

DEVELOPMENT OF A STANDARDIZED LARGE RIVER BIOASSESSMENT PROTOCOL (LR-BP) FOR MACROINVERTEBRATE ASSEMBLAGES[†]

JOSEPH E. FLOTEMERSCH,* KAREN BLOCKSOM, JOHN J. HUTCHENS JR[‡]
and BRADLEY C. AUTREY

National Exposure Research Laboratory, U.S. Environmental Protection Agency, 26 W. Martin Luther King Dr., Cincinnati, OH 45268 USA

ABSTRACT

Efforts to develop benthic macroinvertebrate sampling protocols for the bioassessment of lotic ecosystems have been focused largely on wadeable systems. As these methods became increasingly refined and accepted, a growing number of monitoring agencies expanded their work and are now developing sampling protocols for non-wadeable large rivers. Large rivers can differ from wadeable streams in many ways that preclude the use of some wadeable stream sampling protocols. Hence, resource managers need clear and consistent large river bioassessment protocols for measuring ecological integrity that are cost effective, logistically feasible, and meet or are adaptable to the multi-purpose sampling needs of researchers and managers. We conducted a study using an experimental macroinvertebrate sampling method that was designed to overcome limitations of several methods currently in use. Our objectives were to: (1) determine the appropriate number of sampling points needed; (2) determine an appropriate laboratory subsample size to use and (3) examine how varying reach length affects assemblage characteristics. For six reaches in each of two large rivers, we sampled the macroinvertebrates of both banks at 12 transects separated by increasingly larger distances using a multi-habitat, semi-quantitative technique. Interpretation of results relied on the values attained for nine benthic macroinvertebrate assemblage metrics. Results from Monte Carlo methods indicated that, using the sampling methods described herein, a representative sample of the assemblage was collected by sampling both banks on 6 transects. Across all sites, we did not observe a consistent relationship between transect spacing (i.e. total reach length) and metric values, indicating that our sampling protocol was relatively robust with respect to variation in reach length. Therefore, flexibility exists that permits the study reach length to be dictated by the spatial scale (e.g. repeating geomorphic units) in question. For those preferring to use a fixed reach length, we recommend that transects be spaced at a minimum of 100 m intervals over a 500 m distance. We recommend that the field method be coupled with a fixed laboratory subsample size of 300 organisms for bioassessment purposes, with the recognition that a subsample size of 500 organisms may be needed to meet the objectives of more rigorous studies. It is likely this approach will over-sample sites of uniform composition, but the goal was to develop a robust sampling protocol that would perform well across sites of differing habitat composition. Possible modifications to the method to streamline its future application in the field are provided. Published in 2006 by John Wiley & Sons, Ltd.

KEY WORDS: field method; monitoring; restoration; non-wadeable; laboratory subsampling; reach length; simulations; Monte Carlo

INTRODUCTION

Wadeable streams and smaller rivers are abundant and relatively easy to sample compared to large rivers. As a result, efforts to develop appropriate sampling protocols for the bioassessment of lotic ecosystems have been focused primarily on smaller systems (e.g. Barbour *et al.*, 1999). As these methods have become increasingly refined and accepted, a growing number of government agencies are developing sampling protocols for non-wadeable large rivers (Humphries *et al.*, 1998).

In large rivers, stressor sources are generally more numerous (Sweeting, 1994) and almost certainly more rapidly diffused as a result of greater discharge (Allan, 2000). Individual stressor effects are masked by the presence of other stressors and their impacts are less conspicuous. Biological communities also change with stream size, as do habitat type and quality (Vannote *et al.*, 1980). Assemblages adapted to deeper, wider streams with limited canopy

*Correspondence to: J. E. Flotemersch, National Exposure Research Laboratory, U.S. Environmental Protection Agency, 26 W. Martin Luther King Dr., Cincinnati, OH 45268 USA. E-mail: Flotemersch.Joseph@epa.gov

[‡]Current address: Department of Biology, Coastal Carolina University, Conway, SC 29528 USA.

[†]This article is a U.S. Government work and is in the public domain in the U.S.A.

Received 5 July 2005

Revised 17 November 2005

Accepted 15 December 2005

cover are more likely to occur in downstream, higher order reaches. Thus, expectations for communities in large rivers may be very different from those in smaller systems. In addition, the thalweg of a large river often may not be accessible for sampling because of depth and distance from shore, precluding the use of some wadeable stream sampling protocols. Hence, resource managers need clear and consistent protocols for measuring ecological integrity that are designed specifically for large river systems (Loucks, 2003).

Protocols for sampling fauna of large rivers should be clear, consistent, and reproducible to effectively support a bioassessment programme. To be applicable to a wide audience, they should perform well across numerous non-wadeable habitats and river types, represent site conditions accurately, and ideally, identify the presence of stressors. Protocols should also be cost effective, logistically feasible with only moderate training, and meet or be adaptable to multi-purpose sampling needs of researchers and managers (e.g. trend analysis, point-source and non-point source programmes, habitat rehabilitation and restoration efforts) if they are to be accepted by monitoring and regulatory organizations.

Benthic macroinvertebrates are one of the most common faunal assemblages used in the bioassessment of aquatic ecosystems (Rosenberg and Resh, 1993; Metcalfe-Smith, 1994; Barbour *et al.*, 1999; Karr and Chu, 1999). Many macroinvertebrate collection methods currently used in non-wadeable systems are derived from wadeable methods (Ohio Environmental Protection Agency, 1989; Barbour *et al.*, 1999; Klemm *et al.*, 2000; Flotemersch *et al.*, 2001; Moulton *et al.*, 2002). These methods often involve wading in shallow near-shore areas of larger rivers or sampling from a boat in deep areas without additional modification. The exception is the use of artificial substrates, which were developed largely for non-wadeable invertebrate sampling applications (Cairns, 1982).

Blocksom and Flotemersch (2005) compared six sampling techniques used by three government agencies to assess benthic macroinvertebrate assemblages of large rivers. They found that these methods produced different metric values, that metric response (i.e. positive vs. negative) to certain stressors varied among sampling methods, and that metrics detecting a specific stressor were not consistent among methods (Blocksom and Flotemersch, *in press*). Differences among methods and the relatively poor performance of some were hypothesized to be due partly to the inadequacy of using a single sampling technique (e.g. kick-net, dip-net and artificial substrates) when sampling large rivers. For example, a method that produced a representative sample in a large river with abundant epifaunal substrate and low embeddedness might not be suited for a highly embedded reach. The research of Bartsch *et al.* (1998) and Poulton *et al.* (2003) corroborate this hypothesis, with both concluding that an approach employing multiple sampling techniques was needed to effectively sample all components of a macroinvertebrate assemblage in riverine ecosystems.

To support the development of a more consistent Large River Bioassessment Protocol (LR-BP) for benthic macroinvertebrates, three fundamental issues must be addressed. First, a collection technique is needed that secures a representative sample of benthic macroinvertebrates across the broad range of habitats that occur within and across rivers. Second, an appropriate sampling design will be needed for application of the developed sampling technique. And third, an appropriate laboratory method (e.g. macroinvertebrate subsample size) must be determined. The method must also be logistically feasible for monitoring agencies having limited resources.

To address this issue, we examined data from the different sampling techniques compared by Blocksom and Flotemersch (2005) to identify critical elements of each that could be combined and applied in a standardized manner to support a Large River Bioassessment Protocol (LR-BP). The result was a sampling technique that hypothetically should overcome the limitations of previous approaches and permit standardized sampling across all non-wadeable habitats and river types of varying impoundment status. The approach consisted of features of the Environmental Monitoring and Assessment Programme—Surface Waters kick net sampling method (Klemm *et al.*, 2000) and the multiple habitat dip net methods of the Ohio Environmental Protection Agency (Ohio Environmental Protection Agency, 1989) and the United States Geological Survey (Moulton *et al.*, 2002) that sample all available habitats.

Critical elements of the development of a scientifically sound sampling design include the spatial scale over which samples should be collected (i.e. reach length), the number of samples needed, and the manner in which samples should be distributed within the sample reach. The use of 'reach' in this study follows that of Frissell *et al.* (1986) who defined it as a length of stream between breaks in channel slope, local side-slopes, valley floor width, riparian vegetation, and bank material. For bioassessment purposes, determination of appropriate sample reach lengths are typically linked to measures of geomorphology (e.g. channel widths, meander wavelengths, riffle pool

sequences) (Barbour *et al.*, 1999; Herlihy and Lazorchak, 2000; Moulton *et al.*, 2002) or evaluation of species accumulation curves. Several studies have focused on appropriate reach lengths for macroinvertebrates in wadeable streams (e.g. Li *et al.*, 2001), and for fish in both wadeable and non-wadeable streams (e.g. Lyons, 1992; Hughes *et al.*, 2002). However, an appropriate assessment reach for macroinvertebrates in non-wadeable streams has not been estimated and may vary among rivers in different ecoregions, rivers with different geomorphic structures, or by the question being addressed. One difficulty in answering this question is that benthic macroinvertebrates are usually sampled at specific points, whereas fish are sampled using electrofishing techniques over the whole reach. Hence, determining an appropriate sampling reach length for macroinvertebrates using species accumulation curves as a direct function of distance is logistically impractical because of the large number of contiguous samples that would be required.

Similar challenges are encountered using measures of geomorphology for reach determination on large rivers. The majority of rivers in the U.S. have been anthropogenically altered (especially through dam construction and major channel modification) to the extent that <2% are of a quality worthy of federal protection status (Benke, 1990). For example, using multiples of the natural channel width to set suitable reach lengths for sampling in geomorphically-altered systems may be less appropriate than in natural rivers. Still, even when habitats have been artificially created or altered, they still can have a regular frequency and are inhabited by aquatic organisms. Beyond this issue however the question still remains as to the appropriate number and distribution of samples within the designated reach to effectively represent the reach for bioassessment purposes.

As for development of an appropriate laboratory processing method, the procedures are typically assumed to be readily transferable from wadeable streams to large rivers, but this has not been evaluated. Hence, this study also investigated the efficacy of sample processing in the laboratory. The methods used for laboratory processing of invertebrates can greatly influence sample results and ultimately determine the value of a method for bioassessment purposes. A full count of all invertebrates may provide a more accurate assessment (Doberstein *et al.*, 2000), but is usually not feasible when large numbers of organisms are collected (Barbour and Gerritsen, 1996). As a result, samples are often subsampled in the laboratory using either a fixed-organism count or a fixed proportion (Barbour *et al.*, 1999; Carter and Resh, 2001).

The primary objectives of the study were to: (1) determine the appropriate number of sampling points needed using a new LR-BP for macroinvertebrates in nonwadeable rivers; (2) examine how varying reach lengths affect assemblage characteristics and (3) determine an appropriate laboratory subsample size to accompany this sampling method.

METHODS

We collected data during late July through August 2001 from the Kentucky ($n = 6$ sites) and Great Miami ($n = 6$ sites) rivers, both of which are major tributaries of the Ohio River in east-central United States. The 12 sites were selected from 30 sites that were sampled in a previous study that compared existing large river sampling methods

Table I. Physical characteristics and mean percent (standard deviation) of land use types in the study basins, with means and standard deviations based on sites used in analyses

Parameter	River Basin	
	Great Miami ($n = 6$)	Kentucky ($n = 6$)
Drainage basin (km ²)	13,947 ^a	18,130 ^b
River length (km)	233.4 ^a	410.4 ^b
Average gradient (m/km)	0.74 ^a	0.13 ^b
% Urban land use ^c	5.90 (2.64)	7.81 (0.5)
% Agriculture land use ^c	82.55 (3.21)	80.18 (1.3)
% Forested land use ^c	10.10 (1.44)	10.83 (1.6)

^aOhio Environmental Protection Agency, 1997.

^bKentucky River Authority, 1999.

^cMulti-Resolution Land Characteristics Consortium (Vogelmann *et al.*, 1998).

Table II. Ranges and medians of chemical and physical habitat variables at study sites

	Great Miami River (<i>n</i> = 6)		Kentucky River (<i>n</i> = 6)	
	Range	Median	Range	Median
Physical habitat				
Mean thalweg depth (m)	1.2–2.3	2.03	5.2–9.7	7.0
Mean wetted width (m)	45.8–154.3	94.4	69.9–97.3	80.5
Mean bankfull height (m)	0.7–2.9	1.3	1.7–2.2	2.1
LWD quantity	11–70	30.0	20–43.0	30.5
% Canopy density at bank	0–92	37	71–92	78
% Substrate as large gravel and larger at bank	0–89.5	25.5	0–0.92	0.78
% Urban in riparian	2.1–83.1	42.4	0.4–7.7	1.1
% Agriculture in riparian	9.0–77.1	27.7	9.7–33.6	14.0
% Forest in riparian	6.8–57.5	12.2	65.4–85.8	83.6
Water chemistry				
Mean conductivity ((S/cm)	521.2–857.2	664.6	270.6–435.2	334.2
SO ₄ (mg/L)	33.0–64.4	45.3	33.9–104.6	79.8
NO ₃ (mg/L)	1.37–5.69	1.95	0.37–0.80	0.56
Chloride (mg/L)	25.06–71.64	44.9	3.30–6.61	5.74
Ammonia (mg/L)	0.07–0.23	0.09	0.02–0.07	0.04
Total Kjeldahl Nitrogen (TKN) (mg/L)	0.53–0.95	0.61	0.15–0.29	0.23
Total Phosphorus (µg/L)	0.05–0.28	0.17	0.01–0.04	0.02

(Blocksom and Flotemersch, 2005) (Table I). Sites were selected to partition the sampling effort evenly between impounded and relatively free-flowing sites and across a gradient of habitat conditions within each river. Gradients were based on existing instream and riparian physical habitat data collected using the Environmental Monitoring and Assessment Programme's (EMAP) protocols (Kaufmann, 2000), land use data and best professional judgement.

The Great Miami River flows through several urban and industrial corridors in Ohio (e.g. Dayton, Springfield, Hamilton and Middletown) before reaching the Ohio River. However, the dominant land use in the basin is agriculture (80.3%) (Table I) (Miami Conservancy District, 2004). The river has sections with exposed riffles and rapids and sections with restricted flow associated with low-head dams that temporarily store, rather than regulate, waters.

The Kentucky River has a series of 14 lock-and-dam structures spanning the length of the mainstem, which historically supported commercial traffic. The watershed has some large forested sections and some small areas with mining, agricultural and urban influences (e.g. Lexington) (Table I). As a result of impoundment, all Kentucky River sites sampled in this study were much deeper than those of the Great Miami River (Table II).

Final site selection resulted in sites well-distributed longitudinally along the mainstem of each river and included a mixture of habitat types. Study reaches were positioned so that stream confluences, bridges and obvious stressor sources, such as major outfalls, did not occur within the reach as this might have complicated data analysis within sites and obscured the results.

SAMPLING DESIGN

An appropriate reach length for macroinvertebrates in non-wadeable streams has not been estimated. However, benthic macroinvertebrate and fish assemblage structure are often correlated (e.g. Kilgour and Barton, 1999). Therefore, the available literature on appropriate assessment units for fish in large rivers was used for setting a maximum size for the study reach.

Measures of fish species richness is a function of the number of channel units sampled (Gorman and Karr, 1978; Angermeier and Schlosser, 1989; Lyons, 1992), with the size and spacing of these units a function of stream size (Leopold *et al.*, 1964). The assessment unit length required can also vary by study objectives (Cao *et al.*, 2001;

Hughes *et al.*, 2002). Lyons (1992), working on wadeable streams in Wisconsin, USA, concluded that for assessments of environmental quality or community-level ecological analyses, a distance of 35 times the mean stream width or a length equal to three complete riffle-pool sequences, was sufficient to estimate fish species richness. Pilot studies for the Environmental Monitoring and Assessment Programme suggested that in eastern non-wadeable streams and rivers, a length of 40 channel widths was necessary to characterize the fish community of a site (Herlihy and Lazorchak, 2000). We therefore used a reach length of 40 times the estimated mean wetted width of the channel at each river site. In hydrologically formed channels, this reach length would include approximately four meander wavelengths (Leopold *et al.*, 1964).

The downstream end of the study reach at each site was set at a randomly determined point on one bank and marked with flagging. A systematic sampling design was applied to establish 12 transects within the reach. This design has many desirable features for field studies; and as long as the first point is selected at random, remaining points based on that point can be considered random as well (Cochran, 1977). The simplicity of the design makes it easy to execute without mistakes and results in significant time saving in the field. It also results in the drawn sample being spread more evenly over the population (Cochran, 1977; Manly, 2001).

Proceeding upstream from the initial point, 11 transects spanning the width of the river were identified and flagged. The first four were spaced at a distance equal to the mean wetted width of the channel, followed by two spaced at two times the wetted width, two at four times the wetted width and three at eight times the wetted width (Figure 1). This

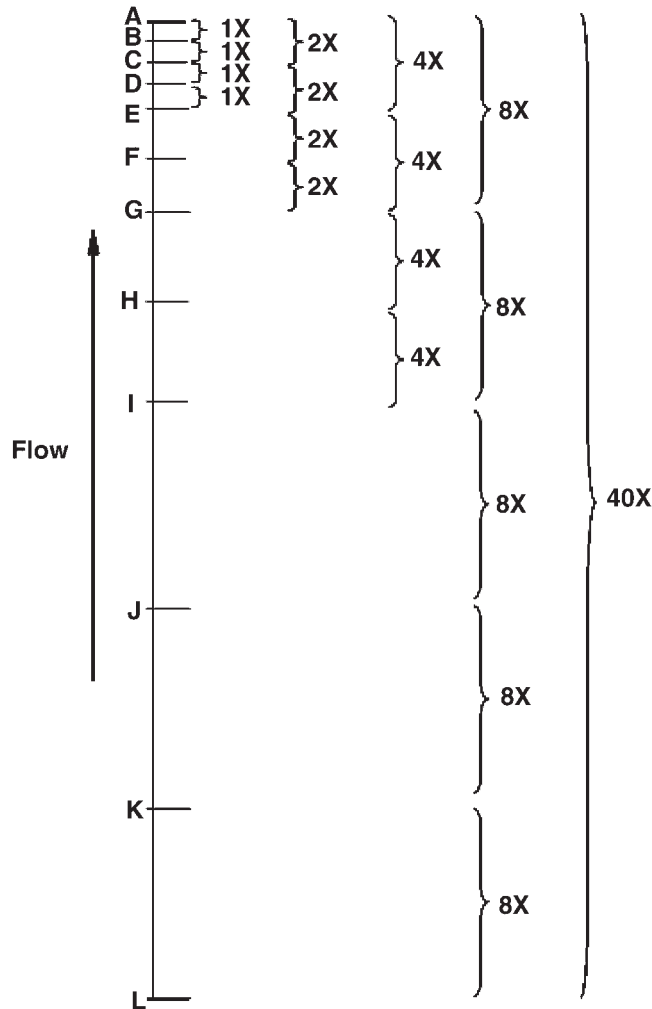


Figure 1. Sampling scheme used to examine the effect of distance on metric values

identified 24 stations in the reach (e.g. two per transect, 12 on each bank) where macroinvertebrates would be collected.

The size of the sampling zone at each sampling station was proportional to the mean wetted width of the river. At each sampling station, a shoreline sampling zone was defined as 0.1 times the estimated wetted width in shoreline length and extended from the shore to the non-wadeable point of the river. Therefore, if the river was 70 m wide, the sampling zone for each station would be 7 m. This served to keep the sampling zone in proportion to the increasing size of habitat features as the size of a river increased. Zone placement was centered on the station transect.

INVERTEBRATE SAMPLING

The technique used to collect benthic macroinvertebrate samples in each zone of both banks was a hybrid of existing techniques and consisted of two distinct sample-collection procedures. The use of two distinct collection approaches provides a representative benthic macroinvertebrate sample from the different types of habitats encountered within and across rivers.

At each sampling zone, two samples were collected with a modified kick-net (50-cm wide \times 30-cm tall \times 60-cm bag-depth; 595- μ m mesh). The area in front of the net equal to the width and length of the net frame (0.5 m; total area = 0.25 m²) was then vigorously kicked for 20 s (see Klemm *et al.*, 2000).

Next, a D-frame net (25.4-cm wide \times 30.5-cm high \times 25.4-cm bag-depth; 595- μ m mesh) was used to sample other available habitats in the sampling zone (e.g. root wads, undercut banks, steep banks and vegetation). These habitats can be difficult to sample with a kick-net procedure and therefore are often underrepresented. At each station the sampling effort was standardized to 3 min per 5 m of sample zone width. While a standardized sampling time may be sufficiently quantitative (Hynes, 1970) and used for quantifying effort, the main purpose of the timed effort was to control for the amount of time field personnel spent at any single station, thus assuring ample time to cover the entire reach or multiple reaches in a day.

At some stations, only one collection procedure was suitable for collection of a representative sample of the fauna from the prevailing habitat. However, both were performed at every sampling zone if safe and practical.

Samples from the kick and D-frame nets from both banks were combined into a single transect sample ($n = 12$ per reach) for use in determining the appropriate number of point-samples that needed to be collected and in examining the effects of reach length on sample results.

Each sample was condensed in the field with a 595- μ m sieve, preserved with 95% ethanol, and diluted to a final concentration approximating 70% ethanol (following Klemm *et al.*, 2000). In the laboratory, individual samples were completely sorted. All organisms were identified to the lowest practical taxonomic level, usually genus or species, given specimen condition and availability of taxonomic keys.

Additional information was collected from each sample station ($n = 24$) to supplement the macroinvertebrate data, characterize each station and study reach and document the gradient of conditions over which samples were collected. Crews collected physical habitat data following EMAP protocols for nonwadeable streams (Kaufmann, 2000). A single depth-integrated water sample was collected from each site and analysed for sulfate, nitrate, total Kjeldahl nitrogen, ammonia, total phosphorus and chloride concentrations. Chemical analyses were conducted using EMAP-SW laboratory protocols (Klemm *et al.*, 1990). Conductivity and water temperature were measured *in situ* using a YSI Model 85 m at the centre of the sampling reach. Land cover data developed by the Multi-Resolution Land Characteristics Consortium (Vogelmann *et al.*, 1998) were overlaid on a riparian corridor 500 m in width on each side of the river for a distance of 4 km upstream of the center of the sampling reach. Proportions of forest, agriculture, and urban (including residential) land uses were then calculated within the riparian corridor.

DATA ANALYSES

The differences in assemblage characteristics between the two banks at a given transect were sometimes quite large. To encompass the spatial variability present at each transect, samples were combined from the two banks at each transect. Thus, all analyses described in this paper use samples from both banks composited at each transect.

Subsample size

Prior to analyses for estimating the minimum number of transects required per site, an appropriate laboratory subsample size was determined. The entire combined sample for each site (all transects combined) was used to simulate fixed-count subsamples of 100 to 1000 organisms in steps of 100. Simulations were run in C++ (Borland C++ Builder 4.0, Inprise Corporation, Scotts Valley, California). It was assumed that organisms were distributed randomly within each sample. Random sampling without replacement was used to simulate each subsample, and 100 subsamples were generated for each site at each fixed count size to estimate laboratory sampling variability.

The effect of subsample size was measured on the quantitative metrics in the Ohio Environmental Protection Agency Invertebrate Community Index (ICI) because these metrics are used to assess the macroinvertebrate assemblage in larger streams and rivers in Ohio (Ohio Environmental Protection Agency, 1988). These metrics included total taxa richness, mayfly taxa richness, caddisfly taxa richness, Diptera richness, % mayflies, % caddisflies, % Tanytarsini, % non-Tanytarsini dipterans and non-insects and % tolerant individuals. Index scores were not calculated for sample sites because it would be misleading since most metrics included in the ICI are calculated from samples derived from artificial substrate samplers, and one is based on a separate qualitative sampling method.

Since taxa richness metrics did not tend to level off with increasing subsample size, the difference in a metric value between sites was used as a way to measure the effect of sample size. The change in the absolute value of the difference in the metric from one sample size (X_i) to the next higher sample size (X_{i+1}) was defined as the 'return', and the percent of the return relative to the maximum value achieved for that metric ($(|X_i - X_{i+1}| / \max(X_{i+1}))$) as the 'relative return'. In this calculation, the maximum value was set as the maximum for the next higher sample size.

The subsample size at which the average relative return leveled off for most metrics was selected for subsequent analyses on the effect of the number of transects on metrics.

Number of transects

After the subsample size was determined, the number of transects needed per site was evaluated. Following the concept of a species area curve, metric values were plotted as a function of the number of transects or samples. This required randomizing the order of transects to ensure that the results were not affected by sequence of samples. However, the nature of the sampling design meant that transects were not equidistant from one another. If there was strong spatial autocorrelation among samples, randomizing the order of transects would not be appropriate. Thus, spatial autocorrelation in assemblage composition was tested by calculating the Coefficient of Community (CC) similarity index (Sorensen, 1948) for each pair of transects within each site. The CC for each pair of transects was plotted against the distance between them (Figure 2). There was no strong trend apparent between the CC and distance, and it was concluded that spatial autocorrelation was not prevalent.

One-hundred randomizations in C++ were then used to determine the number of transects required before metric values leveled off. Transects were ordered randomly within each site for each randomization. Next, transect data for successively larger numbers of transects within each site were combined, beginning with the first transect in the sequence. At each step in each randomization, a simulated subsample was generated based on the *subsample size* results. For each metric and site, the average metric value across the 100 randomizations was plotted against the number of transects. The point at which each metric leveled off was identified by visual inspection. Finally, a similar set of simulations was run for smaller subsample size(s) to examine the influence of subsample size on these plots.

Distance between transects

After determining the number of transects required per site, the effect of distance between transects on metric values was examined. In each site, pairs of transects were grouped by the distance between them. This created up to four groups with five pairs of transects each with inter-transect distances of one two four and eight times the mean wetted width. For each site, data for all samples within a group were combined and 100 simulations of a

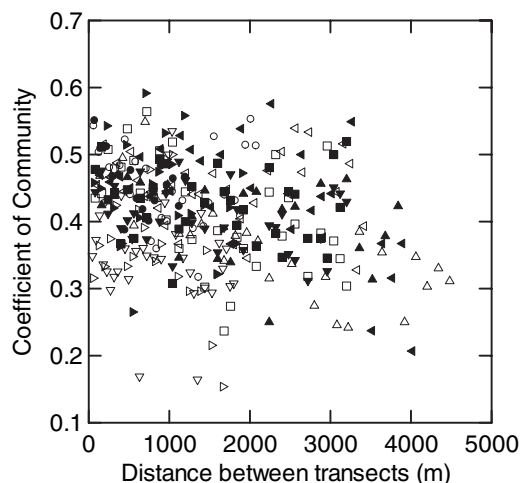


Figure 2. Average Coefficient of Community Index value for each possible pair of transects within each site as a function of the distance between transects. Each type of symbol (shape and fill) represents a different site

500-organism subsample on each group were performed. The within-site differences among groups were assessed qualitatively from plots of the mean metric value (± 1 SE).

RESULTS

Although all 12 selected sites were sampled, fewer than 12 transects were sampled at three sites because of the limited distance between dams, at two sites because of severe weather conditions, and at one site because of loss of daylight. The impact of these logistical limitations on data analysis was negligible. Total reach length sampled at individual sites ranged from 1200 to 4480 m in the Great Miami River and from 1680 to 4000 m in the Kentucky River. The range and median of water chemistry and physical habitat variables at study sites are presented in Table II. The number of organisms per transect sample ranged from 63 to 2369 with a mean of 477.

Subsample size

The metric values for simulated samples quickly leveled off as subsample size increased for percentage metrics, but not for richness metrics (Figure 3). However, the difference in richness metric values between any two sites did not change as rapidly after approximately 500 organisms. In fact, the relative return dropped below about 2% beyond 500 organisms (Figure 4), indicating that additional sorting would not provide sufficient additional information in separating sites from one another. There was also a significant drop between 200 and 300 organisms, resulting in relative returns below 5% for 300 or more organisms. This information is useful to note because most state programmes subsample 300 or fewer organisms for bioassessment samples in streams. Nonetheless, a subsample of 500 organisms was used for further simulation analyses.

Number of transects

There was a strong leveling off of richness metric values at approximately six transects (Figure 5). For percentage metrics, the asymptote typically was reached in fewer transects. When this analysis was rerun using a subsample size of only 300 organisms, similar results were achieved (Figure 6).

Distance between transects

Across all sites, there was no consistent pattern in metric values based on five transects one wetted width apart to five transects eight wetted widths apart (Figure 7). However, within individual sites, there were sometimes very strong differences among the four groups, particularly between the group of transects separated by a distance of

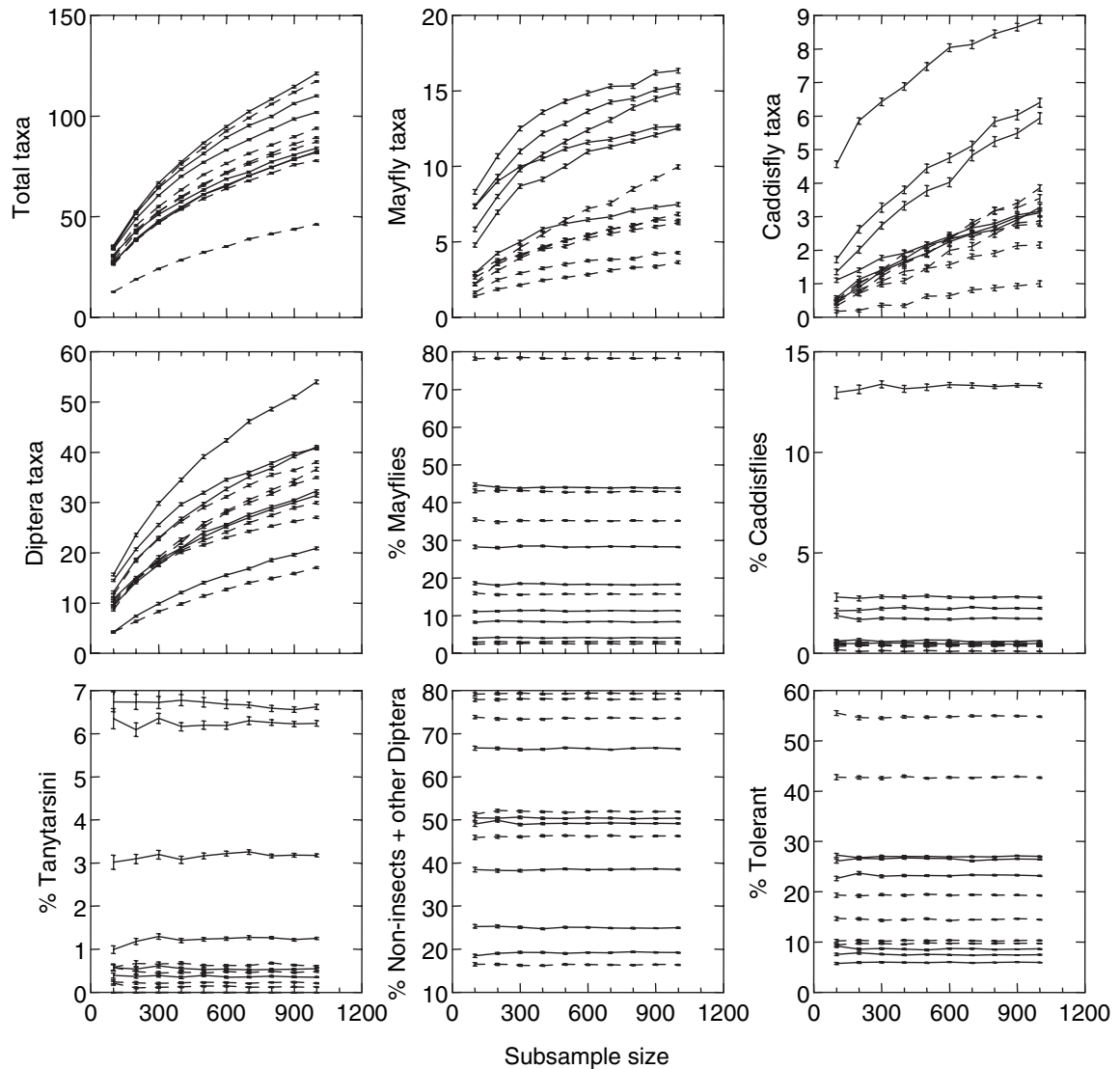


Figure 3. Results of subsample size simulations for each site and metric. Solid lines represent sites in the Great Miami River and dashed lines represent sites in the Kentucky River. Error bars represent 1 standard error of the mean

eight times the wetted width and the other three groups of transects separated by smaller distances (Figure 7). Retrospective analysis of the physical habitat data suggests that at some sites, the likelihood of encountering large variability in one or more coarse physical habitat features (e.g. thalweg depth, substrate composition) increased as the distance between transects increased, although the habitat features causing the variability were not the same across sites.

DISCUSSION

We present here a new protocol for sampling large river macroinvertebrates that is designed to perform well across all shore-line habitats and river types, to integrate different habitats, and thus to represent site conditions accurately. The appropriate fixed count for laboratory subsampling size to use with this method was also determined, based on the ability to separate sites of differing macroinvertebrate composition.

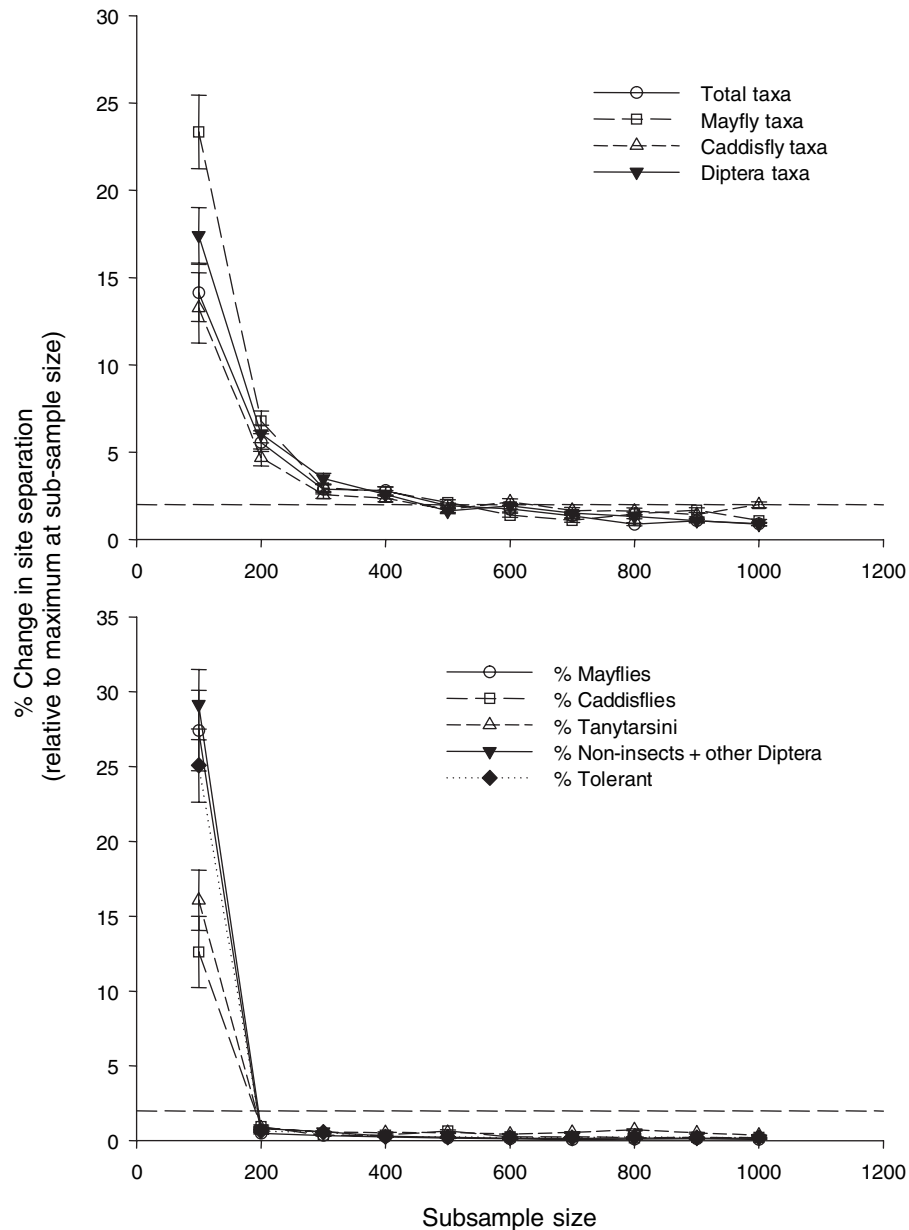


Figure 4. Relative return as a function of subsample size for each richness (top) and percentage (bottom) metric and site. Error bars represent 1 standard error of the mean value

Data collected using our sampling protocol shows that a representative sample of the benthic macroinvertebrate fauna of the study reaches was reached by sampling both banks on six transects. These results were achieved because the sampling method and design effectively sampled the benthic macroinvertebrate fauna of the dominant habitat types within the reach. We did not observe a consistent relationship between transect spacing (i.e. total reach length) and metric values, indicating that our sampling protocol was relatively robust with respect to variation in reach length. Therefore, flexibility exists that permits the study reach length to be dictated by the assessment spatial scale (e.g. repeating geomorphic units) in question. This flexibility in the design increases the utility and applicability of the protocol in that it can be used to meet the multi-purpose sampling needs of researchers and

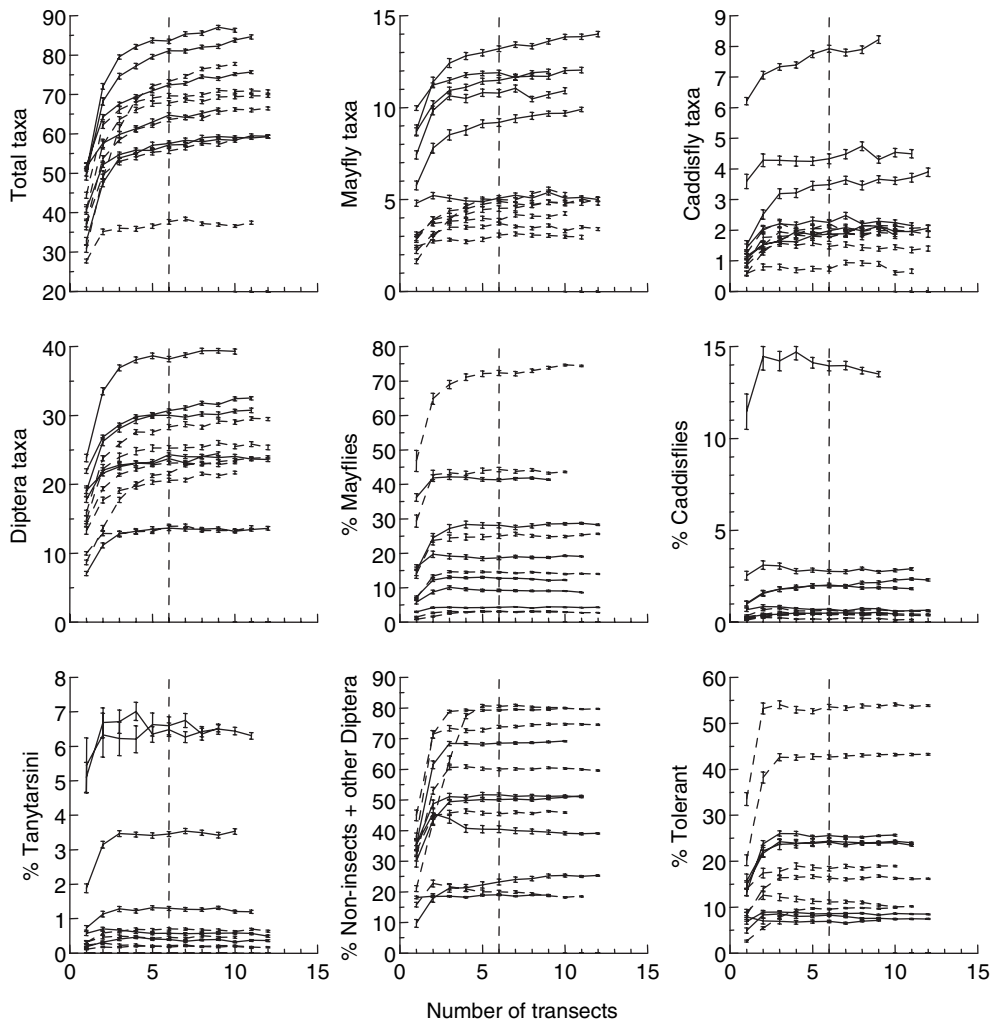


Figure 5. Metric values with increasing number of transects based on 500 organism simulated counts. Solid lines represent Great Miami River and dashed represent Kentucky River sites. Error bars represent 1 standard error. Vertical dashed line represents estimated point at which levelling-off occurs

managers (e.g. trend analysis, point-source and non-point source programmes, habitat rehabilitation and restoration efforts).

For those preferring to use a fixed reach length, we recommend that transects be spaced at a minimum of 100-m intervals over a 500-m reach. This recommendation is based on the observation that, among the sites included in this study, the range of distances covered by six transects was 270 m to 960 m, with a median of 480 m and a mean of 543 m. This distance is equivalent to that used for many years by several state agencies in the USA for sampling riverine benthic macroinvertebrate and fish assemblages (e.g. Ohio Environmental Protection Agency, 1989; Royer *et al.*, 2001). However, caution should be exercised when applying a fixed reach length approach in systems where the reach length selected does not incorporate repeating geomorphic units. Where this occurs, to effectively determine the condition of a reach, it will likely be necessary to establish expectations or a gradient of expectations, for different habitats or combinations of habitats (e.g. a reach containing a riffle versus one that does not).

The recommended design results in a composite sample consisting of 24 20 s kick-net samples and 12-timed samples collected with a D-frame net in habitats complementing those sampled by the kick-net. Therefore, the final composite sample consists of 36 subsamples collected by two complementary sampling techniques. It is likely this approach will over-sample sites of uniform composition, but the goal was to develop a standardized LR-BP that

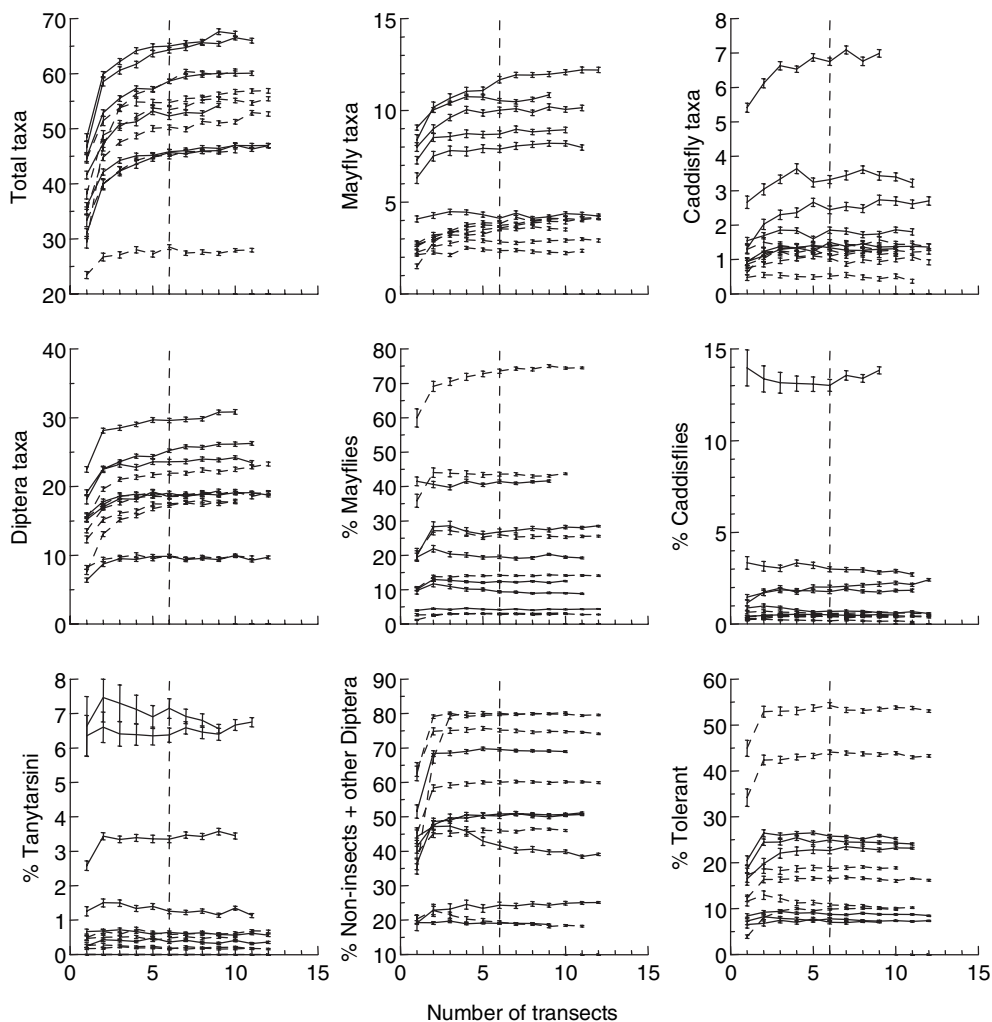


Figure 6. Metric values with increasing number of transects based on 300 organism simulated counts. Solid lines represent Great Miami River and dashed represent Kentucky River sites. Error bars represent 1 standard error. Vertical dashed line represents estimated point at which levelling-off occurs

would perform well across sites of differing habitat composition. These conclusions agree with Bartsch *et al.* (1998), who stated that 18 to 40 subsamples may be required to adequately sample large flood-plain rivers, and that no single sampling technique would efficiently and adequately sample all components of a riverine macroinvertebrate community. It should be noted that this LR-BP for macroinvertebrates has only been tested in main-channel habitats. It may work equally well in off-channel habitats, but this remains to be tested.

We recommend that the field method be coupled with a fixed laboratory subsample size of 300 or 500 organisms to maximize effectiveness of the LR-BP for bioassessment purposes. The fixed laboratory subsample size of 500 organisms does offer lower variability for percentage metrics, but variability for richness metrics was higher. Nonetheless, 300 organisms are likely to be sufficient for most study needs. This recommendation was based on the response of the tested metrics and the observation that the ability to separate sites of different macroinvertebrate composition generally did not increase with larger subsample sizes. Studies on other systems have recommended a broad range of subsample sizes as sufficient (Barbour and Gerritsen, 1996; Vinson and Hawkins, 1996; Grown *et al.*, 1997; Somers *et al.*, 1998). However, a one-size-fits-all subsample size should not be expected, because the quality of information needed by researchers and managers can vary depending on individual studies (Doberstein *et al.*, 2000). The best strategy for determining an appropriate subsample size is to first determine the data quality

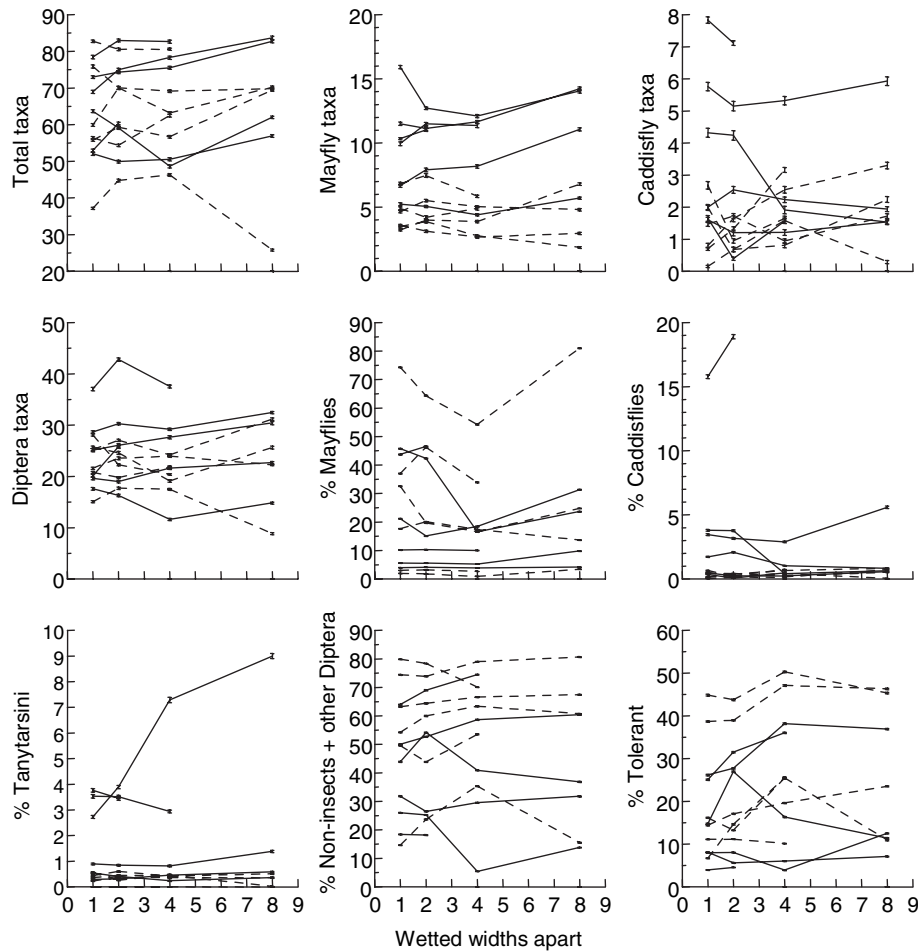


Figure 7. Metric values across groups (define) equidistant by varying numbers of channel widths for each site. Solid lines represent Great Miami River sites and dashed lines represent Kentucky River sites

requirements to meet study objectives and then determine the appropriate subsample size from collected data. This seems especially appropriate when developing new or modifying existing field methods.

The structured nature of the LR-BP provides a standardized sampling protocol that produces a representative sample from the varying habitats and changing impoundment conditions (through time and space) encountered within and across large rivers. Sites in this study varied from free-flowing to those with hydrologic modifications associated with lock-and-dam systems, habitat modifications due to channelization, and the presence of low-head dams. In habitats where both sampling methods could be performed, one method did not supersede the other and both were performed. At others, for example, the banks of a sampling station may have been too steep, rendering the collection of a sample via the kick-net method logistically impossible. However, the benthos could be sampled using the D-frame net from the boat. Hence, a habitat that would have gone unrepresented using a sampling approach that relied purely on kick-net sampling, was still represented in the composite sample of the site.

It is possible that the dominant habitats in some rivers may not be adequately sampled using the combination of sampling methods described in this study. Therefore, it seems reasonable to consider substituting one of our sampling devices with one better suited to sampling the prevailing habitats (e.g. ponar sampling; Poulton *et al.*, 2003). However, the substitution should be consistent across all sites being assessed. Such modifications could result in more accurate and consistent representation of the resource, but may also give rise to data comparability issues when seeking to combine data sets using different sampling methods.

As a result of the additional equipment required to work in non-wadeable streams and rivers (i.e. boats and associated equipment), the effort required to secure a representative sample for bioassessment generally exceeds that required in wadeable streams. Given these realizations, the proposed sampling method is cost effective, logistically feasible, and collects a representative sample for bioassessment purposes. Critical elements of the LR-BP include the complementary sampling techniques, the distance of the sample reach, the number of transects at which both banks are sampled, and the subsample size in the laboratory.

Our objective was to develop a standardized large river bioassessment protocol for macroinvertebrate assemblages. However, we realize that only a certain level of standardization will give ecologically meaningful information for all non-wadeable rivers. Any protocol that is developed may have to be modified to meet some study requirements. Given this, it is our hope that the protocol presented herein serves many as a useful foundation for development of river research and assessment programmes.

Future research

With development of this initial design, additional field sampling has been conducted to allow performance-based testing (Diamond *et al.*, 1996) of the field and laboratory components of the LR-BP. Additional research may also be needed to determine applicability of the LR-BP for use in riverine ecosystems functioning differently than those described in this study (e.g. floodplain-river ecosystems, riverine-influenced reservoirs, fast-flowing rivers). Possible modifications to the method to streamline its application in the field include using the D-frame net configuration for both the kick- and dip-net sampling, setting a depth criterion for kick-net sampling (e.g. 1 m), and using a fixed distance for sample zones (e.g. 10 m). Experimenting with an area quantification of the dip-net sampling may also be considered for use in studies requiring full quantification of sampling effort.

ACKNOWLEDGEMENTS

We thank Robert M. Hughes (Dynamac), James (Sam) Stribling (Tetra Tech), Steve Foster (U.S. ACE), Ken Fritz, Florence Fulk and Brent Johnson (U.S. EPA) and anonymous referees for their critical reviews that greatly improved this manuscript. The United States Environmental Protection Agency through its Office of Research and Development and the Regional Method Initiative funded the research described here. SoBran, Inc. provided laboratory support for sample processing under contract number 68-C6-0019. This paper has been subjected to Agency review and approved for publication. The views expressed in this paper are those of the authors and do not necessarily reflect the views and policies of the U.S. Environmental Protection Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

REFERENCES

- Allan JD. 2000. *Stream Ecology: Structure And Function Of Running Waters*. Kluwer Academic Publishers: Dordrecht.
- Angermeier PL, Schlosser IJ. 1989. Species-area relationships for stream fishes. *Ecology* **70**: 1450–1462.
- Barbour MT, Gerritsen J. 1996. Subsampling of benthic samples: a defense of the fixed-count method. *Journal of the North American Benthological Society* **15**: 386–391.
- Barbour MT, Gerritsen J, Snyder BD, Stribling JB. 1999. *Rapid Bioassessment Protocols For Use In Streams And Wadeable Rivers: Periphyton, Benthic Macroinvertebrates, And Fish* (2nd edn). EPA 841-B-99-002. US Environmental Protection Agency, Office of Water: Washington, DC.
- Bartsch LA, Richardson WB, Naimo TJ. 1998. Sampling benthic macroinvertebrates in a large floodplain river: considerations of study design, sample size, and cost. *Environmental Monitoring and Assessment* **52**: 425–439.
- Benke AC. 1990. A perspective on America's vanishing streams. *Journal of the North American Benthological Society* **9**: 77–88.
- Blocksom KA, Flotemersch JE. 2005. Comparison of macroinvertebrate sampling methods for nonwadeable streams. *Environmental Monitoring and Assessment* **102**: 243–262.
- Cairns J Jr. 1982. *Artificial Substrates*. Ann Arbor Science Publications, Inc.: Ann Arbor.
- Cao Y, Larsen DP, Hughes RM. 2001. Evaluating sampling sufficiency in fish assemblage surveys: a similarity-based approach. *Canadian Journal of Fisheries and Aquatic Sciences* **58**: 1782–1793.
- Carter JL, Resh VH. 2001. After site selection and before data analysis: sampling, sorting, and laboratory procedures used in stream benthic macroinvertebrate monitoring programs by USA state agencies. *Journal of the North American Benthological Society* **20**: 658–682.

- Cochran WG. 1977. *Sampling Techniques*. John Wiley and Sons: New York.
- Diamond JM, Barbour MT, Stribling JB. 1996. Characterizing and comparing bioassessment methods and their results: a perspective. *Journal of the North American Benthological Society* **15**: 713–727.
- Doberstein CP, Karr JR, Conquest LL. 2000. The effect of fixed-count subsampling on macroinvertebrate biomonitoring in small streams. *Freshwater Biology* **44**: 355–371.
- Flotemersch JE, Autrey BC, Cormier SM. 2001. *Comparisons of Boating and Wading Methods Used to Assess the Status of Flowing Waters*. US Environmental Protection Agency: Cincinnati, EPA/600/R-00/108.
- Frissell CA, Liss WJ, Warren CE, Hurley MD. 1986. A hierarchical framework for stream habitat classification: viewing streams in a watershed context. *Environmental Management* **10**: 199–214.
- Gorman OT, Karr JR. 1978. Habitat structure and stream fish communities. *Ecology* **59**: 507–515.
- Growns JE, Chessman BC, Jackson JE, Ross DG. 1997. Rapid assessment of Australian rivers using macroinvertebrates: cost and efficiency of 6 methods of sample processing. *Journal of the North American Benthological Society* **16**: 682–693.
- Herlihy AT, Lazorchak JM. 2000. Initial Site Procedures. In *Environmental Monitoring and Assessment Program-Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Non-Wadeable Rivers and Streams*, Lazorchak JM, Hill BH, Averill DK, Peck DB, Klemm DJ (eds). EPA 620/R-00/007. Office of Research and Development, US Environmental Protection Agency: Cincinnati; 4.1–4.8.
- Hughes RB, Kaufmann PR, Herlihy AT, Intelmann SS, Corbett SC, Arbogast MC, Hjort RC. 2002. Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. *North American Journal of Fisheries Management* **22**: 1229–1240.
- Humphries PJ, Growns E, Serafina LG, Hawking JH, Chick AJ, Lake PS. 1998. Macroinvertebrate sampling methods for lowland Australian rivers. *Hydrobiologia* **364**: 209–218.
- Hynes HBN. 1970. *The Ecology of Running Water*. Liverpool University Press: Liverpool.
- Karr JR, Chu EW. 1999. *Restoring Life in Running Waters: Better Biological Monitoring*. Island Press: Washington, DC.
- Kaufmann PR. 2000. Physical Habitat Characterization—non-wadeable rivers. In *Environmental Monitoring and Assessment Program-Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Non-Wadeable Rivers and Streams*, Lazorchak JM, Hill BH, Averill DK, Peck DB, Klemm DJ (eds). EPA 620/R-00/007. Office of Research and Development, US Environmental Protection Agency: Cincinnati; 6.1–6.29
- Kentucky River Authority. 1999. *Kentucky River Basin Status Report*. <http://water.nr.state.ky.us/dow/pubs/kybasin97/status1.htm>. Kentucky River Authority: Frankfort.
- Kilgour BW, Barton DR. 1999. Associations between stream fish and benthos across environmental gradients in southern Ontario, Canada. *Freshwater Biology* **41**: 553–566.
- Klemm DJ, Lewis PA, Fulk FA, Lazorchak JM. 1990. *Macroinvertebrate Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters*, EPA 600/4-90/030. US Environmental Protection Agency: Cincinnati.
- Klemm DJ, Lazorchak JM, Peck DV. 2000. Benthic macroinvertebrates. In *Environmental Monitoring and Assessment Program-Surface Waters: Field Operations and Methods for Measuring the Ecological Condition of Non-Wadeable Rivers and Streams*, Lazorchak JM, Hill BH, Averill DK, Peck DB, Klemm DJ (eds). EPA 620/R-00/007. Office of Research and Development, US Environmental Protection Agency: Cincinnati; 9.1–9.14.
- Leopold LB, Wolman MG, Miller JP. 1964. *Fluvial Processes in Geomorphology*. W.H. Freeman: San Francisco.
- Li J, Herlihy A, Gerth W, Kaufmann P, Gregory S, Urquhart S, Larsen D. 2001. Variability in stream macroinvertebrates at multiple spatial scales. *Freshwater Biology* **46**: 87–97.
- Loucks DP. 2003. Managing America's rivers: who is doing it? *International Journal of River Basin Management* **1**: 23–31.
- Lyons J. 1992. The length of stream to sample with a towed electrofishing unit when fish species richness is estimated. *North American Journal of Fisheries Management* **12**: 198–203.
- Manly BFJ. 2001. *Statistics for Environmental Science and Management*. Chapman and Hall/CRC: Boca Raton.
- Metcalfe-Smith JL. 1994. Biological water-quality assessment of rivers: use of macroinvertebrate communities. In *The Rivers Handbook: Hydrological and Ecological Principles*, Calow P, Petts GE (eds). Blackwell Scientific: Oxford; 144–170.
- Miami Conservancy District. 2004. <http://www.miamiconservancy.org/> [22 April 2004].
- Moulton SR, Kennen JG, Goldstein RM, Hambrook JA. 2002. *Revised protocols for sampling algal, invertebrate, and fish communities as part of the National Water Quality Assessment Program*. U.S. Geologic Survey Open File Report 02-150.
- Ohio Environmental Protection Agency. 1988. *Biological criteria for the protection of aquatic life: Volume II: users manual for biological field assessment of Ohio surface waters*. Ohio EPA, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus.
- Ohio Environmental Protection Agency. 1989. *Biological criteria for the protection of aquatic life, Volume III: Standardized biological field sampling and laboratory methods for assessing fish and macroinvertebrate communities*. Ohio EPA, Division of Water Quality Monitoring and Assessment, Surface Water Section, Columbus.
- Ohio Environmental Protection Agency. 1997. *Biological and Water Quality Study of the Middle and Lower Great Miami River and Selected Tributaries, 1995*. Division of Surface Water, Monitoring and Assessment Section, Ohio Environmental Protection Agency, Columbus.
- Poulton BC, Wildharber ML, Charbonneau CS, Fairchild JF, Mueller BG, Schmitt CJ. 2003. A longitudinal assessment of the aquatic macroinvertebrate community in the channelized lower Missouri River. *Environmental Monitoring and Assessment* **85**: 23–53.
- Rosenberg DM, Resh VH. 1993. *Freshwater Biomonitoring and Benthic Macroinvertebrates*. Chapman and Hall: London.
- Royer TV, Robinson CT, Minshall GW. 2001. Development of macroinvertebrate-based index for bioassessment of Idaho rivers. *Environmental Management* **27**: 627–636.

- Somers KM, Reid RA, David SM. 1998. Rapid biological assessments: how many animals are enough? *Journal of the North American Benthological Society* **17**: 348–358.
- Sorensen T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *Det Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter, (Copenhagen)* **5**: 1–34.
- Sweeting RA. 1994. River Pollution. In *The Rivers Handbook: Hydrological and Ecological Principles*, Calow P, Petts GE (eds). Blackwell Scientific: Oxford; 23–32.
- Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. 1980. The river continuum concept. *Canadian Journal of Fisheries and Aquatic Sciences* **37**: 130–137.
- Vinson MR, Hawkins CP. 1996. Effects of sampling area and subsampling procedure on comparisons of taxa richness among streams. *Journal of the North American Benthological Society* **15**: 392–399.
- Vogelmann JE, Sohl TL, Campbell PV, Shaw DM. 1998. Regional land cover characterization using landsat thematic mapper data and ancillary data sources. *Environmental Monitoring and Assessment* **51**: 415–428.