

# Long-Term Variability in Bioassessments: A Twenty-Year Study from Two Northern California Streams

Raphael D. Mazor · Alison H. Purcell ·  
Vincent H. Resh

Received: 16 July 2008 / Accepted: 12 March 2009 / Published online: 21 April 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** Long-term variability of bioassessments has not been well evaluated. We analyzed a 20-year data set (1984–2003) from four sites in two northern California streams to examine the variability of bioassessment indices (two multivariate RIVPACS-type O/E scores and one multimetric index of biotic integrity, IBI), as well as eight metrics. All sites were sampled in spring; one site was also sampled in summer. Variability among years was high for most metrics (coefficients of variation, CVs ranging from 16% to 246% in spring) but lower for indices (CVs of 22–26% for the IBI and 21–32% for O/E scores in spring), which resulted in inconsistent assessments of biological condition. Variance components analysis showed that the time component explained variability in all metrics and indices, ranging from 5% to 35% of total variance explained. The site component was large (i.e., >40%) for some metrics (e.g., EPT richness), but nearly absent from others (e.g., Diptera richness). Seasonal analysis at one site showed that variability among seasons was small for some metrics or indices (e.g., Coleoptera richness), but large for others (e.g., EPT richness, O/E scores). Climatic variables did not show consistent trends across all metrics, although several were related to the El Niño Southern Oscillation

Index at some sites. Bioassessments should incorporate temporal variability during index calibration or include climatic variability as predictive variables to improve accuracy and precision. In addition, these approaches may help managers anticipate alterations in reference streams caused by global climate change and high climatic variability.

**Keywords** Annual variability · Biomonitoring · Metrics · Indices · Index of biotic integrity (IBI) · O/E scores · Climate

## Introduction

Although biological monitoring using benthic macroinvertebrates has a long history, only a small number of published studies present more than a few years of benthological data (e.g., Rosenberg and Resh 1993; Jackson and Füreder 2006). Multi-year data sets, however, are essential to characterize long-term variability, detect major trends, and relate local community shifts to worldwide phenomena, such as global climate change (Schmitt and Osenberg 1996; Daufresne and others 2003). As interest in tracking the long-term health of streams and rivers grows, the need to evaluate the performance of metrics and bioassessments over multiple years will likely become more important.

The presumption underlying all biomonitoring studies is that natural variability in biological communities can be measured and controlled through the establishment of appropriate reference conditions (Resh and others 1995; Gebler 2004; Bonada and others 2006a). In biomonitoring applications, the natural variability of metrics (i.e., variability in the absence of impact) is assumed to be less than

---

R. D. Mazor · A. H. Purcell · V. H. Resh  
Department of Environmental Science, Policy, & Management,  
University of California, Berkeley, CA 94720-3114, USA

R. D. Mazor (✉)  
Southern California Coastal Water Research Project, 3535  
Harbor Boulevard, Suite 110, Costa Mesa, CA 92626, USA  
e-mail: raphaelm@sccwrp.org

A. H. Purcell  
Department of Environmental and Natural Resource  
Sciences, Humboldt State University, 1 Harpst Street,  
Arcata, CA 95521, USA

the change caused by a disturbance or restoration project. However, the long-term natural variability in commonly used metrics has not been well evaluated (Jackson and Füreder 2006), and high variability may pose a challenge in using indices to determine the ecological health of a river or stream by reducing precision (Hughes 1995; Bailey and others 2004; Mazor and others 2006).

Variability in benthic community structure can result from spatial and temporal sources. Spatial variability occurs at scales both small (i.e., differences among samples collected within a reach) and large (i.e., differences among reaches within a watershed, or among watersheds). Variability among samples collected at the same site and time has been extensively studied in some classical (e.g., Needham and Usinger 1956; Chutter 1972) as well as recent (e.g., Gebler 2004; Tomanova and Usseglio-Polatera 2007) articles. Large-scale spatial variability in community structure can be caused by differences in the physical or chemical environment among sites, as well as by biogeographical influences. Small-scale spatial variability may result from microhabitat complexity. The effects of both of these features on benthic macroinvertebrates have been well studied, and major environmental gradients that shape biotic communities have been identified (e.g., stream order, pH, riparian vegetation) (Rosenberg and Resh 1989).

Temporal variability, which is the result of changes in community structure over time, may describe changes within years (i.e., seasonal variability) or among years (i.e., annual variability). Seasonal variability has been well studied (e.g., Linke and others 1999; Bonada 2003; Bêche and others 2006) and is driven by short-term climatic factors that vary over the course of a year, such as rainfall (and consequently flooding) or temperature. In regions with mediterranean climates, such as coastal California, flow regimes of streams vary greatly between spring and summer, creating distinct community profiles (Gasith and Resh 1999; Bêche and others 2006; Bonada and others 2006b).

In contrast to other sources of variability, long-term annual variability has not been well studied in stream ecosystems (Jackson and Füreder 2006). Annual variability can include extreme events, such as prolonged droughts or major floods, or more frequent natural phenomena, such as El Niño-related changes in the duration, intensity, and amount of rainfall (Molles and Dahm 1990). Several studies have shown that annual variability is sometimes larger than other sources of variability (e.g., Sandin and Johnson 2004; Bêche and others 2006). More recently, growing concern about global climate change has sparked interest in the effects of climatic variability on stream ecosystems (e.g., Molles and Dahm 1990; Bonada and others 2007).

Unfortunately, without an adequate understanding of annual variability under natural conditions, biomonitoring

programs cannot attribute improvements or deteriorations in ecological condition to human interventions (Schmitt and Osenberg 1996; Scarsbrook and others 2000). Furthermore, as bioassessment programs are increasingly interested in establishing biocriteria for regulatory purposes (e.g., benchmarks for development of total maximum daily loads), they must determine if high temporal variability make static thresholds inappropriate (Reckhow and others 1997).

The goal of this study was to quantify the variability of commonly used bioassessment metrics and indices, and to evaluate the relative contributions of spatial, inter-annual, and seasonal variability to total variability. In this study, we analyzed benthic macroinvertebrate and climate data at 4 sites collected over 20 years. Because we measured these sources of variability at the same sites over the same time period and because collections and identifications were made by the same individuals, many of the problems typically associated with long-term collection and comparisons of data were eliminated. Therefore, we could directly compare each source of variability. We are not aware of other studies where annual and spatial variability have been compared at the same set of sites and over the same time period, nor any studies that used as consistent a sampling methodology over the course of the study. As bioassessment programs continue and long-term data sets accumulate, we anticipate that this study will be one of many to address questions that can only be answered with long-term data. Because the study was conducted in a region of California with a mediterranean climate, which has extreme inter-annual and seasonal variability, the estimates presented here may represent an upper limit for long-term variability when compared to streams in less variable, more mesic climates (Gasith and Resh 1999).

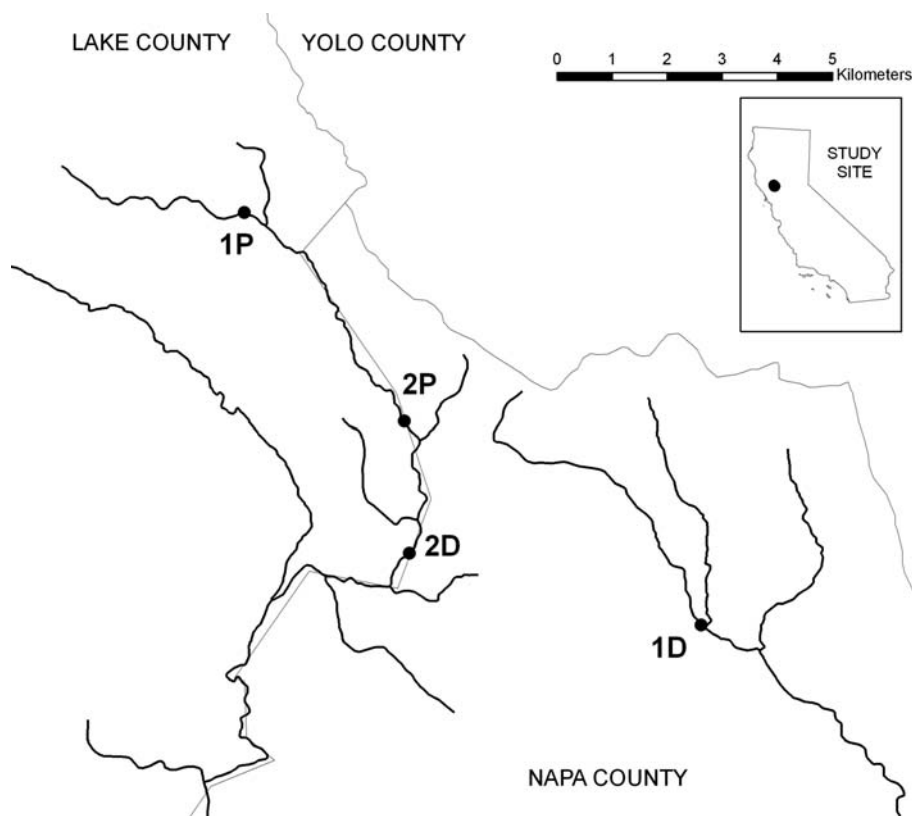
## Methods

### Study Site

Sampling was conducted from 1984–2003 in Knoxville and Hunting Creeks in Lake and Napa Counties, California (Fig. 1). These creeks enter Lake Berryessa, a reservoir on Putah Creek, which drains eastward into the Sacramento River near the San Francisco Bay Delta. The area has a mediterranean climate, with cool, wet winters and hot, dry summers (Gasith and Resh 1999). Nearly all precipitation (>90%) occurs as rain during the wet season, from October through June.

The four sampling sites selected on Knoxville and Hunting Creeks (30-km north of Lake Berryessa) represent a continuum of hydrologic intermittency. The driest site in this continuum (site 1D) was located on a 1st order stream

**Fig. 1** Map of the study area. *Black lines* represent streams, and *gray lines* represent county boundaries



(Knoxville Creek) that consistently went dry every summer (usually July–October). The other sites were located on Hunting Creek. Site 2D was located on a non-perennial side-channel of a 2nd order perennial reach; this side-channel typically flowed from September or October (6 months before the mid-April sampling date) through July. Site 1P was located on a 1st order perennial segment that typically flowed year round, but went dry twice during the summer sampling (mid-August) in 2002 and 2003. Lastly, Site 2P was located on a 2nd order reach and was perennial throughout the entire study period, although flow was greatly reduced in the late summer.

The Hunting Creek sites are located within the University of California McLaughlin Nature Reserve, which is managed to preserve natural resources for conservation and research purposes. The Knoxville Creek site is located on private property. Although this latter site was subjected to a tailings-pond spill in 1996 and a wildfire in 1999, both of these disturbances had little effect on the macroinvertebrate communities (University of California Davis Natural Reserve System 2003; Bêche 2005). Historic mining activity, including mine tailings, potentially affected all sites in the study, and a downstream recreational campground may have affected site 2D.

Both watersheds drain a mixture of volcanic and serpentine soils that are dominated by blue oak (*Quercus douglasii*) woodland and chaparral (University of

California Davis Natural Reserve System 2003). All sites are within the Southern and Central California Chaparral and Oak Woodland Level 3 Ecoregion (Omernik 1995), and are typical of small watersheds in this area. For further descriptions of study sites, see Resh and others (2005), Bêche and others (2006), and Bêche and Resh (2007a, b).

Although nonperennial streams are typically excluded from many bioassessment programs, they are often the primary habitat available for aquatic biota in large regions of the world. Several states, including California, already mandate assessment and regulation of these streams, and thus their inclusion in bioassessment programs will increase.

#### Sampling of Benthic Macroinvertebrates

Macroinvertebrates were sampled at the four sites every spring (post wet-season, 15 April  $\pm$  3 days) between 1984 and 2003. Site 1P was also sampled every summer (post dry-season, 15 August  $\pm$  days). At each sample date, 5 Surber samples (0.093 m<sup>2</sup>, 0.5 mm mesh) were collected in a random design, stratified to riffle areas; the same riffle areas were sampled each year. All individuals in each sample were identified. Most specimens were identified to genus; some non-chironomid Diptera and non-insects were identified to order, family, or sub-family level. To maintain consistency in sampling and

identification procedure throughout the entire course of the 20-year study, all samples were collected by the same person (Vincent H. Resh), and all specimens were identified by the same person (Eric P. McElravy). As a result, this study is based on one of the most consistent long-term benthological data sets available to date. Although macroinvertebrates from each riffle were sorted and identified separately, data from these riffles were pooled and subsequently subsampled to approximate the sampling requirements of the indices and metrics used in this study (described below).

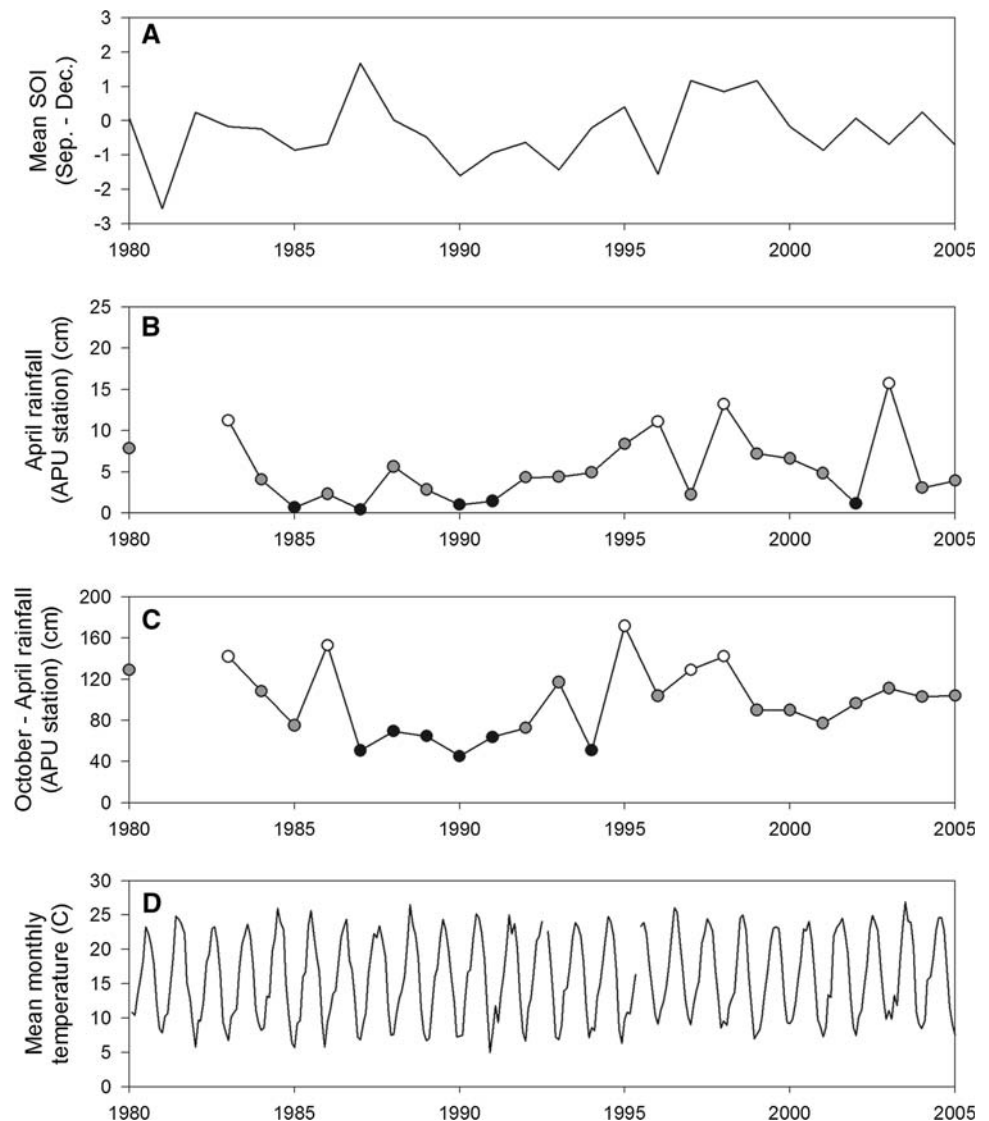
### Climatic Data

Long-term records of precipitation were obtained for a nearby gauge monitored by the California Department of Water Resources (station name: APU, Angwin Pacific Union, [www.water.ca.gov](http://www.water.ca.gov)). The record contained monthly

total precipitation from 1 October 1939 through 30 August 2005 (with a brief interruption for 1981–1982). Climatic condition of each year was calculated in 2 ways: 7-month rainfall (i.e., total rainfall for the 7 months prior to and including the sampling date, October through April), and 1-month rainfall (total rainfall during the month of spring sampling, April). The full record was used to designate years as low (<25th percentile), moderate (between 25th and 75th percentiles), and heavy (>75th percentile) rainfall (Fig. 2).

Mean daily temperature data collected between 1980 and 2008 from Markey Cove in Napa County was downloaded from the National Oceanic and Atmospheric Administration's website (accessed online May 20, 2008: <http://www.nesdis.noaa.gov/>). Mean temperature was calculated for the time periods of April 1–April 15, and October 1–April 15. Mean temperature was also calculated for the period of August 1–August 15 (Fig. 2).

**Fig. 2** Climate variables analyzed in this study. **a** Mean El Niño Southern Oscillation Index (SOI) for the months of September to December; **b** 1-month rainfall (April) at the APU rain gauge. *White circles* indicate years in the 75th percentile (>5.5 cm rain); *gray circles* indicate years between the 25th and 75th percentile (0.5–5.5 cm rain); *black circles* indicate years below the 25th percentile (<0.5 cm rain); **c** 7-months rainfall (October to April) at the APU rain gauge. *White circles* indicate years in the 75th percentile (>82.4 cm rain); *gray circles* indicate years between the 25th and 75th percentile (50.3–82.4 cm rain); *black circles* indicate years below the 25th percentile (<50.2 cm rain). Percentiles were based on the full data series (1939–2005); **d** Mean monthly temperature at the Markey Cove weather station



To investigate the potential impact of the El Niño Southern Oscillation (ENSO) on aquatic communities, we used the Southern Oscillation Index (SOI). The SOI is a measure of the standardized departure in the difference in sea-level pressure in the Pacific Ocean between measurements in Stand Tahiti and Stand Darwin. Because the autumn and early winter ENSO conditions in the tropical Pacific are most likely to affect late winter/early spring climatic patterns in California, we calculated the average SOI for September through December for each year, based on monthly data available from the NOAA Climate Prediction Center (accessed online May 20, 2008: <http://www.cpc.ncep.noaa.gov/data/indices/>) (Fig. 2).

## Data Analysis

### *Calculation of Metrics and the Index of Biotic Integrity (IBI)*

We calculated biological metrics that are widely used in the state of California and other regions of the world to assess long-term variability of bioassessment metrics. All metrics included in the Northern Coastal California Index of Biotic Integrity (IBI, Rehn and others 2005) were used, including three metrics based on richness (Ephemeroptera, Plecoptera, and Trichoptera (EPT) richness; Coleoptera richness; and Diptera richness), and five metrics based on composition (% intolerant individuals, % non-gastropod scraper individuals, % predator individuals, % shredder taxa, and % non-insect taxa). These metrics were then scored and combined to calculate the IBI on a 100-point scale. In addition, we calculated total richness and % EPT because these metrics are widely used in many biomonitoring programs (Resh and Jackson 1993; Bonada and others 2006a).

The invertebrate data were transformed to comply with the requirements of the IBI. For example, taxa were aggregated to conform with operational taxonomic units (OTUs) specified by the standard taxonomic effort (Richards and Rogers 2006) for use in bioassessment throughout California. Semi-aquatic Hemiptera were excluded from all counts. If the samples then contained more than 500 individuals (i.e., the number of individuals required for calculation of the IBI), they were subsampled using a random selection procedure to reduce the size of the sample to 500 individuals. Samples containing fewer than 500 individuals were not subsampled before metric and IBI calculation.

Although IBI scores were calculated for all samples in the study, they were not interpreted to infer biological condition of these sites. The validity of the absolute value of the IBI scores is uncertain because of differences in sample collection and processing, and because of the low representation of nonperennial streams in the calibration set used to develop the IBI (Rehn and others 2005).

However, this study assumes that the relative values and observed variability of IBI scores within each site and season are valid.

### *Calculation of O/E Scores*

To evaluate the long-term variability of multivariate assessments, we calculated the ratio of observed to expected taxa using the California RIVPACS (River InVertebrate Prediction and Classification System) model (described by Ode and others 2008). We calculated scores using both the 100% (O/E<sub>100</sub>) and 50% (O/E<sub>50</sub>) capture probabilities, (i.e., including and excluding rare species, respectively). The invertebrate data were transformed to comply with the requirements of the RIVPACS model. For example, taxa were aggregated to conform with the necessary operational taxonomic units for this model. The samples that contained more than 300 individuals (i.e., the number of individuals required for calculation of O/E scores) were subsampled using a random selection procedure to reduce the size of the sample to 300 individuals. Samples containing fewer than 300 individuals were not subsampled before O/E score calculation.

Several predictor variables for the RIVPACS models were obtained using geographic coordinates (latitude and longitude). From these variables, we determined watershed area (in log km<sup>2</sup>) and percent sedimentary geology in the watershed area upstream of each sampling site using a generalized geologic map of the United States (accessed online December 20, 2007: <http://pubs.usgs.gov/atlas/geologic>). Long-term mean monthly precipitation (log mm) and mean monthly temperature (°C) were estimated from GIS grids of (1961–1990) obtained from the Oregon Climate Center (accessed online December 7, 2007: <http://www.ocs.orst.edu/prism>). Because sites were located in warm (mean monthly temperature >9.9°C) and wet (log mean monthly precipitation >2.952) areas (Table 1), and because Chironomidae were identified to morphospecies, we used RIVPACS submodel 1 that excludes midges. Details of the RIVPACS model can be found at <http://129.123.10.240/wmcportal/DesktopDefault.aspx>.

As with IBI scores, O/E scores were not interpreted to infer biological condition of these sites because of influences resulting from differences in sampling methods and low representation of nonperennial streams in the calibration set used to develop the O/E model. However, this study assumes that the relative values and observed variability of O/E scores within each site and season are valid.

### *Evaluation of Trends*

Metrics and indices were plotted against time to examine trends at each site and season individually. Significant



**Table 1** Characteristics of sites sampled in the study

| Site | Stream          | Stream order | Perenniality              | Watershed area (km <sup>2</sup> ) | Latitude | Longitude | # of years sampled     | Mean annual temperature (°C) | Mean monthly precipitation (mm) |
|------|-----------------|--------------|---------------------------|-----------------------------------|----------|-----------|------------------------|------------------------------|---------------------------------|
| 1D   | Knoxville Creek | 1            | Nonperennial              | 2.1                               | 38.7989  | −122.3147 | 20                     | 15.4                         | 952                             |
| 2D   | Hunting Creek   | 2            | Nonperennial <sup>a</sup> | 29.3                              | 38.8292  | −122.3792 | 20                     | 15.4                         | 954                             |
| 1P   | Hunting Creek   | 1            | Perennial                 | 4.4                               | 38.8656  | −122.4150 | 19 Spring<br>16 Summer | 15.1                         | 1061                            |
| 2P   | Hunting Creek   | 2            | Perennial                 | 22.1                              | 38.8083  | −122.3767 | 19                     | 15.3                         | 961                             |

<sup>a</sup> Riffle was located in a nonperennial side-channel of a perennial reach

trends were identified by regressing metrics against time and comparing slopes to zero. A Bonferroni correction was used to adjust  $\alpha$  to 0.004 to account for multiple comparisons across 11 metrics and indices.

Differences among sites for metrics and indices were tested using crossed ANOVAS, with site and year as factors. To account for multiple comparisons across metrics and indices,  $\alpha$  was set to 0.004 to achieve 95% confidence. Differences between seasons were assessed at site 1P using paired *t*-tests and only years in which data from both seasons were available were included in these tests. Relationships between the indices or metrics and climatic data were evaluated by calculating Spearman's rank correlation ( $\rho$ ) for each site and season independently; correlations with  $\rho^2 \geq 0.2$  were considered strong. Statistical significance of these relationships was not assessed because of low power and the high number of tests required.

### Evaluation of Variability

In order to determine long-term temporal variability, coefficients of variation (CVs) were calculated within each site and season across the entire study period. In addition, we calculated CVs within each year across all sites (excluding summer samples at site 1P) to characterize changes in spatial variability over time. CVs are an intuitively informative and widely used method of characterizing and comparing the variability of metrics and indices (e.g., Resh 1994; Sandin and Johnson 2000). In addition, minimum detectable differences (MDDs) were calculated for metrics and indices at each site and season to determine the amount of change that could be observed after 5 years of monitoring. MDDs were calculated using a 1-sample 2-tailed *t*-test ( $\alpha = 0.05$ ,  $\beta = 0.2$ ). For index scores, MDDs were then compared to established thresholds (i.e., 20 for the IBI, 0.46 for O/E<sub>100</sub>, and 0.32 for O/E<sub>50</sub>; Ode and others 2005, 2008) to determine if the index could detect a change of condition within 5 years.

Because CVs are strongly influenced by the different means among metrics, we also performed a variance components analysis to determine the amount of variability in each metric attributable to year, site, and the interaction of

site and year. In contrast to CVs, variance components are based on the sums of squares that underlie many statistical tests, and are more directly comparable across metrics. Because we had no replication within sites and years, residual variance (the component attributable to variability among samples) was estimated independently from data collected at a different set of sites for a separate study (Rehn and others 2007). Because Rehn and others (2007) analyzed metrics used in the IBI, as well as O/E scores, values from total richness and % EPT were not available. Summer samples were excluded from this analysis. Restricted maximum likelihood (REML) was used to calculate variance components because of the unbalanced design and SAS was used for all calculations (using PROC VARCOMP method = REML, SAS Institute Inc. 2004). Unlike the mean-square method of estimating variance components, REML ensures that all components are greater than or equal to zero (Larsen and others 2001). Because sites were a fixed factor and not a random factor, the variance component attributable to site must be considered a finite, or pseudo variance (Courbois and Urquhart 2004). A second analysis was performed using data from both seasons at site 1P to determine the components of variability attributable to year, season, and their interaction.

## Results

### Overview of the Data Set

Sampling at the four sites over 20 years resulted in 94 samples (with samples missing from site 1P in 1984, 1985, 2002, and 2003, and from site 2P in 1986). Samples contained a total of 206 unique taxa, but converting these taxa to OTUs for metric calculation reduced this number to 137 (largely from aggregating Chironomidae to family and elimination of semiaquatic Hemiptera); conversion for O/E score calculation resulted in 125 OTUs.

The total number of individuals per sample ranged from a low of 161 to a high of 13,952 individuals. Seventeen of the 94 samples contained fewer than 450 organisms (the recommended minimum for calculation of the IBI, Rehn

and others 2005), and 10 contained fewer than 270 (the recommended minimum for calculation of O/E scores). These undersized samples were most frequent at site 2P, where 7 and 5 samples were affected for IBI and O/E score calculations, respectively. Because we were more interested in evaluating the variability of these metrics and indices than assessing the study sites, we retained all samples in all analyses. Furthermore, because the study was designed to establish upper bounds on estimate of long-term variability, inclusion of these samples and the potential increase in variability estimates was consistent with our goals.

## Evaluation of Trends

### Indices of Ecological Condition

Assessment indices varied by both time and year. For example, IBI scores at site 2D ranged from a low of 30 in 1992 to a high of 70 in 1999. Similarly, variability within a year across sites was evident; for example, the range of O/E<sub>100</sub> scores in spring was greater than 0.3 in most years (with a maximum of 0.49 in 1997) (Table 2).

Despite the fact that watershed management generally was constant at sites and disturbances were largely absent, all indices indicated fluctuating condition of these streams over the course of the study (Fig. 3). For example, the IBI suggested that site 2D was in good condition (i.e., IBI  $\geq$  60) in 1987, 1997, and 1999, but in poor condition (i.e., IBI  $<$  40) from 1990 to 1993. Similar variability was seen at other sites and with other indices. Year-to-year fluctuations of O/E<sub>100</sub> scores at site 2P were often as large as 0.40 (Fig. 3). However, no trends were statistically significant at any site, after accounting for multiple comparisons ( $P > 0.004$ ). Even with this high variability, the indices appeared to respond to disturbance; for example the IBI and both O/E scores decreased in 1986 following an accidental sediment spill at site 1D (Fig. 3). Although decreases were observed at other sites in the study that year (described in Resh and Jackson 1993), the decreases were much larger at the site affected by the spill (e.g., IBI scores declined 28 points from 1985 at site 1D, but declined only 15 points at site 1P).

The indices showed improving conditions in the 1990s at most sites in spring. For example, IBI scores at site 1D increased from a low of 22.5 in 1995 to a high of 53.8 in 1999. This increasing trend coincided with a moderately wet period following a drought, as indicated by both 1- and 7-months rainfall (Fig. 2).

Although both of the above indices showed strong and consistent differences among the sites, the differences were stronger for both O/E indices than for the IBI. For example, O/E<sub>50</sub> scores consistently showed that site 2D had the best

**Table 2** Summary of indices at each site and season

| Site and season (# years) | Indices     |                    |                   |
|---------------------------|-------------|--------------------|-------------------|
|                           | IBI         | O/E <sub>100</sub> | O/E <sub>50</sub> |
| 1D Spring (20 years)      |             |                    |                   |
| Mean $\pm$ SD             | 43 $\pm$ 11 | 0.44 $\pm$ 0.10    | 0.33 $\pm$ 0.09   |
| CV                        | 25          | 22                 | 28                |
| Range                     | 23–56       | 0.16–0.57          | 0.14–0.53         |
| MDD                       | 18          | 0.13               | 0.13              |
| 2D Spring (20 years)      |             |                    |                   |
| Mean $\pm$ SD             | 51 $\pm$ 11 | 0.65 $\pm$ 0.15    | 0.49 $\pm$ 0.12   |
| CV                        | 22          | 23                 | 24                |
| Range                     | 30–70       | 0.46–0.94          | 0.27–0.73         |
| MDD                       | 19          | 0.21               | 0.16              |
| 1P Spring (19 years)      |             |                    |                   |
| Mean $\pm$ SD             | 38 $\pm$ 10 | 0.42 $\pm$ 0.09    | 0.28 $\pm$ 0.07   |
| CV                        | 26          | 21                 | 24                |
| Range                     | 20–56       | 0.22–0.58          | 0.15–0.45         |
| MDD                       | 17          | 0.12               | 0.09              |
| 1P Summer (16 years)      |             |                    |                   |
| Mean $\pm$ SD             | 28 $\pm$ 12 | 0.28 $\pm$ 0.06    | 0.11 $\pm$ 0.05   |
| CV                        | 42          | 21                 | 41                |
| Range                     | 8–60        | 0.19–0.41          | 0.05–0.20         |
| MDD                       | 20          | 0.06               | 0.06              |
| 2P Spring (19 years)      |             |                    |                   |
| Mean $\pm$ SD             | 42 $\pm$ 10 | 0.52 $\pm$ 0.12    | 0.37 $\pm$ 0.12   |
| CV                        | 24          | 23                 | 32                |
| Range                     | 20–60       | 0.21–0.67          | 0.14–0.54         |
| MDD                       | 17          | 0.16               | 0.16              |

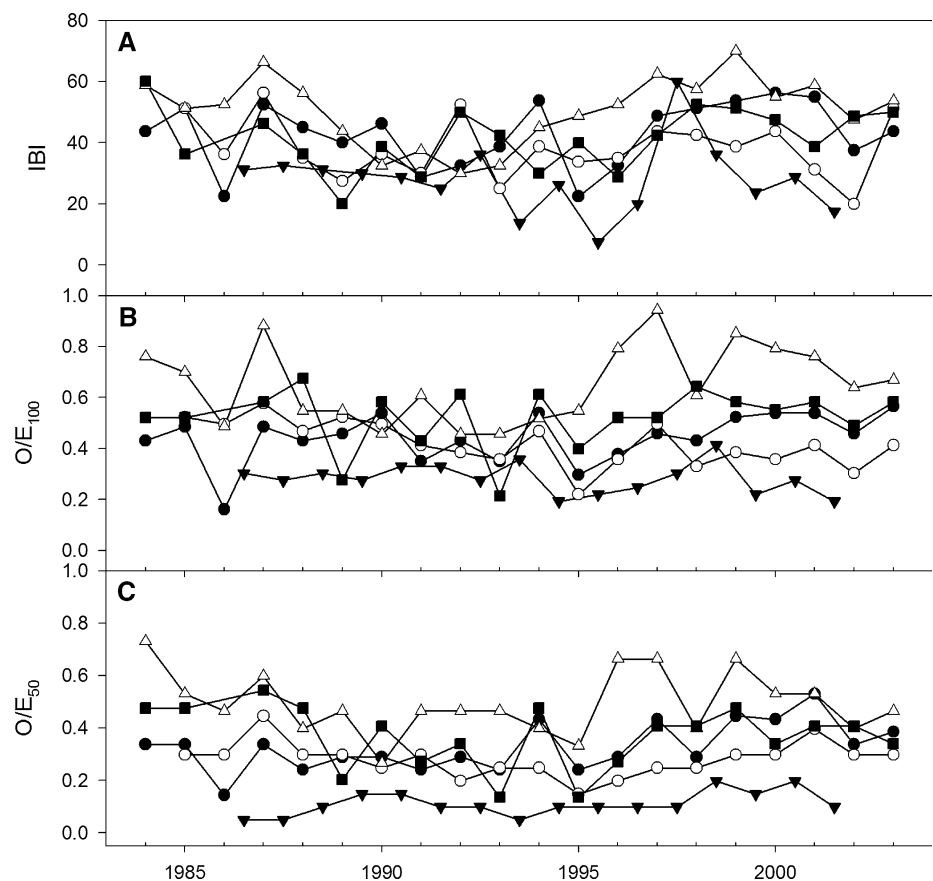
Sites are listed in order of increasing perennality

SD Standard deviation, CV coefficient of variation, *Min* minimum value, *Max* maximum value, *MDD* minimum detectable difference (1-sample 2-tailed *t*-test, 5 years of sampling,  $\alpha = 0.05$ ,  $\beta = 0.2$ )

ecological condition in most years, and that site 1P had the worst condition (Fig. 3c). These differences may reflect variability in watershed disturbance at each site, or sensitivity to natural conditions and variability, such as watershed area and hydrologic regime. Differences among sites were less evident for the IBI (Fig. 3a). However, differences among sites were statistically significant ( $P < 0.004$ ) for all indices.

Samples collected in summer had lower values of indices than those collected in spring at site 1P, as indicated by a paired *t*-test ( $P = 0.0015$  for the IBI, and  $P < 0.0001$  for both O/E scores). This pattern was evident with all indices in most years. In fact, reversals were observed in only one year with the O/E<sub>100</sub> score (i.e., 1998) and two years with the IBI (i.e., 1989 and 1997). No reversals of this pattern were evident with the O/E<sub>50</sub> score (Fig. 3).

**Fig. 3** Values of indices by year. **a** IBI; **b**  $O/E_{100}$ , and **c**  $O/E_{50}$ . Each point represents one sample. *Black circles* represent site 1D. *White triangles* represent site 2D. *White circles* represent spring samples from site 1P. *Black triangles* represent summer samples from site 1P. *Black squares* represent site 2P



### Richness Metrics

Richness metrics varied in their ability to distinguish sites. For example, site 2D consistently had higher EPT richness than other sites in most years, whereas site 1P had the lowest (in both spring and summer) ( $P < 0.0001$ , Table 3, Fig. 4). Total richness showed similar but weaker differences ( $P = 0.0009$ ). In contrast, no consistent trends were evident for Coleoptera or Diptera richness ( $P > 0.004$ ).

Some metrics reflected similar patterns as the assessment indices in showing improving trends in the 1990s. For example, EPT richness increased from 6 taxa in 1993 to 13 taxa in 1999 at site 2P; other sites showed similar increases. However, none of these metrics showed statistically significant changes at any site over the course of the study ( $P > 0.004$ ).

Seasonal differences were strongly evident for some metrics (Table 3, Fig. 4). For example, EPT, Diptera, and total richness were all higher in spring than in summer in most years at site 1P. Paired  $t$ -tests found statistically significant differences between seasons for EPT richness and Coleoptera richness (i.e.,  $P < 0.004$ ). However, differences between seasons for Diptera richness ( $P = 0.0371$ ) and total richness ( $P = 0.0210$ ) were not significant once accounting for multiple comparisons across metrics and indices.

### Composition Metrics

Several composition metrics showed strong temporal consistency among sites. For example, % intolerant and % non-gastropod scrapers showed spikes at all sites in certain years (e.g., 1986 and 1999). However, this consistency may be explained by the fact that these metrics contained many zero or near-zero values at most sites, coinciding with a period of low 1-month rainfall (i.e., 1985–1995) (Table 3, Figs. 2 and 5).

In general, compositional metrics were more similar among sites than richness metrics. For example, no site had consistently higher % shredder taxa than other sites. To some extent, the ability to distinguish sites was strongest with % non-insect taxa, with site 1P (in both spring and summer) having a higher metric value than other sites. Differences among sites were statistically significant for % non-gastropod scrapers ( $P = 0.0004$ ), % predators, % non-insect taxa, and % EPT (all  $P < 0.0001$ ); however, differences were not significant for % intolerant ( $P = 0.0672$ ) and % shredder taxa ( $P = 0.3590$ ).

Some metrics showed consistent differences between seasons at site 1P. For example, % intolerant was higher in the spring than in the summer (paired  $t$ -test  $P = 0.0017$ ). Conversely, % non-insect taxa was higher in the summer than in the spring (paired  $t$ -test  $P = 0.0015$ ). However,



**Table 3** Summary of metrics at each site and season

| Site and season (# years)   | Richness metrics |            |         |        | Composition metrics |                          |             |             |                   |         |  |
|-----------------------------|------------------|------------|---------|--------|---------------------|--------------------------|-------------|-------------|-------------------|---------|--|
|                             | EPT              | Coleoptera | Diptera | Total  | % Intolerant        | % Non-gastropod scrapers | % Predators | % Shredders | % Non-insect taxa | % EPT   |  |
| <b>1D Spring (20 years)</b> |                  |            |         |        |                     |                          |             |             |                   |         |  |
| Mean ± SD                   | 9 ± 4            | 2 ± 2      | 6 ± 2   | 22 ± 6 | 10 ± 7              | 2 ± 3                    | 7 ± 4       | 5 ± 4       | 11 ± 6            | 42 ± 23 |  |
| CV                          | 40               | 73         | 26      | 29     | 69                  | 158                      | 55          | 66          | 51                | 55      |  |
| Range                       | 1–16             | 0–6        | 3–9     | 7–29   | 1–24                | 0–9                      | 1–15        | 0–13        | 0–22              | 5–78    |  |
| MDD                         | 6                | 3          | 3       | 9      | 9                   | 4                        | 5           | 5           | 8                 | 32      |  |
| <b>2D Spring (20 years)</b> |                  |            |         |        |                     |                          |             |             |                   |         |  |
| Mean ± SD                   | 13 ± 3           | 3 ± 2      | 6 ± 1   | 27 ± 6 | 16 ± 8              | 5 ± 6                    | 8 ± 4       | 5 ± 3       | 12 ± 4            | 48 ± 8  |  |
| CV                          | 21               | 59         | 20      | 22     | 53                  | 120                      | 48          | 47          | 34                | 16      |  |
| Range                       | 8                | 0–6        | 3–8     | 16–35  | 5–39                | 0–22                     | 3–19        | 0–10        | 0–19              | 34–63   |  |
| MDD                         | 5                | 3          | 2       | 8      | 11                  | 9                        | 5           | 3           | 6                 | 11      |  |
| <b>1P Spring (19 years)</b> |                  |            |         |        |                     |                          |             |             |                   |         |  |
| Mean ± SD                   | 7 ± 2            | 2 ± 1      | 6 ± 1   | 20 ± 4 | 12 ± 9              | 0 ± 1                    | 17 ± 12     | 4 ± 4       | 20 ± 6            | 20 ± 15 |  |
| CV                          | 30               | 71         | 26      | 19     | 77                  | 246                      | 71          | 107         | 30                | 74      |  |
| Range                       | 4–12             | 0–4        | 3–9     | 15–28  | 1–33                | 0–2                      | 3–44        | 0–12        | 10–32             | 1–57    |  |
| MDD                         | 4                | 2          | 2       | 5      | 12                  | 1                        | 16          | 5           | 8                 | 20      |  |
| <b>1P Summer (16 years)</b> |                  |            |         |        |                     |                          |             |             |                   |         |  |
| Mean ± SD                   | 2 ± 1            | 1 ± 1      | 4 ± 2   | 16 ± 4 | 3 ± 6               | 0 ± 0                    | 22 ± 17     | 5 ± 3       | 31 ± 12           | 16 ± 24 |  |
| CV                          | 51               | 138        | 58      | 23     | 180                 | 302                      | 80          | 63          | 37                | 151     |  |
| Range                       | 1–5              | 0–4        | 1–10    | 11–25  | 0–25                | 0–2                      | 5–64        | 0–14        | 13–55             | 0–78    |  |
| MDD                         | 2                | 2          | 4       | 5      | 8                   | 1                        | 24          | 5           | 16                | 33      |  |
| <b>2P Spring (19 years)</b> |                  |            |         |        |                     |                          |             |             |                   |         |  |
| Mean ± SD                   | 10 ± 2           | 2 ± 1      | 5 ± 2   | 21 ± 5 | 17 ± 12             | 2 ± 2                    | 6 ± 4       | 6 ± 4       | 11 ± 4            | 44 ± 24 |  |
| CV                          | 24               | 75         | 28      | 26     | 70                  | 109                      | 70          | 80          | 42                | 56      |  |
| Range                       | 5–13             | 0–6        | 2–8     | 10–30  | 4–44                | 0–7                      | 1–13        | 0–14        | 4–21              | 12–90   |  |
| MDD                         | 4                | 3          | 3       | 7      | 16                  | 3                        | 6           | 6           | 6                 | 33      |  |

Sites are listed in order of increasing perennality

SD Standard deviation, CV coefficient of variation, Min minimum value, Max maximum value, MDD minimum detectable difference (1-sample 2-tailed *t*-test, 5 years of sampling,  $\alpha = 0.05$ ,  $\beta = 0.2$ )

many metrics showed no significant difference between the seasons. For example, % EPT was on average only 0.642% higher in spring than in summer ( $P = 0.9096$ ).

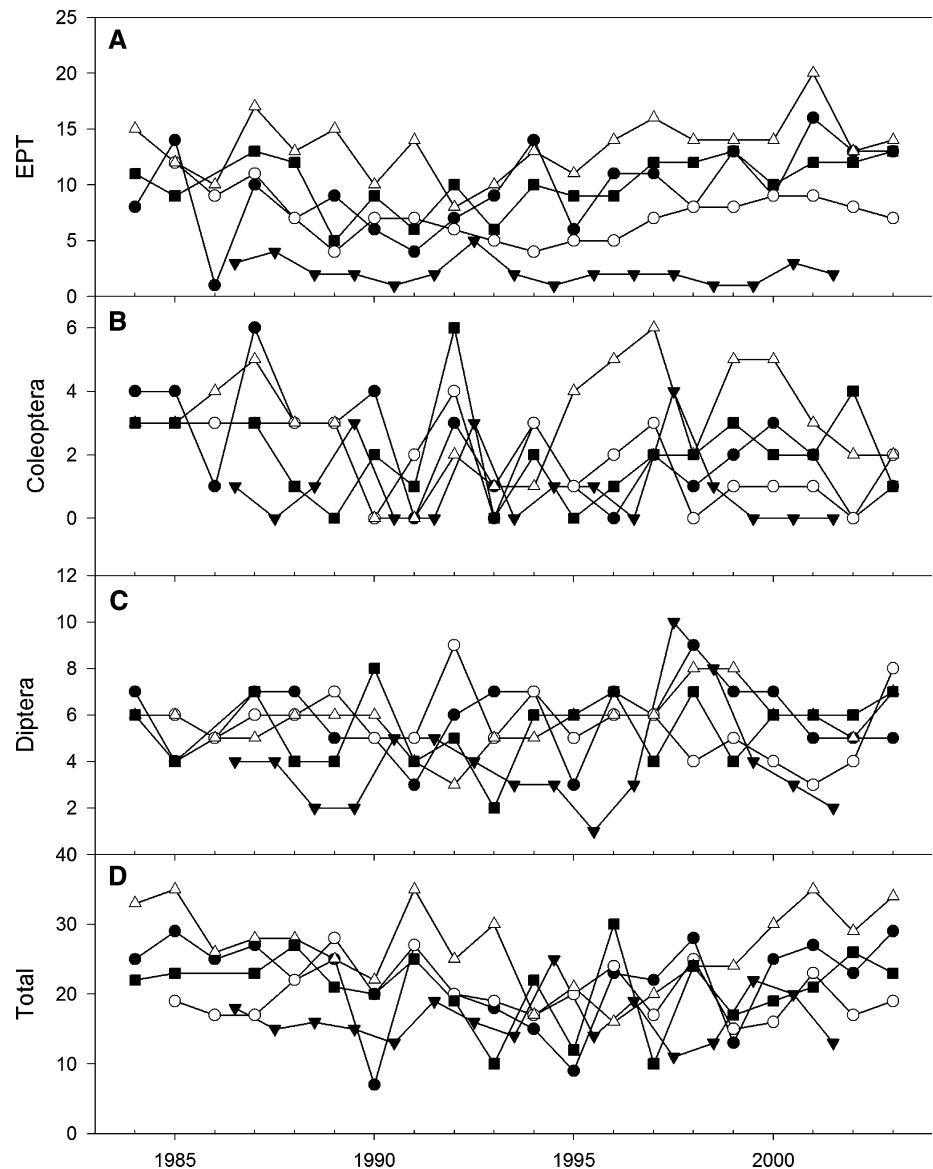
*Relationship Between Indices or Metrics and Climate*

Climate variables did not show consistent relationships with metrics and indices at all sites and seasons, although a few strong ( $\rho^2 > 0.2$ ) correlations were observed. For example, Diptera richness had a strong positive relationship with 1-month rainfall at sites 1D and 2D ( $\rho = 0.53$  and  $0.54$ , respectively), but weak or negative relationships at other sites and seasons ( $\rho$  ranges from  $-0.34$  to  $0.24$ ). Apart from mean temperature from October 1 to April 1, all climate variables had a strong relationship with at least one metric or index at one or more site and season. The

SOI had strong relationships with more metrics or indices—and at more sites—than other variables, specifically the IBI (sites 2P and 2D), EPT richness (site 2P), Coleoptera richness (site 2D), % EPT (site 2D), and % intolerant (site 2P). Two metrics (% non-gastropod scrapers and % predators) did not have a strong relationship with any climatic variables (Table 4).

Some sites were more influenced by climate than others. For example, relationships were observed more often at the two second order sites (6 at site 2D and 5 at site 2P) than the first order sites (3 at site 1D and 2 at site 1P in each season). Metrics and indices at first order sites were often influenced by precipitation, especially 7-months rainfall. In contrast, the SOI only had strong relationships with metrics or indices at the second order sites. No patterns relating to degree of perennality were evident (Table 4).

**Fig. 4** Value for richness metrics by year. **a** EPT richness, **b** Coleoptera richness, **c** Diptera richness, and **d** total richness. Each point represents one sample. Symbols are identical to Fig. 3



## Evaluation of Variability

### Variability Over Time

Several metrics showed low variability (i.e.,  $CV < 50\%$ ) over time at all sites. For example, CVs of total richness were under 30% at every site (Table 3, Fig. 6a). A similar pattern was observed with both the IBI and O/E scores, which had low long-term variability (i.e.,  $CV < 50\%$ ) at all sites in spring (Table 2, Fig. 6a). MDDs for all indices were small enough that changes among condition classes could be detected within 5 years (i.e.,  $MDDs \leq 20$  for IBI scores,  $\leq 0.46$  for  $O/E_{100}$ , and  $\leq 0.32$  for  $O/E_{50}$  scores) (Table 2). Long-term variability was highest for the % non-gastropod scrapers metric compared to other metrics, especially at site 1P (both spring and summer), with the CV being over 100% at every site for this metric. Variability

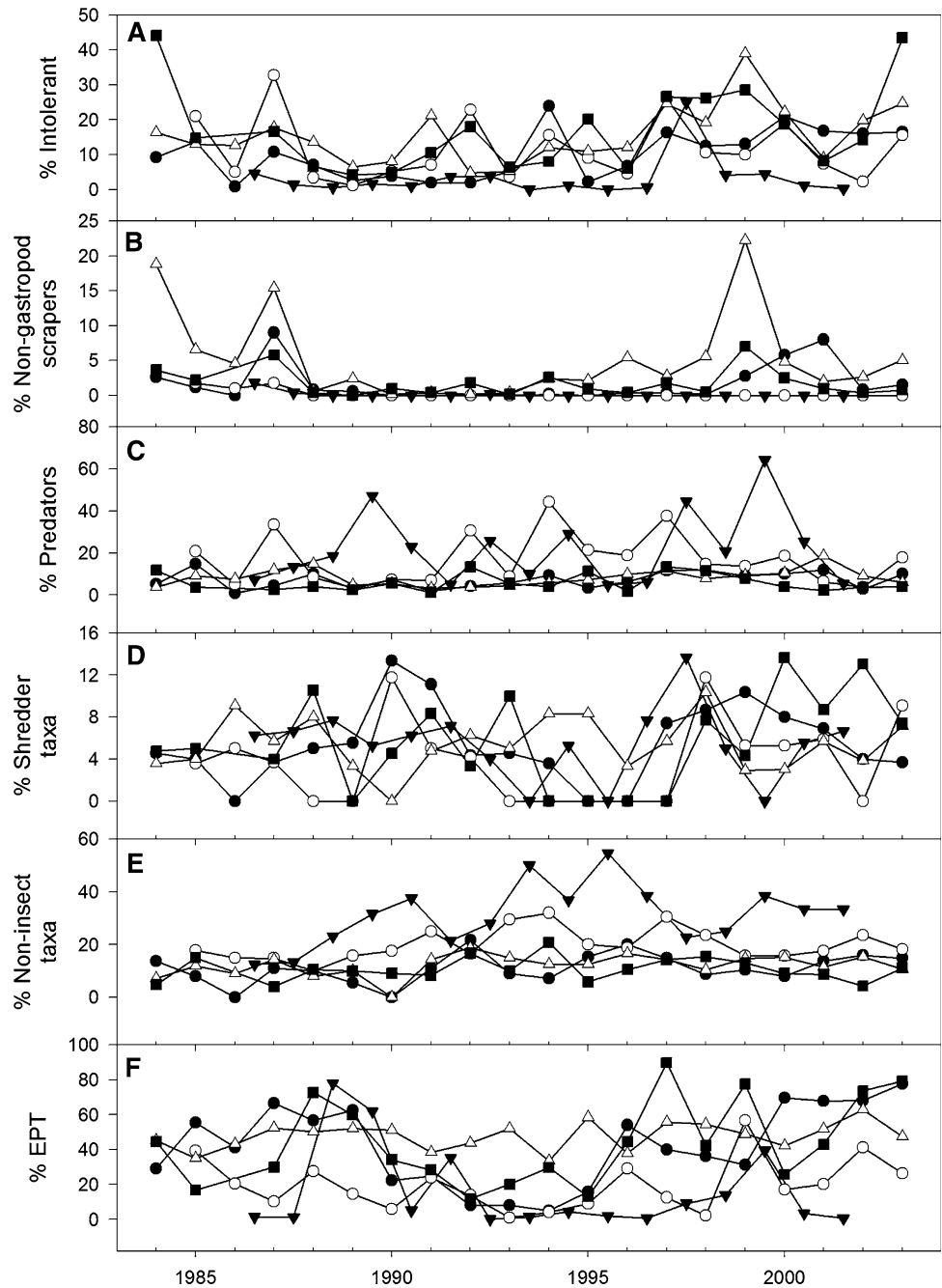
was relatively low (i.e.,  $CV < 50\%$ ) for EPT richness and % non-insect taxa, and CVs were similar at all sites.

Season influenced long-term variability. For example, samples collected in the summer had higher CVs than those collected in the spring. This trend was most evident for Coleoptera richness, Diptera richness, % intolerant, % non-gastropod scrapers, and % EPT (Fig. 6a). For other metrics, differences in CVs between spring and summer were small (i.e.,  $< 25\%$ ) or absent, except for the % shredder taxa metric, which was more variable in spring than summer (CV 107% vs. 63% respectively) (Fig. 6a).

### Variability Over Space

Approximately one-half of the metrics examined showed low spatial variability (i.e.,  $CV < 100\%$ ) in all years. For example, EPT richness, Diptera richness, and all indices

**Fig. 5** Values for composition metrics by year. **a** % intolerant, **b** % non-gastropod scrapers, **c** % predators, **d** % shredder taxa, **e** % non-insect taxa, and **f** % EPT. Each point represents one sample. Symbols are identical to Fig. 3



had CVs across sites below 100% for all years. In contrast, spatial variability was consistently high for % non-gastropod scrapers, with CVs across sites over 100% in most years. Other metrics (e.g. % shredder taxa) showed more complex patterns, with high variability (CV > 100%) in some years, and low variability (CV < 100%) in others (Fig. 6b).

*Components of Variability*

Variance components analysis of samples collected in spring showed different patterns for annual and spatial

components of variability. For example, the annual component explained a portion of variability for all metrics and indices, ranging from a low of 5% of total variance explained (for Diptera richness) to a high of 35% (for % intolerant). In contrast, the spatial component of variability differed strongly among the metrics and indices, and did not always explain a portion of the variability. For example, the spatial component of the two O/E scores, % non-insect taxa, and EPT richness were all over 40%, indicating that these metrics and indices were strongly influenced by the characteristics of the site. However, other metrics had very small spatial components (e.g., % non-insect taxa,

**Table 4** Spearman rank correlations ( $\rho$ ) between metrics or indices and climate variables

| Index or metric     | Climatic variable        | Site      | $\rho$ | $\rho^2$ | <i>P</i> |
|---------------------|--------------------------|-----------|--------|----------|----------|
| IBI                 | SOI                      | 2P        | 0.45   | 0.20     | 0.0521   |
|                     | SOI                      | 2D        | 0.60   | 0.36     | 0.0050   |
| O/E <sub>100</sub>  | 7-months rainfall        | 1D        | -0.46  | 0.21     | 0.0420   |
|                     | 7-months rainfall        | 1P Spring | -0.60  | 0.36     | 0.0071   |
| O/E <sub>50</sub>   | Mean temp (Aug 1–Aug 15) | 1P Spring | -0.56  | 0.31     | 0.0158   |
| EPT richness        | SOI                      | 2P        | 0.63   | 0.40     | 0.0040   |
| Coleoptera richness | SOI                      | 2D        | 0.49   | 0.24     | 0.0266   |
|                     | 7-months rainfall        | 1D        | -0.61  | 0.37     | 0.0044   |
| Diptera richness    | 1-month rainfall         | 1D        | 0.53   | 0.28     | 0.0164   |
|                     | 1-month rainfall         | 2D        | 0.54   | 0.29     | 0.0144   |
|                     | Mean temp (Aug 1–Aug 15) | 2P        | 0.60   | 0.35     | 0.0092   |
| Total richness      | Mean temp (Aug 1–Aug 15) | 2D        | -0.55  | 0.31     | 0.0139   |
| % EPT               | SOI                      | 2D        | 0.49   | 0.24     | 0.0282   |
|                     | Mean temp (Oct 1–Aug 15) | 1P Summer | -0.52  | 0.27     | 0.0397   |
| % Intolerant        | SOI                      | 2D        | 0.50   | 0.25     | 0.0235   |
|                     | SOI                      | 2P        | 0.55   | 0.31     | 0.0137   |
|                     | Mean temp (Apr 1–Apr 15) | 2P        | -0.49  | 0.24     | 0.0313   |
| % Shredder taxa     | Mean temp (Oct 1–Aug 15) | 1P Summer | 0.49   | 0.24     | 0.0534   |

*P*-Values are provided, but were not used to test statistical significance

*Temp* temperature, *Apr* April, *Aug* August, *Oct* October, *SOI* Southern Oscillation Index. Only relationships with  $\rho^2 \geq 0.2$  are shown

Coleoptera richness), or none at all (i.e., Diptera richness), indicating that the location had a minimal influence on these metrics independent from time (Fig. 7a). The spatial and temporal component explained the majority of the variance for all metrics and indices, except for Diptera richness and % non-gastropod scrapers.

The interaction of space and time was the largest component of variability for all metrics, except for EPT richness and the O/E<sub>50</sub> score (Fig. 7a). This interaction term represents the combined effect of site and time, indicating that most metrics varied over time at different sites in different ways. This interaction is evident in most of the plots of metric over year, where changes in value occurred at some sites and not others.

Seasonal analysis of variance components at site 1P yielded mixed results, with some metrics showing a large influence of season, and others showing a large influence of year. For example, year was the largest variance component for Coleoptera richness and % non-gastropod scrapers. However, season was the largest variance component for EPT richness and both O/E scores, and the year component was small or estimated to be zero.

Analysis of seasonal components of variability at site 1P showed more complex patterns, and most metrics did not show similar trends. For example, the annual component of variability was a very large component of EPT richness and the O/E scores, but was a negligible component of several composition metrics (i.e., % non-gastropod scrapers, %

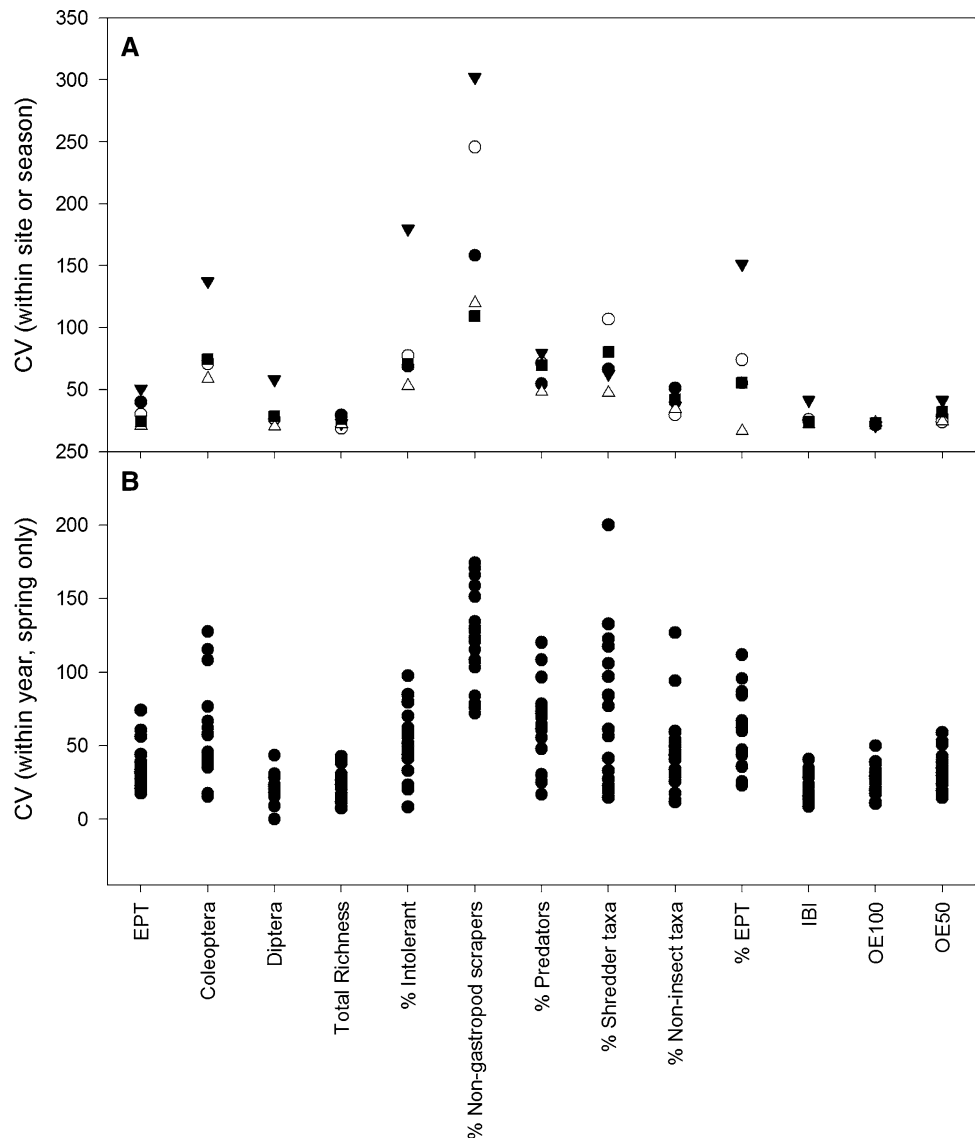
predators, % shredder taxa, and % EPT). The seasonal component was large for Coleoptera richness and % non-gastropod scrapers. As with the spatial variance components analysis, interaction terms were frequently large, often comprising more than half the variance (e.g., Diptera richness, total richness, % predators, and % shredder taxa, and % EPT). For these metrics, seasonal differences waxed and waned from year to year (Fig. 7b). Season and year explained the majority of the variance, except for % non-gastropod scrapers, for which residual variance accounted for nearly all the explained variability.

## Discussion and Conclusions

Although ecologists have long recognized the large spatial variability of benthic macroinvertebrate communities (e.g., Needham and Usinger 1956), consideration of the temporal component of variability has been a more recent development (Jackson and Füreder 2006), and applications to bioassessment lag further still (Resh and others 2005). This study represents one of the first analyses of long-term variability of bioassessment metrics using such an extensive and consistent data set.

Long-term variability was generally larger for metrics than for indices at the four study sites, as indicated by both large variance components and high CVs. Similarly high CVs were observed in other long-term data sets

**Fig. 6** **a** CVs within each site and season for each metric and index. Each point represents a site or season. Symbols are identical to Fig. 3; **b** CVs within each year (spring only) for each site and year. Each symbol represents a year



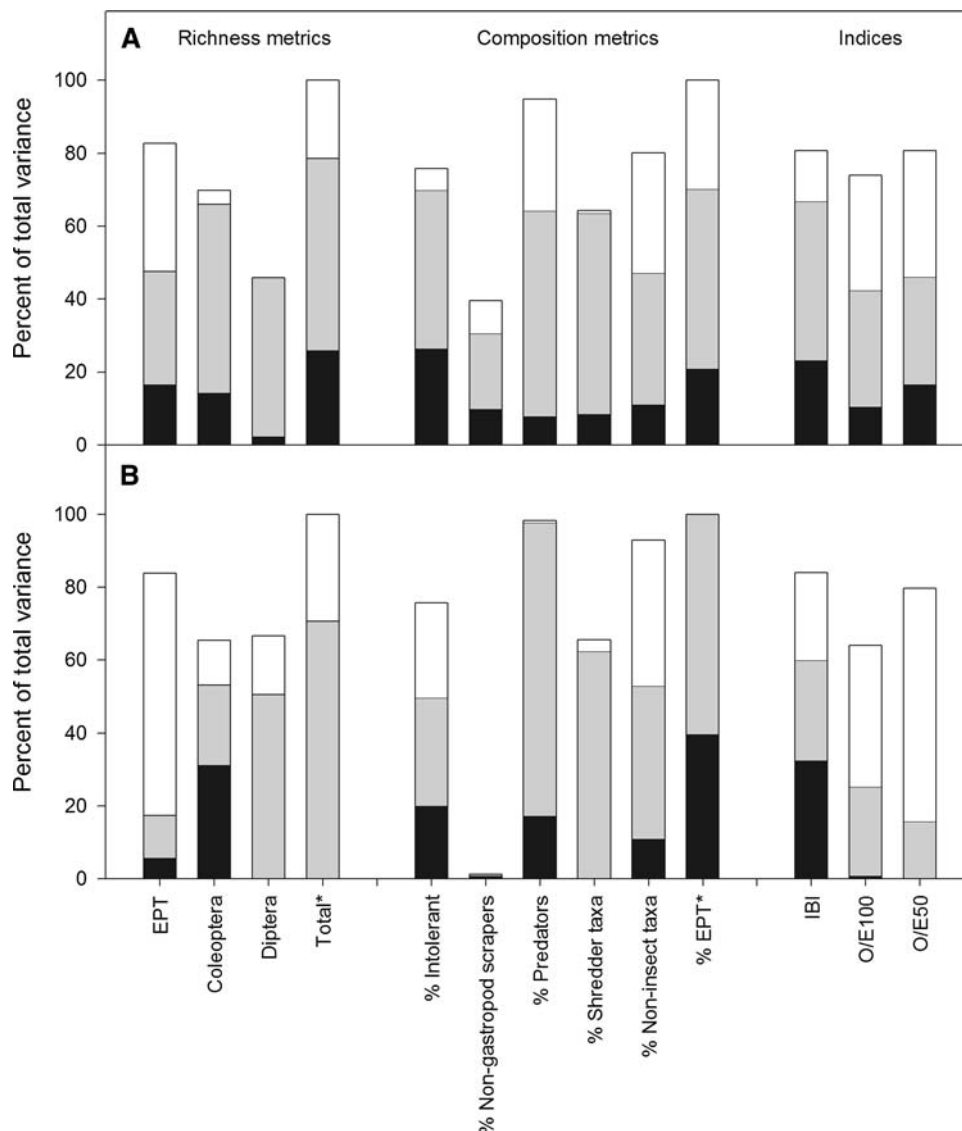
(e.g., Sandin and Johnson 2004). However, both the IBI and the two O/E scores had lower long-term variability than most individual metrics in the present study, indicating that these indices were relatively robust to inter-annual changes, and reflect the local conditions better than most single metrics. By combining metrics into a multimetric index, overall long-term variability may be reduced because metrics with lower variability (e.g., EPT richness) may dampen the influence of highly variable metrics (e.g., % intolerant). Furthermore, highly variable metrics may counteract each other out if they vary independently. The lower variability observed for the O/E scores may result from the fact that these indices are weighted towards taxa that were spatially common in the calibration data set. Studies have shown that spatially common taxa are often temporally common (Resh and others 2005), and therefore may introduce less long-term variability into the index.

Additionally, the use of long-term climatic variables (i.e., mean monthly precipitation and mean monthly temperature) as predictors in the California RIVPACS O/E models may incorporate some long-term variability in their estimates of E (i.e., expected number of taxa), albeit in a non-dynamic way.

However, it is not surprising that such high variability was observed for some of the biological metrics and indices in this study, which was designed to capture a large amount of spatial and temporal variability. Spatial variability was large (i.e., CV > 100%) because study sites represented a gradient of stream order and perenniality. Thus, despite the narrow geographical distribution and the small number of sites examined, considerable variability among sites was captured. Furthermore, annual variability was influenced and likely increased by sampling macro-invertebrates over a long time period that included both



**Fig. 7** Variance components for all metrics and indices. **a** Spatial variance components analysis. *White portions* of the bars represent the component of variability attributable to year. *Black portions* of the bars represent the component of variability attributable to site. *Gray portions* of the bars represent the interaction between site and year. Only spring samples were used to calculate these variance components. Residual variance is indicated by the difference between 100% and the total height of the bars. Residual variance was not estimated for metrics marked with an asterisk; **b** Seasonal variance components analysis. *White portions* of the bars represent the component of variability attributable to year. *Black portions* of the bars represent the component of variability attributable to season. *Gray portions* of the bars represent the interaction between season and year. Residual variance is indicated by the difference between 100% and the total height of the bars. Residual variance was not estimated for metrics marked with an asterisk. Only samples from site 1P were used to calculate these variance components



severe droughts and years of considerable rainfall. This high inter-annual variability of macroinvertebrate communities is typical of streams in mediterranean climates (Gasith and Resh 1999; Bêche and others 2006), and may represent an upper limit of variability for streams in more mesic climates.

Several factors contribute to inter-annual, seasonal, and spatial variability in benthic communities and bioassessment metrics, and perhaps the most important source of inter-annual variability is long-term variability in climate. Bêche and Resh (2007a, b) analyzing this data set found that persistent changes in macroinvertebrate community structure followed long-term droughts. For example, the drought from 1987 to 1991 precipitated major changes in community structure at all sites, particularly in sites 1D and 1P; these changes were associated with encroachment of macrophytes (*Typha* sp.) into the streambed during dry

years that lacked flows to remove vegetation. Likewise, Daufresne and others (2003) showed that rising water temperatures in the Upper Rhône River was correlated with long-term changes in fish and macroinvertebrate communities, including the replacement of cold-water species with thermophilic species.

Inter-annual variability may also arise from biological factors, which are not necessarily directly related to climate. Outbreaks of parasites or disease, and invasions of non-native species can cause short- and long-term changes in benthic community structure. For example, Kohler and Wiley (1997) demonstrated that outbreaks of the microsporidian pathogen *Cougourdella* decimated populations of a dominant caddisfly grazer in streams, shifting the invertebrate community to other grazer species, as well as to filter-feeders. A 15-years study on another microsporidian parasite of caddisflies showed that outbreaks occurred on a

multi-year cycle, causing population collapses approximately every 4 years (Kohler and Hoiland 2001). Other biotic forces, such as predation and competition, and invasion of non-native species, also may affect community structure over long-term cycles (Power and others 1988). For example, Einarsson and others (2002) saw that fluctuations in resource availability and inter-specific competition led to multi-year cycles in the abundance and body size of emerging chironomids in an Icelandic lake, although the authors observed that the fluctuations were ultimately driven by climatic cycles. There was no evidence that biotic interactions were a major source of long-term inter-annual variability in the present study, although such effects may be difficult if not impossible to detect using standard bioassessment protocols. Apart from invasion by non-native species, long-term changes in community structure resulting from biotic interactions have rarely been recorded in bioassessment studies (but see Marten 2001), and may represent an under-recognized cause of variability in benthic macroinvertebrate assemblages.

As with inter-annual variability, seasonal variability in benthic communities arises from both environmental and biological factors. In mediterranean-climate streams, environmental factors are particularly strong, as regular summer droughts results in flow reductions, changes in primary productivity, decreases in dissolved oxygen, and increases in conductivity over the course of the season (Gasith and Resh 1999). These changes may eliminate taxa that are not adapted to the different seasonal conditions. Life history may also dictate which species are found in which season. For example, larvae that are common in spring may emerge and oviposit before the summer sampling date. Additionally, biological factors like predation, parasitism, and competition may be intensified as low summer flows lead to higher densities of individuals, and more frequent opportunities for biotic interactions, such as competition and predation (Power and others 1988). We observed that the number of predators increased in summer samples, occasionally exceeding 50%, suggesting that predatory pressures changed seasonally. Bêche and others (2006) observed many seasonal differences in biological traits in this data set, showing that each season exerted distinct pressures, for which different sets of traits were suitable.

Sources of spatial variability arise from spatial differences in environmental factors that affect benthic communities. Numerous studies have focused on spatial variability of benthic communities at continental (Omernik 1995; Stoddard and others 2006; Ode and others 2008), watershed (e.g., Mazor and others 2006), reach (e.g., Sandin and Johnson 2004), and even micro-habitat (e.g., Needham and Usinger 1956; Gebler 2004) scales. As with inter-annual and seasonal variability, spatial variability

arises from both environmental and biological factors. Environmental factors include differences in geology, geomorphology, and climate, and these factors influence spatial variability at all scales. Biological factors arise from biogeographical differences, such as island neo-endemism (e.g., Polynesian black flies, Craig and others 2001) or range expansion (e.g., Pleistocene expansion into deglaciated regions of Europe, Bonada and others 2005). We would expect that biogeography had no influence on spatial variability among the sites in the study; rather, spatial variability was more likely influenced by differences in watershed area, stream order, and hydrologic regime present at these sites.

Inter-annual, seasonal, and spatial variability do not operate independently, and interactions among them may be large. Indeed, variance components analysis showed that interactions were the largest component of variability for many metrics in this study. Interactions between spatial and inter-annual variability arises from the site-specific manner in which long-term changes affect sites. Despite the fact that all sites in the study experienced a similar climate, climate affected each site differently. As noted earlier, the multi-year drought affected the first order streams most acutely, allowing encroachment of macrophytes into the channel. Several of the biological factors described above may also affect streams in a site-specific manner because streams may vary in their vulnerability to infections by parasites or invasions by non-native species (e.g., Kohler and Wiley 1997; Kohler and Hoiland 2001). Inter-annual variability may interact with seasonal variability by altering emergence times, and hastening or prolonging seasonal changes in flow and water chemistry (e.g., Wagner and others 2000). Although seasonal and spatial interactions were not addressed in this study, they may operate in a similar manner to inter-annual and spatial interactions, with first order sites being more acutely affected by summer drought than sites draining larger watersheds.

The high degree of site-specificity observed in our study may have been a result of the small sizes of the watersheds. Long-term studies in larger watersheds have showed higher consistency among sites, whereas studies in smaller watersheds have found large variability among sites. For example, in their long-term study of mainstem sites on the Rhône River, Daufresne and others (2003) noted a consistent change in species composition over time at all sites. In contrast, a long-term study of small watersheds in Wales found that changes in community structure were larger and more closely related to climate change in streams with neutral chemistry than in acidified streams (Durance and Ormerod 2007). As a result, interactive effects of inter-annual and spatial variability may be stronger in smaller watersheds.

Despite extensive studies showing a strong influence of season on invertebrate communities (Gasith and Resh 1999; Bonada and others 2006b; Bêche and others 2006), we found that seasonal variability was sometimes much lower than annual variability. We found that some metrics, such as EPT richness and O/E scores, were very responsive to seasonal changes. However, most metrics had lower seasonal than annual variability (particularly Coleoptera richness and % non-gastropod scrapers). This pattern suggests that benthic macroinvertebrates may be well adapted to the large yet predictable changes that occur in each season, but not as well adapted to the unpredictable changes that occur in certain years. Bêche and others (2006) found that annual shifts in community composition were much larger between years than between season. For example, in drought years, spring samples more closely resembled summer samples than spring samples taken from other years. Thus, long-term trends and inter-annual climatic factors affecting a stream community can be greater than the effects caused by intra-annual changes in season. In other words, seasonality of benthic macroinvertebrate communities is itself subject to inter-annual variability and is subordinate to the longer-term influence of year-to-year changes in environmental conditions.

High long-term variability in macroinvertebrate communities resulting from climate change or other changes in the natural environment can pose a problem for bioassessment programs. However, collection and analysis of long-term data is extremely useful in addressing these problems. For example, high variability may obscure real changes or may erroneously indicate deteriorating health when conditions actually represent a natural window of variability. Analysis of long-term data has led to insightful biomonitoring studies about the inefficacy of pollution remediation efforts (Linke and others 1999). For example, Scarsbrook and others (2000) demonstrated that improvements seen in impaired streams over 8 years could not be distinguished from similar changes observed in reference streams. Similarly, Marten (2001) showed that the supposed recovery of macroinvertebrate diversity in the Rhine River did not reflect a return to historic conditions but instead a shift to a new community dominated by recent invaders from the Danube River. Durance and Ormerod (2007) used long-term data to show that directional and cyclic changes in climate have distinguishable impacts on macroinvertebrate communities in streams with neutral chemistry. In these studies, long-term collection of data led to a better understanding of historic conditions and natural variability and prevented erroneous conclusions about pollution remediation efforts.

Despite the above examples, the magnitude of long-term variability in stream ecosystems has not been addressed by most bioassessment programs. Only a handful of programs

explicitly monitor sites for long-term trends analysis, although this number is growing (e.g., Stormwater Monitoring Coalition Bioassessment Working Group 2007). Moreover, we are unaware of any program that recalibrates assessment indices to incorporate long-term variability in establishing thresholds. A limited review of bioassessment programs showed that 3–4 years are typically used for index development or model calibration (e.g., Rosenberg and others 1999; Hill and others 2000; Ode and others 2005; Stoddard and others 2006), which may not adequately capture the full range of variability in climate, or in benthic community structure. The IBI in this study was calibrated with 4 years of data (2000–2003) and the O/E scores were calibrated with 2 years (2000–2001).

The fact that all indices and metrics suggested fluctuating conditions over time at these sites, which suffered few obvious disturbances and no changes in management, suggests that a snapshot approach to bioassessment may lead to erroneous conclusions about the health of certain sites. Ramifications for regulatory objectives are of great concern. For example, bioassessment programs may not be able to set reasonable thresholds to establish biocriteria when the indices on which they are based fluctuate greatly under natural conditions. This variability underscores the need for well designed studies that include reference sites and long-term data collection to distinguish the impacts anthropogenic and natural disturbances on benthic communities. Regulatory agencies may be unable to make proper determinations of impairment without the context provided by an adequate understanding of long-term variability.

Bioassessment programs can account for long-term variability in several ways. We observed that climatic variability was associated with metric and index fluctuations, although these associations varied among sites. Indeed, this study joins a growing body of research that supports the idea that bioassessment programs can measure impacts from climate change (e.g., Molles and Dahm 1990; Bonada and others 2007; Durance and Ormerod 2007). Bioassessment programs that invest in long-term monitoring at a network of reference and non-reference sites will be able to identify drivers of variability and prevent erroneous determinations of impairment. This approach may be particularly useful in predicting the effects of climate change on reference and non-reference sites (Bonada and others 2007). Furthermore, bioassessment programs can incorporate temporal variability into index development by using multiple years of data for calibration, perhaps requiring an iterative approach with regular updates to establish new thresholds. However, any approach must address the types of site-by-year interactions observed within this small study. Long-term monitoring at a large numbers of reference sites may identify the factors that drive these interactions.

Clearly, benthic communities experience considerable year-to-year variability. This variability is potentially large, and may lead to inaccurate assessments of specific sites, as well as poor precision in regional assessments. However, as this study demonstrates, there is potential to improve bioassessments by incorporating long-term variability in index calibration, and relating this variability to climatic variability.

In this study, site-by-year interactions were the largest component of variability for nearly all metrics and indices, implying that site-specific approaches may be required to separate these sources of variability. Clearly, benthic communities experience considerable year-to-year variability and because this variability is potentially large, it may lead to inaccurate assessments of specific sites, as well as poor precision in regional assessments. However, as this study demonstrates, there is potential to improve bioassessments by incorporating long-term variability in index calibration, and relating this variability to climatic patterns and changes.

**Acknowledgments** We thank Eric McElravy, Peter Connors, Leah Bêche, Kerry Ritter, Peter Ode, Charles Hawkins, John Gebler, John Van Sickle, Virginia Dale, and three anonymous reviewers for their assistance with this manuscript. We thank Andy Rehn for providing supplemental data. Funding was provided at various times over the course of the 20 years of this study by the Homestake Mining Company, Bodega Research Associates, The University of California at Berkeley, and the University of California Water Resources Center.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Bailey CR, Norris RH, Reynoldson TB (2004) Bioassessment of freshwater ecosystems: using the reference condition approach. Kluwer, The Netherlands
- Bêche LA (2005) Long term variability, disturbance, and biological traits of aquatic invertebrates in Mediterranean climate streams. PhD Dissertation, University of California, Berkeley
- Bêche LA, Resh VH (2007a) Short-term climatic trends affect the temporal variability in California ‘Mediterranean’ streams. *Freshwater Biology* 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. *Fundamental and Applied Limnology* 169(1):1–23
- Bêche LA, McElravy EP, Resh VH (2006) Long-term seasonal variation in biological traits of benthic-macroinvertebrates in two Mediterranean-climate streams in California, U.S.A. *Freshwater Biology* 51:56–75
- Bonada N (2003) Ecology of the macroinvertebrate communities in Mediterranean rivers at different scales and organization levels. PhD. Dissertation, University of Barcelona, Barcelona, Spain.
- Accessed online July 1, 2008: <http://www.tdx.cat/TDX-0722103-091734>
- Bonada N, Zamora C, Rieradevall M, Prat N (2005) Ecological and historical filters constraining spatial caddisfly distribution in Mediterranean rivers. *Freshwater Biology* 50:781–797
- Bonada N, Prat N, Resh VH, Statzner B (2006a) Development in aquatic insect biomonitoring: a comparative analysis of recent approaches. *Annual Review of Entomology* 51:495–523
- Bonada N, Rieradevall M, Prat N, Resh VH (2006b) Benthic macroinvertebrate assemblages and microhabitat connectivity in Mediterranean-climate streams of northern California. *Journal of the North American Benthological Society* 25:32–43
- Bonada N, Dolédec S, Statzner B (2007) Taxonomic and biological trait differences of stream macroinvertebrate communities between mediterranean and temperate regions: implications for future climate scenarios. *Global Change Biology* 13:1658–1671
- Chutter FM (1972) A reappraisal of Needham and Usinger’s data on the variability of a stream fauna when sampled with a Surber sampler. *Limnology and Oceanography* 17:139–141
- Courbois J-YP, Urquhart NS (2004) Comparison of survey estimates of the finite population variance. *Journal of Agricultural, Biological, and Environmental Statistics* 9:236–251
- Craig DA, Currie DC, Deirdre AJ (2001) Geographical history of the central-western Pacific black fly sub genus *Inselliellum* (Diptera: Simuliidae: *Simulium*) based on reconstructed phylogeny of the species, hot-spot archipelagoes and hydrological considerations. *Journal of Biogeography* 28:110–1127
- Daufresne M, Roger MC, Capra H, Lamouroux N (2003) Long-term changes within the invertebrate and fish communities of the Upper Rhône River: effects of climatic factors. *Global Change Biology* 10:124–140
- Durance I, Ormerod SJ (2007) Climate change effects on upland stream macroinvertebrates over a 25-year period. *Global Change Biology* 13:942–957
- Einarsson A, Gardarsson A, Gíslason GM, Ives AR (2002) Consumer-resource interactions and cyclic population dynamics of *Tanytarsus gracilentus* (Diptera: Chironomidae). *Journal of Animal Ecology* 71:832–845
- Gasith A, Resh VH (1999) Streams in mediterranean climate regions: abiotic influences and biotic responses to predictable seasonal events. *Annual Review of Ecology and Systematics* 30:51–81
- Gebler JB (2004) Mesoscale spatial variability of selected aquatic invertebrate community metrics from a minimally impaired stream segment. *Journal of the North American Benthological Society* 23:616–633
- Hill B, Herlihy AT, Kaufman PR, Stevenson RJ, McCormick FH, Burch Johnson C (2000) Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society* 19:50–67
- Hughes RM (1995) Defining acceptable biological status by comparing with reference conditions. In: Davies WS, Simon TP (eds) *Biological assessment and criteria: tools for water resource planning and decision making*. Lewis Publishers, Ann Arbor, pp 31–47
- Jackson JK, Füreder L (2006) Long-term studies of freshwater macroinvertebrates: a review of the frequency, duration and ecological significance. *Freshwater Biology* 51:591–603
- Kohler SL, Hoiland WK (2001) Population regulation in an aquatic insect: the role of disease. *Ecology* 82:2294–2305
- Kohler SL, Wiley MJ (1997) Pathogen outbreaks reveal large-scale effects of competition in stream communities. *Ecology* 78:2164–2176
- Larsen DP, Kincaid TM, Jacobs SE, Urquhart NS (2001) Designs for evaluating local and regional scale trends. *Bioscience* 51:1069–1078



- Linke S, Bailey RC, Schwindt J (1999) Temporal variability of stream bioassessments using benthic macroinvertebrates. *Freshwater Biology* 42:575–584
- Marten M (2001) Environmental monitoring in Baden-Württemberg with special reference to biocoenotic trend-monitoring of macrozoobenthos in rivers and methodical requirements for evaluation of long-term biocoenotic changes. *Aquatic Ecology* 35:159–171
- Mazor RD, Reynoldson TB, Rosenberg DM, Resh VH (2006) Effects of biotic assemblage, classification, and assessment method on bioassessment performance. *Canadian Journal of Fisheries and Aquatic Sciences* 63:394–411
- Molles MC, Dahm CN (1990) A perspective on El Niño and La Niña: global implications for stream ecology. *Journal of the North American Benthological Society* 9:68–76
- Needham PR, Usinger RL (1956) Variability in the macrofauna of a single riffle in Prosser Creek, California, as indicated by the Surber sampler. *Hilgardia* 14:383–489
- Ode PR, Rehn AC, May JT (2005) A quantitative tool for assessing the integrity of southern coastal California streams. *Environmental Management* 35:493–504
- Ode PR, Hawkins CP, Mazor RD (2008) Comparability of biological assessments derived from predictive models and multimetric indices of increasing geographic scope. *Journal of the North American Benthological Society* 27:967–985
- 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
- Power ME, Stout RJ, Cushing CE, Harper PP, Hauer FR, Matthews WJ, Moyle PB, Stazner B, Wais De Badgen IR (1988) Biotic and abiotic controls in river and stream communities. *Journal of the North American Benthological Society* 7:456–479
- Reckhow KH, Warren-Hicks W, Gibson G Jr (1997) *Biological criteria: technical guidance for survey design and statistical evaluation of biosurvey data*. Report #822-B-97-002. Health and Ecological Criteria Division, Office of Water, Environmental Protection Agency, Washington, DC
- Rehn AC, Ode PR, May JT (2005) Development of a benthic index of biotic integrity (B-IBI) for Wadeable streams in northern coastal California and its application to regional 305(b) assessment. Report to the State Water Resources Control Board. California Department of Fish and Game Aquatic Bioassessment Laboratory, Rancho Cordova, CA
- Rehn AC, Ode PR, Hawkins CP (2007) Comparison of targeted-riffle and reach-wide benthic macroinvertebrate samples: implications for data sharing in stream-condition assessments. *Journal of the North American Benthological Society* 26:332–348
- Resh VH (1994) Variability, accuracy, and taxonomic costs of rapid assessment approaches in benthic biomonitoring. *Bolletino di Zoologia* 61:375–383
- Resh VH, Jackson JK (1993) Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. In: Rosenberg DM, Resh VH (eds) *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman & Hall, New York, pp 195–233
- Resh VH, Norris RH, Barbour MT (1995) Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. *Australian Journal of Ecology* 20:108–121
- Resh VH, Bêche LA, McElravy EP (2005) How common are rare taxa in long-term macroinvertebrate surveys? *Journal of the North American Benthological Society* 24:976–989
- Richards AB, Rogers DC (2006) List of freshwater macroinvertebrate taxa from California and adjacent states including standard taxonomic effort levels. Southwest Association of Freshwater Invertebrate Taxonomists. Accessed online March 1, 2008: [http://www.swrcb.ca.gov/swamp/docs/safit/ste\\_list.pdf](http://www.swrcb.ca.gov/swamp/docs/safit/ste_list.pdf)
- Rosenberg DM, Resh VH (1989) Spatial temporal variability and the study of aquatic insects. *Canadian Entomologist* 121:941–963
- Rosenberg DM, Resh VH (1993) *Freshwater biomonitoring and benthic macroinvertebrates*. Chapman and Hall, New York, 488 pp
- Rosenberg DM, Reynoldson TB, Resh VH (1999) Establishing reference conditions for benthic invertebrate monitoring in the Fraser River catchment, British Columbia, Canada. FRAP Rep. No. DOE-FRAP 1998-32. Fraser River Action Plan, Environment Canada, Vancouver, BC CA. Accessed online July 1, 2006: <http://www.rem.sfu.ca/FRAP/9832.pdf>
- Sandin L, Johnson RK (2000) The statistical power of selected indicator metrics using macroinvertebrates for assessing acidification and eutrophication of running waters. *Hydrobiologia* 422(423):233–243
- Sandin L, Johnson RK (2004) Local, landscape and regional factors structuring benthic macroinvertebrate assemblages in Swedish streams. *Landscape Ecology* 19:501–514
- SAS Institute Inc (2004) SAS OnlineDoc 9.1.3. Cary, NC
- Scarsbrook MR, Boothroyd IKG, Quinn JM (2000) New Zealand's national river water quality network: long-term trends in macroinvertebrate communities. *New Zealand Journal of Marine and Freshwater Research* 34:289–302
- Schmitt RJ, Osenberg CW (1996) *Detecting ecological impacts: concepts and applications in coastal habitats*. Academic Press, San Diego
- Stoddard JL, Peck DV, Olsen AR, Larsen DP, Van Sickle J, Hawkins CP, Hughes RM, Whittier TR, Lomnický G, Herlihy AT, Kaufmann PR (2006) *Environmental monitoring and assessment program (EMAP) western streams and rivers statistical summary*. EPA 620/R-05/006. U.S. Environmental Protection Agency, Washington, DC
- Stormwater Monitoring Coalition Bioassessment Working Group (2007) *Regional monitoring of southern California's coastal watersheds*. Technical Report 539. Southern California Coastal Water Research Project, Costa Mesa, CA
- Tomanova S, Usseglio-Polatera P (2007) Patterns of benthic community traits in neotropical streams: relationship to meso-scale spatial variability. *Fundamental and Applied Limnology* 170(3):243–255
- University of California Davis Natural Reserve System (2003) *Natural history of the McLaughlin Reserve*. Napa, Lake, and Yolo Counties, California, 2nd edn. University of California, Davis
- Wagner R, Dapper T, Schmidt H-H (2000) The influence of environmental variables on the abundance of aquatic insects: a comparison of ordination and artificial neural networks. *Hydrobiologia* 422(423):143–152