

**National Water-Quality Assessment Program**

**Prepared in cooperation with the U.S. Environmental Protection Agency**

# **Design Tradeoffs for Trend Assessment in Aquatic Biological Monitoring Programs**

Open-File Report 2013–1222



National Water-Quality Assessment Program

Prepared in cooperation with the U.S. Environmental Protection Agency

# **Design Tradeoffs for Trend Assessment in Aquatic Biological Monitoring Programs**

By Martin E. Gurtz, John Van Sickle, Daren M. Carlisle, and Steven G. Paulsen

Open-File Report 2013–1222

**U.S. Department of the Interior**  
**U.S. Geological Survey**

**U.S. Department of the Interior**  
SALLY JEWELL, Secretary

**U.S. Geological Survey**  
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2013

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit <http://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod>

To order this and other USGS information products, visit <http://store.usgs.gov>

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Gurtz, M.E., Van Sickle, J., Carlisle, D.M., and Paulsen, S.G., 2013, Design tradeoffs for trend assessment in aquatic biological monitoring programs: U.S. Geological Survey Open-File Report 2013–1222, 15 p. plus 1 app.

# Contents

Abstract.....	1
Introduction.....	1
Purpose and Scope .....	2
Methods.....	2
Study Area and Sampling Methods.....	2
Sample Processing and Metric Calculation.....	5
Estimation of Variance Components and Minimum Detectable Trend .....	6
Trend Design Choices .....	7
Major Findings.....	7
Metrics.....	7
Design Elements.....	8
Sampling Period.....	8
Number of Samples Per Year.....	8
Sampling Interval.....	9
Number of Sites in Multiple-Site Designs .....	9
Balancing the Tradeoffs.....	10
Design [1]: Site-Specific Trend .....	10
Design [2]: Multiple Sites—Same Sites Resampled.....	10
Design [3]: Multiple Sites—New Sites Sampled.....	11
Discussion.....	12
Acknowledgments.....	13
References Cited.....	13
Appendix 1. Variance Component Equations.....	16
Design [1]: Site-Specific Trend .....	16
Design [2]: Trend in Multiple Sites—Same Sites Resampled.....	16
Design [3]: Trend in Multiple Sites—New Sites Sampled.....	17

# Figures

1. Locations of 32 streams and rivers sampled in this study.....	3
2. Variance components as a percentage of total variance for 17 metrics and indices commonly used in bioassessments.....	7
3. Minimum detectable trend at a site in Total Richness or MMI for designs with annual sampling and one, two, or three samples per year.....	8
4. Minimum detectable trend at a site in Total Richness and MMI for 10-year, 20-year, and 30-year sampling periods and different sampling intervals .....	9
5. Minimum detectable trend for multiple sites in Total Richness and MMI for different numbers of sites and sampling intervals.....	9

- 6. Percentage decrease in minimum detectable trend for an individual site, resulting from a 100-percent increase in sampling effort for different design choices and different metrics.....10
- 7. Percentage reduction in minimum detectable trend for multiple sites, resulting from a 100-percent increase in sampling effort for different design choices and different metrics.....11
- 8. Percentage reduction in minimum detectable trend for multiple sites, resulting from a 100-percent increase in sampling effort for different design choices and different metrics.....11

## Tables

- 1. Watershed characteristics of the 32 sites in this study.....3
- 2. Terms used in this report.....5
- 3. Invertebrate metrics and indices, with mean and range for all samples, grouped by type of data. ....6



# Design Tradeoffs for Trend Assessment in Aquatic Biological Monitoring Programs

By Martin E. Gurtz<sup>1</sup>, John Van Sickle<sup>2</sup>, Daren M. Carlisle<sup>1</sup>, and Steven G. Paulsen<sup>3</sup>

## Abstract

Assessments of long-term (multiyear) temporal trends in biological monitoring programs are generally undertaken without an adequate understanding of the temporal variability of biological communities. When the sources and levels of variability are unknown, managers cannot make informed choices in sampling design to achieve monitoring goals in a cost-effective manner. We evaluated different trend sampling designs by estimating components of both short- and long-term variability in biological indicators of water quality in streams. Invertebrate samples were collected from 32 sites—9 urban, 6 agricultural, and 17 relatively undisturbed (reference) streams—distributed throughout the United States. Between 5 and 12 yearly samples were collected at each site during the period 1993–2008, plus 2 samples within a 10-week index period during either 2007 or 2008. These data allowed calculation of four sources of variance for invertebrate indicators: among sites, among years within sites, interaction among sites and years (site-specific annual variation), and among samples collected within an index period at a site (residual). When estimates of these variance components are known, changes to sampling design can be made to improve trend detection. Design modifications that result in the ability to detect the smallest trend with the fewest samples are, from most to least effective: (1) increasing the number of years in the sampling period (duration of the monitoring program), (2) decreasing the interval between samples, and (3) increasing the number of repeat-visit samples per year (within an index period). This order of improvement in trend detection, which achieves the greatest gain for the fewest samples, is the same whether trends are assessed at an individual site or an average trend of multiple sites. In multiple-site surveys, increasing the number of sites has an effect similar to that of decreasing the sampling interval; the benefit of adding sites is greater when a new set of different sites is selected for each sampling effort than when the same sites are sampled each time. Understanding variance components of the ecological attributes of interest can lead to more cost-effective monitoring designs to detect trends.

## Introduction

Long-term monitoring studies contribute valuable information toward understanding the condition of our natural resources and how they respond to human influences and natural variation (Hirsch and others, 2006; Jackson and Füreder, 2006; Peterson and others, 2011). Such studies are rare, however, and are susceptible to budget constraints that may jeopardize the continuity of monitoring networks. Because of limited data on the sources of variation in trends and factors affecting the variability of ecological attributes of interest, design choices are often made primarily to reduce costs, with minimal consideration of effects on the goals of trend detection. This study estimated temporal components of variance in long-term biological monitoring data to evaluate how different monitoring design choices affect the ability to detect trends.

Biological monitoring programs that assess the ecological condition of streams and rivers typically focus on macroinvertebrates, although sometimes supplemented by collections of fish or algae. Long-term studies of freshwater invertebrates are uncommon, particularly those that span 10 or more years—with a few notable exceptions, including Robinson and others (2000); Beche and others (2006); Mazor and others (2009); Ormerod and Durance (2009); Webb and King (2009); Peterson and others (2011). Even fewer studies have used a common approach and methods over time at more than a local scale (Davies-Colley and others, 2011; Kennen and others, 2012; Miller and others, 2012). More long-term investigations are needed to enable researchers and resource managers to relate changes in stream health to local, regional, and global phenomena, including climate change (Daufresne and others, 2004; Durance and Ormerod, 2007; Mazor and others, 2009; Lawrence and others, 2010). Such data also would provide information about sources of temporal variation that could be used in the design of long-term monitoring programs.

Monitoring programs—whether intended to assess status and change at individual sites or a group of sites—can be designed in several ways that influence trend detection,

<sup>1</sup> U.S. Geological Survey, National Water-Quality Assessment Program.

<sup>2</sup> Oregon State University, Department of Fisheries and Wildlife

<sup>3</sup> U.S. Environmental Protection Agency, Office of Research and Development



with tradeoffs in the program cost (number of samples) and duration (number of years in the sampling period) required to achieve the program's objectives. These design choices include: (1) increasing the sampling period (the total elapsed time, in years, over which samples are collected), (2) decreasing the sampling interval (number of years between sampling efforts), and (3) increasing the number of samples collected each year. When the goal is to assess the average trend throughout a stream network, a fourth design element—the number of sites to be sampled—can be adjusted. This report focuses on these four elements of the sampling design—the allocation of sampling effort within and among years and across sites (Stevens and Urquhart, 2000; Larsen and others, 2001).

Sampling design decisions for monitoring programs are ideally founded on an understanding of the temporal variability of biological communities and other ecological endpoints of interest. Program designers should consider sources of variation that affect the ability of the sampling program to detect change in some attribute of interest over time. One framework for understanding these sources of variation (Urquhart and others, 1998; Larsen and others, 2001) identifies four major variance components: (1) site-to-site geographic variation, (2) coherent (or synchronous) year-to-year variation expressed by all of the sites together, (3) site-by-year interaction with differing temporal patterns across sites, due to site-specific factors, and (4) residual variation, composed of temporal variation within the sampling season as well as sampling error. If coherent temporal variation is large, the only option is to sample over a longer time period; however, other sources of variation can be managed through design and sample size choices (Kincaid and others, 2004).

Information about how variance components influence design tradeoffs—for example, number of replicates, number of sites, and number of times to sample—is often limited (Philippi and others, 1998). Estimation of variance components for trend assessments requires data collection that is designed to incorporate both short-term (for example, within year) and long-term (among years) temporal sampling. Such studies have been done for evaluating trends in the ecological condition of lakes (Larsen and others, 1995; Urquhart and others, 1998; Larsen and others, 2001; Kincaid and others, 2004); related studies of streams have been limited to a small number of sites (Mazor and others, 2009). Most studies that have estimated variance components in monitoring programs have examined how these components affect statistical power for detecting trends (Wagner and others, 2007; Al-Chokhachy and others, 2011; Anlauf and others, 2011). The emphasis in this report is not on statistical power but on the magnitude of trend that can be detected, so that the findings can be more directly applied to management decisions on the design of monitoring programs.

## Purpose and Scope

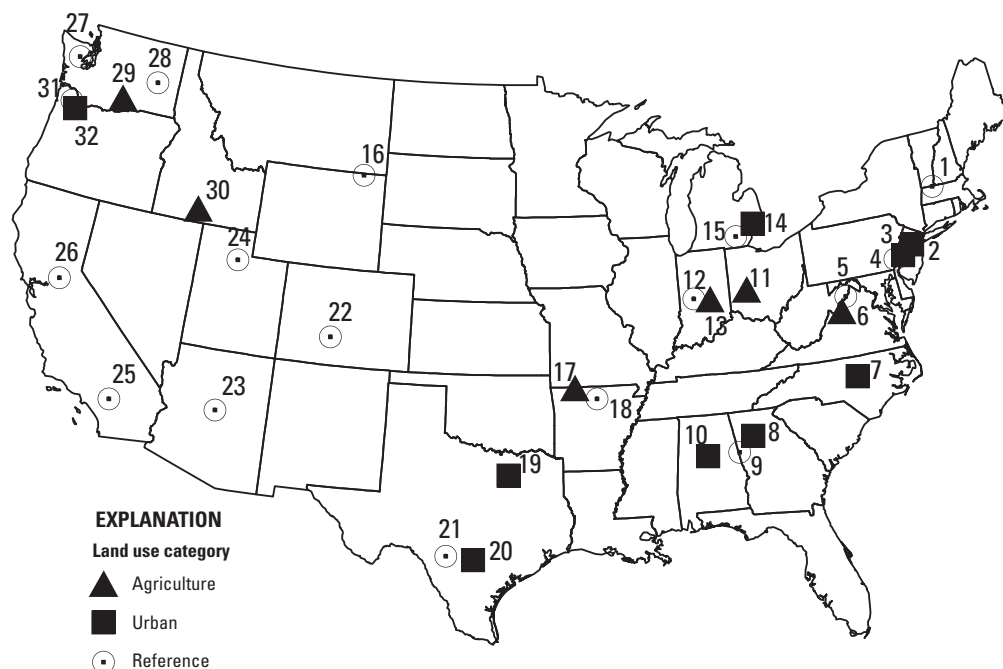
This report presents the research and findings of a joint effort of the U.S. Geological Survey National Water-Quality Assessment (USGS NAWQA) Program and the U.S. Environmental Protection Agency (EPA) National Rivers and Streams Assessment. The objective of this study was to evaluate how different monitoring program designs affect trend detection in invertebrate communities in streams throughout the United States at sites sampled by the NAWQA Program. We follow an approach similar to Larsen and others (1995, 2001) and estimate spatial and temporal variance components to quantify the magnitude of trend that could be detected for different monitoring designs—varying the number of samples collected, the number of years over which the monitoring was conducted, and the number of sites—and evaluate tradeoffs associated with different trend design choices.

## Methods

### Study Area and Sampling Methods

We selected 32 sites that were sampled as part of the NAWQA Program (<http://water.usgs.gov/nawqa/>), distributed across a broad range of geographic regions and a wide range of human influences (fig. 1, table 1). Seventeen sites represented least-disturbed conditions, referred to in this paper as “reference” sites, 9 sites were dominated by urban land use, and 6 sites represented agricultural land use. All sites were Wadeable, draining 19 to 3,204 km<sup>2</sup>, with mean annual discharge ranging from 0.1 to 14.5 m<sup>3</sup>/s (table 1).

Invertebrate community samples were collected during 1993–2008 during a 6-week index period specified for each site, generally during summer low flow. Index periods were selected so that annual sampling was conducted under similar seasonal conditions from one year to the next (“index period” and other terms used in this report are defined in table 2). During either 2007 or 2008, two samples were collected at each site within an extended 10-week index period to allow estimation of variation within the index period; for that year, the time between samples at a site ranged from 27 to 64 days, with one exception—a site with only 11 days between samples. The span of calendar days at any site, across all sample years, ranged from 36 to 67. The number of samples at a site ranged from 6 to 13, representing from 5 to 12 distinct years; 24 of 32 sites had at least 8 years of biological data.



**Figure 1.** Locations of 32 streams and rivers sampled in this study (6 agriculture, 9 urban, and 17 reference sites). Land use refers to the dominant land use in the watershed as defined by Gilliom and others (2006).

**Table 1.** Watershed characteristics of the 32 sites in this study.

[Site locations are shown in figure 1. Ecoregions are Omernik Level III (Omernik, 1987). km<sup>2</sup>, square kilometers; m, meters; m<sup>3</sup>/s, cubic meters per second]

Site	Station number	Station name	Land use	Drainage area (km <sup>2</sup> )	Elevation (m)	Mean annual discharge (m <sup>3</sup> /s)	Ecoregion at sampling reach
1	01170100	Green R nr Colrain, Massachusetts	Reference	107	133	2.6	Northeastern Highlands
2	01403900	Bound Brook at Middlesex, New Jersey	Urban	125	8	2.2	Northern Piedmont
3	01464907	Little Neshaminy Cr nr Neshaminy, Pennsylvania	Urban	69	58	1.5	Northern Piedmont
4	01472157	French Cr nr Phoenixville, Pennsylvania	Reference	153	49	2.6	Northern Piedmont
5	01610400	Waites Run nr Wardensville, West Virginia	Reference	33	378	0.5	Middle Atlantic Coastal Plain
6	01621050	Muddy Cr at Mount Clinton, Virginia	Agriculture	37	402	0.3	Ridge and Valley
7	02087580	Swift Cr nr Apex, North Carolina	Urban	54	89	0.7	Southeastern Plains
8	02335870	Sope Cr nr Marietta, Georgia	Urban	80	269	1.4	Piedmont
9	02338523	Hillabahatchie Cr nr Franklin, Georgia	Reference	44	277	0.6	Piedmont
10	0242354750	Cahaba Valley Cr at Pelham, Alabama	Urban	66	134	1.2	Ridge and Valley
11	03267900	Mad R nr Springfield, Ohio	Agriculture	803	276	9.3	Eastern Corn Belt Plains

#### 4 Design Tradeoffs for Trend Assessment in Aquatic Biological Monitoring Programs

**Table 1.** Watershed characteristics of the 32 sites in this study.—Continued

[Site locations are shown in figure 1. Ecoregions are Omernik Level III (Omernik, 1987). km<sup>2</sup>, square kilometers; m, meters; m<sup>3</sup>/s, cubic meters per second]

Site	Station number	Station name	Land use	Drainage area (km <sup>2</sup> )	Elevation (m)	Mean annual discharge (m <sup>3</sup> /s)	Ecoregion at sampling reach
12	03357330	Big Walnut Cr nr Roachdale, Indiana	Reference	339	244	5.1	Eastern Corn Belt Plains
13	03361650	Sugar Cr at New Palestine, Indiana	Agriculture	243	240	3.1	Eastern Corn Belt Plains
14	04161820	Clinton R at Sterling Heights, Michigan	Urban	800	184	7.8	Southern Michigan/ Northern Indiana Drift Plains
15	04175600	River Raisin nr Manchester, Michigan	Reference	342	274	3.0	Southern Michigan/ Northern Indiana Drift Plains
16	06324970	Little Powder R nr Weston, Wyoming	Reference	3,204	1,039	0.6	Northwestern Great Plains
17	07053250	Yocum Cr nr Oak Grove, Arkansas	Agriculture	137	305	1.4	Ozark Highlands
18	07060710	North Sylamore Cr nr Fifty Six, Arkansas	Reference	150	133	1.4	Ozark Highlands
19	08057200	White Rock Cr at Dallas, Texas	Urban	172	122	2.5	Texas Blackland Prairies
20	08178800	Salado Cr at San Antonio, Texas	Urban	490	161	1.3	Texas Blackland Prairies
21	08195000	Frio R at Concan, Texas	Reference	1,008	367	3.5	Edwards Plateau
22	08227000	Saguache Cr nr Saguache, Colorado	Reference	1,541	2,448	1.8	Southern Rockies
23	09505800	West Clear Cr nr Camp Verde, Arizona	Reference	624	1106	1.7	Arizona/New Mexico Mountains
24	10172200	Red Butte Cr nr Salt Lake City, Utah	Reference	19	1646	0.1	Wasatch and Uinta Mountains
25	11051500	Santa Ana R nr Mentone, California	Reference	544	594	0.7	Southern California Mountains
26	11335000	Cosumnes R at Michigan Bar, California	Reference	1388	51	14.1	Southern and Central California Chaparral and Oak Woodlands
27	12056500	North Fork Skokomish R nr Hoodspont, Washington	Reference	148	232	14.5	Coast Range
28	12464770	Crab Cr nr Ritzville, Washington	Reference	995	506	1.3	Columbia Plateau
29	12505450	Granger Drain at Granger, Washington	Agriculture	161	219	1.0	Columbia Plateau
30	13092747	Rock Cr at Twin Falls, Idaho	Agriculture	671	1106	3.1	Snake River Plain
31	14205400	East Fork Dairy Cr nr Meacham Corner, Oregon	Reference	88	88	2.4	Willamette Valley
32	14206950	Fanno Cr at Durham, Oregon	Urban	82	35	1.2	Willamette Valley

**Table 2.** Terms used in this report.

Term	Definition
Index period	Interval of time such as 6 weeks or 10 weeks within which sampling occurs during a calendar year; for example, June 1–July 15.
Variance components	Sources of variability.
SITE	Site-to-site geographic variation.
YEAR	Coherent (or synchronous) year-to-year variation expressed by all sites together.
SITE:YEAR	Year-to-year variation that differs among sites due to site-specific factors (site-by-year interaction).
WITHIN INDEX PERIOD	Temporal variation within the sampling (index) period; this variance component includes sampling error—for example, due to differences in field crews and measurement error in field and laboratory analyses (residual variation).
Minimum Detectable Trend (MDT)	Smallest trend that can be detected within a defined confidence interval; for example, 95%.
Sampling period	Total elapsed time (years) over which samples are collected; for example, a 10-year sampling period.
Sampling interval	Number of years between sampling efforts (Index Periods); for example, a 1-year sampling interval is the same as annual sampling, although multiple samples may be collected within a given year (Index Period).
Number of samples per year	Number of (repeat-visit) samples collected within the same year (within the Index Period).
Multiple-site designs	Trend monitoring designs for estimating average trends among multiple sites over a period of time (sampling period).
Same sites resampled	A trend monitoring design in which each site is sampled during each sampling effort; for example, each site is sampled (“revisited”) every year.
New sites sampled	A trend monitoring design in which a new set of sites is selected for each sampling effort.

Invertebrate samples were collected following the Richest-Targeted Habitat (RTH) method of the NAWQA Program (Cuffney and others, 1993; Moulton and others, 2002). Five discrete collections of 0.25 m<sup>2</sup> each were taken from riffle habitat with a Slack sampler equipped with 500- $\mu$  mesh within a predefined reach. Collections were composited into a single sample, passed through a 500- $\mu$ -mesh sieve, preserved in a 10-percent buffered formaldehyde solution in the field, and transported to the laboratory.

### Sample Processing and Metric Calculation

All samples were processed at the USGS National Water-Quality Laboratory in Denver, Colorado, according to standard procedures (Moulton and others, 2000). Most taxa were identified to genus, but some non-insect taxa were identified to family or higher levels of taxonomic resolution. All pupae and non-aquatic taxa or lifestages were removed from the dataset.

Taxonomic identifications, mostly at the genus level or higher, were made at the USGS laboratory and harmonized with operational taxonomic units (OTUs) used by EPA’s

Wadeable Streams Assessment (WSA) (U.S. Environmental Protection Agency, 2006; Yuan and others, 2008); this ensured that individual taxa were consistently counted as a single taxonomic level across all samples. We standardized samples to a computerized 300 count by randomly resampling (without replacement) individuals in each sample.

Decisions on ambiguous taxa (identifications that were not made to the resolution of the OTU; Cuffney and others, 2007) were made independently for each site. Specimens that could not be identified to the target level were assigned a taxon only if the assignment could be made unambiguously (for example, when only one taxon was found in any of the samples from that station); otherwise these specimens were removed from the dataset.

We examined a set of invertebrate metrics and indices commonly used in monitoring programs (table 3). Metrics—including functional, structural, and tolerance characteristics of invertebrate communities—were all candidates for the national-scale Multimetric Index (MMI) used in the WSA (U.S. Environmental Protection Agency, 2006; Stoddard and others, 2008).

## 6 Design Tradeoffs for Trend Assessment in Aquatic Biological Monitoring Programs

**Table 3.** Invertebrate metrics and indices (from Stoddard and others, 2008), with mean and range for all samples, grouped by type of data.

[EPT=Ephemeroptera, Plecoptera, Trichoptera; MMI=Multimetric Index; PTV=Pollution Tolerance Value (values range from 0 to 10 and were assigned to categories); WSA=Wadeable Streams Assessment]

Metric	Definition	Mean	Range
Richness			
EPT Richness	Number of EPT taxa	9.0	0–28
Scraper Richness	Number of scraper taxa	4.9	0–14
Clinger Richness	Number of clinger taxa	13.9	5–29
Mayfly Richness	Number of mayfly taxa	4.3	0–11
Shredder Richness	Number of shredder taxa	3.5	0–8
Intolerant Richness	Number of intolerant taxa	5.1	0–20
Total Richness	Total number of taxa	26.0	11–46
Percent taxa			
% EPT Taxa	Percentage of EPT taxa	33.8	0–68
% Mayfly Taxa	Percentage of mayfly taxa	16.0	0–38
% Chironomid Taxa	Percentage of chironomid taxa	25.9	0–61
% Non-Insect Taxa	Percentage of non-insect taxa	18.9	0–59
% Clinger Taxa	Percentage of clinger taxa	53.3	24–78
% Burrower Taxa	Percentage of burrower taxa	11.9	0–30
% Taxa with PTV 0–5.9	Percentage of taxa with PTV 0–5.9	66.1	31–89
% Taxa with PTV 8–10	Percentage of taxa with PTV 8–10	7.4	0–32
Percent individuals			
% EPT Individuals	Percentage of EPT individuals	44.3	0–91
% Non-Insect Individuals	Percentage of non-insect individuals	16.9	0–91
% Individuals Top 3 Taxa	Percentage of individuals in the three most-dominant (most-abundant) taxa	53.9	25–95
% Individuals Top 5 Taxa	Percentage of individuals in the five most-dominant (most-abundant) taxa	67.5	38–97
% Tolerant Individuals	Percentage of tolerant individuals	10.4	0–75
Indices			
Shannon Diversity	Shannon diversity index	2.4	0.7–3.3
MMI	WSA National Multi-Metric Index	46.7	5.6– 90.1

### Estimation of Variance Components and Minimum Detectable Trend

We estimated four components of variance for each metric, using restricted maximum likelihood estimation (REML; Littell and others, 2006): (1) among **SITES**, (2) among **YEARS** within sites, (3) the **SITE:YEAR** interaction, and (4) among samples collected within the index period of 1 year at each site (**WITHIN INDEX PERIOD**) (Larsen and others, 2001). The **YEAR** component represents the “coherent” (or concordant) annual variability that is common to all sites; this is the degree to which a dry year that occurs nationwide would affect invertebrates similarly at all sites. The **SITE:YEAR** interaction expresses any additional **YEAR** effects that differ

among sites, such as wet or dry years that occur in different parts of the country. The **WITHIN INDEX PERIOD** component is the residual variance among multiple samples collected at one site within the index period and year, including other local sources of error such as differences in field crews and measurement error in field and laboratory analyses. We also tested whether each variance component differed from zero ( $P < 0.05$ ; Littell and others, 2006). Components are expressed as proportions of their sum, defined as the total variance. REML estimation occasionally yields negative estimates for small components, which we set equal to zero prior to calculating totals and proportions.

We assume that the goal is to detect a linear trend of the form  $Y = A + B*YEAR$ , where  $Y$  is an invertebrate index

or metric score, and the trend slope ( $B$ ) has units of change per year in the invertebrate score. If  $b$  is the estimate of  $B$  based on sample data, then its variance,  $Var(b)$ , determines the magnitude of trend that would be declared significantly ( $P < 0.05$ ) different from zero. Specifically, a 95 percent confidence interval (CI) for  $B$  is centered on  $b$ , with a half-width of  $1.96 * \sqrt{Var(b)}$ . Thus, the estimated Minimum Detectable Trend (MDT, the slope of the linear trend line) is  $MDT = 1.96 * \sqrt{Var(b)}$ , because the 95 percent CI centered on any trend estimate that exceeds MDT will not include the value  $B = 0$  and thus will be declared significant (Larsen and others, 1995).

## Trend Design Choices

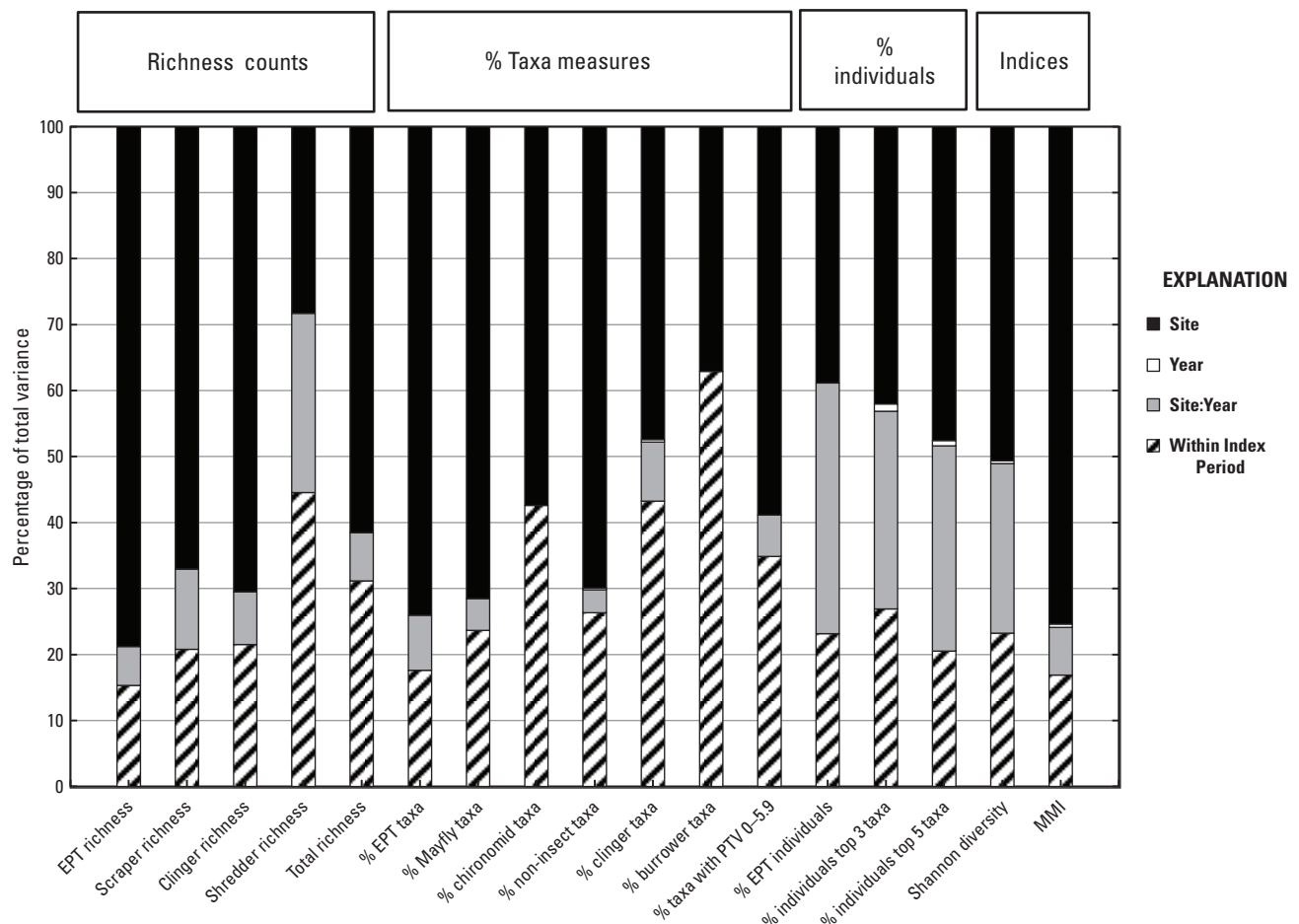
We calculated  $Var(b)$  and MDT for different scenarios to illustrate tradeoffs of alternative design choices for detection of a trend at a single site and an average trend of multiple sites. For a single site (Design [1]), we examined the effect of changing the sampling period (increasing the number of years), the sampling interval (number of years between sampling efforts), and the number of samples collected at a site each year (within the restricted index period). For multiple

sites, we also examined the effect of changing the number of sites for two designs: all sites revisited for each sampling effort (Design [2]: “same sites”), and a new set of different sites selected for each sampling effort (Design [3]: “new sites”). Equations adapted from Larsen and others (1995, 2001) for calculating  $Var(b)$  from variance components for these three trend designs are included in Appendix 1.

## Major Findings

### Metrics

Of the 22 metrics and indices for which variance components were calculated, 5 were eliminated from further analysis because of nonnormal distributions (determined by visual inspection): Mayfly Richness, % Non-Insect Individuals, Intolerant Richness, % Tolerant Individuals, and % Taxa with a Pollution Tolerance Value (PTV) of 8–10. Of the 4 variance components, **SITE** represented the highest percentage (39–75 percent) of total variance for 15 of the 17 metrics (fig. 2); because these data are from sites that span the entire United States and different land uses, **SITE** variance estimates



**Figure 2.** Variance components as a percentage of total variance for 17 metrics and indices commonly used in bioassessments. (EPT, Ephemeroptera, Plecoptera, Trichoptera; PTV, Pollution Tolerance Value; MMI, Multimetric Index).

are probably much larger than would be observed in any single region or basin. **WITHIN INDEX PERIOD** accounted for 7–63 percent of total variance. **SITE:YEAR** interaction was a higher proportion of total variance for metrics that were based on percent individuals than for other metric categories. Variance due to **YEAR** was zero or negligible for nearly all metrics, which is not surprising, because this variance component represents consistent annual variation for all sites, and sites were distributed among diverse settings throughout the United States. Other than **YEAR** variance, at least two of the remaining three variance components were significantly greater than zero for all metrics; significance levels may be overestimated because we did not do familywise corrections for these tests.

Two commonly used measures—Total Richness and the MMI—were selected for examining responses to changes in trend design elements. Responses of selected other metrics are contrasted to highlight the effects of different relative proportions of variance components. Our goal is not to ascertain the “best” metric; we assume that choice of metrics for trend monitoring is based on the objectives of the program.

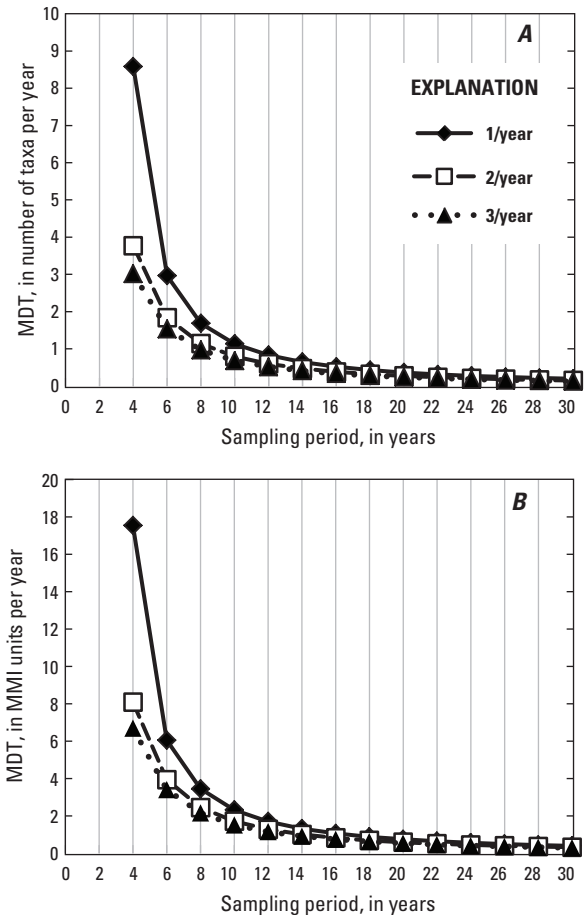
## Design Elements

Design elements that can be modified to improve trend detection, such as a reduction in the minimum detectable trend, include the sampling period, the number of samples collected each year, the sampling interval, and, for multiple-site designs, the number of sites. These design choices affect the MDT in different ways that depend on a combination of mathematical relations, ecological attributes that are measured, and characteristics of the sites.

## Sampling Period

Sampling period affects trend detection in mathematically predictable ways, according to the equations of Larsen and others (2001; see Appendix 1). Only the term for the sum of squares in each of the three equations is affected by sampling period, and this term (the denominator) does not contain any variance components. Thus, the proportional effect of a change in period on the variance of trend slope is the same, regardless of variance component values. Data from our study provide context to these equations to demonstrate how trend detection changes when certain design elements are altered.

As predicted by the variance slope equations, the site-specific MDT for Total Richness (fig. 3A) and MMI (fig. 3B) declined sharply as the sampling period increased. Based on our estimated variances, the rates of reductions in MDT were much lower after about 10 years. The change in metric that is detectable over the sampling period can be obtained by multiplying the MDT (trend slope) by the number of years; thus, the change in the number of taxa that is detectable after 4 years is very large at 34 ( $8.6 \times 4$ ), but the number falls to 11 taxa after 10 years ( $1.1 \times 10$ ), 7 taxa after 20 years, and 6 taxa after 30 years.



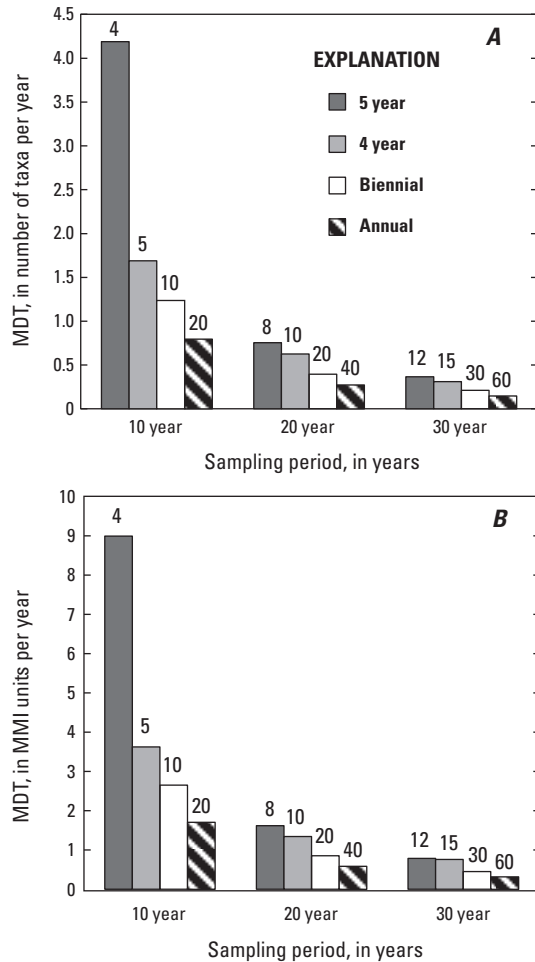
**Figure 3.** Minimum detectable trend (MDT) at a site in Total Richness (A) or MMI (B) for designs with annual sampling and one, two, or three samples per year. Units for the Y-axis are number of taxa per year (A) and MMI units per year (B). (MDT, Minimum Detectable Trend; MMI, Multimetric Index).

## Number of Samples Per Year

Increasing the number of samples per year (collected within the index period) strongly reduced the MDT at a site, but only for short sampling periods (fig. 3). The results showed that the MDT for Total Richness for a single-site trend with a 4-year sampling period dropped sharply from nine to four taxa/year as the number of samples/year increased from one to two, with a smaller decline in MDT to three taxa/year if three samples/year were collected; reductions in MDT were negligible after about 10 years (fig. 3A). MMI showed a similar pattern with increasing samples/year (fig. 3B). Reductions in MDT for all trend designs from increasing the number of samples/year are higher for metrics with relatively high **WITHIN INDEX PERIOD** variance, and(or) low **YEAR** variance, according to the equations of Larsen and others (2001; see Appendix 1).

## Sampling Interval

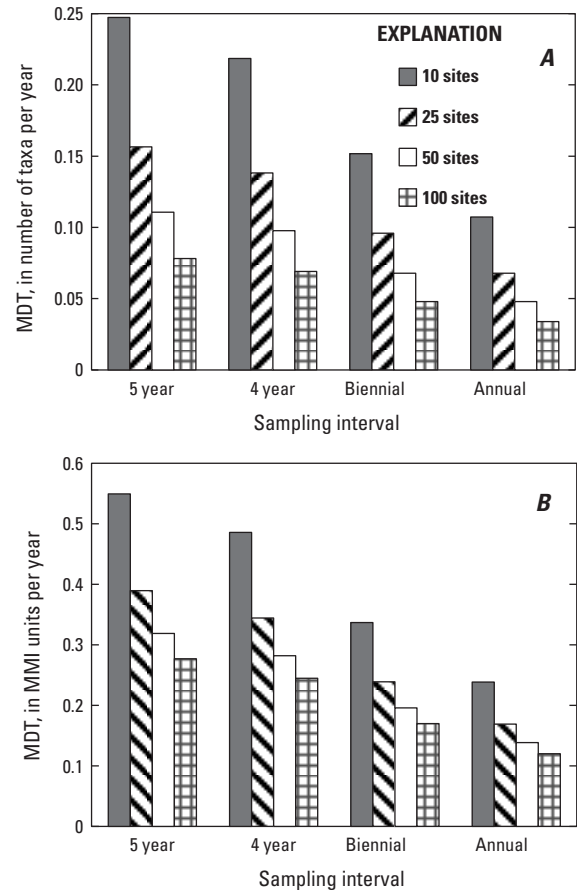
As with sampling period, declines in MDT as the sampling interval decreases are also predicted by the variance slope equations of Larsen and others (2001; see Appendix 1). Based on our data, the reduction in MDT for a site-specific trend was lower for longer sampling periods (fig. 4). For example, a decrease in sampling interval from 5 years to annual for a 10-year sampling period, which increases the number of samples from 4 to 20, reduced the MDT for Total Richness from 4.2 to 0.8 taxa/year (fig. 4A). However, over a 30-year sampling period, which provided an increase from 12 to 60 samples, MDT was reduced much less, decreasing from 0.4 to 0.2 taxa/year. The pattern was the same for MMI (fig. 4B). As with sampling period, a decrease in sampling interval has the same proportional effect on MDT for all metrics, because only the denominator in the equation for the variance of trend slope is modified (see equations in Appendix 1).



**Figure 4.** Minimum detectable trend (MDT) at a site in Total Richness (A) and MMI (B) for 10-year, 20-year, and 30-year sampling periods and different sampling intervals; all cases assume two samples in each sampled year. Total numbers of samples collected over the sampling period are shown above each bar. (MDT, Minimum Detectable Trend; MMI, Multimetric Index).

## Number of Sites in Multiple-Site Designs

Multiple-site designs have the additional option to modify the number of sites. For our data and a same-site design (same sites resampled for each effort), the MDT for Total Richness decreased 68 percent as the number of sites in the sampling design was increased from 10 to 100 sites, regardless of the sampling interval (fig. 5A). The MDT is the average trend slope for the total number of sites; the magnitude of change that can be detected in multiple-site designs is lower than for single sites—for example, 2.2 taxa ( $0.11 \times 20$ ) for 10 sites with annual sampling for 20 years. The pattern was the same for MMI (fig. 5B), with a corresponding decrease of 50 percent. In general, MDT decreased as the number of sites increased for both same-site and new-site designs, with proportional differences in reductions in both designs dependent on variance components of the metrics.



**Figure 5.** Minimum detectable trend (MDT) for multiple sites (same sites resampled) in Total Richness (A) and MMI (B) for different numbers of sites and sampling intervals (all designs with 1 sample/year and a 20-year sampling period). (MMI, Multimetric Index).



## Balancing the Tradeoffs

Any design change to improve trend detection incurs a cost, in terms of sampling effort (number of samples), or time (sampling period), or both. For example, increasing the sampling interval from annual to biennial can reduce the sampling effort in half for each year, but achieving the same level of trend detection would require extending the sampling period for twice as long, thus incurring a cost in terms of time or a delay in trend detection. Although tradeoffs can be evaluated in terms of the proportional gain in trend detection (in other words, reduction in MDT) relative to the increase in sampling effort, the actual MDT should determine the merit of any particular design for meeting program goals.

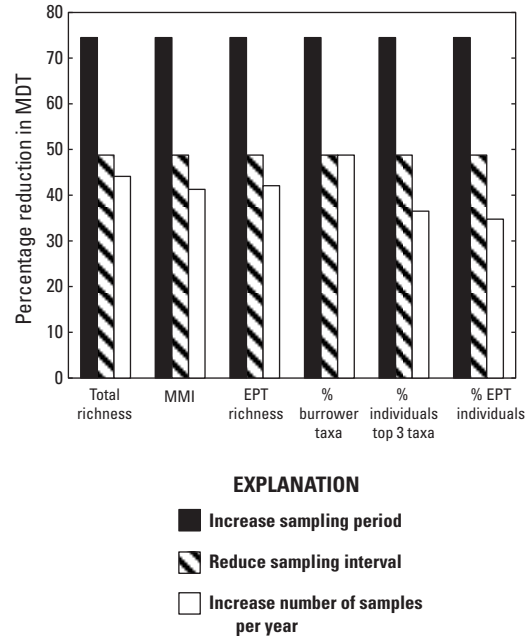
### Design [1]: Site-Specific Trend

Tradeoffs among design choices can be expressed as the proportional reduction in MDT for a specified increase in sampling effort. The results show that the greatest reduction in MDT from a doubling of sampling effort is achieved by increasing the sampling period, followed by reducing the sampling interval, with the least reduction from increasing the number of samples/year (fig. 6). Given a starting design of biennial sampling and 1 sample/year for 10 years, doubling the sampling period yielded a 74-percent reduction in MDT for Total Richness. In contrast, doubling the number of samples by reducing the interval in half from biennial to annual yielded only a 49-percent reduction in MDT, and doubling the number of samples/year resulted in a 44-percent reduction in MDT.

For reasons given earlier, increasing the sampling period, or decreasing the sampling interval, yields the same proportional reductions in MDT at an individual site for all metrics (fig. 6). In contrast, increasing the number of samples per year affects MDT differently among metrics, depending on their variance component characteristics. The greatest reduction in MDT from increasing the number of samples per year occurs for metrics with high **WITHIN INDEX PERIOD** variance (relative to other metrics), particularly if **YEAR** variance is also relatively low (for example, % Burrower Taxa). The reduction in MDT is least for metrics with low **WITHIN INDEX PERIOD** (% EPT Individuals) and/or high **YEAR** variance (for example, % Individuals Top 3 Taxa).

### Design [2]: Multiple Sites—Same Sites Resampled

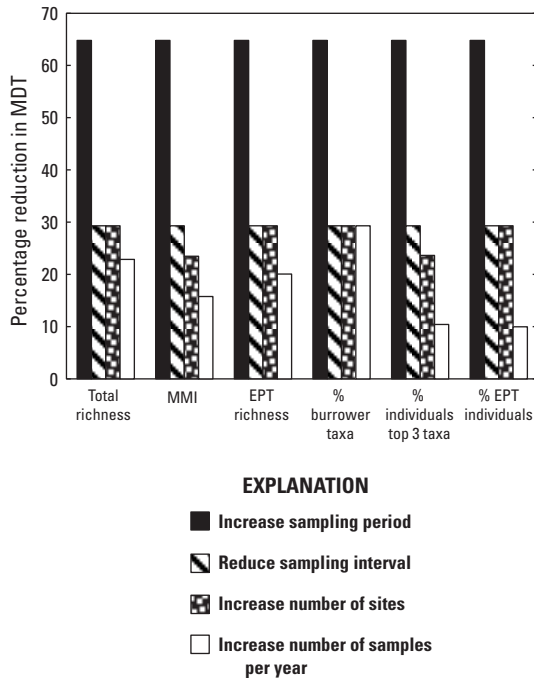
The order of greatest reduction in MDT is the same for multiple sites as for an individual site for the same design options: the greatest reduction is achieved by increasing the sampling period, then by decreasing sampling interval, and least by increasing the number of samples per year; increasing the number of sites yields reductions between the latter two design elements (fig. 7). Doubling the sampling effort from the initial design of biennial sampling, 1 sample/year, 10 years,



**Figure 6.** Percentage decrease in minimum detectable trend for an individual site, resulting from a 100-percent increase in sampling effort for different design choices and different metrics. Each bar represents a design change that doubles the total number of samples collected at a site, from 5 to 10; the starting design sampling period is 10 years, the interval is biennial, and 1 sample/year is collected. Metrics shown represent different relative contributions of the variance components. (MDT, Minimum Detectable Trend; MMI, Multimetric Index; EPT, Ephemeroptera, Plecoptera, Trichoptera).

and 10 sites by increasing the sampling period to 20 years yielded a 65-percent reduction in MDT for Total Richness, whereas reducing the sampling interval to annual or increasing the number of sites to 20 sites yielded only a 29-percent reduction. Increasing the number of samples/year yielded the least reduction in MDT of 23 percent. These reductions from a doubling of effort are proportionally less than those for an individual site, because of the additional variance component of **SITE:YEAR** interaction for multiple sites (see equations in Appendix 1).

The percentage reductions resulting from changes in sampling period or sampling interval are the same for all metrics, as predicted by the variance equations (see Appendix 1). However, increasing the number of sites yields reductions in MDT that vary among metrics. Least proportional reduction occurs for metrics that have relatively high **YEAR** variance (for example, MMI and % Individuals in Top 3 Taxa); the number of sites affects the terms for **SITE:YEAR** interaction and **WITHIN INDEX PERIOD** variance, but not **YEAR** variance (see Appendix 1, equation 2). Doubling the number of samples per year yielded reductions in MDT from 10 to



**Figure 7.** Percentage reduction in minimum detectable trend for multiple sites (same sites resampled each year), resulting from a 100-percent increase in sampling effort for different design choices and different metrics. Each bar represents a design change that doubles the total number of samples collected, from 50 to 100; the starting design sampling period is 10 years, the interval is biennial, 10 sites are sampled, and 1 sample/year is collected. Metrics as in figure 6. (MDT, Minimum Detectable Trend; MMI, Multimetric Index; EPT, Ephemeroptera, Plecoptera, Trichoptera).

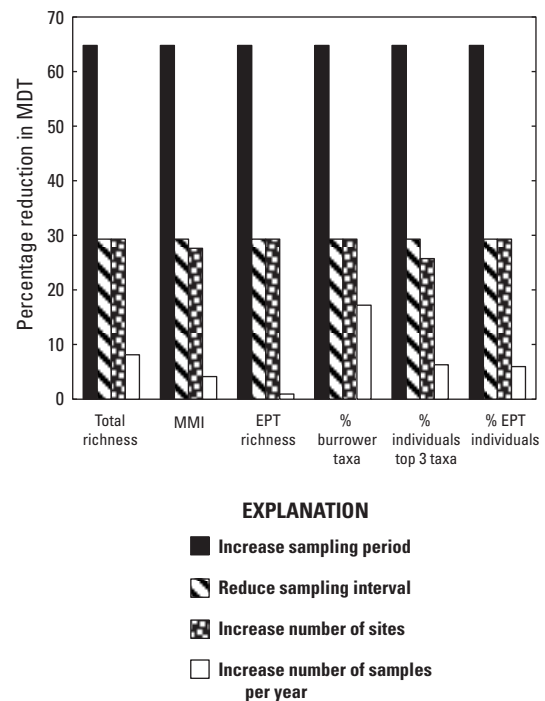
29 percent among six metrics; the greatest reduction was for % Burrower Taxa, because of the negligible variance due to **YEAR**, and relatively high **WITHIN INDEX PERIOD** variance, for this metric.

### Design [3]: Multiple Sites—New Sites Sampled

A multiple-site trend monitoring program in which a new set of sites is selected for each sampling effort (“new-site” design) has the same relative importance among design choices as one in which each site is sampled during each effort (“same-site” design). Thus, the sampling period has the greatest effect on trend detection for same-site design, followed by the sampling interval, the number of sites, and the number of samples per year (fig. 8). The principal difference in these designs is that **SITE** variance contributes more to the trend slope variance in the new-site design, causing MDTs to be higher than a comparable same-site design (Larsen and others, 2001; see equations in Appendix 1). The proportional reduction in MDT resulting from reducing the sampling interval or

increasing the sampling period is the same for both designs, because both of these changes affect only the denominator of the variance equation (Appendix 1). However, an increase in the number of samples per year yields a greater proportional reduction in MDT for all metrics for same-site than for new-site designs (fig. 8), because the **SITE** variance contributes to total variance only when different sites are sampled each time.

Differences in the effect of increasing the number of sites on MDT between the two types of multiple-site designs depend on the variance components of the metric. For example, for metrics with negligible **YEAR** variance, such as Total Richness and EPT Richness, the effect on MDT of increasing the number of sites is the same for either type of network. However, for metrics with a larger **YEAR** variance, such as MMI, a “new-site” network will have a greater reduction in MDT from increasing the number of sites than a “same-site” network; this is because the number of sites affects two terms in the slope variance equation for new sites, but only one term for same sites (see Appendix 1, equations 2 and 3). Similarly, new-site designs benefit more than do same-site networks from increasing the number of sites for metrics with relatively high **SITE** variance, such as MMI.



**Figure 8.** Percentage reduction in minimum detectable trend for multiple sites (new sites sampled each year), resulting from a 100-percent increase in sampling effort for different design choices and different metrics. Each bar represents a design change that doubles the total number of samples collected, from 50 to 100; the starting design sampling period is 10 years, the interval is biennial, 10 sites are sampled, and 1 sample/year is collected. Metrics as in figure 6. (MDT, Minimum Detectable Trend; MMI, Multimetric Index; EPT, Ephemeroptera, Plecoptera, Trichoptera).

## Discussion

Most previous studies comparing the effects of different design choices on trend detection have focused on statistical power for detecting trends—in other words, the probability of detecting a trend if one is present (Wagner and others, 2007; Dauwalter and others, 2009; Al-Chokhachy and others, 2011; Anlauf and others, 2011). While statistical properties are crucial for understanding temporal variations in ecological characteristics, managers of environmental monitoring programs typically have an interest in practical questions, such as the magnitude of trend that can be detected, the length of time or the number of sampling visits required before a trend can be detected, or the cost or sampling effort required to achieve some trend detection objective. In this report, we followed an approach similar to that of Larsen and others (1995, 2001), and derived the minimum detectable trend (MDT) from components of variance for ecological attributes of interest and evaluated tradeoffs associated with different trend design choices. This approach can be applied to any response attribute used in long-term monitoring, including indices and metrics of biological communities, habitat, or chemistry, and can be applied to terrestrial as well as aquatic systems.

The results of this study support the conclusions of Larsen and others (1995)—that extending the sampling period of a consistent long-term monitoring program is the most cost-effective way to improve trend detection at a site, that is, to reduce the magnitude of trend that can be detected. Delays in trend detection, however, are not always consistent with monitoring program goals; to achieve results sooner, a reduction in sampling interval—for example, from once every 4 or 5 years to biennial or annual—yields the next best reduction for a given increase in sampling effort. To further reduce the detectable trend, the effects of within-year variation must be reduced by increasing the number of samples collected within the index period—perhaps in combination with minimizing measurement errors (for example, through improved training or sampling protocols), or redefining the index period to a narrower temporal window.

Although site-specific trends provide insight into ecological patterns and relations to changes in land use and specific stressors at the watershed scale, multiple-site trends address questions about larger systems such as a regional population (Urquhart and others, 1998). The order of improvement in trend detection is the same for both site-specific and multiple-site trends: increasing the sampling period is most effective, decreasing the sampling interval is the next best option, and increasing the number of samples per year is least effective. Although the order of improvement is the same, there is less gain from decreased sampling interval for regional trends than for an individual site. Multiple-site trend monitoring has the additional option of increasing the number of sites, the benefits of which depend on the regional scale. Adding sites yields greater improvement in trend detection for networks that encompass broad heterogeneous areas such as the NAWQA Program national network, which may have larger variance

due to site and site-specific annual variation (site-year interaction) components than networks that cover a smaller region; for example, a single state or ecoregion. Costs associated with expanding the number of sites, however, go beyond the additional costs of sample collection and include reconnaissance, obtaining landowner permissions, and establishing a sampling reach.

A monitoring design with repeated visits of the same sites over time is often considered the preferred design for regional trends (Philippi and others, 1998; Al-Chokhachy and others, 2011) because of the reduced influence of site-to-site variation on trend detection. Although regional trends can be inferred from a network of fixed sites, any bias in the selection of those sites reduces their representativeness of the regional population. Networks in which new sites are selected for each sampling effort—or “panel” combinations of repeat-visit and new sites—are used when trends representing a population of sites are desired; (for example, Urquhart and others, 1998). The effect of the additional site variation in a network of new sites is reduced only by increasing the number of sites. Increasing the number of samples per year yields a better reduction in MDT for same-site than new-site networks.

The choice of metrics affects detection and interpretation of trends and the characteristics of monitoring design required to detect a change (Ferris and Lehman, 2008). There are substantial differences in variance components among the metrics examined herein, as has been found in other studies (Mazor and others, 2009); the effect of these differences on trend detection depends on the design. Although the effects on trend detection of two design options—sampling period and sampling interval—are the same for all metrics, other options vary among metrics. The number of samples per year can be increased to reduce the MDT for those metrics with relatively high within-index period variance. For multiple-site networks, increasing the number of sites reduces the MDT for metrics where among-year variance is low.

Benefits in trend detection from increasing the number of years of monitoring have other complementary advantages. Long-term monitoring data contribute to understanding of natural temporal variation, which is influenced by climatic variations in temperature and flow (Durance and Ormerod, 2007; Kaushal and others, 2010; Lawrence and others, 2010; Kennen and others, 2012), and species-specific characteristics such as life history and traits that can vary seasonally; for example, Beche and others (2006). Long-term data, therefore, are needed so that changes due to anthropogenic influence can be distinguished from those due to natural variability (Jackson and Füreder, 2006). Some sources of natural variation can be minimized in trend monitoring by restricting sampling to a seasonal temporal window (index period) or by scheduling annual sampling to match an annual population cycle (Larsen and others, 2001). Variance is expected to increase as length of monitoring continues because of the increased likelihood of encountering stochastic changes or disturbance events (Al-Chokhachy and others, 2011). Other considerations in design for long-term monitoring include the need for more

frequent sampling if the goals include distinguishing effects of multiple factors, particularly in dynamic systems where natural and human factors are rapidly changing. Changes in water quality and biological diversity may not be linear (Magurran and others, 2010), but rather reflect short-term increases and decreases that are a function of both natural factors such as the proportion of wet years to dry years, cool years to warm years, or differences in antecedent seasonal conditions, and human-caused factors such as changes in rates of chemical application to the land, changes in management policies, or stream restoration efforts.

Future research needs for improving the design of trend monitoring programs include evaluating variance components for reference sites to better characterize natural variability for commonly measured ecological attributes; such a study should be tailored to the appropriate geographic scope for the program. Although the present study characterized variance components for sites representing a broad range of natural and anthropogenic characteristics over a large geographic area, studies of sites in more homogeneous regional networks would be useful to compare variance components and to select metrics appropriate for different monitoring objectives.

The approach presented here provides managers with tools for adjusting the design of trend monitoring programs in the most cost-effective manner. Studies to quantify sources of variability are particularly valuable when conducted early in a trend monitoring program. Tradeoffs among design elements can be evaluated so that monitoring goals can be achieved with the lowest cost (fewest samples) or within the expected duration of the monitoring program. The potential benefits of improving understanding of long-term consequences of human activities on aquatic ecosystems are increased when managers have the tools to design cost-effective monitoring programs.

## Acknowledgments

This research was a collaboration between the U.S. Geological Survey and the U.S. Environmental Protection Agency; the authors thank Charlie Crawford, Alan Herlihy, Phil Kaufmann, Tony Olsen, Dave Peck, John Stoddard, and Ian Waite for contributing to the design of this study. We are especially grateful to the many biologists in the National Water-Quality Assessment (NAWQA) Program who collected the samples for this study and whose hard work and dedication over the past 2 decades has made this long-term study possible, and to Keith Slack, who designed the original Slack samplers that helped create a basis for NAWQA's use of nationally consistent methods. We also thank Pete Ruhl for assisting with data retrievals and the data management structure of BioTDB and BioData, and James Falcone, who provided ancillary data and cartographic assistance.

## References Cited

- Al-Chokhachy, R., Roper, B.B., Archer, E.K., and Miller, S., 2011, Quantifying the extent of and factors associated with the temporal variability of physical stream habitat in headwater streams in the interior Columbia River basin: *Transactions of the American Fisheries Society*, v. 140, no. 2, p. 399-414.
- Anlauf, K.J., Gaeuman, W., and Jones, K.K., 2011, Detection of regional trends in salmonid habitat in coastal streams, Oregon: *Transactions of the American Fisheries Society*, v. 140, no. 1, p. 52-66.
- Beche, L.A., McElravy, E.P., and Resh, V.H., 2006, Long-term seasonal variation in the biological traits of benthic-macroinvertebrates in two Mediterranean-climate streams in California, USA: *Freshwater Biology*, v. 51, no. 1, p. 56-75.
- Cuffney, T.E., Bilger, M.D., and Haigler, A.M., 2007, Ambiguous taxa: effects on the characterization and interpretation of invertebrate assemblages: *Journal of the North American Benthological Society*, v. 26, no. 2, p. 286-307.
- Cuffney, T.F., Gurtz, M.E., and Meador, M.R., 1993, Methods for collecting benthic invertebrate samples as part of the National Water-Quality Assessment Program: U.S. Geological Survey 93-406 [Open-File Report].
- Daufresne, M., Roger, M.C., Capra, H., and Lamouroux, N., 2004, Long-term changes within the invertebrate and fish communities of the Upper Rhone River: effects of climatic factors: *Global Change Biology*, v. 10, no. 1, p. 124-140.
- Dauwalter, D.C., Rahel, F.J., and Gerow, K.G., 2009, Temporal variation in trout populations: implications for monitoring and trend detection: *Transactions of the American Fisheries Society*, v. 138, no. 1, p. 38-51.
- Davies-Colley, R.J., Smith, D.G., Ward, R.C., Bryers, G.G., McBride, G.B., Quinn, J.M., and Scarsbrook, M.R., 2011, Twenty years of New Zealand's National Rivers Water Quality Network: benefits of careful design and consistent operation: *Journal of the American Water Resources Association*, v. 47, no. 4, p. 750-771.
- Durance, I., and Ormerod, S.J., 2007, Climate change effects on upland stream macroinvertebrates over a 25-year period: *Global Change Biology*, v. 13, no. 5, p. 942-957.
- Ferris, J.A., and Lehman, J.T., 2008, Nutrient budgets and river impoundments: Interannual variation and implications for detecting future changes: *Lake and Reservoir Management*, v. 24, no. 3, p. 273-281.

- Gilliom, R.J., Barbash, J.E., Crawford, C.G., Hamilton, P.A., Martin, J.D., Nakagaki, N., Nowell, L.H., Scott, J.C., Stackelberg, P.E., Thelin, G.P., and Wolock, D.M., 2006, The Quality of our Nation's Waters—Pesticides in the Nation's Streams and Ground Water, 1992–2001. U.S. Geological Survey Circular 1291: U.S. Geological Survey, 172 p.
- Hirsch, R.M., Hamilton, P.A., and Miller, T.L., 2006, U.S. Geological Survey perspective on water-quality monitoring and assessment: *Journal of Environmental Monitoring*, v. 8, no. 5, p. 512–518.
- Jackson, J.K., and Füreder, L., 2006, Long-term studies of freshwater macroinvertebrates: a review of the frequency, duration and ecological significance: *Freshwater Biology*, v. 51, no. 3, p. 591–603.
- Kaushal, S.S., Likens, G.E., Jaworski, N.A., Pace, M.L., Sides, A.M., Seekell, D., Belt, K.T., Secor, D.H., and Wingate, R.L., 2010, Rising stream and river temperatures in the United States: *Frontiers in Ecology and the Environment*, v. 8, no. 9, p. 461–466.
- Kennen, J.G., Sullivan, D.J., May, J.T., Bell, A.H., Beaulieu, K.M., and Rice, D.E., 2012, Temporal changes in aquatic-invertebrate and fish assemblages in streams of the north-central and northeastern U.S.: *Ecological Indicators*, v. 18, p. 312–329.
- Kincaid, T.M., Larsen, D.P., and Urquhart, N.S., 2004, The structure of variation and its influence on the estimation of status: Indicators of condition of lakes in the Northeast, U.S.A: *Environmental Monitoring and Assessment*, v. 98, no. 1–3, p. 1–21.
- Larsen, D.P., Kincaid, T.M., Jacobs, S.E., and Urquhart, N.S., 2001, Designs for evaluating local and regional scale trends: *BioScience*, v. 51, no. 12, p. 1069–1078.
- Larsen, D.P., Urquhart, N.S., and Kugler, D.L., 1995, Regional scale trend monitoring of indicators of trophic condition of lakes: *Water Resources Bulletin*, v. 31, no. 1, p. 117–140.
- Lawrence, J.E., Lunde, K.B., Mazor, R.D., Beche, L.A., McElravy, E.P., and Resh, V.H., 2010, Long-term macroinvertebrate responses to climate change: implications for biological assessment in mediterranean-climate streams: *Journal of the North American Benthological Society*, v. 29, no. 4, p. 1424–1440.
- Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., and Schabenberger, O., 2006, SAS for mixed models (2nd ed): Cary, North Carolina, SAS Institute.
- Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M., Smith, R.I., Somerfield, P.J., and Watt, A.D., 2010, Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time: *Trends in Ecology & Evolution*, v. 25, no. 10, p. 574–582.
- Mazor, R.D., Purcell, A.H., and Resh, V.H., 2009, Long-term variability in bioassessments: a twenty-year study from two northern California streams: *Environmental Management*, v. 43, no. 6, p. 1269–1286.
- Miller, M.P., Kennen, J.G., Mabe, J.A., and Mize, S.V., 2012, Temporal trends in algae, benthic invertebrate, and fish assemblages in streams and rivers draining basins of varying land use in the south-central United States, 1993–2007: *Hydrobiologia*, v. 684, no. 1, p. 15–33.
- Moulton, S.R., II, Carter, J.L., Grotheer, S.A., Cuffney, T.F., and Short, T.M., 2000, Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory – Processing, taxonomy, and quality control of benthic macroinvertebrate samples: U.S. Geological Survey Open-File Report 00–212, 49 p.
- Moulton, S.R., II, Kennen, J.G., Goldstein, R.M., and Ham-brook, J.A., 2002, Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water-Quality Assessment Program: U.S. Geological Survey Open-File Report 02–150, 75 p.
- Omernik, J.M., 1987, Ecoregions of the conterminous United States: *Annals of the Association of American Geographers*, v. 77, p. 118–125.
- Ormerod, S.J., and Durance, I., 2009, Restoration and recovery from acidification in upland Welsh streams over 25 years: *Journal of Applied Ecology*, v. 46, no. 1, p. 164–174.
- Peterson, M.J., Efroymson, R.A., and Adams, S.M., 2011, Long-term biological monitoring of an impaired stream: synthesis and environmental management implications: *Environmental Management*, v. 47, no. 6, p. 1125–1140.
- Philippi, T.E., Dixon, P.M., and Taylor, B.E., 1998, Detecting trends in species composition: *Ecological Applications*, v. 8, no. 2, p. 300–308.

- Robinson, C.T., Minshall, G.W., and Royer, T.V., 2000, Inter-annual patterns in macroinvertebrate communities of wilderness streams in Idaho, U.S.A.: *Hydrobiologia*, v. 421, p. 187–198.
- Stevens, D.L., and Urquhart, N.S., 2000, Response designs and support regions in sampling continuous domains: *Environmetrics*, v. 11, no. 1, p. 13–41.
- Stoddard, J.L., Herlihy, A.T., Peck, D.V., Hughes, R.M., Whittier, T.R., and Tarquinio, E., 2008, A process for creating multimetric indices for large-scale aquatic surveys: *Journal of the North American Benthological Society*, v. 27, no. 4, p. 878–891.
- Urquhart, N.S., Paulsen, S.G., and Larsen, D.P., 1998, Monitoring for policy-relevant regional trends over time: *Ecological Applications*, v. 8, no. 2, p. 246–257.
- U.S. Environmental Protection Agency, 2006, Wadeable Streams Assessment: a collaborative survey of the Nation's streams. EPA 841/B-06/002: Office of Water, U.S. Environmental Protection Agency EPA 841-B-06-002, 98 p.
- Wagner, T., Bence, J.R., Bremigan, M.T., Hayes, D.B., and Wilberg, M.J., 2007, Regional trends in fish mean length at age: components of variance and the statistical power to detect trends: *Canadian Journal of Fisheries and Aquatic Sciences*, v. 64, no. 7, p. 968–978.
- Webb, J.A., and King, E.L., 2009, A Bayesian hierarchical trend analysis finds strong evidence for large-scale temporal declines in stream ecological condition around Melbourne, Australia: *Ecography*, v. 32, no. 2, p. 215–225.
- Yuan, L.L., Hawkins, C.P., and Van Sickle, J., 2008, Effects of regionalization decisions on an O/E index for the US national assessment: *Journal of the North American Benthological Society*, v. 27, no. 4, p. 892–905.

## Appendix 1. Variance Component Equations

Before the variance components were calculated, metric scores were first “detrended” by regressing all scores on **YEAR**; only one metric had a significant (and numerically small) trend, and all other metrics had essentially zero trend. Variance components were then estimated for the residuals from this regression. Contributions of the variance components to the uncertainty of the estimated trend slope,  $b$ , can be expressed in the linear trend model  $Y = A + B \cdot \text{YEAR}$ , where  $Y$  is the value of a metric in different years (Larsen and others, 2001),  $A$  is a constant (y-axis intercept), and  $B$  is the trend slope with units of change per year in the invertebrate score. Specific equations for estimating trends at a single site, trends for multiple sites when the same sites are sampled each year, such as a regional trend, and trends for multiple sites when a new set of sites is selected each year are described below (Larsen and others, 1995, 2001).

### Design [1]: Site-Specific Trend

For simplicity, assume that invertebrates are sampled at regular yearly intervals over a sampling period and that each yearly sample has the same number ( $N_{seas}$ ) of repeat-visit samples within the sampling season (index period). Then the variance of the estimated trend slope,  $b$ , is given by:

$$\text{Var}(b) = \frac{\text{Var}(\text{YEAR}) + \frac{\text{Var}(\text{WITHIN INDEX PERIOD})}{N_{seas}}}{SSQ(\text{sampling period})} \quad (1)$$

In equation 1,  $\text{Var}(\text{YEAR})$  and  $\text{Var}(\text{WITHIN INDEX PERIOD})$  are the **YEAR** and **WITHIN INDEX PERIOD** variance components. For trends at a single site, the variance of the trend slope decreases as the number of samples within a season ( $N_{seas}$ ) increases, or as the number of annual samples, which affects  $SSQ$ , increases. The  $SSQ$  in the denominator is the sum of squares of the (centered) years over the entire sampling period (the sum of squared deviations of each sampling year from the mean year of the whole period).  $SSQ$  can be expressed in terms of sampling period and sampling interval, as follows:

$$SSQ = p^2 k(k-1)/12,$$

where

- $p$  = sampling interval (years)=number of years between samples, and  
 $k$  = number of years in the sampling period in which samples are collected.

For example, for biennial sampling for a 20-year period,  $k$  is 10 (samples collected every other year), and  $p$  is 2.

Because equation 1 is for trends at a single site,  $\text{Var}(\text{YEAR})$  is equal to the sum of the **YEAR** and **SITE:YEAR** components. Thus, for trends at a single site, the variance of the trend slope decreases as the number of samples within a season ( $N_{seas}$ ) increases, or as the number of annual samples, which affects  $SSQ$ , increases.

### Design [2]: Trend in Multiple Sites—Same Sites Resampled

For a monitoring program of multiple sites in which all sites are visited for each sampling effort (“same sites”), a separate linear time trend can be estimated at each site, based on its own multiyear data. The trend slope,  $B$ , then can be defined as the average of these site-specific trend slopes. The estimation uncertainty of  $b$  is given by Larsen and others (1995, 2001):

$$\text{Var}(b) = \frac{\text{Var}(\text{YEAR}) + \frac{\text{Var}(\text{SITE:YEAR}) + \frac{\text{Var}(\text{WITHIN YEAR})}{N_{seas}}}{N_{sites}}}{SSQ(\text{sampling period})} \quad (2)$$

In equation 2,  $N_{sites}$  is the number of (same) sites, and  $\text{Var}(\text{SITE:YEAR})$  is the **SITE:YEAR** interaction variance component. The interaction variance component appears because it represents differential variation in **YEAR** effects at different **SITES**, leading to greater variability in the estimated trend. There is no  $\text{Var}(\text{SITE})$  component in equation 2 because temporally consistent, between-site differences (**SITE** effects) have been removed by estimating the trends separately for each site using only that site’s data, essentially treating sites as “blocks” as in an analysis of variance. Equation 2 shows that  $\text{Var}(b)$  is reduced by increasing  $N_{seas}$  or the number of annual samples, strategies that are also effective for monitoring a single site. However,  $\text{Var}(b)$  can now also be reduced by increasing the number ( $N_{sites}$ ) of monitored sites.

### Design [3]: Trend in Multiple Sites—New Sites Sampled

For a monitoring program of multiple sites in which a new set of sites is selected each year, a mean metric value can be estimated in Year 1 across the sites that are chosen in that year. As new sets of sites are chosen in Years 2, 3, 4, etc., mean metric values are estimated using only the data for the sites sampled in each year. The trend slope,  $B$ , is then defined as the slope of the linear time regression fitted to these yearly means.

For this sampling and estimation approach, the estimation uncertainty of  $b$  is given by:

$$Var(b) = \frac{\frac{Var(SITE)}{N_{sites}} + Var(YEAR) + \frac{Var(SITE:YEAR) + \frac{Var(WITHIN YEAR)}{N_{seas}}}{N_{sites}}}{SSQ(sampling\ period)} \quad (3)$$

This is the same  $Var(b)$  as for equation 2 (same sites), plus a substantial term of new uncertainty due to  $Var(SITE)/N_{sites}$ . This term appears because now only new sites are sampled each year, thereby losing the “blocking” advantages of resampling the same sites over multiple years. To reduce the effect of  $Var(SITE)$ , the number of sites can be increased, or metrics can be chosen for which the  $Var(SITE)$  variance component is relatively low. The effects of all variance components, other than  $Var(YEAR)$ , can be reduced by increasing  $N_{sites}$ .



For additional information about this report write to:

Martin E. Gurtz

Daren M. Carlisle

U.S. Geological Survey, MS 413

National Water-Quality Assessment Program

12201 Sunrise Valley Drive

Reston, VA 20192

John Van Sickle

Department of Fisheries and Wildlife

Oregon State University

Corvallis, OR 97331

Steven G. Paulsen

U.S. Environmental Protection Agency

200 SW 35<sup>th</sup> Street

Corvallis, OR 97330



