Water Qual. Res. J. Canada, 2007 · Volume 42, No. 3, 184-193

An Evaluation of Rapid Bioassessment Protocols for Stream Benthic Invertebrates in Southern Ontario, Canada

Jeff P. Borisko,[1,5]* Bruce W. Kilgour,[2] Les W. Stanfield,[3] and F. Chris Jones[4]

[1] *Toronto and Region Conservation Authority, Downsview, Ontario M3N 1S4*
[2] *Stantec Consulting, Ottawa, Ontario K1Z 7T1*
[3] *Ontario Ministry of Natural Resources, Picton, Ontario K0K 2T0*
[4] *Ontario Ministry of the Environment, Dorset, Ontario P0A 1E0*
[5] *Current address: Lower Trent Conservation, 714 Murray Street, R.R. 1, Trenton, Ontario K8V 5P4*

In this study, we sampled stream benthos using rapid bioassessment methods (i.e., D-nets, coarse taxonomy) from sites in and around Toronto, Ontario that represented a range of stream characteristics (e.g., drainage areas). The protocols were developed by or adapted from: (1) Toronto and Region Conservation Authority, (2) Ontario Ministry of Natural Resources, (3) Ontario Benthos Biomonitoring Network/Ministry of the Environment, and (4) Canadian Aquatic Biomonitoring Network/ Environment Canada. Summary indices (Hilsenhoff's modified Biotic Index [HBI]; percent of fauna as Ephemeroptera, Plecoptera, and Trichoptera [% EPT]; number of taxa [*S*]; and Shannon's Diversity Index [*H'*]) calculated from the resulting data were used to evaluate differences between protocols in a series of analyses including analysis of variance and variance components analysis. The study found that sampling method was an unimportant source of variation in summary index values relative to other factors such as the stream or year sampled. The largest percent variance amongst stream or year factors (main and interactions) was at least three times greater than the corresponding method related percent variances. The results of this study suggest that these four bioassessment methods are interchangeable within the context of large geographic scales or for the detection of major impacts. In cases where impacts to the benthic community are subtle, protocols should include one or more of discrete habitat sampling, replication, and lower taxonomic resolution. The data and findings here may help agencies integrate different sampling protocols into their biomonitoring and assessment programs.

Key words: **benthic invertebrates, stream sampling, biomonitoring programs, rapid bioassessment protocols, methods standardization, aquatic ecology**

Introduction

Benthic invertebrates are used to assess stream condition and, in some cases, to detect changes in biological condition over time. They are effect-based indicators which are useful for making environmental management decisions. Benthos are relatively easy and inexpensive to collect, are variously sensitive to ambient environmental conditions, and are sedentary. They thus integrate and accumulate conditions at a site and can provide a level of diagnosis as to the nature of observed effects. Some variables that have been shown to directly impact benthos include: stream water temperature (Sponseller et al.; 2001Wang and Kanehl 2003), dissolved oxygen (i.e., organic enrichment) and/or siltation (Lemly 1982; McClelland and Brusven 1980), and pesticides (Schulz and Liess 1999; Cuppen et al. 2000).

In Canada, the use of benthic invertebrates for bioassessments in stream systems extends across all geographic and political landscapes. Federally, Environment Canada has incorporated benthos surveys in Environmental Effects Monitoring (EEM) programs for pulp and paper mills and mines (Dumaresq et al. 2002; Walker et al. 2002), and as part of the proposed approach to monitoring municipal wastewaters (Kilgour et al. 2005). The Canadian Aquatic Biomonitoring Network (CABIN), also part of Environment Canada, uses benthos on larger scales in partnership with various agencies to understand the quality of Canada's freshwater and aquatic biodiversity. CABIN emphasizes finer taxonomic detail for regional applications to both water quality evaluations and risk assessment. The Environmental Monitoring and Assessment Network (EMAN) of Environment Canada recommends two protocols; one authored by the Department of Fisheries and Oceans (Rosenberg et al. 1998) and a second by the Ontario Benthos Biomonitoring Network/Ministry of the Environment (Jones et al. 2004).

In Ontario, various agencies such as the Ministries of the Environment, Natural Resources, and Municipal Affairs and Housing collected or currently collect benthic invertebrates as part of their monitoring and assessment activities. The protocols used by these agencies originated from efforts by the United States Environmental Protection Agency to develop rapid bioassessment protocols (RBPs) in the late 1980s and 1990s (Bode 1988; Plafkin et al. 1989; Barbour et al. 1999). These RBPs have been used extensively in state and federal monitoring programs (Iowa DNR 2004; Ohio EPA 2006) as well as in studies on impacts such as acid mine drainage (Rhea et al. 2006) and wastewater treatment facilities (Winger et al. 2005). A RBP can be described as a cost-effective and scientifically valid procedure that allows multiple biological surveys in

* Corresponding author; jborisko@bqrap.ca

a single field season with a quick turn-around of results for management decisions (Barbour et al. 1999). The use of RBPs for stream systems in Ontario is extensive. But the literature citing their application and resulting data is rare (Kilgour and Barton 1999; Stanfield and Kilgour 2005) or not readily available particularly for large-scale (i.e., provincial/regional) monitoring programs and studies.

At the municipal level in Ontario, most collections are undertaken by watershed-based agencies called Conservation Authorities, however some municipalities also independently sample benthos (e.g., City of Ottawa). There are two new initiatives in Ontario that are attempting to standardize benthos sampling methods. CABIN represents national interests, while the Ontario Benthos Biomonitoring Network (OBBN) is a collective of provincial (Ontario Ministry of the Environment) and municipal agencies (e.g., Nottawasaga Valley Conservation Authority) that is attempting to standardize stream, lake, and wetland sampling methods to facilitate data sharing. The OBBN protocols are derived from the federal CABIN protocols, and share a common database. Ideally, benthos data would be shared among provincial and municipal agencies, and could then feed the larger scale initiatives (i.e., CABIN, EMAN) to facilitate reporting at national scales.

A challenge, however, lies with agencies that wish to be part of the larger networks, but have developed substantive historic databases and used methods that are different than those being proposed under OBBN and CABIN. Information is needed to evaluate if and how these historic datasets can be used in conjunction with these newer protocols, particularly since the differences between these RBPs may be quite subtle. There are also issues with respect to applying new and historic datasets to answer questions of impact using different protocols. If, however, there was information on how the results (i.e., summary indices) compared between sampling methods, then older historical data could contribute to the larger national network database, and agencies might be more comfortable in adopting or integrating different protocols.

The objective of this study was to evaluate differences between sampling protocols within the context of spatial and temporal variation, which is part of any bioassessment. This was ultimately accomplished by assessing the relative importance of variation in summary indices between methods versus other factors (i.e., sample year, stream). A means to account for variance sources (i.e., confounding factors) becomes important when trying to detect impacts to benthic communities, and has been done elsewhere through variance components analysis (Reid et al. 1995; Somers et al. 1998). To date, there have been numerous studies that have compared benthic invertebrate sampling methods using various measures and aquatic habitats (Mackey et al. 1984; Barton and Metcalfe-Smith 1992; Kerans et al. 1992; Lenz and Miller 1996; Muzaffar and Colbo 2002; Scarsbrook and Halliday 2002; O'Connor et al. 2004; Blocksom and Flotemersch 2005). Other studies have used approaches that are more empirical (Cao et al. 2005; Herbst and Silldorf 2006) or suggested performance-based methodologies (Diamond et al. 1996). This study used a complementary approach where the variation associated with common bioassessment factors was explicitly measured and, in turn, could be related to other protocol issues such as analytical power and precision.

In this study, we sampled stream benthos from several minimally-exposed locations in and around Toronto that represented a range of stream characteristics. The protocols were developed by or adapted from: (1) Toronto and Region Conservation Authority (TRCA) (Stanfield 2005); (2) Ontario Ministry of Natural Resources (MNR) (Stanfield 2005); (3) OBBN (Jones et al. 2004); and (4) CABIN (Reynoldson et al. 2002). Benthic taxa counts and summary biological indices calculated from the resulting data were used in a series of analyses related to the study objective.

Materials and Methods

Study Sites

Sampling was carried out in the fall of 2003 with the TRCA, MNR, and CABIN methods, and repeated in 2004 with the addition of the OBBN method. All 11 sites (Fig. 1) were minimally-exposed with no obvious point source impacts, and varied in upstream catchment physiographies. Sites were located in the three larger watersheds in TRCA jurisdiction (Humber River, Rouge River, and Duffins Creek).

Drainage areas for the study sites ranged from small (e.g., 3.5 km^2 for Rouge25) to large (e.g., 810.2 km^2 for Humber07). Drainage areas were derived through use of a Geographic Information System (GIS) application with a digital elevation model of 25-m resolution following the principles in Stanfield and Koyvenhoven (2003). The same GIS estimated site slope from elevations at 100-m upstream and 100-m downstream of each sampling site. On a finer scale, there was variability in baseflow index, water depths, and water/habitat quality (Table 1). The baseflow index calculated here has been shown to relate to in-situ baseflow (Piggott et al. 2002) and was calculated by ranking each quaternary surficial geology unit based on its depth to bedrock and hydraulic conductivity, and then summing the ranked areas for each catchment. Water depths were calculated through the averaging of 40 to 60 single-point water depths measured along a series of transects as per Stanfield (2005). Water quality was measured by Hydrolab and was generally good, with roughly neutral pH and dissolved oxygen levels near saturation. Substrate ranged from silt/sand to cobble/boulder. Streams exhibited cool water conditions.

Field and Laboratory

The TRCA method involved sampling benthos with a D-framed net (500-μm mesh). The stream substrate was "kicked and swept" along each of 10 to 20 transects (number depends on stream width) situated between two

TABLE 1. Summary of stream characteristics for sites surveyed in this study

Site Characteristics	*Minimum*	*Median*	*Maximum*
Drainage area (km^2)	3.5	31.8	810.2
Site slope (m)	0	1.2	2.87
Baseflow Index (no units)	38.9	42.7	59.7
During summer months			
Average water depth (mm)	64.1	126.5	236.3
Daily minimum temp (^{O}C)	13.9	17.2	19.5
Daily mean temp (^{O}C)	15.0	18.3	20.3
Daily maximum temp (^{O}C)	16.1	20	22.1
At time of benthic survey			
Temp (^{O}C)	2.5	8	9.3
pH	7.94	8.33	8.59
Conductivity (?S/cm)	221.3	561.2	792.4
Dissolved oxygen (mg/L)	7.7	9.7	13.4

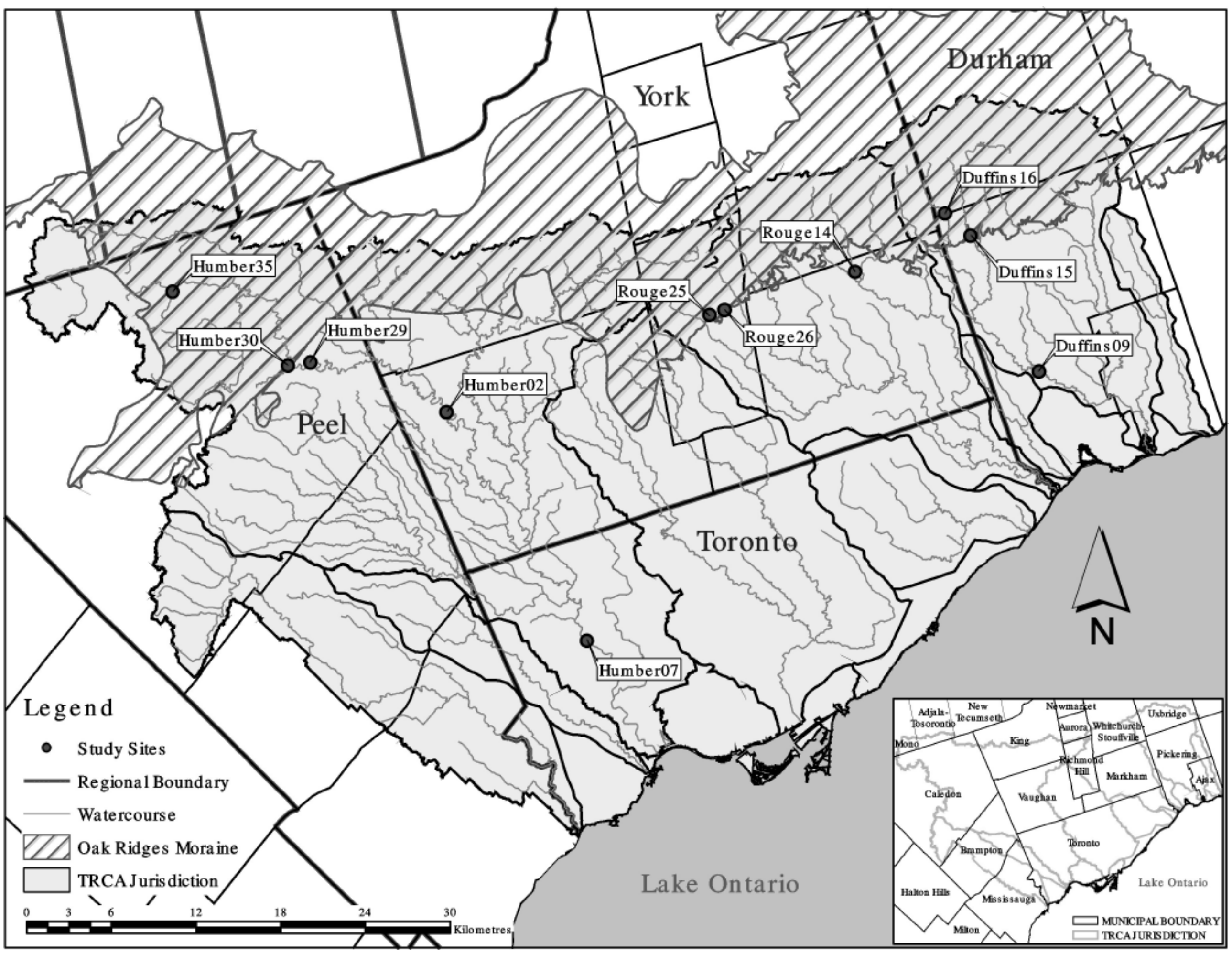

Fig. 1. Map of study area showing 11 sampling sites in TRCA jurisdiction, Toronto, Ontario.

crossovers (i.e., riffles) (Stanfield 2005). The sample was a composite from all transects. The OBBN method also used a 500-μm D-framed net and collected a sample along three transects: one in pool habitat and two in riffle habitats (Jones et al. 2004). Each transect sample (i.e., collection) was processed separately with the resulting count data averaged across the three transects to calculate index values. The CABIN method used a 3-minute, zigzag, bank-to-bank traveling kick primarily within riffle/glide habitat (Reynoldson et al. 2002). This method specifies the use of a 400-μm kick-net, however this study used a 500-μm D-framed net. Finally, the MNR sample consisted of a fixed-area kick sample collected within a crossover (i.e., riffle) (Stanfield 2005). The area kicked was approximately 1 m^2, while the mesh of the kick net was approximately 1 mm (window screen). Samples for each protocol were collected in a downstream to upstream direction and coordinated so that no area was disturbed upstream prior to collection. All samples at a given site were collected within the same 2- to 3-h time period, and were preserved in the field with 80% ethanol. Animals were sorted from the debris with the assistance of a dissecting microscope at 10 to 15x magnification. A minimum of 100 animals from each collection were randomly picked and identified to major group (27 groups, combination of Phylum, Order, and Family) as per Stanfield (2005).

Statistical Analyses

This analysis was designed to produce three lines of evidence to compare differences between methods in the description of the benthic community and ultimately measure the importance of spatial and temporal factors versus sampling method. For the first line of evidence, we simply compared taxa counts across all sites for 2003 and 2004 surveys separately to determine if any taxa were missed or added between the different protocols. Boxplots were also used to graphically represent the calculated index values to show potential differences between protocols. We then used a complete randomized-block design analysis of variance (ANOVA) with Bonneferroni post-hoc tests to determine whether differences (if any) between mean index values were statistically significant ($\alpha = 0.05$). The analysis was done separately for each year since the OBBN method was not used in 2003. A block design was used to remove some of the expected variation in index values associated with the sites. Removing this variation source provided greater analytical power to detect differences between collection methods.

In the second line of evidence, community index values for each method were compared relative to variation among stream sites for each index. Specifically, the average index value was calculated for each method-by-year combination. The difference in average values for each paired method (e.g., CABIN 2003 versus MNR 2003) was then divided by the root mean square calculated from the residual (i.e., error term) in the ANOVA described above. This calculated number represented an "effect size" (δ) that was used as a more objective measure for comparing methods. Following the work of Kilgour et al. (1998), an effect size greater than 2 ($\delta > 2$) was adopted as a criterion to detect the presence of a method effect. Again, the analysis was done separately for each year since the OBBN method was not included in the 2003 collection.

For the third line of evidence, the relative importance of collection method in benthic community index values was assessed by variance components analysis (VCA). The VCA used the expected mean square values from an ANOVA to estimate variance components, which were then represented as percent variances for each factor/interaction (Quinn and Keough 2002). The variance components were sampling site (random), sampling method (fixed), sampling year (fixed), and their interactions. The ANOVA required that the variance associated with the three-way interaction (site by method by year) remained in the residual (i.e., model error). This allowed the estimation of variance contributions for the main factors (e.g., year) and simple interactions (e.g., year by site) given the limited replication in this study. In addition, only three of the studied methods (CABIN, MNR, TRCA) were included in the VCA as they were used in both 2003 and 2004.

These analyses were conducted for a suite of basic benthic summary metrics that included (1) number of taxa (*S*), (2) Shannon's Diversity Index (*H'*), (3) percent of fauna as Ephemeroptera, Plecoptera, and Trichoptera (% EPT), and (4) Hilsenhoff's modified Biotic Index (HBI) based on Stanfield and Kilgour (2005). The choice of which index and the number of indices to include was somewhat arbitrary. Each of the chosen indices is commonly reported in the literature, easy to calculate, and represents a different type as described in David et al. (1998). For example, *S* is described as a "richness measure" whereas % EPT is a "compositional index". These four indices were also selected given their commonness in agency biomonitoring programs, particularly those that have a larger geographic focus. In one such large-scale study in the United States, the authors only used the number of EPT taxa to study the effect of sampling different habitat types in l regional bioassessment surveys (Gerth and Herlihy 2006).

Results

In 2003 and 2004, no sampling protocol obviously missed or added major taxonomic groups (Table 2). The average counts for the surveyed sites were comparable between protocols. For example, common taxa such as Coleoptera were equally abundant for all methods, whereas uncommon taxa such as Culicidae were equally rare

In general, the calculated index scores demonstrated that the study sites were minimally impaired (Fig. 2). The majority of HBI scores were less than 6, typically indicating an unimpaired benthic community (Kilgour 1998; Hilsenhoff 1987). Similarly, the sensitive EPT taxa were found in all samples with the median % EPT score greater than 20% in all cases.

Overall, there was limited evidence to indicate that

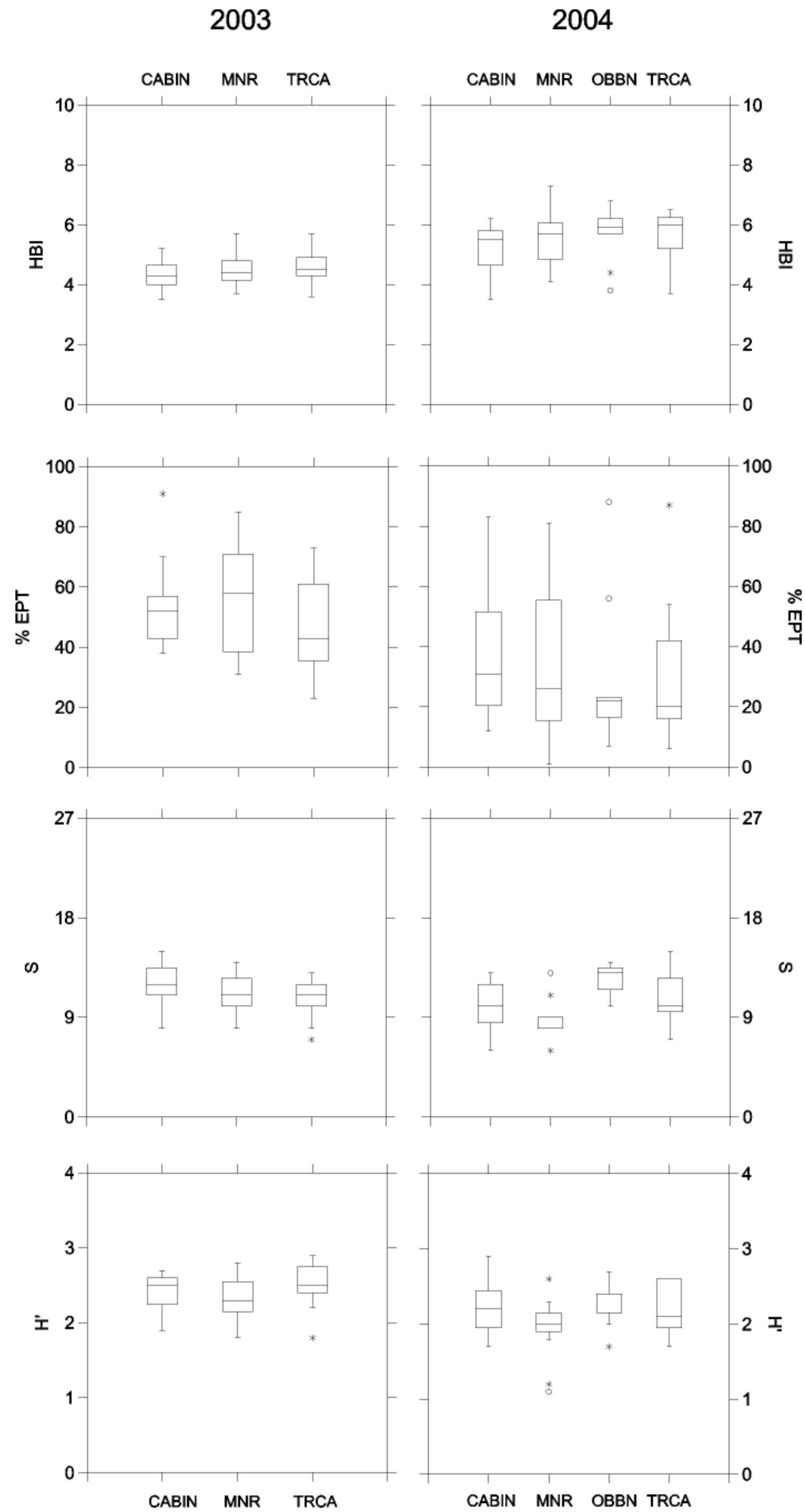

Fig. 2. Boxplots for HBI, % EPT, *S*, and *H'* indices grouped by method for 2003 and 2004. The center vertical line marks the median and the length of each box shows the range within which the central 50% of the values fall (i.e., interquartile range). Asterisks indicate values between 1.5 to 3 times the interquartile range. Empty circles indicate values greater than 3 times the interquartile range.

TABLE 2. Average number of individuals per taxa of a possible 27 taxa [a]

	2003			2004			
Taxa	*CABIN*	*MNR*	*TRCA*	*CABIN*	*MNR*	*OBBN*	*TRCA*
Acarina	0	0	0	0	0	0	1
Amphipoda	3	2	5	2	1	4	2
Anisoptera	0	0	1	0	0	0	0
Ceratopogonidae	0	0	0	0	0	0	1
Chironomidae	8	18	16	56	73	53	76
Coleoptera	25	20	21	21	18	10	16
Misc. Diptera	1	1	2	1	1	2	3
Ephemeroptera	14	17	9	20	16	13	16
Gastropoda	1	1	2	1	0	1	0
Hemiptera	1	0	0	0	0	0	0
Megaloptera	0	1	2	0	0	0	1
Oligochaeta	2	2	3	5	9	16	10
Pelecypoda	4	3	4	4	0	2	2
Plecoptera	9	8	10	11	5	10	8
Simuliidae	1	0	0	1	0	1	0
Tipulidae	6	4	4	5	3	4	5
Trichoptera	36	41	34	21	28	15	19

[a] As per Stanfield et al. (2005), and Jones et al. (2004) for each sampling method in 2003 and 2004 (n = 11).

different sampling methods would yield different community index values. Only taxa richness, (S) and to a lesser degree diversity (H'), visibly differed among methods (in 2004), with the OBBN method producing higher values than the MNR method (Fig. 2). From the ANOVA results in Table 3, these same cases showed either a significant method effect (S in 2004) or a slightly greater F value (i.e., signal) for the method factor versus the site/block factor (H' in 2004). But, the results from more objective criteria (i.e., second line of evidence) showed no cases with $\delta > 2$ (including S and H' in 2004), which indicated the absence of a method effect (Table 4).

Sampling method was a relatively unimportant source of variation for the indices of benthic community composition (Table 5). The percent variance was consistently greater for year and sampling site factors as compared to sampling method. The largest one- and two-way nonmethod variance sources (explained by the model) for each index were at least three times greater than the corresponding method variance sources. Although sampling method significantly ($P < 0.05$) influenced values for two of the indices (HBI, % EPT) as measured by the ANOVA used in the VCA, the percent variance associated with

TABLE 3. F values for method & site/block factor from ANOVA [a]

	2003		2004	
Index	*method*	*site/block*	*method*	*site/block*
HBI	2.808	8.275	2.883	14.129
% EPT	3.441	9.230	2.294	21.727
S	1.063	1.230	8.699 *	2.643
H'	1.493	2.240	2.398	1.976

[a] A significant method effect indicated by asterisk (α=0.05).

TABLE 4. Calculated effect sizes (δ) for each method/index comparison by sample year [a]

Year	*Method/Index*	*HBI*	*% EPT*	*S*	*H'*
2003	CABIN vs MNR	0.26	0.08	0.36	0.20
	CABIN vs TRCA	0.35	0.28	0.57	0.33
	MNR vs TRCA	0.09	0.35	0.21	0.53
2004	CABIN vs MNR	0.22	0.08	0.41	0.61
	CABIN vs OBBN	0.32	0.22	0.91	0.14
	CABIN vs TRCA	0.24	0.17	0.38	0.01
	MNR vs OBBN	0.10	0.14	1.32	0.75
	MNR vs TRCA	0.01	0.09	0.78	0.60
	OBBN vs TRCA	0.65	0.05	0.53	0.15

[a] A value of $\delta > 2$ indicates a method effect (Kilgour et al. 1998).

TABLE 5. Summary of VCA of benthic invertebrate stream sampling protocols for 2003 and 2004 [a]

	% Variance			
Factor	*HBI*	*% EPT*	*S*	*H'*
Year	42	23	10	19
Method	2	2	2	6
Site	35	55	6	8
Year by Site	9	8	24	15
Method by Site	0	1	0	0
Method by Year	0	0	10	0
Year by Method by Site [b]	12	11	48	52

[a] Only CABIN, MNR, & TRCA methods included in analysis.
[b] Residual.

method was relatively small (i.e., less than 3%). Amongst the indices used in this study, the largest method related variance was *S* in the interaction with year (10%). In addition, percent variances were comparable within, but not between, the HBI and % EPT pair and the *S* and *H'* pair.

Discussion

The emergence of programs such as the OBBN and CABIN present opportunities such as data sharing for agencies and organizations involved in benthic invertebrate monitoring. Groups with historic data sets and existing monitoring protocols, however, are faced with two separate but related issues. The first issue concerns the utility of data collected by protocols other than those included in the larger monitoring network. If historic and network collection methods are not deemed comparable, data gaps could result, compromising long-term monitoring and reporting. The second issue is demonstrating the need for replacing or integrating network protocols with current benthic invertebrate monitoring methods. If historic and network collection methods vary in their ability to meet study objectives, changes may be warranted. Furthermore, the existence of useful comparative datasets would make the decision to replace or adopt alternate protocols less arbitrary.

In this study, sampling method was an unimportant source of variation for indices of benthic community composition relative to other factors such as the stream and year sampled (and their interactions). The variance associated with the main method factor was never more than half of the next largest nonmethod variance source, in most cases. The results here demonstrated the utility of VCA, and the importance of spatial and temporal factors in the evaluation of alternate sampling methods. Although not explicitly examined using VCA, other studies also suggested that sampling method was not as important as other factors such as site and/or year sampled (Scarsbrook and Halliday 2002; O'Connor et al. 2004). Studies that do not consider these other factors risk limitations on the applicability of their findings (Mackey et al. 1984; Muzaffar and Colbo 2002), particularly when such methods are applied over broader scales where spatial and temporal variation is likely important. A related consequence of not looking at variance sources is that methods thought to be different may in fact be comparable.

If sampling method represents a relatively small portion of all measured variation, as found here, there are two important implications. First, the methods can be considered coarse. As applied in this study, the sampling methods did not include any spatial replication, used a mix of Order/Family taxonomy, and utilized 500 micron D-nets. They are rapid bioassessment protocol (RBP) tools. Second, these RBPs are unlikely to detect subtle impacts to the benthic community. This is a consequence of the first point and is related to the insufficient statistical power to detect subtle differences (Kerans et al. 1992). These protocols are better suited to the detection of major impacts or gross impairment (Kilgour et al. 2005). Unlike previous studies, the comparative approach used here did not measure how well they detect impacts (Barton and Metcalfe-Smith 1992; Blocksom and Flotemersch 2005) or meet method objectives (Mackey et al. 1984; Kerans et al.1992; Scarsbrook and Halliday 2002; O'Connor et al. 2004). These previous studies found variation among methods and were more detailed in terms of taxonomy, spatial replication, and discrete habitat sampling. Given the coarseness of the methods in this study, it would be expected they would all perform similarly, even with a suitable impact gradient or a-priori sampling objectives. The results here indicate the importance of considering whether a chosen sampling method is appropriate for the impact to be studied or the a-priori sampling objective.

Subtle method effects were found and could be related to the different habitats that the samples came from. Differences among methods may not have been as pronounced due to the coarseness of the protocols and/or the relative homogeneity of the streams sampled in terms of habitat types (i.e., riffles and pools). Since this study did not quantify habitat types, the role of this factor could not be fully investigated. Such a study, however, would be impractical given one of the four RBPs used in this study (MNR) was a single habitat method, whereas the other three were multiple habitat methods, of which one had three separate field collections (OBBN). These details would also make an investigation on the role of habitat type somewhat meaningless within the respective agency monitoring programs/objectives or the larger network. Yet, discrete sampling from different habitats may be an important factor, particularly in terms of detecting subtle impacts. Subtle impacts are thought to only be detectable in single habitats, or are obscured by variation in the physical structure of pool and riffle habitats among streams (Kerans et al. 1992). The rationale for discrete, multiple habitat collections also applies to the use of more detailed taxonomy, which should increase the ability to detect impacts (Lenat and Resh 2001; Waite et al. 2004). Unfortunately, the cost to conduct more detailed protocols is often prohibitive to most agencies.

Subtle impacts could be investigated in cases where more detailed protocols are used, and therefore the selection of benthic community metrics would become more important. Summary indices can independently influence the likelihood of detecting impacts (i.e., index sensitivity) to the benthic community (Carlisle and Clements 1999; Kilgour et al. 2004). Some indices are designed for detecting certain impacts such as HBI for organic enrichment (Hilsenhoff 1987), and others are impractical for certain types of sampling methods (e.g., artificial substrates and percent Chironomidae) (Barton and Metcalfe-Smith 1992). The temporal stability of a given index and/or sampling method can also affect bioassessments (Barton and Metcalfe-Smith 1992; Reid et al. 1995; Linke et al. 1999) where shifts in the benthic community are due to seasonal changes (a confounding factor) rather than the impact of concern, thereby leading to incorrect inferences on cause. Previous stream benthos studies have also noted the importance of accuracy where different index/method combinations yielded variable bioassessment results (Barton and Metcalfe-Smith 1992; Kerans et al. 1992; Resh 1994).

There was some evidence that indices performed differently in this study, which may have contributed to observed subtle method effects in spite of the relatively coarse methods. In practice, these basic measures were an extension of the RBP or the complete method (sampling and data), and would ultimately represent an unimportant source of variation. The results from this study suggest that these basic measures of community composition can be compared, even when different protocols are used to collect the data. This finding would be applicable in site specific assessments. Any of the RBPs could be used in a before-after-control impact study design and provide the same likelihood of detecting an impact (or not). The findings here would also be applicable in large-scale assessments. None of the protocols would relatively increase or decrease the influence of spatial and/or temporal variation, and therefore would not likely change the results of any such assessment. Again, this applies only to the coarse protocols and measures applied in this study (i.e., D-nets, major taxonomic groups, no spatial replication). For more detailed protocols where method factor variance could be problematic, recent studies demonstrated that different methods were comparable in terms of bioassessment results (Cao et al. 2005; Herbst and Silldorf 2006). These findings could enable data sharing between agencies, and bioassessments across large geographic areas. But, in studies where the detection of specific stressors is the purpose, methods (sampling and data) may not necessarily be interchangeable (Blocksom and Flotemersch 2005).

In conclusion, this study demonstrated that sampling method was a relatively unimportant source of variation for community index values, and different benthic invertebrate collection methods yield comparable results. Agencies with historic datasets could use basic measures of benthic community composition derived from any of the applied protocols for use in large-scale assessments. In terms of choosing whether to adopt or integrate any of the studied protocols, agencies with new or existing programs could do so but only with the understanding that these are coarse methods and are not likely to detect subtle impacts. This decision would be better guided by clearly defined a-priori program objectives. This would include knowledge on what impacts are to be measured, in addition to equally important practical considerations such as reporting requirements, resource availability, and funding.

Acknowledgments

Funding for this work was provided by Toronto and Region Conservation Authority, Ontario Ministry of Natural Resources and the City of Toronto. We acknowledge the support of the Lake Ontario Modeling Team and its various members. We thank two anonymous referees and numerous reviewers for their helpful comments on earlier versions of this manuscript and a special thanks to Cheryl Goncalves at TRCA for her help in field collection and invertebrate identification.

References

Barbour MT, Gerritsen J, Snyder BD, Stribling JB. 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates, and fish, second edition. EPA 841-B-99-002.

Barton DR, Metcalfe-Smith JL. 1992. A comparison of sampling techniques and summary indices for assessment of water quality in the Yamaska River, Québec, based on benthic macroinvertebrates. Environmental Monitoring and Assessment **21:** 225-244.

Bode RW. 1988. Quality assurance workplan for biological stream monitoring in New York State. New York State Department of Environmental Conservation, Albany, New York.

Blocksom KA, Flotemersch JE. 2005. Comparison of macroinvertebrate sampling methods for nonwadeable streams. Environmental Monitoring and Assessment **102**:243-262.

Cao Y, Hawkins CP, Storey AW. 2005. A method for measuring the comparability of different sampling methods used in biological surveys: implications for data integration and synthesis Freshwater Biology **50**:1105-1115.

Carlisle DM, Clements WH. 1999. Sensitivity and variability of metrics used in biological assessments of running waters. *Environmental Toxiology Chemistry* **18**:285–291.

Cuppen JGM, Van den Brink PJ, Camps E, Uil KF, Brock TCM. 2000. Impact of the fungicide carbendazim in freshwater microcosms: I. Water quality, breakdown of particulate organic matter and responses of macroinvertebrates. Aquat. Toxicol. **48**:233–250.

David SM, Somers KM, Reid RA, Hall RJ, Girard RE. 1998. Sampling Protocols for the Rapid Bioassessment of Streams and Lakes using Benthic Macroinvertebrates, second edition. Ontario Ministry of the Environment, Dorset, Ontario.

Diamond JM, Barbour MT, Stribling JB. 1996. Characterizing and comparing bioassessment methods and their results: a perspective. J. N. Am. Benthol. Soc. **15:**713-727.

Dumaresq C, Hedley K, Michelutti R. 2002. Overview of the Metal Mining Environmental Effects Monitoring Program. Water Quality Research Journal of Canada **37:**213-218.

Gerth WJ, Herlihy AT. 2006. Effect of sampling different habitat types in regional macroinvertebrate bioassessment surveys. Journal of the North American Benthological Society **25:** 501-512.

Hilsenhoff WL. 1987. An improved biotic index of organic stream pollution: Great Lakes Entomologist **20:**31-39.

Herbst DB, Silldorf EL. 2006. Comparison of the performance of different bioassessment methods: similar evaluations of biotic integrity from separate programs and procedures. Journal of the North American Benthological Society **25:** 513-530.

Iowa DNR (Iowa Department of Natural Resources). 2004. Biological Assessment of Iowa's Wadeable Streams. Iowa Department of Natural Resources, Environmental Services Division. Des Moines, Iowa.

Jones C, Somers KM, Craig B, Reynoldson TB. 2004. Ontario Benthos Biomonitoring Network protocol manual, version 1.0. Ontario Ministry of Environment.

Kerans BL, Karr JR, Ahlstedt SA. 1992. Aquatic invertebrate assemblages: spataial and temporal differences among sampling protocols. J. N. Am. Benthol. Soc. **11**:377-390.

Kilgour BW. 1998. Developing an index of nutrient status based on rapid assessment methodology for collecting benthic macroinvertebrates. Prepared by Water Systems Analysts, Guelph for the Ontario Ministry of Natural Resources, Picton.

Kilgour BW, Barton DR. 1999. Associations between stream fish and benthos across environmental gradients in southern Ontario, Canada. Freshwater Biology **41:** 553-566.

Kilgour BW, Munkittrick KR, Portt CB, Hedley K, Culp J, Dixit S, Pastershank G. 2005. Biological criteria for municipal wastewater effluent monitoring programs. Water Quality Research Journal of Canada **40:**374-387.

Kilgour BW, Somers KM, Barton DR. 2004. A comparison of the sensitivity of stream benthic community indices to effects associated with mines, pulp and paper mills, and urbanization. Environmental Toxicology and Chemistry **23:**212-221.

Kilgour BW, Somers KM, Matthews DE. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. Ecoscience **5**:542–550.

Lemly AD. 1982. Modification of Benthic Insect Communities in Polluted Streams: Combined Effects of Sedimentation and Nutrient Enrichment. Hydrobiologia **87:** 229-245.

Lenat DR, Resh VH. 2001. Taxonomy and stream ecology – The benefits of genus- and species-level identifications. J. N. Amer. Benthol. Soc. **20:** 280-286.

Lenz BN, Miller MA. 1996. United States Geological Survey-National Water-Quality Assessment. Comparison of aquatic macroinvertebrate samples collected using different field methods. Fact Sheet FS-216-96.

Linke S, Bailey RC, Schwindt J. 1999. Temporal variability of stream bioassessments using benthic macroinvertebrates. Freshwater Biology **42:** 575-584.

Mackey AP, Cooling DA, Berrie AD. 1984. An evaluation of sampling strategies for qualitative surveys of macroinvertebrates in rivers, using pond nets. Journal of Applied Ecology **21:** 515-534.

McClelland WT, Brusven MA. 1980. Effects of sedimentation on the behavior and distribution of riffle insects in a laboratory stream. Aquatic Insects **2:**161-169.

Muzaffar SB, Colbo MH. 2002. The effects of sampling technique on the ecological characterization of shallow, benthic macroinvertebrate communities in two Newfoundland ponds. Hydrobiologia **477:** 31 – 39.

O'Connor A, Bradish S, Reed T, Moran J, Regan E, Visser M, Gormally M, Sheehy Skeffington M. 2004. A comparison of the efficacy of pond-net and box sampling methods in turloughs – Irish ephemeral aquatic systems. Hydrobiologia **524:**133–144.

Ohio EPA (Ohio Environmental Protection Agency). 2006. Ohio 2006: Integrated Water Quality Monitoring and Assessment Report. State of Ohio Environmental Protection Agency, Division of Surface Water. Columbus, Ohio.

Piggott A, Brown D, Moin S. 2002. Calculating a groundwater legend for existing geological mapping data. Environment Canada, internal publication, Burlington, Ontario.

Plafkin JL, Barbour KD, Porter KD, Gross SK, Hughes RM. 1989. Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. US EPA/440/4-89/001. 125 pp.

Quinn G, Keough M. 2002. Experimental design and data analysis for biologists. Cambridge University Press. Cambridge, United Kingdom.

Reid RA, Somers KM, David SM. 1995. Spatial and Temporal Variation in littoral-zone Benthic Invertebrates from 3 south-central Ontario lakes. Canadian Journal of Fisheries and Aquatic Sciences **52:**1406-1420.

Resh VH. 1994. Variability, accuracy, and taxonomic costs of rapid assessment approaches in benthic macroinvertebrate biomonitorng. Italian Journal of Zoology **61**: 375-383.

Reynoldson TB, Logan C, Pascoe T, Thompson SP. 2002. CABIN (Canadian Aquatic Biomonitoring Network) Invertebrate Biomonitoring Field and Laboratory Manual National Water Research Institute Environment Canada, Burlington, Ontario.

Rhea DT, Harper DD, Farag AM, Brumbaugh WG. 2006. Biomonitoring in the Boulder River watershed, Montana, USA: Metal concentrations in biofilm and benthic macroinvertebrates, and relations with macroinvertebrate assemblage. Environment Monitoring and Assessment **115:** 381-393.

Rosenberg DM, Davies IJ, Cobb DG, Wiens AP. 1998. Protocols for measuring biodiversity: Benthic Macroinvertebrates in Fresh Waters. Freshwater Institute, Department of Fisheries and Oceans, Winnipeg, Manitoba.

Scarsbrook MR, Halliday J. 2002. Detecting patterns in hyporheic community structure: does sampling method alter the story? New Zealand Journal of Marine and Freshwater Research **36:** 443–453.

Schulz R, Liess M. 1999. A field study of the effects of agriculturally derived insecticide input on stream macroinvertebrate dynamics. Aquatic Toxicology **46:** 155-176.

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.

Sponseller RA, Benfield EF, Valett HM. 2001. Relationships between land use, spatial scale and stream macroinvertebrate communities. Freshwater Biology **46:** 1409-1424.

Stanfield LW (ed.). 2005. Ontario Stream Assessment Protocol, Version 7. Fish and Wildlife Branch, Ontario Ministry of Natural Resources, Peterborough, Ontario.

Stanfield LW, Kilgour BW. 2005. Effects of percent impervious cover on fish and benthos assemblages and in-stream habitats in Lake Ontario Tributaries. Proceedings of Special Symposium: Influences of landscape on stream habitat and biological communities, American Fisheries Society, Madison, Wisconsin.

Stanfield LW, Kuyvenhoven R. 2003. Protocol for applications used in the aquatic landscape inventory software application: A tool for delineating, characterizing and classifying valley segments within the Great Lakes basin. Ontario Ministry of Natural Resources, Peterborough, Ontario.

Waite IR, Herlihy AT, Larsen DP, Urquhart NS, Klemm DJ. 2004. The effects of macroinvertebrate taxonomic resolution in large landscape bioassessments: an example from the Mid-Atlantic Highlands, U.S.A. Freshwater Biology **49:** 474-489.

Walker SL, Hedley K, Porter E. 2002. Pulp and Paper Environmental Effects Monitoring in Canada: An Overview. Water Quality Research Journal of Canada **37:**7-19.

Wang L, Kanehl P. 2003. Influences of Watershed Urbanization and Instream Habitat on Macroinvertebrates in Cold Water Streams. Journal of the American Water Resources Association **39:** 1181-1196.

Winger PV, Lasier PJ, Bogenrieder KJ. 2005. Combined use of rapid bioassessment protocols and sediment quality triad to assess stream quality. Environmental Monitoring and Assessment. **100:** 267-295.

Received: 8 May 2006; accepted: 9 September 2007