Freshwater Biology (2005) **50,** 1105–1115 doi:10.1111/j.1365-2427.2005.01377.x

APPLIED ISSUES

A method for measuring the comparability of different sampling methods used in biological surveys: implications for data integration and synthesis

YONG CAO,* CHARLES P. HAWKINS* & ANDREW W. STOREY†

**Western Center for Monitoring and Assessment of Freshwater Ecosystems, Department of Aquatic, Watershed and Earth Resources, Utah State University, Logan, UT, U.S.A.*

†*Aquatic Research Laboratory, School of Animal Biology, The University of Western Australia, Nedlands, WA, Australia*

SUMMARY

1. Numerous methods have been developed to sample the biota occurring in different ecosystems. However, the comparability of data derived from different sampling methods is generally unknown and is a major concern when integrating data from different studies.
2. Examination of assemblage-level attributes such as taxa richness and biotic index scores is generally inappropriate for evaluating the degree to which different sampling methods produce comparable descriptions of entire assemblages, because these measures provide no information regarding taxonomic composition. Multivariate methods are generally more appropriate for this purpose, but some of the methods previously used are not satisfactory and others have not been tested. A useful measure of sampling-method comparability (SMC) should be independent of sampling effort, independent of the sites sampled and have an explicit biological interpretation.
3. We used simulated data to compare two potential methods of assessing SMC, the *R*-value produced by ANOSIM and a modified version of classification strength (CS-SMC) derived from Van Sickle's Mean Similarity Analysis. Analyses were based on similarities between the assemblages captured by two different sampling methods (electrofishing and seining) employed at the same sites. Similarities were calculated two different ways: the Bray–Curtis index and the Jaccard coefficient.
4. Based on simulated data, ANOSIM *R*-values were strongly affected by sampling effort, highly variable across sites and difficult to interpret biologically. In contrast, CS-SMC values were highly stable over a range of sampling effort, across sites and easy to interpret biologically.
5. Application of CS-SMC to field data showed that seining and electrofishing produced highly comparable samples of fish in small streams: 97% comparable on average for species lists and 94% comparable for relative abundances. Kicknet and Surber samples of benthic invertebrates were also comparable after being standardised to a fixed count, but to a lesser extent than fish samples: 77% comparable on average for the taxa lists and 93% comparable for relative abundances. CS-SMC should be of general use when integrating and synthesising assemblage data from a variety of assemblages.

Keywords: ANOSIM, bioassessment, community similarity, data synthesis, sampling methods, stream assemblages

Correspondence: Yong Cao, Western Center for Monitoring and Assessment of Freshwater Ecosystems, Department of Aquatic, Watershed and Earth Resources, Utah State University, 5210 Old Main Hill, Logan, UT 84322-5210, U.S.A. E-mail: yong.cao@usu.edu

Introduction

Biotic surveys provide basic information to address a variety of questions in ecology and environmental management. Many sampling methods are often available for collecting biological samples, particularly for benthic macroinvertebrates (e.g. Hellawell, 1978; Merritt, Cummins & Resh, 1984) and fish (e.g. Hellawell, 1978; Sutherland, 1996). When conducting assemblage surveys, the choice of sampling method often is based on the particular objective of a project, sampler availability, cost and familiarity with a method (Resh & McElravy, 1993; Carter & Resh, 2001; Bonar & Hubert, 2002). However, when assemblage data collected in conjunction with different studies are compiled and used to test general ecological hypotheses or describe ecological patterns (e.g. Karlson & Cornell, 1998; Van Rensburg *et al*., 2000; Gurevitch, Curtis & Jones, 2001), the comparability of different sampling methods may compromise analyses and inferences (McGeoch & Gaston, 2002). This problem is significant enough that agencies involved in biological monitoring are considering or requiring that users of multi-source data specify the degree to which data collected in different ways are comparable (Intergovernmental Task Force on Monitoring Water Quality, 1995; Carter & Resh, 2001; Houston *et al*., 2002).

Sampling-method comparability (SMC) can be generally defined as how comparable different sampling methods are in characterising biological assemblages. However, assemblages can be characterised in numerous ways. Many studies have compared how estimates of individual variables, such as taxa richness, density, biotic index values and diversity index values vary among sampling methods (e.g. Elliott & Drake, 1981; Stark, 1993; Humphries *et al*., 1998; Muzaffar & Colbo, 2002; Trenkel *et al*., 2004). Unfortunately, this univariate approach is subject to a major limitation. Because individual variables often differ with respect to how comparable they are (Solimini *et al*., 2000; Houston *et al*., 2002), the overall comparability of different sampling methods is uncertain. Answering questions in community ecology (ter Braak, 1987; Legendre & Legendre, 1998) and estimating the overall biological condition of ecosystems (e.g. Hawkins *et al*., 2000; Wright, Sutcliffe & Furse, 2000) often requires that we adequately characterise entire assemblages.

Other studies have used multivariate approaches to evaluate SMC (e.g. Furse *et al*., 1981; De Pauw, Roels & Fontoura, 1986; Somerfield & Clarke, 1996). SMC has been assessed by calculating a similarity index between the samples collected with two methods (e.g. Storey, Edward & Gazey, 1991). However, this method is problematic for at least three reasons. First, replicate samples collected with the same method should be regarded as fully comparable although they will rarely be identical (Schleier & Van Bernem, 1998; Diserud & Aagaard, 2002; Plotkin & Muller-Landau, 2002). SMC based solely on the similarity between samples collected in different ways will therefore be underestimated. Second, the degree of similarity among replicate samples often varies with site because of differences in species richness and species-abundance distributions (Cao, Williams & Larsen, 2002; Plotkin & Muller-Landau, 2002). Third, the similarity among replicates typically depends on sampling effort (Wolda, 1981; Plotkin & Muller-Landau, 2002), defined either as the number of sample units pooled or the number of individuals counted. These three difficulties can all be attributed to incomplete assemblage characterisation (Cao *et al*., 2002).

The SMC has also been evaluated with ordination or cluster analysis by examining if replicate samples collected with different methods group by site rather than by sampling method (Furse *et al*., 1981; De Pauw *et al*., 1986; Storey *et al*., 1991; Somerfield & Clarke, 1996; Turner & Trexler, 1997). However, the grouping of replicate samples will be affected by differences in assemblage structure among sites as well as by the sampling effort applied. In other words, if those sites being compared are biologically distinct, samples are likely to be grouped by site rather than by sampling method. In contrast, if the sites compared are biologically similar, samples could be grouped by sampling method. The choice of the particular multivariate method used and the options selected in each step of multivariate analysis may also yield different results (James & McCulloch, 1990). Several nonparametric multivariate methods are potentially more useful for assessing SMC, including analysis of similarity or ANOSIM (Clarke & Green, 1988), the Mean-Similarity Method (Van Sickle, 1997) and the Multi-Response Permutation Procedures or MRPP (Mielke, Berry & Johnson, 1976; McCune & Grace, 2002). However, to our knowledge, no one has conducted such an analysis. These three methods all measure the

difference between within-group and between-group similarities, but in different ways.

Whatever the method used, a robust measure of SMC should meet three criteria: (i) it should be independent of sampling effort, (ii) it should be stable across sites (i.e. it should not be affected by differences in assemblage structure among sites) and (iii) it should have an explicit biological interpretation. In this paper, we examined how well two multivariate indices performed as measures of SMC with respect to these three criteria. One index was based on a version of classification strength (CS) described by Van Sickle (1997). The other index was the *R*-value produced by ANOSIM (Clarke & Green, 1988). We applied these measures to samples of fish collected by electrofishing and seining and samples of macroinvertebrates collected with kicknet and Surber samplers. Our specific objectives were to determine: (i) how the values of the two SMC measures varied across sampling sites and with different sampling effort, (ii) if estimates of SMC differed when measured with different similarity indices and (iii) how comparable the two pairs of commonly used sampling methods were in characterising taxonomic composition and assemblage structure.

Methods

Two similarity-based, nonparametric approaches for measuring SMC

Clarke & Green (1988) described a randomisation test on a similarity matrix, ANOSIM, to test for differences in assemblage structure among groups of samples. The difference between two groups is measured by *R*, which is calculated as

$$R = \frac{4(\bar{r}_B - \bar{r}_W)}{n(n-1)},$$

where $\bar{r}_B$ = mean ranked between-group similarity, $\bar{r}_W$ = mean ranked within-group similarity and n = the total number of samples. *R*-values range between −1 and 1. A positive value indicates differences exist among groups and 0 means complete random grouping. Negative values occur when the samples within a group are less similar to one another than to the samples of other groups, which may be caused by inappropriate sampling designs (Chapman & Underwood, 1999). The ANOSIM function in the Vegan-library of *R* statistical routines (Dixon, 2003; Oksanen, 2004) calculates ANOSIM *R* and reports statistical significance levels based on 1000 runs.

Van Sickle (1997) proposed a conceptually similar, but different way of measuring the difference among groups, which he called classification strength (CS). CS is calculated as

$$CS = \frac{2\bar{S}_b}{\bar{S}_{w1} + \bar{S}_{w2}}$$

for comparison of two groups, which was the focus of the present study, where $\bar{S}_b$ = mean between-group similarity, $\bar{S}_{w1}$ = the mean within-group similarity for group 1 and $\bar{S}_{w2}$ = the mean within-group similarity for group 2. We used the FORTRAN software provided by Van Sickle (1998) to calculate CS. Mielke *et al.* (1976) described a similar method, called MRPP. Because the statistic describing differences between groups in MRPP, A, is expected to vary in a similar way as CS (Van Sickle, 1998), we did not include MRPP in this comparison. For convenience, we use 100% × CS as a measure of SMC, hereafter referred to as CS-SMC. When samples collected with two methods share no species (i.e. $\bar{S}_b = 0$), but replicate samples collected with the same method share at least one species (i.e. $\bar{S}_{w1} > 0$ and $\bar{S}_{w2} > 0$), then CS-SMC = 0%. If the average within-method similarity ($\bar{S}_{w12}$) is equal to $\bar{S}_b$, CS-SMC = 100%.

Two commonly used similarity indices

The values of both within-group and between-group similarities will be dependent on which similarity measure is used (e.g. Storey *et al.*, 1991; Cao, Williams & Bark, 1997), and so will ANOSIM *R* and CS-SMC values. Among the numerous similarity measures available (Legendre & Legendre, 1998), the Jaccard coefficient and the Bray–Curtis index appear to be the most frequently used binary and abundance-based indices, respectively (Krebs, 1989; Legendre & Legendre, 1998) and were thus selected for quantifying SMC in our study. We followed Krebs (1989) when calculating the two indices. Data were transformed by $\log(x + 1)$ to down-weight the effect of abundant taxa on Bray–Curtis values (Palmer, 1993).

Data sets

We used data from two field surveys to evaluate how ANOSIM *R* and CS-SMC values quantified

sample comparability. In both field surveys, stream fauna were collected with two different sampling methods.

Wyoming stream fish. Patton *et al.* (2000) surveyed nine stream sites (2.5–10.2 m wide) in the Missouri River drainage of eastern Wyoming, U.S.A. They delineated 16 contiguous 50-m long subreaches at each site. Two sampling methods, electrofishing and seining, were used to sample alternate subreaches. The number of individuals in each species was recorded for each subreach (Table 1).

Australian stream macroinvertebrates. Storey *et al.* (1991) collected six Surber samples (0.0625 m^2, 250 μm mesh) and one 3-min kicknet sample (approximately 10 m transect, 250 μm mesh) from riffle habitats at multiple sites in 1987. Most taxa were identified to genus or species level. A subset of 15 samples by each method that contained at least 1000 individuals was drawn from the data set (Table 1) to examine how SMC varied over a wide range of fixed counts and across sampling sites.

Table 1 Summary of two assemblage surveys: fish at nine stream sites, Wyoming, U.S.A. (Patton *et al.*, 2000) and macroinvertebrates on 15 sampling occasions in Australia streams (Storey *et al.*, 1991)

Sampling sites or occasions	Sampling methods			
	Electrofishing		Seining	
Wyoming fish	Total individuals	Richness	Total individuals	Richness
1	2189	7	3304	7
2	5352	11	2052	11
3	1070	13	651	13
4	450	9	1249	9
5	1188	14	3206	14
6	556	7	413	7
7	1959	8	935	8
8	790	9	425	9
9	278	10	251	10
Australian macroinvertebrates	Surber samplers		3-min Kick-net	
1	3597	73	3656	49
2	11 212	40	2933	38
3	2361	36	1056	16
4	4253	58	2063	45
5	3815	36	1359	30
6	3591	44	1402	23
7	6353	31	1596	27
8	5510	42	3610	22
9	18 280	26	5132	21
10	20 357	32	8120	21
11	13 299	41	2740	23
12	8215	32	2215	14
13	30 738	32	2306	21
14	7881	38	3412	25
15	18 179	32	8149	23

Simulated data. The two field data sets were based on much greater sampling efforts than those used in routine stream assemblage surveys. However, both ANOSIM and the Mean-Similarity Method require that samples of each group being compared be independent of one another, and these two data sets were not large enough to examine the effect of sampling effort with much rigor. For example, electrofishing samples from eight subreaches can be combined into only two distinct samples with a sampling effort of four times the size of the original samples. We therefore used simulation to overcome this difficulty (Warton & Hudson, 2004). The Wyoming fish data set is unique in that species-sample curves reached an asymptote for both sampling methods at all sites (Patton *et al.*, 2000), i.e. collecting more replicates would not add new species, and is therefore ideal for simulation. We used this data set to first derive models that best described the distribution of each species among the original samples and then generate several, large sets of simulated replicates of different size.

To describe the spatial distribution of fish species within the stream, we used generalised linear models to fit the distributions of abundances for each fish species across the eight subreaches at each site that were sampled with each method. The spatial variation in abundances of these species were best described by either a negative binomial distribution (usually abundant species) or a Poisson distribution (usually rare species), patterns that agree with many previous observations (e.g. Green & Young 1993; Plotkin & Muller-Landau, 2002; McArdle & Anderson 2004). We then used the derived model parameters (mean and an over-dispersion measure) to simulate 80 random replicates for each species at each of the nine sites and for each sampling method. To check the realism of the simulation, we compared how the simulated replicates clustered in non-metric multidimensional scaling (NMDS) ordination space (PC-ORD; McCune & Grace, 2002) relative to the actual field replicates. In all cases, the eight field replicates appeared to be randomly distributed among the 80 simulated repli-

cates in ordination space, confirming the realism of our simulated samples.

Data analysis

Data analysis consisted of examination of simulated data to evaluate the variation of ANOSIM *R* and CS-SMC values across sampling sites and at different sampling efforts followed by application of CS-SMC to evaluate the comparability of the two different fish (eletrofishing and seining) and benthos (Surber and kicknet) sampling methods.

To determine if ANOSIM *R* and CS-SMC were robust measures of sample comparability, we incrementally pooled sets of the 80 replicates (subreaches) to vary sampling effort. For example, each of the original 80 replicates represented a sampling effort of one; pooling sets of two replicates created 40 samples having a sampling effort of two, and so on. We then calculated ANOSIM *R* and CS-SMC values for each sampling effort at each site.

Because results of the simulation analysis showed ANOSIM *R* to be sensitive to sampling effort and differences in assemblage structure between sites, we only used CS-SMC to quantify comparability of the fish and benthos sampling methods. Because the number of replicate samples per site for both the fish and invertebrate data was small, we used a random re-sampling procedure to generate more accurate estimates of CS-SMC values than would have been otherwise possible. Also, although the original spatially defined sampling units were used in the fish sampling analyses, we used fixed-counts as the measure of sampling effort for the invertebrate samples, a procedure typically used in bioassessment surveys (e.g. Carter & Resh, 2001). Such computer-based subsamples were very similar in both richness and taxonomic composition to those obtained by manual subsampling (Cao, Hawkins & Vinson, 2003).

To conduct the CS-SMC analyses, we randomly drew m replicates or individuals for each sampling method (where m ranged between 1 and $N/2$) from a set of N replicate samples or individuals and pooled them into a single sample. Random, individual-based sampling is commonly used to standardise sampling effort in studies of stream macroinvertebrate ecology and for bioassessment purposes (Carter & Resh, 2001). Moreover, fixed-count subsampling is simply a manual form of rarefaction, a procedure that has long been used to standardise comparisons of species richness (Heck, Van Belle & Simberloff, 1975; Gotelli & Graves, 1996; Gotelli & Colwell, 2001). We then calculated the similarity (S_b) between these two samples. This process was repeated 1000 times to estimate the average similarity ($\bar{S}_b$) between samples collected by two methods. We then randomly drew an even number (n) of replicates or individuals ($2 \le n \le N$) from the total set of N replicates or individuals available for each sampling method, evenly divided them into two samples and calculated the within-method similarity for the pair of samples ($\bar{S}_{w1}$ or $\bar{S}_{w2}$). This process was also repeated 1000 times to obtain an average within-method similarity for each method ($\bar{S}_{w1}$ or $\bar{S}_{w2}$) at each site. CS was then calculated as described earlier. However, we do not report significance levels for these analyses because the 1000 pairs of samples generated by re-sampling were not truly independent from one another. Our goal in this study was to estimate CS-SMC as accurately as possible, not test for significant differences between groups of samples. Likewise, because the statistical distributions of CS-SMC values are not known, we did not use standard deviations or 95% confidence limits to quantify variation in these two variables. Instead, we plotted the values of CS-SMC for each site against sampling effort to show their variation across sites and with sampling effort. When plotting data, we slightly jittered completely overlapping data points horizontally to avoid obscuring trends.

Results

Simulated data and ANOSIM *R and CS-SMC values*

When applied to simulated data, ANOSIM *R* and CS-SMC differed markedly in their sensitivity to sampling effort and sites used (Figs 1 & 2). ANOSIM *R*-values based on Bray–Curtis index increased substantially with increasing sampling effort at all nine sites (Fig. 1a) and were always statistically significant ($P < 0.01$). On average, ANOSIM *R*-values increased by 0.6 units when sampling effort increased from one to eight pooled subreaches. This result implied that the comparability between samples collected by electrofishing and seining decreased with higher sampling effort, although $\bar{S}_b$ actually increased with increasing sampling effort (Fig. 1c). In contrast, CS-SMC values based on the Bray–Curtis index varied little with sampling effort

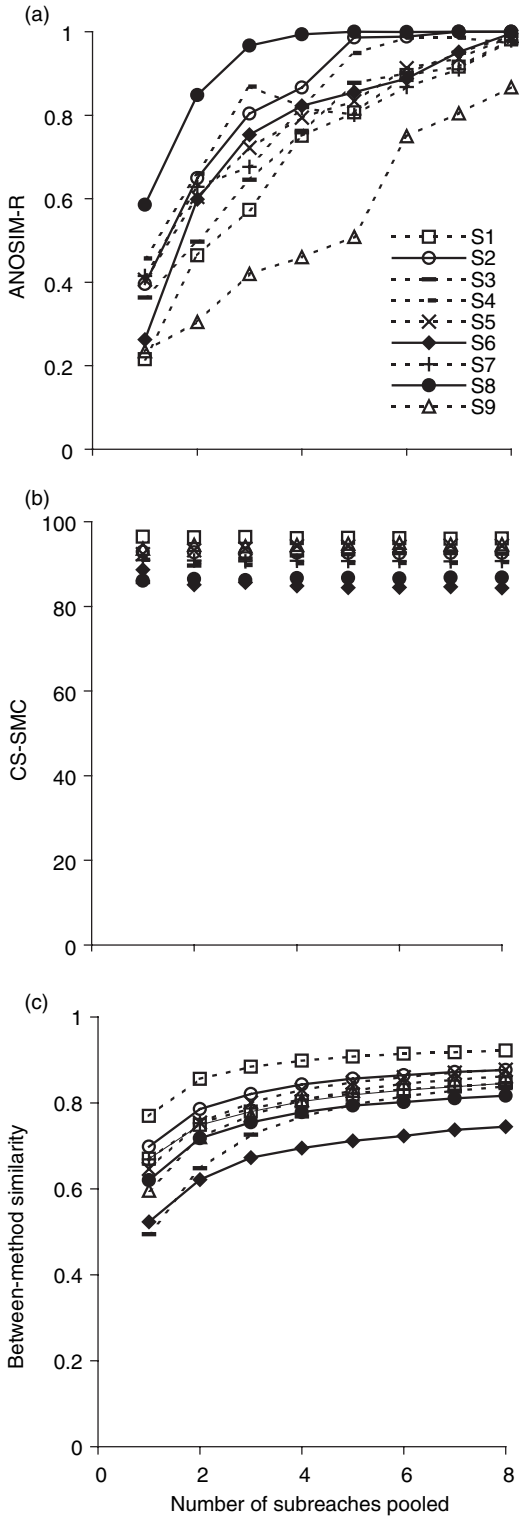

Fig. 1 Changes in values of ANOSIM *R* (a), CS-SMC (b) and between-method similarity (c) measured with the Bray–Curtis index with increasing sampling effort (i.e. the number of subreaches pooled) and across nine stream sites based on simulated fish samples.

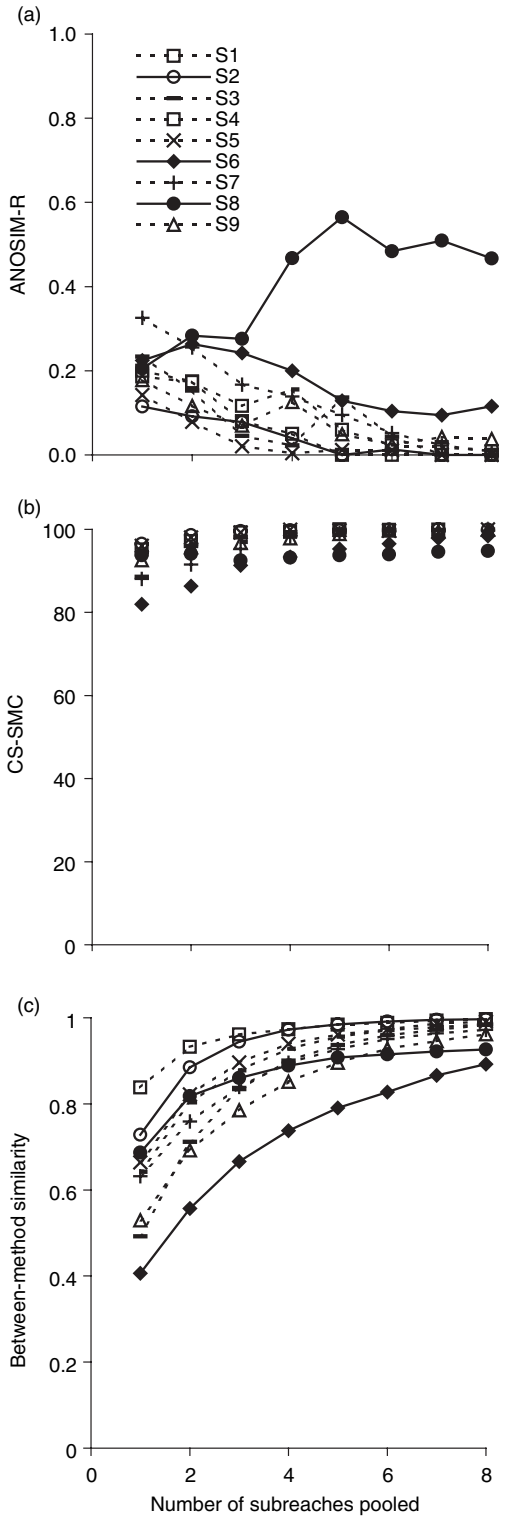

Fig. 2 Changes in ANOSIM *R* (a), CS-SMC (b) and the between-method similarity (c) measured with the Jaccard coefficient with increasing sampling effort (i.e. the number of subreaches pooled) and across nine stream sites based on simulated fish samples.

(Fig. 1b), although they were always significant ($P < 0.05$). Both measures of SMC were affected by differences in assemblage structure between sites, but the differences among sites in CS-SMC values (up to 0.11) were much smaller than for ANOSIM *R*-values (up to 0.55).

When similarities were measured with the Jaccard Coefficient, ANOSIM *R*-values usually decreased with increasing sampling effort (Fig. 2a), indicating that the differences between electrofishing and seining samples decreased with increasing sampling effort. These ANOSIM *R*-values were usually significant ($P < 0.01$) at low sampling effort, but not at high sampling effort ($P > 0.05$). With the exception of one stream, the trend for decreasing *R*-values with increasing sampling effort generally paralleled the increase in between-method Jaccard coefficient values that occurred with increasing sampling effort (Fig. 2c). At the highest sampling effort (eight pooled subreaches), ANOSIM *R*-values were zero at several sites because all species were captured at these sites with both methods. CS-SMC values based on the Jaccard coefficient were much less affected by sampling effort (Fig. 2b), although significance levels associated with the CS-SMC values were affected similarly as in ANOSIM *R*. As with Bray–Curtis based measures of SMC, both ANOSIM *R* and CS-SMC based on the Jaccard coefficient were also affected by site, but the differences among sites in CS-SMC values (up to 0.15) were again much smaller than for ANOSIM *R*-values (up to 0.56).

Applying CS-SMC to field data

The CS-SMC based on the field samples showed that electrofishing and seining were highly comparable at all sampling efforts and at all sites (Fig. 3a,b). CS-SMC values ranged between 70 and 100% (mean = 97%) when based on the Jaccard coefficient (Fig. 3a), indicating the two sampling methods almost always captured the same set of species. CS-SMC varied from 83 to 99% (mean = 91%) when based on the Bray–Curtis index (Fig. 3b), indicating the two sampling methods were also highly comparable in characterising the relative abundances of species within the assemblage.

Fixed-count subsamples of benthic macroinvertebrates drawn from kicknet and Surber samples were somewhat less comparable with one another than the fish sampling methods, especially when based on the Jaccard coefficient (Fig. 3c,d). CS-SMC based on the Jaccard coefficient ranged between 61 and 99% across all 15 sites and five sampling efforts (100–500 counts), with a mean of 77% (Fig. 3c). However, CS-SMC based on the Bray–Curtis index ranged between 85 and 100% across all sites and sampling efforts (mean = 93%) (Fig. 3d). CS-SMC values were slightly >100% in a few cases, which is equivalent to small negative ANOSIM *R*-values.

Discussion

Sharing data on assemblage-level surveys among different environmental agencies, monitoring programmes and researchers promotes data syntheses and facilitates the testing of ecological hypotheses through meta-analysis or related techniques. However, the variability in use of different sampling methods, among other inconsistencies (Resh & McElravy, 1993; Carter & Resh, 2001; Cao *et al*., 2003), presents a major challenge to such efforts. CS-SMC provides a means for evaluating the compatibility of different data sets prior to analysis in terms of four desirable properties: (i) it is based on the whole assemblage, (ii) it provides much better control over sampling effort and site effects than does either simple measures of between-method similarity or ANOSIM *R*-values, (iii) it has an explicit biological interpretation and (iv) it appears less sensitive to the similarity index used than were ANOSIM *R*-values.

In contrast to CS-SMC, ANOSIM *R*-values do not meet the three criteria that characterise a useful measure of SMC. It was strongly dependent on both sampling effort (also see Growns *et al*., 1997) and differences in assemblage structure between sites. It can also be potentially misleading. For example, when the similarity between samples collected with two methods increased at higher sampling efforts, ANOSIM *R*-values implied that the two sets of samples became increasingly dissimilar, which contradicted the increase in between-method similarity ($\bar{S}_b$). This behaviour likely occurred because (i) when similarity values are ranked in ANOSIM, the magnitude of the real difference between samples is lost and (ii) as within-group similarities increased with higher sampling effort, slight, but consistently lower between-group similarities could lead to high *R*-values and higher significance levels. As is true of statistical tests in general (Johnson, 1999), biological significance

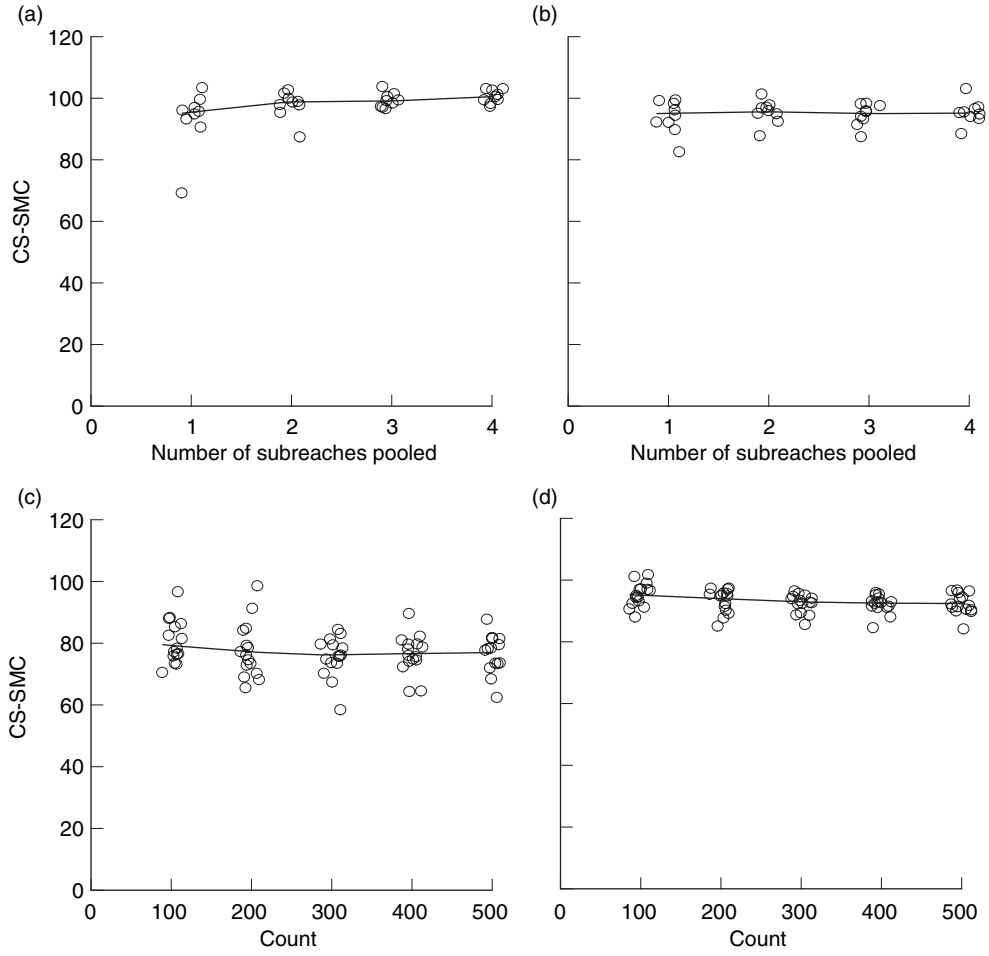

Fig. 3 Changes in the comparability (CS-SMC) of electrofishing and seining samples measured with the Jaccard coefficient (a) and Bray–Curtis index (b) with increasing sampling effort (one to eight subreaches pooled) and across nine stream sites in Wyoming, and changes in the comparability of kicknet and Surber samples after being standardised to a fixed count, measured with the Jaccard coefficient (c) and Bray–Curtis index (d) with increasing sampling effort (100–500 fixed counts) and across 15 stream sites in Australia.

should not be inferred solely from a significant statistical test. For these reasons, we do not recommend use of ANOSIM *R*-values as a measure of SMC.

Use of CS-SMC implied that samples of both fish and macroinvertebrates collected by different methods can be highly comparable if species composition is the primary variable of interest. Even for comparisons based on relative abundance, the >90% agreement may be sufficient to allow combining of data without serious confounding of results. Even for the invertebrate survey data, in which raw sample data may not be comparable because of gross differences in counts, comparable data were derived by standardising the samples to a common fixed-count of individuals. The few cases in which CS-SMC was slightly >100% was likely a consequence of random sampling error and should not compromise use of CS-SMC.

We believe that CS-SMC should be applicable to many other taxonomic groups and ecosystems. In practice, the evaluation of SMC should include enough replicate samples, e.g. >10 taken at enough sites (e.g. >10) to ensure that results are both accurate and can be generalised to other data sets. The sampling sites chosen should represent a range of habitat characteristics and biodiversity. When integrating existing assemblage data from multiple surveys, criteria should be set regarding the minimum level of comparability that must be met. However, the exact level of CS-SMC to be used will likely vary depending on the specific objectives of a study. For

example, a relatively low CS-SMC might be appropriate for describing trends in species richness at large spatial scales, but a higher CS-SMC may be needed when developing and applying predictive models used to assess the biological condition of individual sites, e.g. RIVPACS (Wright *et al*., 2000). We emphasise that CS-SMC only measures the comparability of sampling methods. The extent to which a given sampling method characterises the biological assemblage of interest (i.e. sample representativeness) is a different question and must be evaluated separately with different methods. The mean similarity across multiple pairs of replicates is one way of measuring sample representativeness (e.g. Cao *et al*., 2002). Finally, SMC is a major, but not the only concern in data synthesis. Other concerns include differences in taxonomic resolution used across data sets, sampling frequency and the survey design used.

Data sharing can be of great value to programmes that have limited sampling budgets. Although it is unlikely that different agencies, states or nations will adopt standard sampling methods as a means of increasing data sharing, many of the benefits of data sharing might be realised if we knew which data sets can be appropriately combined and which ones can not. We believe use of the CS-SMC measure described here will allow more robust and meaningful assessment of the comparability of assemblage-level data than previously possible.

Acknowledgments

We thank T. Patton for allowing us to use his survey data and R. Hughes, M. Vinson and three anonymous reviewers for constructive comments on earlier versions of the manuscript. Recommendations by L. Yuan improved the modelling of fish distributions across subreaches, and J. Oksanen provided R-language programs to calculate ANOSIM *R*-values. This research was funded by the U.S. EPA – Science to Achieve Results (STAR) Program Grant no. R-82863701.

References

Bonar S.A. & Hubert W.A. (2002) Standard sampling of inland fish: benefits, challenges, and a call for action. *Fisheries*, **27**, 10–16.

ter Braak C.J.F. (1987) Ordination. In: *Data analysis in community and landscape ecology* (Eds R.H. Jongman, C.J.F. ter Braak & O.F.R. van Tongeren), pp. 91–169. Pudoc, Wageningen, the Netherlands.

Cao Y., Williams W.P. & Bark A.W. (1997) Similarity measure bias in river benthic Aufwuchs community analysis. *Water Environment Research*, **69**, 95–106.

Cao Y., Williams D.D. & Larsen D.P. (2002) Comparison of ecological communities - the problem of sample representativeness. *Ecological Monographs*, **72**, 41–56.

Cao Y., Hawkins C.P. & Vinson M. (2003) Measuring and controlling data quality in assemblage surveys with a special reference to stream macroinvertebrates. *Freshwater Biology*, **48**, 1898–1911.

Carter J.L. & Resh V.H. (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.

Chapman M.G. & Underwood A.J. (1999) Ecological patterns in multivariate assemblages: information and interpretation of negative values in Anosim tests. *Marine Ecology Progress Series*, **180**, 257–265.

Clarke K.R. & Green R.H. (1988) Statistical design and analysis for a 'biological effects' study. *Marine Ecology Progress Series*, **46**, 213–226.

De Pauw N., Roels D. & Fontoura A.P. (1986) Use of artificial substrates for standardized sampling of macroinvertebrates in the assessment of water quality by the Belgian Biotic Index. *Hydrobiologia*, **133**, 237–285.

Diserud O.H. & Aagaard K. (2002) Testing for changes in community structure based on repeated sampling. *Ecology*, **83**, 2271–2277.

Dixon P. (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Elliott J.M. & Drake C.M. (1981) A comparative study of seven grabs used for sampling benthic macroinvertebrates in rivers. *Freshwater Biology*, **11**, 99–120.

Furse M.T., Wright J.F., Armitage P.D. & Moss D. (1981) An appraisal of pond-net samples for biological monitoring of lotic macroinvertebrates. *Water Research*, **15**, 679–689.

Gotelli N.J. & Colwell R.K. (2001) Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, **4**, 379–391.

Gotelli N.J. & Graves G.R. (1996) *Null Models in Ecology*. Smithsonian Institution Press, London.

Green R.H. & Young R.C. (1993) Sampling to detect rare species. *Ecological Applications*, **3**, 351–356.

Growns J.E., Chessman B.C., Jackson J.E. & Ross D.G. (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.

Gurevitch J., Curtis P.S. & Jones M.H. (2001) Meta-analysis in ecology. *Advances in Ecological Research*, **32**, 199–247.

Hawkins C.P., Norris R.H., Hogue J.N. & Feminella J.W. (2000) Development and evaluation of predictive models for measuring the biological integrity of streams. *Ecological Applications*, **10**, 1456–1477.

Heck K.L. Jr, Van Belle G. & Simberloff D. (1975) Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology*, **56**, 1459–1461.

Hellawell J.M. (1978) *Biological Surveillance of Rivers. A Biological Monitoring Handbook*. Water Research Centre, Medmenham and Stevenage, England.

Houston L., Barbour M.T., Lenat D. & Penrose D. (2002) A multi-agency comparison of aquatic macroinvertebrate-based stream bioassessment methodologies. *Ecological Indicators*, **1**, 279–292.

Humphries P., Growns J.E., Serafini L.G., Hawking J.H., Chick A.J. & Lake P.S. (1998) Macroinvertebrate sampling methods for lowland Australian rivers. *Hydrobiologia*, **364**, 209–218.

Intergovernmental Task Force on Monitoring Water Quality (1995) *The Strategy for Improving Water-Quality Monitoring in the United States – Final Report of the Intergovernmental Task Force on Monitoring Water Quality*. Office of Water Data Coordination, USGS, Reston, Virginia, Open-File Report 95–742.

James F.C. & McCulloch C.E. (1990) Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annual Review of Ecology and Systematics*, **21**, 129–166.

Johnson D.H. (1999) The insignificance of statistical significant testing. *Journal of Wildlife Management*, **63**, 763–772.

Karlson R.H. & Cornell H.V. (1998) Scale-dependent variation vs. regional effects on coral species richness. *Ecological Monographs*, **68**, 259–274.

Krebs C. (1989) *Ecological Methodology*. Harper and Row Publishers, New York, U.S.A.

Legendre P. & Legendre L. (1998) *Numerical Ecology*, 2nd edn. Elsevier, New York, U.S.A.

McArdle B.H. & Anderson M.J. (2004) Variance heterogeneity, transformation, and models of species abundance: a cautionary tale. *Canadian Journal of Fisheries and Aquatic Science*, **61**, 1294–1302.

McCune B. & Grace J.B. (2002) *Analysis of Ecological Communities*. MjM Software, Gleneden Beach, Oregon, U.S.A.

McGeoch M.A. & Gaston K.J. (2002) Occupancy frequency distributions: patterns, artifact, and mechanisms. *Biological Review of Cambridge Philosophical Society*, **77**, 311–331.

Merritt R.W., Cummins K.W. & Resh V.H. (1984) Collecting, Sampling, and Rearing Methods for Aquatic Insects. In: *An introduction to the Aquatic insects of North America* (Eds R.W. Merritt & K.W. Cummins), 2nd edn, pp. 11–26. Kendall/Hunt Publishing, Dubuque, IA.

Mielke P.W., Berry K.J. & Johnson E.S. (1976) Multi-response permutation procedures for a priori classifications. *Communications in Statistics – Theory and Methods A*, **5**, 1409–1424.

Muzaffar S.B. & Colbo M.H. (2002) The effect of sampling techniques on the ecological characterization of shallow, benthic macroinvertebrate communities in two Newfoundland ponds. *Hydrobiologia*, **477**, 31–39.

Oksanen J. (2004) *Vegan: R functions for Vegetation Ecologists*. Department of Biology, University of Oulu, PO Box 3000, 90014 Oulu, Finland (http://cc.oulu.fi/~jarioksa/softhelp/vegan.html) [accessed on 11 April 2005].

Palmer M.W. (1993) Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology*, **74**, 2215–2230.

Patton T.M., Hubert M.A., Rahel F.J. & Gerow K.G. (2000) Effort needed to evaluate species richness in small streams on the Great Plains in Wyoming. *North American Journal of Fisheries and Management*, **20**, 394–398.

Plotkin J.B. & Muller-Landau H.C. (2002) Sampling the species composition of a landscape. *Ecology*, **83**, 3344–3356.

Resh V.H. & McElravy E.P. (1993) Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. In: *Freshwater Biomonitoring and Benthic Macroinvertebrates* (Eds D.M. Rosenberg & V.H. Resh), pp. 159–194. Chapman and Hall, London, U.K.

Schleier U. & Van Bernem K.H. (1998) Statistical methods to determine sample size for the estimation of species richness and species' abundances in benthic communities. *Archive of Fishery and Marine Research*, **46**, 205–223.

Solimini A.G., Gulia P., Monfrinotti M. & Carchini G. (2000) Performance of different biotic indices and sampling methods in assessing water quality in the lowland stretch of the Biber River. *Hydrobiologia*, **422/423**, 197–208.

Somerfield P.J. & Clarke K.R. (1996) A comparison of some methods commonly used for the collection of sublittoral sediments and their associated fauna. *Marine Environmental Research*, **43**, 145–156.

Stark J.D. (1993) Performance of the macroinvertebrate community Index: effects of sampling methods, sample replicates, water depth, current velocity, and substratum on index values. *New Zealand Journal of Marine and Freshwater Research*, **27**, 463–478.

Storey A.W., Edward D.H.D. & Gazey P. (1991) Surber and kick sampling: a comparison for the assessment of macroinvertebrate community structure in streams of south-western Australia. *Hydrobiologia*, **211**, 111–121.

Sutherland W.J. (1996) *Ecological Census Techniques*. Cambridge University Press, Cambridge.

Trenkel V.M., Pinnegar J.K., Rochet M.-J. & Rackham B.D. (2004) Different surveys provide similar pictures of trends in a marine fish community but not of individual fish populations. *ICES Journal of Marine Science*, **61**, 351–362.

Turner A.M. & Trexler J.C. (1997) Sampling aquatic invertebrates from marshes: evaluating the options. *Journal of the North American Benthological Society*, **16**, 694–709.

Van Rensburg B.J., McGeoch M.A., Mathew W., Chown S.L. & Van Jaarsveld A.S. (2000) Testing generalities in the shape of patch occupancy frequency distributions. *Ecology*, **81**, 3163–3177.

Van Sickle J. (1997) Using mean similarity dendrograms to evaluate classification. *Journal of Agricultural, Biological, and Environmental Statistics*, **2**, 370–388.

Van Sickle J. (1998) *Documentation for MEANSIM, Version 6.0*. Western Ecology Division, USEPA-NHEERL, Corvallis, Oregon (http://www.epa.gov/wed/pages/models/dendro/meansim6.htm) [accessed on 11 April 2005].

Warton D.I. & Hudson H.M. (2004) A manova statistic is just as powerful as distance-based statistics, for multivariate abundances. *Ecology*, **85**, 858–874.

Wolda H. (1981) Similarity indices, sample size and diversity. *Oecologia*, **50**, 296–302.

Wright J.F., Sutcliffe D.W. & Furse M.T. (2000) *Assessing the Biological Quality of Fresh Waters: RIVPACS and other Techniques*. Freshwater Biological Association, Ambleside, Cumbria, U.K.

(Manuscript accepted 11 March 2005)