S. Missimer on study design and analysis were valuable. W. Arthurs assisted with formatting, data analysis, and the preparation of Figure 1; and we appreciate editing and formatting assistance from S. Easthouse. Feedback from W. Minshall and P. Wiegand improved earlier versions of the manuscript. Thorough and thoughtful reviews and suggestions provided by two anonymous reviewers further improved the manuscript.

References

- Aloi JE (1990) A critical review of recent freshwater periphyton field methods. *Can J Fish Aquat Sci* 47:656–670
- American Public Health Association, American Water Works Association and Water Environment Federation (APHA, AWWA and WEF) (2000) Standard methods for the examination of water and wastewater, 20th edn. APHA, Washington, DC
- Association of Clean Water Administrators (ACWA) (2012) Use of biological assessment in state water programs: focus on nutrients. ACWA, Washington, DC
- Barbour M, Gerritsen J, Snyder BD, Stribling JB (1999) Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. EPA 841-B-99-002, 3rd edn. U.S. Environmental Protection Agency, Office of Water, Washington, DC
- Barton DR, Metcalfe-Smith JL (1992) A comparison of sampling techniques and summary indices for assessment of water quality in the Yamaska River, Québec, based on benthic macroinvertebrates. *Environ Monit Assess* 21:225–244

- Bêche LA, Resh VH (2007a) Short-term climatic trends affect the temporal variability in California ‘Mediterranean’ streams. *Freshw Biol* 52:2317–2339
- Bêche LA, Resh VH (2007b) Biological traits of benthic macroinvertebrates in California Mediterranean-climate streams: long term annual variability and trait diversity patterns. *Fundam Appl Limnol* 169:1–23
- Bêche LA, Connors PG, Resh VH, Merenlender AM (2009) Resilience of fishes and invertebrates to prolonged drought in two California streams. *Ecography* 32:778–788
- Biggs BJJ, Kilroy C (1996) Stream periphyton monitoring manual. National Institute of Water and Atmospheric Research, Christchurch
- Brown BL (2003) Spatial heterogeneity reduces temporal variability in stream insect communities. *Ecol Lett* 6:316–325
- Cao Y, Hawkins CP, Olson J (2007) Modeling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *J N Am Benthol Soc* 26:566–585
- Cardinale BJ, Palmer MA, Swan CM, Brooks S, Poff NL (2002) The influence of substrate heterogeneity on biofilm metabolism in a stream ecosystem. *Ecology* 83:412–422
- Carter JL, Resh VH (2001) After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. *J N Am Benthol Soc* 20:658–682
- Cooper SD, Barmuta L, Sarnelle O, Kratz K, Diehl S (1997) Quantifying spatial heterogeneity in streams. *J N Am Benthol Soc* 16:174–188
- Dodds WK, Robinson CT, Gaiser EE, Hansen GJA, Powell H, Smith JM, Morse NB, Johnson SL, Gregory SV, Bell T, Kratz TK, McDowell WH (2012) Surprises and insights from long-term aquatic data sets and experiments. *Bioscience* 62:709–721
- Downes BJ, Hindell JS, Bond NR (2000) What is a site? Variation in lotic macroinvertebrate density and diversity in a spatially replicated experiment. *Aust J Ecol* 25:128–139
- Dudley JL, Arthurs W, Hall TJ (2001) A comparison of methods used to estimate river rock surface areas. *J Freshw Ecol* 16:257–261
- Fayram AH, Miller MA, Colby AC (2005) Effects of stream order and ecoregion on variability in coldwater fish index of biotic integrity scores within streams in Wisconsin. *J Freshw Ecol* 20:17–25
- Flinders CA, Ragsdale RL, Hall TJ (2009) Patterns of fish community structure in a long-term watershed-scale study to address the aquatic ecosystem effects of pulp and paper mill discharges in four U.S. receiving streams. *Integr Environ Assess Manag* 5:219–233
- Florida Department of Environmental Protection (FDEP) (2011) Sampling and use of the stream condition index (SCI) for assessing flowing waters: a primer. Standards and Assessment Section Bureau of Assessment and Restoration Support DEP-SAS-001/11. FDEP, Tallahassee
- Franklin JF (1988) Importance and justification of long-term studies in ecology. In: Likens GE (ed) *Long-term studies in ecology: approaches and alternatives*. Springer, New York, pp 3–19
- Gebler JB (2004) Mesoscale spatial variability of selected aquatic invertebrate community metrics from a minimally impaired stream site. *J N Am Benthol Soc* 23:616–633
- Gregg DC, Stednick JD (2000) Variability in measures of macroinvertebrate community structure by stream reach and stream class. *J Am Water Resour Assoc* 36:95–103
- Hall LW, Scott MC, Killen WD (1996) Development of biological indicators based on fish assemblages in Maryland coastal plain streams. CBWP-MANTA-EA-96-1. Maryland Department of Natural Resources, Chesapeake Bay and Watershed Programs, Annapolis
- Hall TJ, Fisher RP, Rodgers JL, Minshall GW, Landis WG, Kovacs TG, Firth BK, Dubé MG, Deardorff TL, Borton DL (2009a) A long-term multitrophic level study to assess pulp and paper mill effluent effects on aquatic communities in four United States receiving waters: background and status. *Integr Environ Assess Manag* 5:189–198
- Hall TJ, Ragsdale RL, Arthurs WJ, Ikoma J, Borton DL, Cook DL (2009b) A long-term multi-trophic level study to assess pulp and paper mill effluent effects on aquatic communities in 4 U.S. receiving waters: characteristics of the study streams, sample sites, mills, and mill effluents. *Integr Environ Assess Manag* 5:199–218
- Hart DD, Finelli CM (1999) Physical-biological coupling in streams: the pervasive effects of flow on benthic organisms. *Annu Rev Ecol Syst* 30:363–395
- Hawkins CP, Olson JR, Hill RA (2010) The reference condition: predicting benchmarks for ecological and water-quality assessments. *J N Am Benthol Soc* 29:312–343
- Hester FE, Dendy JS (1962) A multiple sampler for aquatic macroinvertebrates. *Trans Am Fish Soc* 91:420–421
- Hill AM, Lodge DM (1999) Replacement of resident crayfishes by an exotic crayfish: the roles of competition and predation. *Ecol Appl* 9:678–690
- Hilsenhoff WL (1987) An improved biotic index of organic stream pollution. *Gt Lakes Entomol* 20:31–39
- Hilsenhoff WL (1988) Rapid field assessment of organic pollution with a family level biotic index. *J N Am Benthol Soc* 7:65–68
- Hitt NP, Angermeier PL (2008) River-stream connectivity affects fish bioassessment performance. *Environ Manag* 42:132–150
- Idaho Department of Health and Welfare (1993) Water quality monitoring protocols report no. 5. Protocols for assessment of biotic integrity (macroinvertebrates) for Wadeable Idaho streams. Idaho Department of Health and Welfare, Washington, DC
- Jackson JK, Füreder L (2006) Long-term studies of freshwater macroinvertebrates: a review of the frequency, duration and ecological significance. *Freshw Biol* 51:591–603
- Johnson RC, Carreiro MM, Jin H-S, Jack JD (2012) Within-year temporal variation and life-cycle seasonality affect stream macroinvertebrate community structure and biotic metrics. *Ecol Indic* 13:206–214
- Kohler SL, Hoiland WK (2001) Population regulation in an aquatic insect: the role of disease. *Ecology* 82:2294–2305
- Larsen DP, Kincaid TM, Jacobs SE, Urquhart NS (2001) Designs for evaluating local and regional scale trends. *Bioscience* 51:1069–1078
- Legendre P, Legendre L (1998) *Numerical ecology*. Elsevier, New York
- Leung ASL, Dudgeon D (2011) Scales of spatiotemporal variability in macroinvertebrate abundance and diversity in monsoonal streams: detecting environmental change. *Freshw Biol* 56:1193–1208
- Li H, Reynolds JE (1995) On definition and quantification of heterogeneity. *Oikos* 73:280–284
- Li J, Herlihy A, Gerth W, Kaufmann P, Gregory S, Urquhart S, Larsen DP (2001) Variability in stream macroinvertebrates at multiple spatial scales. *Freshw Biol* 46:87–97
- Lindenmayer DB, Likens GE (2009) Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends Ecol Evol* 24:482–486
- Lindenmayer DB, Likens GE, Krebs CJ, Hobbs RJ (2010) Improved probability of detection of ecological “surprises”. *Proc Natl Acad Sci* 107:21957–21962
- Linke S, Bailey RC, Schwindt J (1999) Temporal variability of stream bioassessments using benthic macroinvertebrates. *Freshw Biol* 42:575–584

- Lunde KB, Cover MR, Mazor RD, Sommers CA, Resh VH (2013) Identifying reference conditions and quantifying biological variability within benthic macroinvertebrate communities in perennial and non-perennial northern California streams. *Environ Manag* 51:1262–1273
- Maxted JR, Barbour MT, Gerritsen J, Poretti V, Primrose N, Silvia A, Penrose D, Renfrow R (2000) Assessment framework for mid-Atlantic coastal plain streams using benthic macroinvertebrates. *J N Am Benthol Soc* 19:128–144
- Mazor RD, Purcell AH, Resh VH (2009) Long-term variability in bioassessments: a twenty-year study from two northern California streams. *Environ Manag* 43:1269–1286
- Mazor RD, Stein ED, Ode PR, Schiff K (2014) Integrating intermittent streams into watershed assessments: applicability of an index of biotic integrity. *Freshwater Sci* 33:459–474
- Merritt RW, Cummins KW (1996) An introduction to the aquatic insects of North America, 3rd edn. Kendall/Hunt, Dubuque
- Mykrä H, Heino J, Muotka T (2004) Variability of lotic macroinvertebrate assemblages and stream habitat characteristics across hierarchical landscape classifications. *Environ Manag* 34:341–352
- National Climatic Data Center (NOAA) (2002) State of the climate: drought for August 2002. <http://www.ncdc.noaa.gov/sotc/drought/2002/8>. Accessed 2 March 2015
- Nelson JS, Crossman EJ, Espinosa-Pérez H, Findley LT, Gilbert CR, Lea RN, Williams JD (eds) (2004) Common and scientific names of fishes from the United States, Canada and Mexico. American Fisheries Society, Bethesda
- Omernik JM (1995) Ecoregions: a spatial framework for environmental management. In: Davis WS, Simon TP (eds) Biological assessment and criteria. Tools for water resource planning and decision making. Lewis Publishers, Boca Raton, pp 49–62
- Palmer MA, Hakenkamp CC, Nelson-Baker K (1997) Ecological heterogeneity in streams: why variance matters. *J N Am Benthol Soc* 16:189–202
- Peterson JT, Paukert CP (2009) Chapter 12: converting nonstandard fish sampling data to standardized data. In: Bonar SA, Hubert WA, Willis DW (eds) Standard methods for sampling North American freshwater fishes. American Fisheries Society, Washington, DC
- Pyron M, Lauer TE, Gammon JR (2006) Stability of the Wabash River fish assemblages from 1974 to 1998. *Freshw Biol* 51:1789–1797
- Pyron M, Lauer TE, LeBlanc D, Weitzel D, Gammon JR (2008) Temporal and spatial variation in an index of biotic integrity for the middle Wabash River, Indiana. *Hydrobiologia* 600:205–214
- Rabeni CF, Wang N, Sarver RJ (1999) Evaluating adequacy of the representative stream reach used in invertebrate monitoring programs. *J N Am Benthol Soc* 18:284–291
- Resh VH, Bêche LA, Lawrence JE, Mazor RD, McElravy EP, O'Dowd AP, Rudnick D, Carlson SM (2013) Long-term population and community patterns of benthic macroinvertebrates and fishes in Northern California Mediterranean-climate streams. *Hydrobiologia* 719:93–118
- Robson BJ, Hogan M, Forrester T (2005) Hierarchical patterns of invertebrate assemblage structure in stony upland streams change with time and flow permanence. *Freshw Biol* 50:944–953
- Sandin L, Johnson RK (2000) The statistical power of selected indicator metrics using macroinvertebrates for assessing acidification and eutrophication of running waters. *Hydrobiologia* 422:233–243
- Schaefer J, Mickle P, Spaeth J, Kreiser BR, Adams S, Matamoros W, Zuber B, Vigueira (2006) Effects of hurricane Katrina on the fish fauna of the Pascagoula River drainage. In: Proceedings from the 36th annual Mississippi water resources conference, pp 62–68
- Schlösser IJ (1990) Environmental variation, life history attributes, and community structure in stream fishes: implications for environmental management and assessment. *Environ Manag* 14:621–628
- Schoolmaster DR Jr, Grace JB, Schweiger EW (2012) A general theory of multimetric indices and their properties. *Methods Ecol Evol* 3:773–781
- Simon TP (1991) Development of ecoregion expectations for the index of biotic integrity. I. Central corn belt plain. EPA 905-9-91-025. U.S. Environmental Protection Agency, Region 5, Chicago
- Simpson JC, Norris RH (2000) Biological assessment of water quality: development of AUSRIVAS models and outputs. In: Wright JF, Sutcliffe DW, Furse MT (eds) Assessing the biological quality of freshwaters: RIVPACS and similar techniques. Freshwater Biological Association, Ambleside, pp 125–142
- Sokal RJ, Rohlf FJ (1995) Biometry, 3rd edn. W.H. Freeman and Company, New York
- Stevenson RJ, Pan Y, Manoylov KM, Parker CA, Larsen DP, Herlihy AT (2008) Development of diatom indicators of ecological conditions for streams of the western US. *J N Am Benthol Soc* 27:1000–1016
- United States Environmental Protection Agency (USEPA) (2002) Summary of biological assessment programs and biocriteria development for states, tribes, territories, and interstate commissions: streams and Wadeable Rivers. EPA-822-R-02-048. U.S. Environmental Protection Agency, Washington, DC
- Wang H, Fu R, Kumar A, Li W (2010) Intensification of summer rainfall variability in the southeastern United States during recent decades. *J Hydrometeorol* 11:1007–1018
- West Virginia Department of Environmental Protection (WV DEP) (2009) 2009 standard operating procedures V01 SOP. WV DEP, Charleston
- Wieczorek ME, LaMotte AE (2010) Attributes for NHDPlus catchments (version 1.1) for the conterminous United States: NLCD 2001 land use and land cover. U.S. Geological Survey Digital Data Series DDS-490-15. http://water.usgs.gov/GIS/metadata/usgswrd/XML/nhd_nlcd01.xml. Accessed January 2011
- Zuellig RE, Carlisle DM, Meador MR, Potapova M (2012) Variance partitioning of stream diatom, fish, and invertebrate indicators of biological condition. *Freshw Sci* 31:182–190