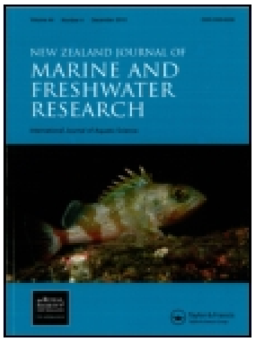

Taylor & Francis
Taylor & Francis Group

New Zealand Journal of Marine and Freshwater Research

ISSN: 0028-8330 (Print) 1175-8805 (Online) Journal homepage: https://www.tandfonline.com/loi/tnzm20

Average score per metric: An alternative metric aggregation method for assessing wadeable stream health

Kevin J. Collier

New Zealand Journal of Marine and Freshwater Research, 2008, Vol. 42: 367–378
0028–8330/08/4204–0367

Average score per metric: an alternative metric aggregation method for assessing wadeable stream health

KEVIN J. COLLIER
Environment Waikato
P. O. Box 4010
Hamilton East
Hamilton, New Zealand
email: kevin.collier@ew.govt.nz

Centre for Biodiversity and Ecology Research
The University of Waikato
P. O. Box 3105
Hamilton, New Zealand

Abstract Integrating multiple metrics derived from stream macroinvertebrate communities into single scores that reflect ecological condition can bridge the needs of multiple groups using biomonitoring data. Macroinvertebrate metrics from 511 Waikato, New Zealand, stream samples were standardised by their maximum observed value. Metric redundancy and optimisation processes identified three standardised metrics from an initial set of 17 that provided greatest discrimination between reference sites and those influenced by urbanisation or high levels of pastoral development: the richness of Ephemeroptera, Plecoptera and Trichoptera excluding Hydroptilidae (EPT*), % EPT* and the macroinvertebrate community index (MCI). The mean of these standardised metrics was used to calculate the average score per metric (ASPM). Narrative condition bands were developed based on the lower SD of individual metrics for reference samples to distinguish "very high" values, and metric quartiles between this point and a hypothetical worst-case community were used to define "high", "moderate", "low", and "very low" bands. When compared with its component metrics, the ASPM distinguished reference conditions and low–moderate levels of catchment modification and local habitat degradation more accurately than EPT* richness or MCI, and displayed lower temporal variability within reference sites than either EPT* metric. The ASPM was calculated for an independent test data set of urban and reference sites which, respectively, were allocated to low–very low and high–very high narrative condition bands. This analysis suggests that prudent application of the ASPM can provide a parsimonious and effective screening tool for assessing the condition of wadeable streams where more complex methods are not practical.

Keywords macroinvertebrate; bioassessment; biomonitoring; Waikato; New Zealand

INTRODUCTION

Biomonitoring information needs to be communicated in a simple yet effective manner that can bridge the needs of a variety of users ranging from policy makers to community groups. There have been many attempts to condense biological data from streams into single numbers intended to accurately represent ecological condition, and debate continues over the best approach (e.g., Suter 1993; Gerritsen 1995; Norris 1995; Reynoldson et al. 1997; Karr 1999; Norris & Hawkins 2000). Criticism has commonly been directed at the use of single metrics because some, like functional feeding group metrics, may vary with developmental stage (although see Weigel 2003), and others may lead to loss of information or suffer from incomplete knowledge of statistical dispersion properties and ecological relationships (Norris 1995; Reynoldson et al. 1997). Integration of several metrics into multimetric indices intuitively provides a more robust assessment of ecological condition than individual metrics by increasing the probability of capturing a wider range of response trajectories to different environmental stressors (Karr 1999). However, indices using additive metrics have also been criticised because of the potential loss of ecologically meaningful data, and the finding that

M08034; Online publication date 31 October 2008
Received 25 June 2008; accepted 8 September 2008

some metrics may be redundant and compound errors (Gerritsen 1995; Reynoldson et al. 1997; Norris & Hawkins 2000).

In New Zealand, most biological assessment is currently based on analysis of single metrics and few studies have aggregated metrics into indices that reflect a range of response variables (although see Quinn et al. 2004), as has been done for fish (Joy & Death 2004). The New Zealand regional authority Environment Waikato is responsible for ensuring that development of resources is conducted in a sustainable way over a large (25 000 km^2) and ecologically diverse area of the central North Island. As part of this responsibility, surveys of habitat and invertebrate communities are conducted at a network of over 100 sites annually to document the state of the region's stream environments (Collier 2005). The specific sites included in this monitoring network can vary from year to year, although a core group of sites has been retained in multiple years. This survey information needs to be communicated to a range of audiences, but the dilemma arises of how to communicate this information with the current approach of using single metrics that may not provide consistent interpretations where they respond differently to particular stressors.

To assist with bridging the needs of contrasting user groups and facilitating the translation of biological data into an easily communicable summary of stream condition, I compared the use of a simple metric aggregation approach with the individual component metrics drawn from a subset of *a priori*-defined structural, compositional and tolerance metrics for stream macroinvertebrate communities. This approach involved the averaging of standardised key metrics to derive an average score per metric (ASPM) for each sample. The specific aim of this study was to assess the performance of the ASPM compared with its component metrics. The overall goal was to develop a parsimonious ecological condition score for Waikato wadeable streams that can be communicated effectively at multiple levels to a wide range of end-user audiences.

MATERIALS AND METHODS

Study area and sampling sites

The Waikato region covers a diverse area of New Zealand's central North Island between latitudes 36°S and 39°S (Fig. 1). Mean annual air temperatures in most of this region are in the range 12.5 to 15.0°C, but decline to <8.0°C on southern mountaintops (2797 m a.s.l.) (Kilpatrick 1999). Average annual rainfall is variable, being lower in the north (1000–1500 mm yr^{-1}) than in the east (Coromandel Peninsula) or west (up to 2500 mm yr^{-1}), and exceeding 5000 mm yr^{-1} on southern mountaintops where some falls as snow (ew.govt.nz; metservice.com). Landforms range from active volcanoes and upland plateaus (c. 600 m a.s.l.) in the south of the region, to steep hill-country (300–600 m a.s.l.) along the west coast, through the central parts of the region and along Coromandel Peninsula, and extensive lowland wetlands and plains towards the north.

The region is geologically diverse with extensive areas of volcanic rock (rhyolites, andesites, basalts, and dacites) around the southern and western volcanoes and along Coromandel Peninsula (McCraw 1971). Tertiary sedimentary rocks comprising limestone, siltstone (mudstone), and sandstone occur predominantly along the western side of the region, along with areas of much older Mesozoic sedimentary rocks (greywackes and argillites). Extensive flows of ignimbrite occur in the central and eastern parts of the region.

Pre-European vegetation cover was mainly podocarp-hardwood forest in hill-country and western areas, with extensive areas of beech in the south (Clarkson et al. 2002; ew.govt.nz). Fernland/scrubland also occurred over wide areas before European colonisation, and sub-alpine grassland and scrubland still occur at higher altitudes on the southern mountains. Most of the Waikato region has now been developed for pastoral agriculture with some pine forestry, although extensive remnants of original vegetation persist in upland parts of the southern and central region, and along Coromandel Peninsula (Clarkson et al. 2002; ew.govt.nz). Urban areas represent <1% of land cover and include New Zealand's seventh most populous city, Hamilton (stats.govt.nz).

Invertebrate sample collection and processing

Stream macroinvertebrates were collected in summer 2002 to 2005 over 50–100 m long reaches from stable substrates in flowing water, where invertebrate numbers were expected to be relatively high, using a D-frame net (0.5 mm mesh). Sampling of hard-bottomed streams (<50% bottom substrates sand/silt/clay) was conducted mainly in riffles by kicking stones and brushing larger substrate elements at several points (typically 3–5) along the reach until around 0.5–1.0 m^2 of habitat had been sampled. Where riffle habitat was rare or non-existent, sampling involved brushing wood and jabbing the

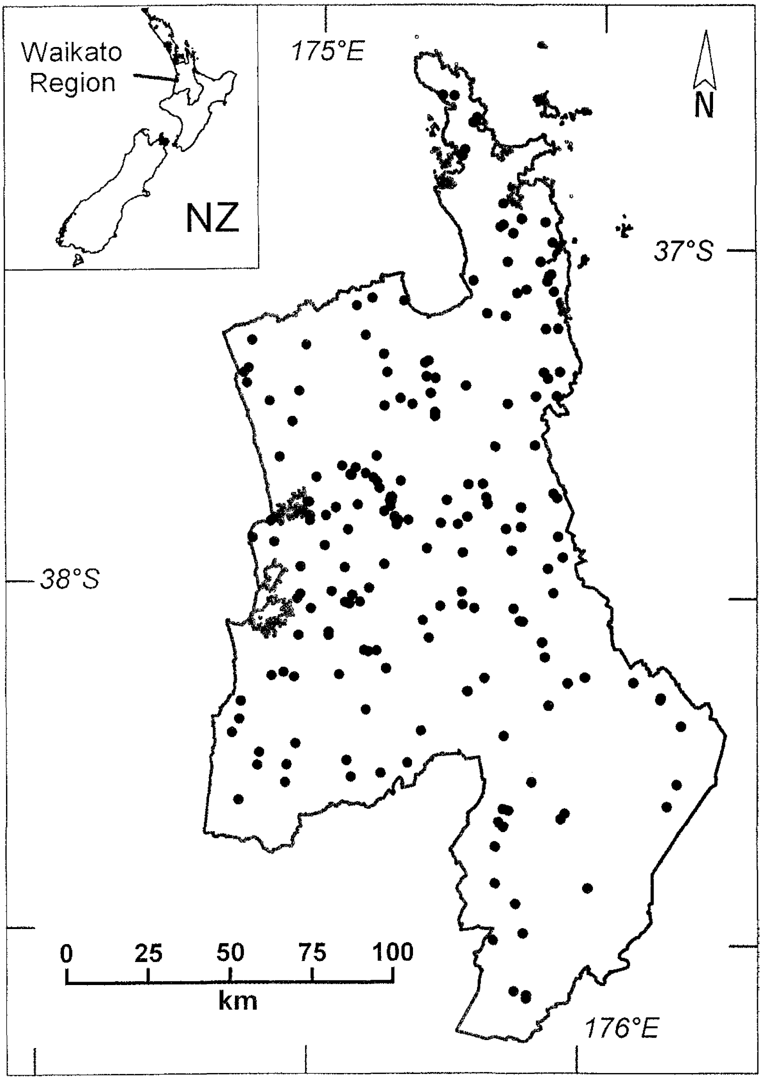

Fig. 1 Map showing location of sampling sites comprising the calibration data set in the Waikato region, New Zealand.

net along banks and among macrophytes, at up to 10 locations along the reach so that approximately 3 m^2 of habitat were sampled in runs (see Collier & Kelly 2005).

Samples were preserved in c. 70% isopropanol, and were subsequently processed by identifying at least 200 invertebrates (excluding pupae) from randomly selected grid squares in a sorting tray, following protocol P2 in Stark et al. (2001). Most insects and molluscs were identified to genus, and identification of other groups ranged from family to phylum, as recommended by Stark et al. (2001) for wadeable stream monitoring and calculation of the macroinvertebrate community index (MCI). Final invertebrate numbers ranged from 200 to 300 in this data set which was used to develop the ASPM (referred to as the "calibration" data set). Data obtained in the above manner were available from 511 samples collected at 192 sites, including 37 reference samples from sites with upstream catchments and riparian areas entirely in native vegetation (see Table 1 for site characteristics)

Selection of candidate metrics

A suite of 17 macroinvertebrate community metrics was selected *a priori* for screening based on empirical studies and conceptual understanding of stream invertebrate responses to environmental stressors. This initial set comprised eight structural or diversity metrics (Margalef diversity, total taxa richness, % native taxa, richness of Ephemeroptera (E), Plecoptera (P) and Trichoptera (T), combined EPT, and % EPT richness); seven compositional metrics (% E, T, EPT, insects, non-"worms" (i.e., excluding Oligochaeta, Platyhelminthes, Nemertea, Nematoda and Hirudinea), native abundance, and dominant taxon), and presence/absence and abundance-based versions of the MCI (Stark 1985). Hydroptilidae was excluded from all EPT metrics as recommended by Maxted et al. (2003; denoted by "*") because the most common hydroptilid taxon proliferates in algal blooms. For calculation of the MCI, tolerance scores were the same as those listed in Collier & Kelly (2005), except the combined Chironomidae taxon was designated a score of 5 based on the average value for all Chironomidae sub-families, as these were not discriminated before 2002. The complement of % dominant taxon (i.e., 100 – % dominant taxon) was used so that higher values reflected better ecological condition.

Metrics were initially screened for redundancy within similar metric groupings (i.e., structural, compositional, or tolerance) with a correlation analysis (Pearson or Spearman; Sokal & Rolf 1981); only those with correlation coefficients <0.7 were retained for the next step of the evaluation. If two metrics were highly correlated, the one with more parsimonious data requirements was selected. For example, QMCI and MCI were highly correlated (r = 0.86), but the latter metric was selected because it did not involve counting individuals of all taxa. This procedure reduced the initial set of 17 candidate metrics to eight metrics for use in further development of the ASPM. These metrics were: EPT* richness, % native richness, % Trichoptera abundance, % EPT* abundance, % non-worms, % native abundance, the complement of % dominant taxon, and MCI.

Calculation of the ASPM

Initially, the average of the eight metrics standardised by the maximum measured value for each metric was calculated, resulting in values between 0 and 1. The metrics were then optimised to determine if a set of metrics existed amongst these that maximised discrimination between reference samples and those from sites in catchments influenced by pastoral and urban development. Optimisation involved comparing the average scores for reference samples from native forested catchments (n = 37) versus samples from sites with >95% of upstream catchment and segment land cover in pasture (n = 42), or samples from Hamilton urban streams (n = 17). Starting with the subset of eight metrics, individual metrics were removed sequentially in order of their correlations with the primary axis of a principal components analysis of metrics using a cross-product correlation matrix and a Euclidean distance-based biplot (see Collier 2008 for further details on this approach). Removal started with the lowest correlated metric (% non-worms), and the average standardised value was re-calculated (see Table 2). The next lowest correlated metric was then removed and this step was repeated until only the MCI remained. This procedure identified a combination of metrics that provided the greatest discrimination between reference and both pastoral and urban sites. Individual metrics were then added to this combination to determine if discrimination between urban and pastoral stressors was enhanced. The average standardised value of the optimised set of metrics was then used to calculate the ASPM. Triangulation plots were used to visualise the relationships between these standardised metrics.

To provide narrative condition bands for communication purposes, four thresholds were derived by calculating quartiles of each metric between the lower SD of the reference mean and a hypothetical worst-

Table 1 Selected site characteristics of the calibration data set. For land cover, % = percentage of total upstream catchment area above the stream segment on which the sampling site was located.

	Median	Mean	SD	Minimum	Maximum
Stream order	3.0	2.9	1.3	1.0	7.0
Average segment elevation (m a.s.l.)	72.8	139.6	160.3	4.0	791.0
Segment slope (m/m)	0.01	0.02	0.03	0.00	0.21
Indigenous forest (%)	0.0	23.0	34.2	0.0	100.0
Exotic forest (%)	0.0	2.3	12.1	0.0	100.0
Scrub (%)	0.0	9.1	20.2	0.0	100.0
Pasture (%)	74.2	59.8	39.9	0.0	100.0
Urban (%)	0.0	3.3	13.3	0.0	80.0

case community comprising 100% Oligochaeta, *Chironomus*, Psychodidae or Syrphidae which all have MCI tolerance scores of 1 (Stark et al. 2001). The resulting values distinguished "very low" (<0.20), "low" (0.20–0.36), "moderate" (0.37–0.52), and "high" (0.53–0.68) condition bands. Scores above the high band marker (i.e., above the lower reference site SD) were classified as "very high" (>0.68).

Performance of the ASPM

Performance of the ASPM relative to its component metrics was investigated by assessing: (1) sensitivity to different levels of assumed low–moderate impact based on upstream land cover and habitat quality; (2) temporal sensitivity at reference sites over multiple years; and (3) responses to an independent test data set of reference and urban samples. Impact sensitivity was determined for hard-bottomed streams sampled in 2005 by assessing the ability of the ASPM and component metrics to differentiate between reference sites and three approximately equal-sized groups of sites with different levels of reach-scale habitat quality and percentage of indigenous vegetation cover in upstream catchments. Habitat quality was assessed using a standard procedure that evaluates nine attributes related to riparian, bank, channel, and instream conditions by scoring each attribute on a scale of 1 to 20 and summing these to provide an integrated score (see Collier & Kelly 2005). Habitat quality scores were then expressed as a percentage of the maximum possible. The three impact groups were classified as follows: "1"—upstream catchment area in native forest >85% and habitat scores 72–89% of maximum ($n = 4$); "2"—native forest 72–83% and habitat scores 66–71% ($n = 5$); "3"—native forest 52–59% and habitat scores 57–64% ($n = 4$). These impacted sites were sampled over the same period as the reference sites ($n = 23$) using identical methods, and were considered low–moderate impact because all had upstream catchments predominantly in native forest and habitat quality scores close to or greater

Table 2 Mean (1 SD) average score per metric (ASPM) values following sequential removal or addition of metrics for samples collected from reference (catchment 100% indigenous vegetation; $n = 37$), urban ($n = 17$) and pastoral (≥ 95% upstream and segment land cover; $n = 42$) sites. (Diff, difference between mean values for reference and pasture or urban samples (greatest differences in bold); 8 metrics, % non-worms, % native abundance, % Trichoptera, complement of % dominant taxon, % native richness, % EPT*, EPT* taxa, MCI; 3 metrics, EPT* richness, % EPT*, and MCI, EPT*, Ephemeroptera, Plecoptera, Trichoptera excluding Hydroptilidae; MCI, macroinvertebrate community index.)

	Reference	Pasture		Urban	
	Mean	Mean	Diff.	Mean	Diff.
8 metrics	0.80 (0.04)	0.51 (0.06)	0.29	0.46 (0.09)	0.34
7 metrics (– % non-worms)	0.77 (0.05)	0.45 (0.06)	0.32	0.43 (0.06)	0.35
6 metrics (– % native abundance)	0.73 (0.06)	0.37 (0.07)	0.36	0.33 (0.07)	0.40
5 metrics (– % Trichoptera)	0.83 (0.06)	0.42 (0.07)	0.41	0.38 (0.08)	0.45
4 metrics (– complement % dom. taxon)	0.84 (0.06)	0.39 (0.06)	0.45	0.38 (0.05)	0.46
3 metrics (– % native richness)	0.79 (0.08)	0.22 (0.07)	**0.57**	0.21 (0.05)	**0.58**
2 metrics (– % EPT*)	0.78 (0.07)	0.30 (0.07)	0.48	0.29 (0.06)	0.49
MCI	0.88 (0.06)	0.51 (0.05)	0.37	0.49 (0.06)	0.38
3 metrics + complement % dom. taxon	0.79 (0.01)	0.29 (0.08)	0.50	0.25 (0.08)	0.54
3 metrics + % non-worms	0.84 (0.07)	0.39 (0.06)	0.45	0.33 (0.12)	0.51
3 metrics + % Trichoptera	0.66 (0.06)	0.19 (0.08)	0.47	0.17 (0.05)	0.49
3 metrics + % native abundance	0.84 (0.07)	0.40 (0.05)	0.44	0.40 (0.04)	0.44

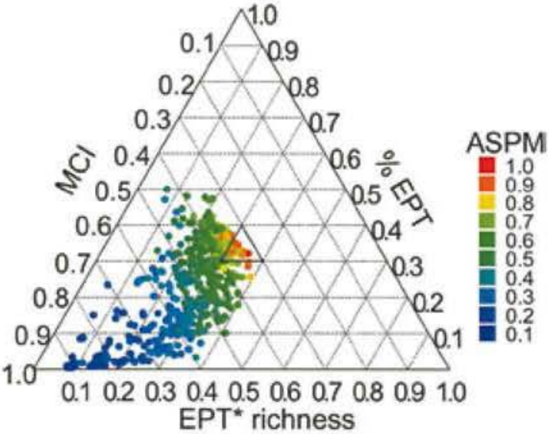

Fig. 2 Triangle plot of the three component metrics standardised by the maximum observed value showing their relationship with the average score per metric (ASPM). Solid triangle in the centre shows the concentration of highest ASPM values indicative of values >0.6 for all metrics. EPT*, Ephemeroptera, Plecoptera, Trichoptera excluding Hydroptilidae; MCI, macroinvertebrate community index.

than those typical of sites in the calibration dataset (median = 63%). Differences among categories were tested using one-way ANOVA with sites nested within impact group (Sokal & Rolf 1981), except for % EPT* which appeared to deviate from normality and was analysed using ranked data (Conover & Iman 1981). Post-hoc analyses were conducted using Bonferroni tests (Miller 1981).

To assess temporal sensitivity, reference sites were chosen that had at least 3 years of data. Only three reference sites met this criterion, with two sites sampled in 4 consecutive years and one site for 3 years. Reference sites were used for this analysis to investigate interannual variability in the absence of anthropogenic stressors. The ability of the ASPM and MCI to assign reference sites to the highest condition band in all years was evaluated.

Response of the ASPM to an independent test data set of samples at the upper and lower ends of a land-cover disturbance gradient (urban versus indigenous forest) was investigated using samples collected in 2006. These test samples were collected using identical protocols to those used for the calibration samples. Maximum metric values from the calibration data set were used to standardise metrics for calculation of the ASPM in test samples.

Effect of subsample size

Before the development of standardised macroinvertebrate protocols (Stark et al. 2001), Environment Waikato used 100 count samples for stream bioassessment. Information on the relationship between these data and the current 200+ counts is important for retrospectively calculating and comparing bioassessment indices for trend analysis. To enable direct assessment of the effects of different subsample sizes on metrics, the Species Diversity module in the computer program ECOSIM (Gotelli & Entsminger 2005) was used to generate communities of 100 individuals from the calibration dataset. Relationships between metrics based 100 and 200+ counts were assessed by simple linear regression (Sokal & Rolf 1981).

RESULTS

Development of the ASPM

Sequential removal or addition of metrics demonstrated that the MCI in isolation yielded highest reference scores (Table 2). The combination of EPT* richness, % EPT*, MCI, and % Trichoptera provided lowest scores for urban and pastoral stressors. However, EPT* richness, % EPT*, and MCI metrics together provided greatest discrimination between reference and impact conditions for test sites, and addition of other core metrics did not enhance this difference (Table 2). The use of these three metrics was also tested with the replacement of MCI with QMCI which increased the difference between reference and urban samples from 0.58 to 0.63. However, the MCI was used in calculating the ASPM because: (1) ASPM values using MCI or QMCI were highly correlated ($r = 0.99$); (2) the use of QMCI did not enhance the sensitivity analysis; and (3) QMCI would require increased effort in correctly identifying small instars for some taxa which was not consistent with the goal of developing a parsimonious metric aggregation tool. Maximum values used to standardise metrics for the ASPM in the calibration data set were 22, 96.6%, and 163.3 for EPT* richness, % EPT*, and MCI, respectively.

Triangulation of metrics in relation to the ASPM demonstrated a concentration of high ASPM values where all standardised scores exceeded 0.6 (Fig. 2). This analysis also demonstrated that different combinations of metric scores can result in similar ASPM values. For example, the combination of intermediate standardised MCI values of 0.5–0.6, % EPT* of 0.8–0.9, and EPT* richness of 0.6–0.7 can

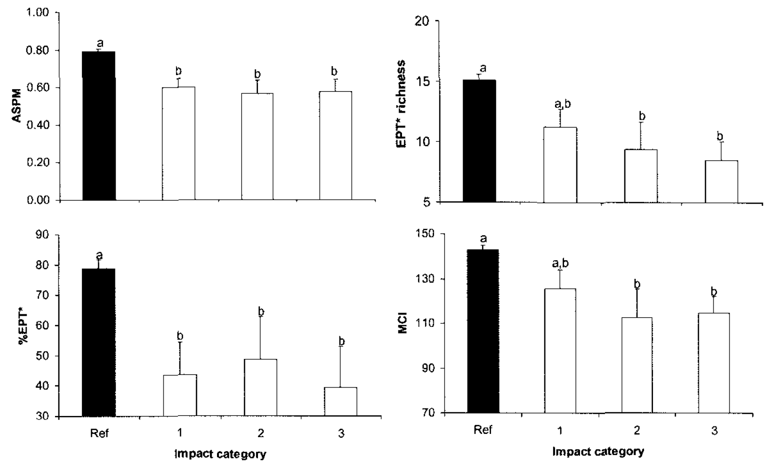

Fig. 3 Mean (+1 SE) values for the average score per metric (ASPM) and component metrics for reference sites (Ref, n = 23) and three categories of low impact test sites (1–3, n = 4–5) sampled in 2005. EPT*, Ephemeroptera, Plecoptera, Trichoptera excluding Hydroptilidae; MCI, macroinvertebrate community index. Bars with same letter above are not significantly different (Bonferroni post-hoc test, $P < 0.05$).

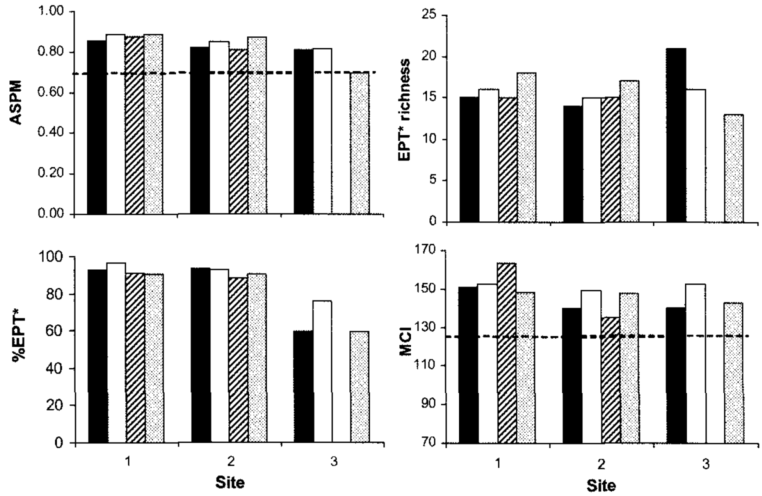

Fig. 4 Variation in metric and index scores at three reference sites (1–3) over 3–4 consecutive years (2002–05). Closed bars, 2002; open bars, 2003; diagonal striped bars, 2004; stippled bars, 2005. Dashed horizontal lines indicate the start of the highest condition bands for the macroinvertebrate community index (MCI) following Wright-Stow & Winterbourn (2003), and the average score per metric (ASPM). EPT*, Ephemeroptera, Plecoptera, Trichoptera excluding Hydroptilidae.

yield high ASPM values, as can standardised values of 0.8–0.9 for MCI, and 0.5–0.6 for % EPT* and EPT* richness. However, samples characterised by few taxa comprising mostly EPT, thereby producing high MCI and % EPT* values, produced low ASPM values (Fig. 2).

Performance of the ASPM

The ASPM showed similar sensitivity to % EPT* in terms of its ability to discriminate between reference samples and low–moderate impact samples based on upstream catchment land use and local habitat quality ($F_{1,3}$ = 13.47 for ASPM and 9.39 for % EPT*, $P < 0.001$) (Fig. 3). However, MCI ($F_{1,3}$ = 9.84, $P < 0.001$) and EPT* richness ($F_{1,3}$ = 9.17, $P < 0.001$) did not discriminate between reference sites and the lowest impact group (Fig. 3). Neither the ASPM nor its component metrics discriminated between the three low impact categories used. The site term was significant only for % EPT* ($F_{1,3}$ = 4.61, $P < 0.05$) suggesting that this metric may have been influenced by site differences other than habitat quality and catchment land cover.

Temporal variability within reference sites among consecutive years (i.e., the maximum difference between all years) varied between 0.03 and 0.12 for the ASPM (Fig. 4), equivalent to a coefficient of variation (CV) of 63%. This CV was similar to CVs for the two EPT* metrics (62–65%) but much higher than that for the MCI (8%). Nevertheless, all ASPM values among this data set were allocated to the very high condition band, although in one year

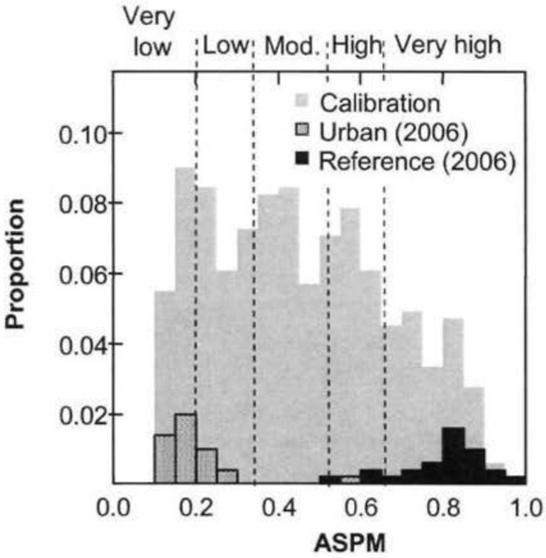

Fig. 5 Distribution of average score per metric (ASPM) values from the calibration data set (2002–05; grey bars) and test samples from independent urban and reference stream test data sets from 2006.

this allocation was at the lower end of the very high range.

Metric calculations for the 2006 reference test data set yielded higher maxima for EPT* taxa richness and MCI (23 and 164.8, respectively) than the calibration data set. As the calibration maxima were used to calculate the ASPM in the test data set, standardised values >1.0 were set to 1.0. Calculation of ASPM values for the independent test data set allocated most 2006 reference samples to the very high class, although there was considerable spread (Fig. 5). Urban test sites were allocated to the lower end of the scoring range (equivalent to low or very low bands), with the exception of one high value urban site with relatively low upstream impervious area which had an intermediate ASPM.

Effects of subsample size

Comparison of the 200+ fixed-counts and the computer-generated 100 fixed-counts derived from the 200+ data indicated strong linear relationships for the ASPM, % EPT* and MCI (Fig. 6). Slopes for these relationships were ≥0.93 when the intercept was forced through zero. However, the relationship for EPT* richness using the two processing protocols displayed considerable variability and had a slope markedly different to 1 (Fig. 6).

DISCUSSION

Development and performance of the ASPM

The ASPM approach provided temporally consistent reference site allocation to high condition bands, and was effective at discriminating reference condition from low–moderate levels of catchment- and reach-scale habitat degradation. Comparison with the ratio of observed species at Waikato test sites to expected species at reference sites (O/E), derived from an expanded data set incorporating other regions (R. G. Death, Massey University unpubl. data), yielded results that were strongly correlated with the ASPM ($r^2 = 0.81$) (author's unpubl. data). This relationship suggests that both multimetric and O/E methods can respond in a similar way to environmental conditions, as also noted by Stribling et al. (2008), who reported that a multimetric index derived from averaging six metrics provided consistent and repeatable results compared with O/E.

The three metrics (EPT* richness, % EPT*, and MCI) that optimised the difference between reference and pastoral or urban impact sites are widely recognised as being sensitive to a range of anthropogenic stressors (e.g., Lenat 1984; Collier 1995; Angradi 1999; Collier & Smith 2005; Rois & Bailey 2006; Walsh 2006), and are commonly used in multimetric bioassessments (see Plafkin et al. 1989; Barbour et al. 1999; Maxted et al. 2000; Quinn et al. 2004). Although these metrics were strongly intercorrelated ($r = 0.79$–0.90; author's unpubl. data), all were retained because they summarised different aspects of macroinvertebrate communities (structure, composition, and tolerance), and they enhanced discrimination between impact and reference samples when aggregated. Collier (2006) concluded that these three metrics are powerful at discriminating temporal changes attributable to anthropogenic impacts from background variability.

The comparison of 200+ and 100 counts derived from the same data set indicates that two of the component metrics, % EPT* and MCI, were robust to variations in sample size. However, EPT* richness was sensitive to sample size, reflecting species-number relationships and also the search of the entire sample for rare taxa carried out as part of the 200+ count protocol. The averaging approach used to calculate the ASPM reduced this variability considerably suggesting that, although variable count sizes may influence some component metrics, the ASPM can be used with reasonable accuracy to calculate scores retrospectively for trend analysis where sample counts have changed over the years.

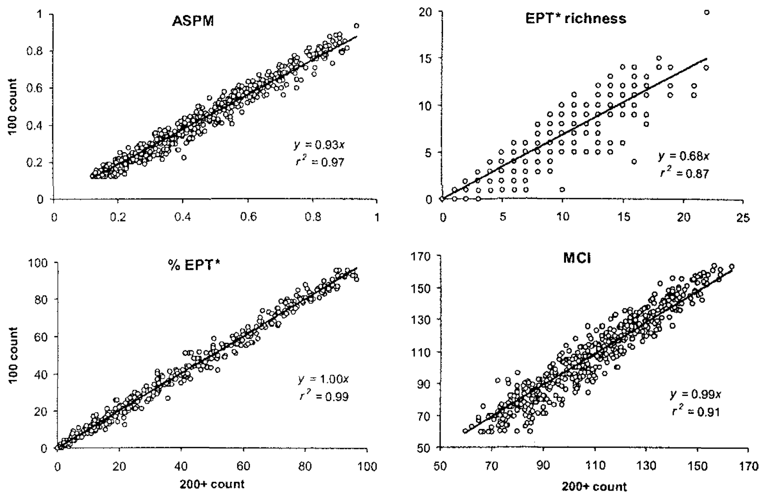

Fig. 6 Relationships between 100 count and 200+ count macroinvertebrate processing protocols for the average score per metric (ASPM) and its component metrics. EPT*, Ephemeroptera, Plecoptera, Trichoptera excluding Hydroptilidae; MCI, macroinvertebrate community index.

The approach described in this study involves: (1) collecting and processing a sample using standard protocols, placing taxa initially into two groups—EPT and non-EPT; (2) identifying taxa present in each group to a pre-determined level of taxonomy to calculate EPT* richness and MCI; (3) enumerating the EPT as a single group to calculate % EPT*; (4) standardising metrics by a maximum value, typically derived from a reference site data set; and (5) taking the average of the standardised metric values to produce the ASPM. This procedure does not require identification of all early instars if mature specimens are found (e.g., small leptophlebiid genera could be allocated to Ephemeroptera if larger instars are present), as would be required if quantitative indices such as the QMCI were included. Moreover, although the effect of combining all Chironomidae into a single group was not evaluated, any loss of sensitivity caused by taxonomic consolidation during retrospective calculation is likely to be reduced through the averaging process. Some researchers have suggested that eliminating Chironomidae for bioassessment can enhance metric sensitivity and the efficiency of biomonitoring programmes (Rabeni & Wang 2001). However, quantitative data at a high level of taxonomy can provide greater diagnostic power if species tolerances are known, and also provide information on community structure and evenness that may be useful for other analyses (e.g., biodiversity assessment).

Advantages and disadvantages of metric aggregation

Reliance on individual metrics to assess stream health can produce misleading results. For example, EPT taxa may persist in streams affected by mine drainage where they can comprise significant proportions of otherwise depauperate communities and result in high MCI values (Hickey & Clements 1998; Hickey & Golding 2002). The ASPM reduced the influence of MCI on health assessments through the averaging process, and resulted in low scores for samples with high MCI and % EPT* values but low numbers of EPT* taxa (see Fig. 2). As noted by Yuan & Norton (2003), when patterns of responses between metrics are markedly different, it may be beneficial to aggregate metrics to determine the overall severity of ecological effects. Interpretation of aggregated indices in association with key individual metrics, for example through triangulation plots, may provide some diagnostic power when attempting to identify causes of impairment. Changes in metric aggregation scores resulting from the addition of other metrics that reflect particular stressor responses may also enhance diagnostic power, and provide a useful avenue for future development of metric aggregation methods such as the ASPM.

Bressler et al. (2006) considered metrics based on EPT faunas to be useful only where EPT can be expected to comprise significant proportions of the fauna naturally. Although EPT taxa may not always be common, for example in large rivers

(Collier & Lill 2008), they can comprise substantial proportions of the fauna colonising wood in Waikato reference streams where fine substrates dominate benthic habitats (Collier et al. 2007). A refinement of the ASPM could be to apply the version of the MCI developed for soft-bottomed streams (Maxted et al. 2003), but this version may be inappropriate in situations where the dominance of benthic substrates by fine material and proliferation of macrophytes reflects anthropogenic modification rather than natural processes. As with other metric aggregation approaches, the ASPM is also influenced by natural environmental variables as well as human impacts, and relies on comparable reference sites being available to interpret condition. In the present study, responses of the ASPM were tested mainly to changes in land cover, although (as noted above) it seems likely to also respond effectively to mining impacts which may be overlooked using some single metrics.

Development of condition bands

Several methods have been applied to develop condition band thresholds when deriving multimetric scores, with most involving selection of percentiles of final scores or range splitting into multiple equally-spaced groupings (see Maxted et al. 2000). Usually, impairment or condition thresholds are established arbitrarily by dividing the data set into groups based on percentiles of the reference site range (e.g., <25th percentile of reference site samples; Hannaford & Resh 1995). Smith et al. (2005) suggested that the use of tolerance intervals, that take data variability into account, would remove some of the subjectivity of using standard percentiles. In contrast, van Sickle et al. (2005) used 2 SD of the mean of reference sites to define impairment in an O/E model for mid-Atlantic and North Carolina streams, United States. The approach used in the present study derived condition bands based on individual metric quartile values below the lower SD of the reference mean. Although increasing the number of reference sites could change the condition bands, the magnitude of change is likely to be small because of the stability of reference site metric values among years and across seasons (author's unpubl. data). Since the Environment Waikato monitoring network includes the same population of reference sites on each occasion, increases in variation caused by spatial differences in temporal variability are not likely to be a concern. The 2006 test data included reference site samples that exceeded the best-case condition (i.e., maximum recorded value for each metric) used to standardise metrics in the calibration data set. In these situations any exceedences can be allocated the maximum score, or alternatively annual maxima from a reference site network could be used to account for temporal variations in climatic conditions. There is also the potential to use different reference site data sets for standardisation, such as for different stream types or geographic areas, to make the ASPM more specific to particular situations, as proposed by Stribling et al. (2008) for mountain, low valley and plain streams in Montana, United States.

CONCLUSIONS

There are many ways to analyse macroinvertebrate community data to provide an indication of ecological condition, with each approach having strengths and weaknesses. Although metric aggregation methods may not fully satisfy all of the criteria required for stream condition indices (e.g., Norris & Hawkins 2000), the prudent application of simple methods, coupled with an understanding of limitations, may help overcome some of the barriers that perpetuate the continued use of single metrics for making stream management decisions. The ASPM approach provides an alternative aggregation method that can circumvent some of the uncertainties associated with taxonomic allocation of small instars, is readily calculated without specialised software, and can be communicated easily to a range of stakeholders at multiple levels. As such, the ASPM may provide a useful summary of the ecological condition of wadeable streams in situations where more complex methods are not practical. Furthermore, it may provide an effective screening tool in a tiered assessment approach that also considers the relative responses of core metrics and the influence of additional stressor-specific metrics on the overall score.

ACKNOWLEDGMENTS

Johlene Kelly, Myles Hill, Nathan Singleton, and Steve Moore (Landcare Research) provided valuable contributions to data acquisition. Thanks to John Quinn, Bill Vant, and two anonymous referees for reviewing the manuscript.

REFERENCES

Angradi TR 1999. Fine sediment and macroinvertebrate assemblages in Appalacian streams: a field experiment with biomonitoring applications. Journal of the North American Benthological Society 18: 49–66.

Barbour M, Gerritsen J, Snyder BD, Stribling JB 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish. 2nd ed. EPA 841-B-99-002. US Environmental Protection Agency, Office of Water; Washington DC, United States. http://www.epa.gov/owowwtr1/monitoring/rbp/index.html [accessed 6 September 2008].

Bressler DW, Stribling JB, Paul MJ, Hicks MB 2006. Stressor tolerance values for benthic macroinvertebrates in Mississippi. Hydrobiologia 573: 155–172.

Clarkson B, Merrett M, Downs, T 2002. Botany of the Waikato. Hamilton, New Zealand, Waikato Botanical Society. 136 p.

Collier KJ 1995. Environmental factors affecting the taxonomic composition of aquatic macroinvertebrate communities in lowland waterways of Northland, New Zealand. New Zealand Journal of Marine and Freshwater Research 29: 453–465.

Collier KJ 2005. Review of Environment Waikato's regional ecological monitoring of streams (REMS) programme: past practices and future directions. Environment Waikato Technical Report TR05/48. Environment Waikato, Hamilton, New Zealand. http://www.ew.govt.nz/PageFiles/2934/tr05-48.pdf [accessed 20 June 2008].

Collier KJ 2006. Temporal trends in macroinvertebrate metrics for some Waikato streams. New Zealand Natural Sciences 31: 79–91. Erratum: New Zealand Natural Sciences 33: 51.

Collier KJ 2008. Linking multimetric and multivariate approaches to assess the ecological condition of streams. Environmental Monitoring and Assessment in press.

Collier KJ, Kelly J 2005. Regional guidelines for ecological assessments of freshwater environments: macroinvertebrate sampling in wadeable streams. Environment Waikato Technical Report TR05/02. Environment Waikato, Hamilton, New Zealand. http://www.ew.govt.nz/PageFiles/3114/tr05-02.pdf [accessed 20 June 2008].

Collier KJ, Lill A 2008. Spatial patterns in the composition of shallow-water macroinvertebrate communities of a large New Zealand river. New Zealand Journal of Marine and Freshwater Research 42: 129–141.

Collier KJ, Smith BJ 2005. Effects of progressive catchment harvesting on stream invertebrates in two contrasting regions of New Zealand's North Island. Marine and Freshwater Research 56: 57–68.

Collier KJ, Haigh A, Kelly J 2007. Coupling GIS and multivariate approaches to select reference sites for wadeable stream monitoring. Environmental Monitoring and Assessment 127: 29–45.

Conover WJ, Iman RL 1981. Rank transformations as a bridge between parametric and non-parametric statistics. American Statistician 35: 124–133.

Gerritsen J 1995. Additive biological indices for resource management. Journal of the North American Benthological Society 14: 451–457.

Gotelli NJ, Entsminger GL 2005. EcoSim: Null models software for ecology. Version 7. Acquired Intelligence Inc. & Kesey-Bear. Jericho, VT 05465, United States. http://garyentsminger.com/ecosim/index.htm [accessed 20 June 2008].

Hannaford MJ, Resh VH 1995. Variability in macroinvertebrate rapid-bioassessment surveys and habitat assessments in a northern Californian stream. Journal of the North American Benthological Society 14: 430–439.

Hickey CW, Clements WH 1998. Effects of heavy metals on benthic macroinvertebrate communities in New Zealand streams. Environmental Toxicology and Chemistry 17: 2338–2346.

Hickey CW, Golding LA 2002. Response of macroinvertebrates to copper and zinc in a stream mesocosm. Environmental Toxicology and Chemistry 21: 1854–1863.

Joy MK, Death RG 2004. Application of the index of biotic integrity methodology to New Zealand freshwater fish communities. Environmental Management 34: 415–428.

Karr JR 1999. Defining river health. Freshwater Biology 41: 221–234.

Kilpatrick R 1999. Bateman contemporary atlas of New Zealand. Auckland, New Zealand, David Bateman Ltd.

Lenat DR 1984. Agriculture and stream water quality: a biological evaluation of erosion control processes. Environmental Management 8: 333–344.

Maxted, JR, Barbour MT, Gerritsen J, Poretti V, Primrose N, Silvia A, Penrose D, Renfrow R 2000. Assessment framework for mid-Atlantic coastal plain streams using benthic macroinvertebrates. Journal of the North American Benthological Society 19: 128–144.

Maxted JR, Evans BF, Scarsbrook MR 2003. Development of standard protocols for macroinvertebrate assessment of soft-bottomed streams in New Zealand. New Zealand Journal of Marine and Freshwater Research 37: 793–807.

McCraw JD 1971. The geological history of the Waikato River basin. In: Duncan C ed. The waters of the Waikato. The University of Waikato, Hamilton, New Zealand. Pp. 11–23.

Miller RG 1981. Simultaneous statistical inference. New York, Springer. 299 p.

Norris RH 1995. Biological monitoring: the dilemma of data analysis. Journal of the North American Benthological Society 14: 440–450.

Norris RH, Hawkins CP 2000. Monitoring river health. Hydrobiologia 435: 5–17.

Plafkin JL, Barbour MT, Porter KD, Gross SK, Hughes RM 1989. Rapid bioassessment protocols for use in streams and rivers: benthic invertebrates and fish. US Environmental Protection Agency, Office of Water Regulations and Standards, Washington DC, United States.

Quinn JM, Boothroyd IKG, Smith BJ 2004. Riparian buffers mitigate effects of pine plantation logging on New Zealand streams 2. Invertebrate communities. Forest Ecology and Management 191: 129–146.

Rabeni CF, Wang N 2001. Bioassessment of streams using macroinvertebrates: are the Chironomidae necessary? Environmental Monitoring and Assessment 71: 177–185.

Reynoldson TB, Norris RH, Resh VH, Day KE, Rosenberg DM 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society 16: 833–852.

Rois SL, Bailey RC 2006. Relationship between riparian vegetation and stream benthic communities at three spatial scales. Hydrobiologia 553: 153–160.

Smith JG, Beauchamp JJ, Stewart AJ 2005. Alternative approach for establishing acceptable thresholds on macroinvertebrate community metrics. Journal of the North American Benthological Society 24: 428–440.

Sokal RR, Rolf, FJ 1981. Biometry. 2nd ed. New York, WH Freeman & Co. 859 p.

Stark JD 1985. A macroinvertebrate community index of water quality for stony streams. Water and Soil Miscellaneous publication 87. Ministry of Works and Development, Wellington, New Zealand.

Stark JD, Boothroyd IKG, Harding JS, Maxted JR, Scarsbrook MR 2001. Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate Working Group report no. 1. Ministry for the Environment, Wellington, New Zealand.

Stribling JB, Jessup BK, Feldman DL 2008. Precision of benthic macroinvertebrate indicators of stream condition in Montana. Journal of the North American Benthological Society 27: 58–67.

Suter GW II 1993. A critique of ecosystem health concepts and indexes. Environmental Toxicology and Chemistry 12: 1533–1539.

van Sickle J, Hawkins CP, Larsen DP, Herlihy AT 2005. A null model for the expected macroinvertebrate assemblage in streams. Journal of the North American Benthological Society 24: 178–191.

Walsh CJ 2006. Biological indicators of stream health using macroinvertebrate assemblage composition: a comparison of sensitivity to an urban gradient. Marine and Freshwater Research 57: 37–47.

Weigel BM 2003. Development of stream macroinvertebrate models that predict watershed and local stressors in Wisconsin. Journal of the North American Benthological Society 22: 123–142.

Wright-Stow AE, Winterbourn MJ 2003. How well do New Zealand's stream-monitoring indicators, the Macroinvertebrate Community Index and its quantitative variant, correspond? New Zealand Journal of Marine and Freshwater Research 37: 461–470.

Yuan LL, Norton SB 2003. Comparing responses of macroinvertebrate metrics to increasing stress. Journal of the North American Benthological Society 22: 308–322.