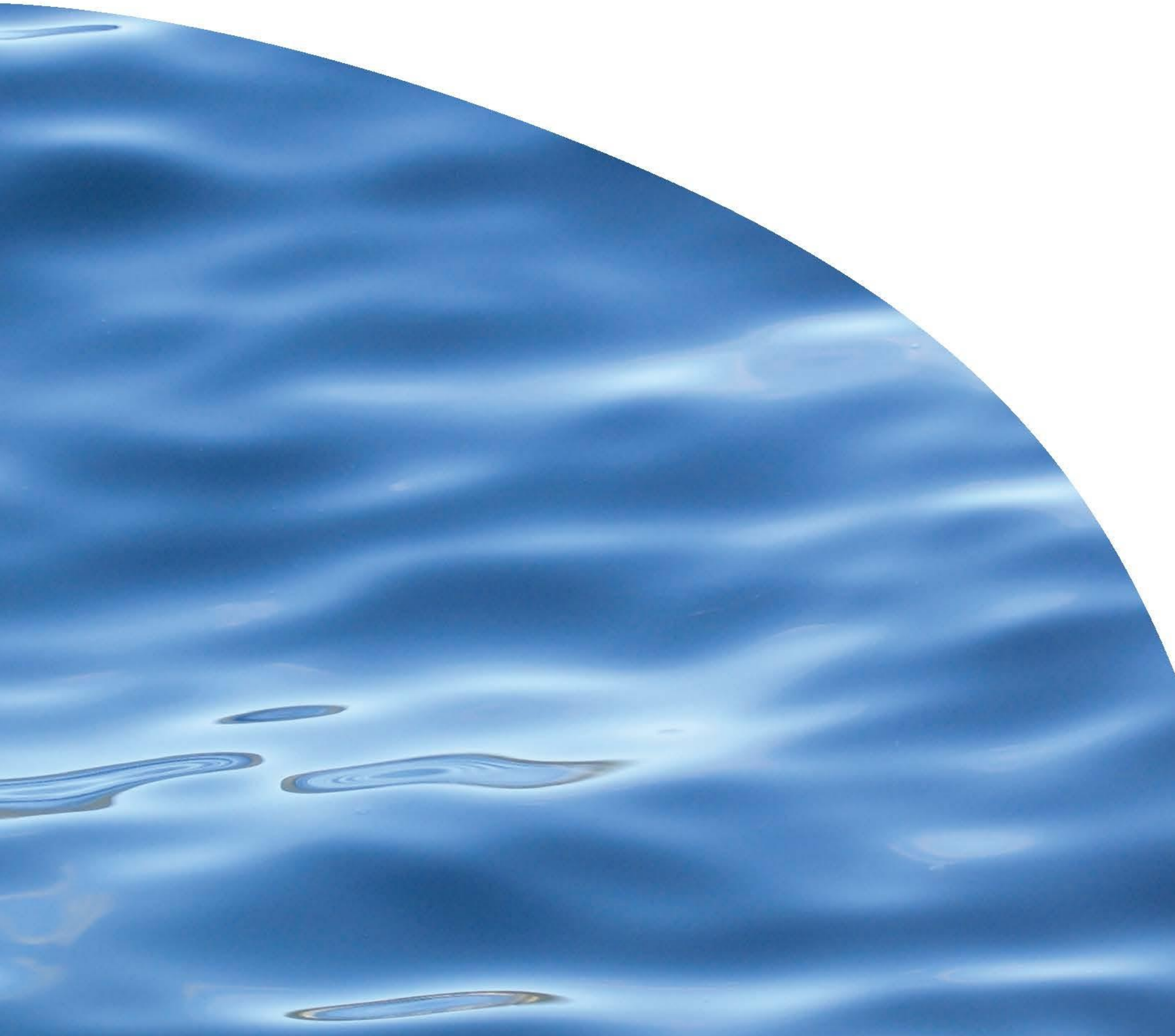
CAWTHRON INSTITUTE

REPORT NO. 2852

A REVIEW OF BENTHIC MACROINVERTEBRATE METRICS FOR ASSESSING STREAM ECOSYSTEM HEALTH



A REVIEW OF BENTHIC MACROINVERTEBRATE METRICS FOR ASSESSING STREAM ECOSYSTEM HEALTH

ANNIKA WAGENHOFF, KAREN SHEARER, JOANNE CLAPCOTT

Prepared for Environment Southland

CAWTHRON INSTITUTE
98 Halifax Street East, Nelson 7010 | Private Bag 2, Nelson 7042 | New Zealand
Ph. +64 3 548 2319 | Fax. +64 3 546 9464
www.cawthron.org.nz

REVIEWED BY:
John Stark

APPROVED FOR RELEASE BY:
Roger Young

ISSUE DATE: 12 May 2016

RECOMMENDED CITATION: Wagenhoff A, Shearer K, Clapcott J 2016. A review of benthic macroinvertebrate metrics for assessing stream ecosystem health. Prepared for Environment Southland. Cawthron Report No. 2852. 49 p. plus appendices.

EXECUTIVE SUMMARY

Benthic macroinvertebrates have been used for decades as biological indicators of river and stream health. The National Policy Statement for Freshwater Management (2014) requires councils to set enforceable water quality and quantity limits that must reflect local and national values. Therefore, if macroinvertebrate indicators are to be included in the NPS-FM framework then they need to have connection to defined values. Indicators also need to be quantitatively linked to management options. The currently-used invertebrate indicators in New Zealand do not have these links. At present this is the major constraint for use of macroinvertebrate information in NPS-FM limit-setting framework, particularly when managing the effects of multiple stressors.

European countries and the USA use benthic macroinvertebrates to assess the ecological and biological status of streams within their respective water policy frameworks. The metrics used reflect the component properties of healthy waterways and, when used together in a multi-metric index, provide a holistic assessment of ecosystem health. Specific metrics can be used diagnostically to identify primary stressors and provide focus for mitigation and/or restoration. The metrics used overseas are applicable to New Zealand data sets and could be used (alongside new metrics currently under development) to ultimately assess the Te Hauora o te Wai value. Component metrics could also contribute to the assessment of the NPS-FM Mahinga kai value alongside non-macroinvertebrate metrics.

Analysis of benthic macroinvertebrate metrics, existing or new, is required to quantify the stressor-response relationship. Significant progress has been made in this research area recently and now it is a matter of pairing invertebrate data with site-specific stressor measures to develop new metrics and to build national or regional models that quantify the relationship between stressors and invertebrate metrics using a space-for-time approach. Furthermore, knowledge of reference condition is limited and a greater reference data set could be used to develop and validate models that predict reference conditions, and hence facilitate the adoption of a reference model approach to assigning management thresholds for invertebrate indicators.

We recommend that a nationally consistent approach to collecting and processing benthic macroinvertebrates is adopted in New Zealand. We further recommend that proximate stressor and habitat data is collected from invertebrate monitoring sites allowing for development of new metrics and quantification of the relationship between stressors and invertebrate metrics. National or regional models can be used for determining broadly applicable thresholds for instream stressors ('freshwater objectives'), which can ultimately inform limits on resource use for individual catchments. Finally, we recommend that more reference data is required to facilitate a reference condition approach for determining biological thresholds that are linked to defined values.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION & BACKGROUND

1.1. Current use of benthic macroinvertebrates for freshwater management in New Zealand

Benthic macroinvertebrates are important organisms in streams and used worldwide as indicators of stream ecosystem health as they respond to human pressures and are taxonomically diverse and easy to sample (Bonada et al. 2006). In New Zealand, regional and unitary councils routinely collect invertebrate samples for State-of-the-Environment (SoE) monitoring and reporting required by the Resource Management Act 1991. Sample collection by most councils started between the late 1990s and early 2000s, hence a significant amount of data has since been collected nationwide.

Each council designed their own monitoring programme according to their specific aims. Typically, the major aim is to identify short and long term trends in water quality and ecosystem health of streams or rivers that are under significant human pressure and/or that are of significant ecological value. Water quality and ecosystem health is currently assessed using taxonomic information that has been condensed into various single metrics or indices, such as taxon richness (the number of taxa per site) and the widely used New Zealand-specific Macroinvertebrate Community Index (MCI).

The terms 'metric' and 'index' are often used interchangeably. A metric can be defined as a measurable aspect of a biological community that responds in a predictable manner to increasing anthropogenic stress. Combination of multiple metrics that respond to a broad range of human impacts (e.g. sedimentation, organic/nutrient enrichment, toxic chemicals, or flow alteration) into a single, so-called 'multi-metric index' is an approach for integrative assessment. The Average Score Per Metric (ASPM) is the only multi-metric invertebrate index currently used, and by a single council only. A multi-metric index was developed including invertebrates, fish and ecosystem process metrics but has not found application in New Zealand (Clapcott et al. 2014). Equally, an adaptation of the Index of Biotic Integrity (IBI) of the US EPA Rapid Bioassessment Protocol III (Plafkin et al. 1989) incorporating seven metrics has not found application (e.g. Quinn et al. 1997). Finally, multivariate rather than metric approaches to stream health assessment such as the AUSRIVAS have been trialled but not routinely applied in New Zealand.

1.2. The National Policy Statement for Freshwater Management (NPS-FM)

The recent National Policy Statement for Freshwater Management (NPS-FM) is a response to the threats that freshwaters are facing with respect to the ecological health, quality, availability and economic value of interest to all New Zealanders (MfE

2014). The NPS-FM stipulates that freshwaters need to be managed sustainably so that human values are provided for, while also providing for economic growth. In other words, the NPS-FM calls for values-based freshwater management.

In this report we address the compulsory national value Te Hauora o te Wai (ecosystem health) and a further national value, fishing, which is part of the Mahinga kai value. These values are described in Appendix 1 of the NPS-FM (MfE 2014) as follows:

Te Hauora o te Wai (ecosystem health)

- 'In a **healthy freshwater ecosystem** ecological processes are maintained, there is a range and diversity of indigenous flora and fauna, and there is resilience to change.'
- 'Matters to take into account for a healthy freshwater ecosystem include the management of adverse effects on flora and fauna of contaminants, changes in freshwater chemistry, excessive nutrients, algal blooms, high sediment levels, high temperatures, low oxygen, invasive species, and changes in flow regime. Other matters to take into account include the essential habitat needs of flora and fauna and the connections between water bodies.'
- 'The health of flora and fauna may be indicated by measures of macroinvertebrates.'

Fishing (Mahinga kai)

- 'For freshwater management units **valued for fishing**, the numbers of fish would be sufficient and suitable for human consumption. In some areas, fish abundance and diversity would provide a range in species and size of fish, and algal growth, water clarity and safety would be satisfactory for fishers. Attributes will need to be specific for fish species such as salmon, trout, eels, lamprey, or whitebait.'

The NPS-FM sets out objectives and policies to direct local government and provides a National Objectives Framework (NOF) to assist regional councils and communities to more consistently and transparently plan for 'freshwater objectives' (desired environmental outcomes).

Freshwater objectives for a freshwater unit can be set at different states which will be agreed upon by iwi and communities. The NOF defines three acceptable states: A, B and C, the latter being the minimum acceptable state. State D falls below the National Bottom Line and would need management action for improvement. These states can be described narratively and numerically for specific attributes which are measureable characteristics of freshwater including physical, chemical and biological properties (MfE 2014). A NOF attribute table (appendix 2 of the NPS-FM 2014) currently provides four attributes with associated thresholds for Te Hauora o te Wai. These are periphyton biomass (trophic state), nitrate and ammonia (toxicity) and dissolved oxygen. Clearly, further attributes will need to be developed to protect ecosystem health from human impacts. There are currently no measurable attributes provided for the fishing value (part of Mahinga kai).

Implementation of the NPS-FM requires councils to set enforceable limits on water quality and quantity in their management plans in order to meet the 'freshwater objectives' with regards to compulsory national human values and additional national or regional values.

2. AIMS OF THIS REPORT

Given the existing large nationwide invertebrate data set, the overarching aim of this report is to assess how invertebrate information, collected and processed using standard methods, could be used for attribute development to support sustainable management of rivers and streams for the national compulsory value Te Hauora o te Wai (ecosystem health) as required by the NPS-FM. Invertebrates are the main food source for the indigenous and introduced fish species valued for their fisheries. Hence, this report will also assess how invertebrate information could inform attribute development for the national value fishing (as part of the Mahinga kai value) set out in the NPS-FM.

More specifically, this report aims to:

1. present the extent and type of existing invertebrate data in New Zealand, including a summary of the way how councils collect and process invertebrates
2. review the invertebrate metrics currently used to assess stream ecosystem health with respect to the ability to inform the national values Te Hauora o te Wai and fishing contained in the NPS-FM
3. review additional invertebrate metrics for ecosystem health assessment featuring in the international literature (including New Zealand)
4. summarise the above review and recommend a proposed pathway for the development of metrics suitable for working within the NPS-FM framework. In addition, recommendations will be made for priority data collection and/or potential changes to the sampling, collection and processing to optimise current sampling effort, so that historical compatibility is provided for.

3. EXISTING MACROINVERTEBRATE DATA AND REVIEW OF THE METRICS CURRENTLY USED IN NEW ZEALAND

3.1. Overview

To evaluate whether existing invertebrate data can be used for the development of certain metrics, it is important to understand how invertebrates have been collected and processed, and whether it has been done consistently across councils. Most important concerns are the timing of the sampling, type(s) of habitat sampled, sampling effort, quantitative or semi-quantitative sampling, and level of taxonomic resolution. This section will briefly describe the purpose of the development of standard protocols in 2001, the current practice as of 2011 followed by recommendations for future collection made in 2011, and a summary of existing invertebrate data. Finally, the invertebrate metrics currently used for stream health assessment will be reviewed.

3.2. Invertebrate collection and processing

3.2.1. Standard protocols

In the early days of invertebrate monitoring, councils used different methodologies for invertebrate collection and processing. Due to directives from new legislation and the demands from the public to improve the environment, there was a need to develop standard protocols to ensure that resource management decisions were transparent, objective and based on reliable and quality data (Stark et al. 2001). The protocols in Stark et al. (2001), applicable to wadeable streams, were developed for three main activities by councils: SoE monitoring, Assessment of Environmental Effects (AEE) and compliance monitoring. The most widely applied protocols will be summarised below.

3.2.2. Councils' current practice (as of 2011)

In 2011, the Ministry for the Environment (MfE) instigated the National Environmental Monitoring and Reporting (NEMaR) project with the main goal of achieving regional freshwater monitoring as a basis for national SoE reporting. To reach this goal, one part of the project aimed at recommending sample collection and processing protocols for future use by all councils to ensure nationwide consistency in the type and quality of the data produced (Davies-Colley et al. 2012).

According to the questionnaire-based survey in 2011, current practice is largely consistent across councils with respect to invertebrate collection in wadeable hard-bottomed streams (Davies-Colley et al. 2012). The majority of 15 councils has been following the standard semi-quantitative sampling protocol C1 within riffle or run habitat (Stark et al. 2001). At five different locations within the riffle/run, a standardised area of the streambed (0.1–0.2 m^2) is disturbed by foot-kicking and dislodged invertebrates are collected with a D-net (mesh size 0.5 mm). All animals are combined into a single composited sample. Less consistency was found in dealing with soft-bottomed streams. Only six councils have been following standard invertebrate collection protocols for soft-bottomed streams, other councils did not have soft-bottomed streams or did not use separate soft-bottomed protocols (Davies-Colley et al*.* 2012).

All 15 councils have been identifying invertebrates to the taxonomic level for which MCI scores have been assigned (genus for most insects and molluscs, various levels for crustaceans, annelids and other phyla). However, about half of the councils have been using coded abundance methods while the rest have been using fixed counts or fully quantitative methods (see details in Davies-Colley et al*.* 2012).

3.2.3. Recommendations made for National Environmental Monitoring and Reporting (NEMaR)

Some of the recommendations for consistent future SoE monitoring in the NEMaR report (Davies-Colley et al*.* 2012) aligned with the practice already in use by the majority of the councils. However, a change was suggested with regards to the sampling method in the field; the semi-quantitative sampling should be substituted by a quantitative method. The following bullet points summarise the recommendations put forward (Davies-Colley et al*.* 2012):

- Annual sampling during late summer (targeting January to March) at a fixed set of wadeable sites
- Sampling to take place at least 2 weeks (but aim for 3 weeks) after a flood exceeding three times the long-term median flow
- For hard-bottomed streams (dominated by gravel, cobble, boulder and bedrock substrates, i.e. coverage of > 50%):
    - Use of standard sample collection protocol C3 (Stark et al*.* 2001) for quantitative sampling within riffle habitat (or run habitat if no riffles are available). At five different locations, animals are collected with a Surber sampler from a defined area (typically 0.1 m^2) of the streambed by disturbing the streambed up to a depth of 5–10 cm and collecting the dislodged animals (mesh size 0.5 mm). All animals are then combined into a single composite sample.
- For soft-bottomed streams (dominated by sand, silt, mud, clay, macrophytes and woody debris, i.e. coverage of > 50%):
    - Use of standard sample collection protocol C2 (Stark et al*.* 2001) for semi-quantitative sampling of a variety of stable substrates (e.g., bank margins, woody debris and macrophytes). The different habitats are sampled in proportion to their occurrence. At each of ten locations, an area of 0.3 m^2 is disturbed and animals are collected with cleaning

sweeps with a D-net or animals are dislodged by hand brushing into the net. All animals are combined into a single composite sample.

- Use of standard sample processing protocol P2 (Stark et al*.* 2001). A fixed count of 200 animals (i.e. a subset of the entire sample) is followed by a scan for rare taxa.
- Identification to the taxonomic level at which MCI scores had been assigned or higher
- Field and lab Quality Assurance/Quality Control (see details in the report).

Quantitative sampling was recommended by the NEMaR report because (1) Surber sampling is more likely to collect strongly attached and heavy taxa compared to kicknet sampling, providing a more accurate estimate of community composition, (2) Surber sampling, due to a highly standardised sampling area, provides a more precise estimate of richness measures and the MCI, and (3) it provides information on taxon densities (animals per m^2) that can be used to estimate the amount of invertebrate food supplied to fish and the level of invertebrate grazing as a control of periphyton biomass.

3.2.4. Existing SoE monitoring invertebrate data

In 2011, fifteen councils collected invertebrates as part of their SoE sampling programmes at a total of over 1000 sampling sites across the country. The majority of councils collected samples annually but two councils sampled twice a year. Semi-quantitative sampling, which was current practice by 13 councils in 2011 (Davies-Colley et al*.* 2012), provides information on the relative abundances of taxa. This type of information is sufficient to calculate the MCI and its semi-quantitative version, the SQMCI as well as total taxon richness, EPT richness, and %EPT richness. The former two MCI-type metrics have been recommended for SoE reporting by Davies-Colley et al. (2012).

Since 2011, Gisborne District Council have sampled invertebrates as part of the development of catchment management plans. Furthermore, several councils are currently in the process of reviewing their monitoring networks and sample collection protocols.

Notwithstanding the NEMaR report, semi-quantitative methods used by councils are considered sufficient for SoE monitoring because (J.D. Stark pers. comm.):

- D-nets are much more versatile in coping with different habitats than Surbers, and quicker to use
- Strongly attached taxa can be collected using a D-net by scrubbing rocks with a brush
- Given sufficient experience with the method, semi-quantitative processing is more cost-effective than 200 fixed count or full counts

- Not all samples have 200 taxa
- Precise estimates of taxon richness may not be that useful as an indicator of stream health
- Estimates of taxon densities may not be useful for indicating stream health because they can vary by several orders of magnitude in relation to flow stability
- The improved precision in MCI using a Surber sampler compared to a D-net is 1 MCI unit.

In this report, we will assess whether the data generated following the semi-quantitative protocols (current practice) will be sufficient for each of the invertebrate metrics reviewed.

3.3. Invertebrate metrics

3.3.1. Macroinvertebrate Community Index (MCI) and its variants

The Macroinvertebrate Community Index (MCI) and the Quantitative MCI (QMCI) were developed to assess organic enrichment in hard-bottomed streams by sampling in stony riffles, where the greatest variety of the most sensitive macroinvertebrates may be expected (Stark 1985, Stark 1993). The MCI is calculated from taxon-specific tolerance values (TVs) ranging from 1-10 that were assigned to individual taxa (at mainly genus level) using a weighting procedure. Weighting is based on the relative occurrence of taxa across three groups of sites differing in organic enrichment levels (*expected* not *measured* nutrient levels, determined from local or upstream agricultural activities). TVs for rare taxa (and those added subsequently) were assigned by professional judgement. A recent report was commissioned to explore the potential to reassign TVs based on an objective computational approach (Greenwood et al*.* 2015). On average new TVs were higher than the previous TVs resulting in significantly higher MCI site scores. As a consequence, Greenwood et al. (2015) recommended further research to develop regional TVs and management boundaries.

A more cost-effective variant of the QMCI, the SQMCI, was developed in 1998 (Stark 1998). More recently, Stark & Maxted (2004; 2007b) developed new biotic indices for assessing the health of soft-bottomed streams. These indices are analogous to the MCI, SQMCI and QMCI, and are denoted by the addition of “-sb” to the respective index names (i.e. MCI-sb, SQMCI-sb and QMCI-sb). These indices are designed to be used with samples collected according to the standard protocols (Stark et al*.* 2001).

The concept of an MCI index was derived from the United Kingdom’s BMWP Score System (BMWP 1978), with the MCI analogous to the ASPT (Average Score Per Taxon) variant of the BMWP Score System (Armitage et al*.* 1983). New Zealand appears to be the only country with qualitative, semi-quantitative and quantitative versions of the same biotic index, and different versions for hard-bottomed and soft-bottomed streams (Stark 1985; Stark 1993; Stark 1998; Stark & Maxted 2007a) and South Island wetlands too (Suren et al. 2010).

A guideline for the interpretation of MCI and QMCI and SQMCI site scores in stony streams is given below in Table 1. The index values corresponding to divisions between the four quality classes were selected initially by Stark (1998) based on professional judgement. However Stark and Maxted (2004, 2007a) used an objective procedure based on the statistical distribution of biotic index values at reference sites, together with an estimation of the lowest practical index value, to determine divisions between quality classes (see more details in Stark & Maxted 2007b). Wright-Stow and Winterbourn (2003) found inconsistencies in the classification of sites between the use of the MCI and QMCI, the former which seemed to generally be more conservative. The authors suggested different boundaries for these quality classes. There is some evidence that the interpretation given in Table 1 may require tweaking to apply in different regions of New Zealand (J.D. Stark pers. comm.). The same authors found that low-level identification (order, class, phylum) of fixed-count data ranked sites in a similar way to conventional taxonomic resolution (Wright-Stow & Winterbourn 2003) in contrast with Stark (1985) who found that indices based on family-level data for Taranaki streams could detect grossly enriched streams but could not distinguish between unenriched, and slightly to moderately enriched conditions.

Table 1. Thresholds for interpretation of MCI-type biotic indices.

Stark & Maxted (2004, 2007a) Quality Class	Stark (1998) description	MCI MCI-sb	QMCI & SQMCI QMCI-sb & SQMCI-sb
Excellent	Clean water	>119	>5.99
Good	Doubtful quality or possible mild pollution	100-119	5.00-5.98
Fair	Probable moderate pollution	80-99	4.00-4.99
Poor	Probable severe pollution	<80	<4

The MCI and its variants are the only comprehensive biotic indices based on the pollution tolerance of New Zealand macroinvertebrate taxa for assessing stream ecosystem health (Boothroyd & Stark 2000). Consequently, all regional councils that undertake SoE monitoring use the MCI and/or SQMCI/QMCI for reporting purposes. The MCI (and variants) is recognised by councils as a good metric for assessing ecosystem health with respect to nutrient enrichment and organic pollution, (for which it was developed), as well as sedimentation (Stark & Maxted 2007b). The scientific literature supports the ability of the MCI to respond to these instream stressors (e.g. Quinn & Hickey 1990; Niyogi et al. 2007; Wagenhoff et al. 2011; Lange et al. 2014; Piggott et al. 2015) or to catchment land-use change (e.g. Collier 1995; Quinn et al. 1997; Death & Collier 2010; Clapcott et al. 2012; Lange et al. 2014). Lack of response of the QMCI to heavy metals in a study by Hickey and Clements (1998) is a reminder that biotic indices have been designed for certain stressors and may not be able to detect the effects on invertebrate communities of all types of stressors. Hence, a single biotic index is unlikely to be able to accurately assess stream ecosystem health if stressors other than those that were used for development are present.

Furthermore, while a biotic index like the MCI (as well as other single metrics) responds to a suite of stressors, there is an inevitable loss of information compared with the raw data (taxon list and abundances) from which index values are derived (Boothroyd & Stark 2000). Hence, biotic indices should not be the sole means of analysing or depicting biological data (Boothroyd & Stark 2000).

Two reports have been commissioned to inform discussion on the inclusion of the MCI as an attribute for ecosystem health under the National Objectives Framework (Clapcott & Goodwin 2014; Collier et al*.* 2014). The MCI has not been accepted for the current NPS-FM because it responds to a suite of stressors and as yet values have not been translated into catchment loads of pollutants such as nutrients and fine sediment. However, since the MCI (and variants) can detect the effects on invertebrate communities of excessive nutrients, algal blooms and high sediment levels as either a result of runoff or changes in flow regime (e.g. Booker et al*.* 2014), it is a metric that could inform the compulsory national value Te Hauora o te Wai (ecosystem health) contained in the NPS-FM (see Section 1.2 for description of the value). It is also highly correlated with the Cultural Health Index that assesses freshwater habitats from a Māori perspective (Townsend et al*.* 2004; Harmsworth et al*.* 2011).

The MCI and QMCI have also been suggested as surrogate tools to assess the adverse effects of nutrient enrichment on higher trophic levels (e.g. fish). However, the interpretation of the MCI/QMCI regarding the quality of the food resources for fish has shown to be ambiguous (Hayes 2014). More specifically, the MCI/QMCI or EPT metrics do not consider food web linkages (energy transfer) between invertebrates and fish, nor is there any science underpinning a relationship between these indices and fish. They were not designed to be used in this manner. A major problem with using the MCI and QMCI to infer effects on higher trophic levels is that they do not incorporate fundamental information on invertebrate size and availability as prey to fish (Hayes 2014).

3.3.2. EPT metrics

In general, the three aquatic insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies), and Trichoptera (caddis flies), known collectively as EPT taxa, are sensitive to the pollution or degradation of stream ecosystems. For this reason, EPT metrics calculated from an invertebrate taxon list are commonly used in New Zealand and worldwide to assess stream ecosystem health. The typical EPT metrics for SoE reporting are:

- EPT richness (number of EPT taxa)
- % EPT richness (percentage of taxa belonging to the EPT orders)
- % EPT abundance (percentage of individuals belonging to the EPT orders).

The last requires information on the relative abundances of the taxa whereas the former two require only knowledge on the presence/absence of the taxa. In New Zealand, the caddis fly family Hydroptilidae is less sensitive to pollution, mainly because of its strong association with high algal biomass (often an indicator of nutrient enrichment). Hence, genera (*viz*. *Oxyethira* and *Paroxyethira*) in this family are sometimes excluded from EPT metric calculations.

High EPT metric values generally indicate clean water and undisturbed, structurally complex invertebrate habitat. In turn, EPT metrics have been shown to respond negatively to impairment of water and habitat quality (instream stressors) in New Zealand, such as nutrient pollution, elevated water temperature, flow reduction and sedimentation (e.g. Quinn & Hickey 1990; Matthaei et al. 2006; Niyogi et al. 2007; Townsend et al. 2008; Matthaei et al. 2010; Piggott et al. 2015). Values decline in response to catchment land-use intensification (e.g. Matthaei et al. 2006; Death & Collier 2010; Clapcott et al. 2012; Lange et al. 2014), although EPT metrics have also shown to initially respond positively to nutrient enrichment (Wagenhoff et al. 2011; Wagenhoff et al. 2012).

Since EPT metrics can detect the effects on invertebrate communities of pollutants or changes in flow regime (e.g. water abstraction), they are metrics that can at least partially inform the compulsory national value Te Hauora o te Wai. Like the MCI, EPT metrics may not respond to all types of stressors and therefore may not accurately assess stream ecosystem health on their own. EPT taxon richness is furthermore a measure of the diversity of indigenous fauna, which also defines a healthy ecosystem according to the NPS-FM. High taxonomic richness indicates a diverse habitat. And highly diverse ecosystems may be desirable because they can be more resilient to environmental disturbance and support a broader range of ecosystem functions (Elmqvist et al. 2003). Both ‘resilience’ and ‘maintenance of ecological processes’ are characteristics of healthy ecosystem according to the NPS-FM.

Large EPT taxa are also preferred prey items for drift-feeding fish and because they are prone to drifting in the water column, thus EPT metrics are considered a coarse indicator of food quality for drift-feeding fish. However, more specific invertebrate food quality indicators for fish, based on foraging strategies (see Trout Prey Index described in Section 5.4), are likely to be more appropriate for assessment of fisheries quality within the NPS-FM framework.

3.3.3. Taxon richness

Taxon richness is a metric potentially highly dependent on sample area size (area and depth of stream bed sampled), the diversity of habitats sampled, antecedent flow, as well as the number of individuals identified and the taxonomic level of identification. Hence, consistency in collection and processing is crucial for comparison across sites or years. The best way to obtain comparable taxon richness across sites or samples is to use rarefaction to estimate taxon richness for a given number of individuals counted (Heck et al*.* 1975).

Taxon richness is commonly reported by councils to complement the MCI and EPT metrics in stream health assessment. While taxon richness could be considered as a direct measure of the diversity of indigenous fauna, it is a less reliable metric than EPT richness. Taxon richness often poorly indicates perturbation as it combines all taxa, including those that become more prevalent under perturbed conditions, while EPT taxa are generally pollution-sensitive. In summary, although commonly calculated, taxon richness is considered a very coarse measure of the state of invertebrate communities. It is not recommended that this metric is interpreted singly when assessing stream ecosystem health. Taxon richness is also unlikely to be a suitable metric for fisheries quality assessment.

3.3.4. Average Score Per Metric (ASPM)

Collier (2008) introduced an aggregation method for assessing wadeable stream ecosystem health considering the relative responses of core metrics. The Average Score Per Metric (ASPM) is composed of three individual metrics, the MCI, EPT richness and % EPT abundance (both excluding Hydroptilidae), which were selected from a suite of 17 candidate metrics for their ability to discriminate between reference sites and sites influenced by urbanisation or high levels of pastoral development in the Waikato region. Aggregation involved standardisation of the observed metric values by their respective observed maximum value. The mean of these standardised metrics was used to calculate the ASPM.

Five condition bands for a tiered assessment (ranging from ‘very high’ to ‘very low’) were defined based on calculations for each individual metric as follows: the lower standard deviation of the metric values of ‘reference’ samples was calculated to distinguish the ‘very high’ band. Metric quartiles between this point and a hypothetical worst-case value were calculated to distinguish between the other bands. Compared with its component metrics, the ASPM distinguished reference conditions and low-moderate levels of catchment modification and local habitat degradation more accurately and showed lower temporal variability within reference sites (Collier 2008).

The ASPM is currently only used by Waikato Regional Council. However, the aggregation approach could be equally applied in other regions or at a national level. The ASPM has, equally to its component metrics (MCI, EPT richness and % EPT), the ability to inform the national value Te Hauora o te Wai contained in the NPS-FM but is likely to be less suitable to inform the fishing value.

3.3.5. Invertebrate density and biomass

Invertebrate density and biomass require quantitative sample collection (Barbour et al. 1999; Carter & Resh 2013). Invertebrate density is the number of individuals per square metre of stream bed sampled. Biomass is calculated by multiplying density (No. / m^2) by total dry weight (mg) of the animals. Total dry weight can be determined by drying and weighing all individuals within a sample (or doing this for individual taxa separately and calculating the sum). However, this process is time-consuming and destroys the animals. Alternatively, known size-specific dry weights for invertebrate taxa can be used to estimate biomass. For example, size-specific dry weights, based on relationships between invertebrate body length and dry weight, from the scientific literature (e.g. Towers et al. 1994) and from unpublished data (K. Shearer, Cawthron Institute), can be used to efficiently calculate biomass from densities and invertebrate dry weights. Compared to the former method involving dry weighting (hence destruction) of the invertebrates in a sample, the latter method requires only taxonomic identification and measurement of body sizes (which can be done in size classes, e.g. 1-mm or 3-mm size class categories).

Invertebrate density is a broad metric for measuring the state of an invertebrate community under the assumption that a healthy environment supports abundant aquatic life. However, streams receiving nutrients from anthropogenic sources often have higher invertebrate densities, at least at intermediate enrichment levels, than streams with naturally low nutrient concentrations (e.g. Wagenhoff et al. 2012). Sedimentation has also shown to initially increase invertebrate density due to proliferation of primarily pollution-tolerant taxa such as oligochaetes and nematodes (Wagenhoff et al. 2012). For these reasons, invertebrate density (or biomass) is on its own not a suitable metric to assess Te Hauora o te Wai. Current practice of semi-quantitative sampling for SoE reporting prevents calculation of invertebrate density.

However, consent monitoring often involves quantitative sampling and in some cases, invertebrate density can be a useful indicator. For example, sudden changes in density between sites immediately up and downstream of a discharge (and in the absence of any flow-varying factor) can indicate an environmental impact. Hence this metric finds application in the assessment of environmental effects of point discharges. Bear in mind that density at a site can be highly variable over time and over space in a heterogeneous environment, mostly due to preceding flow events (floods and droughts) and to heterogeneity in habitat and availability of food resources.

Total invertebrate density, and to a lesser extent biomass, have been used to assess the potential productivity of a trout fishery in a general sense (e.g. Shearer et al. 2007; Shearer & Hayes 2010). The underlying assumption is that high density and biomass is good, and low is bad. However, total density and biomass do not account for other aspects of the invertebrate community that are important for supporting a good fishery, such as size structure. As fish increase in size, they require larger invertebrate prey to grow or maintain their condition. For example, a river may have a high invertebrate density (such as 10,000 / m^2), but if the community comprises mainly small animals the larger fish are likely to expend more energy capturing small prey than they can gain from the food. This may prevent fish growing to an optimum size. It becomes clear that including analysis of the size structure of invertebrate communities and density can provide a clearer picture of the quality and quantity of the food resources for fish populations. Another important aspect is the functional identity of species, e.g. drifting taxa. For example, an ideal invertebrate community for supporting a good trout fishery will have a comparatively large proportion of large, drifting individuals. Due to the limited ability of invertebrate density to indicate trout fisheries quality (part of national value Mahinga kai (fishing) contained in the NPS-FM, see Section 1.2), the Trout Prey Index, which incorporates knowledge of taxon and hence functional identity and size structure, is currently in the process of development (see Section 5.4).

4. INTERNATIONAL APPROACHES TO ASSESSING STREAM HEALTH

4.1. Overview

Stream benthic invertebrate monitoring has a long tradition in overseas industrialised countries reaching back to the early 1900s (Bonada et al. 2006). From the beginning of the 1970s, invertebrate monitoring became the most commonly used tool to assess stream ecosystem health, which was increasingly seen as an important human value (Bonada et al. 2006). For example, the Biological Monitoring Working Party (BMWP) Score System, which was set up by the Department of the Environment (United Kingdom) in 1976, recommended the development of a score system based on benthic macroinvertebrates. Trial and reassessment over many years resulted in the development of an Average Score Per Taxon metric which has been used to report stream health in the United Kingdom since the 1990 National Biological Survey (Hawkes 1997).

Driven by the European Union's Water Framework Directive 2000/60/EC (EU WFD 2000) and the Clean Water Act (CWA 1972) of the United States, most European countries and the states of the USA have developed integrative and sophisticated systems to assess stream health using benthic invertebrates. While the EU WFD prescribes certain broader characteristics for development of an assessment system (see section 4.2.1), member states were left to decide upon more specific aspects such as the metrics to be used or the approach to quality class boundary setting (Birk et al. 2012). Consequently, this resulted in a vast range of systems since many countries tried to adapt their existing systems (reviewed in Birk et al. 2012).

In this section, we present the basics of the German, as an EU example, and the United States assessment systems, including a review on the invertebrate metrics used. Australia primarily uses a multivariate approach to assessment (reviewed in Section 5) but, like New Zealand, also has developed a general stream health index. Many other countries have developed approaches to assessing stream health but they are not reviewed here; in general systems and metrics used internationally are captured in detail by the European Union, United States and Australian examples.

We furthermore searched the literature for invertebrate metrics potentially able to inform the national values Te Hauora o te Wai and Fishing. Among these, we found several scientific papers presenting the development of pressure-specific invertebrate metrics and some testing of their performance (reviewed in Section 5.2) but it appears that they largely are not used in the practice of stream assessment. This may be due to their development after stream assessment systems, including core metric selection, had been established.

4.2. Metrics used in Germany to fulfil the requirements of the EU WFD 2000

4.2.1. WFD requirements for assessment of the stream ecological status

The European Union Water Framework Directive (EU WFD 2000) requires member states to assess streams and rivers based on ‘Biological Quality Elements’ (e.g. invertebrates, among others) by comparing the observed status with stream-type specific reference conditions. Assessment criteria should cover the different aspects of biological elements, composition and abundance, the share of sensitive/tolerant taxa, and the degree of diversity as well as functional aspects. Furthermore, assessment methods should be capable of identifying the cause of degradation which requires stressor-specific approaches. A final assessment of the ecological status involves classification into one of five ‘Ecological Quality Classes’: ‘high’, ‘good’, ‘moderate’, ‘poor’ or ‘bad’. The WFD demands a ‘good ecological status’ by 2015 and any assessment of less than ‘good’ requires management action. The stressor-specific approach to assessment should be able to discriminate between the impacts of individual perturbations and hence help guide restoration measures. Recommended sample collection and processing protocols are very intensive and include 20 Surber samples per site across multiple habitats with possible sub-sampling and full count of the composite sample.

4.2.2. Overview of the German assessment system

Modular structure

The German assessment system encompasses three modules aimed at assessing the impact of three different types of anthropogenic stressors: (1) organic pollution evaluated with the German Saprobic Index, (2) general degradation using a multi-metric index, and (3) acidification for two stream types (Meier et al*.* 2006). The result of each module is a classification of the stream into one of five quality classes. These results are combined to lead to the final assessment of the overall ecological status. All metrics are calculated from a taxa list containing semi-quantitative information on benthic invertebrates.

The modular structure of the German assessment system allows assessment of the output on different levels (Meier et al*.* 2006):

- Level 1: Ecological Quality Class
- Level 2: Causes of degradation (organic pollution, acidification, general degradation)
- Level 3: Results of the Core Metrics (useful for data interpretation purposes)
- Level 4: Results of all metrics, including those, which have not been used for the multi-metric Index

A multi-metric index for general degradation

A multi-metric index combines the results from several individual metrics (e.g. diversity indices and percentages of certain taxa or traits) and therefore has the potential to provide a more complete picture of the ecological status than a single metric. The multi-metric index to assess the impact of general degradation (second module) is composed of a set of 'Core Metrics' chosen from an initial list including 76 metrics. The core metrics reflect hydro-morphological degradation, stagnation, residual flow, land use in the catchment, pesticides, hormone-equivalent substances; however, hydro-morphological degradation is the usually most important stressor. This set of metrics meets the WFD requirements to include measures of

1. richness/ diversity,
2. composition/ abundance,
3. tolerance/ intolerance, and
4. functional aspects.

The number and individual set of core metrics differs for each group of stream types (selection described below) and was chosen to obtain a robust result and enable profound data interpretation with respect to the potential effects of different stressors. See Table 2 for the metrics that were chosen for at least one group of stream types.

Table 2. Core metrics for assessment of the general degradation of streams and rivers in Germany developed for WFD purposes (Meier et al. 2006). Each metric was chosen for at least one group of stream types. All metrics expressed as percentages were calculated from abundances. # = number of; E=Ephemeroptera, P=Plecoptera, T=Trichoptera, C= Coleoptera, B=Bivalvia, O= Odonata; LTI_{quan}=Lake outlet typology index (quantitative).

Richness/ diversity	Composition/ abundance	Tolerance	Functional aspects
# Trichoptera # EPT taxa # EPTCBO	% EPT	German Fauna Index (separate for various stream types) LTI_{quan} Potamon-Typie-Index % Oligosaprobes	% Epirhithral % Metarhithral % Hyporhithral % Litoral % Epipotamal % Metapotamal Rheoindex after Banning % Pelal % Phytal

Selection process for the core metrics for general degradation

Abiotic data on habitat quality and land use were used to select the core metrics for general degradation via a multi-step selection process. As a result, all core metrics are likely to have some sort of a relationship with habitat quality and land use. Firstly the degradation gradient was determined. Multivariate analyses were used to select parameters out of a whole set of available individual abiotic parameters (> 29 for habitat quality and several for land use), that spanned a wide gradient within the data set (i.e. covered the entire spectrum from natural to very degraded hydro-morphological conditions) and that have at least a theoretical relationship with invertebrate communities. These analyses were performed separately for the habitat quality and the land-use data sets and also for the different groups of similar stream types. The urban, crop, pasture and forest land uses were selected and another land-use index was created combining the former three into a single index[1] for further analysis. Thirteen habitat quality parameters were selected for at least one group of stream type including characteristics such as variance in depths or widths, flow or substrate diversity in current velocities, woody debris structures etc. Similarly to the land-use index, a single structural index was created from the selected habitat quality parameters.

The second major step involved the actual selection of core metrics, which was based on bivariate correlation analyses between the selected abiotic parameters and the invertebrate metrics initially considered. This selection process itself involved multiple steps but the final product was a set of core metrics for each group of stream types including at least one metric per metric 'type' (Table 2). Below we review the most important invertebrate metrics developed for the German assessment system (Meier et al*.* 2006).

4.2.3. German Saprobic Index

The German Saprobic Index can be classified as a 'tolerance metric'. The Saprobic Index assesses the impacts of organic pollution, based on an indicator taxa list taking into account the relative abundances of and a weighting factor for each taxon. The higher the organic pollution, the higher the rate of degradation processes, which are associated with a lowering of the oxygen content in the water. Hence, the saprobic indicator values for a total of 741 taxa have been assigned according to the autecological knowledge of the oxygen requirements. The weighting factors have been assigned according to the specificity of each taxon for its indicator value. The Saprobic Index values increase with increasing organic pollution but may also positively respond to other perturbations (e.g. sedimentation) and negatively respond to acidification.

Development of the Saprobic Index was based on a large data set (1621 sampling sites) spanning a wide gradient in organic pollution for the various stream types. Data were collected by regional authorities, universities and private companies to meet certain criteria, e.g. semi-quantitative sampling to determine the relative or coded abundances of taxa, identification to species level where possible.

Reference conditions for each stream type had been calculated as the mean value of the 10th percentile of all data minus two times the standard deviation. Several stream

[1] Index = 4x urban + 2x crop + 1x pasture (Meier et al. 2006)

types with similar reference Saprobic Index values were then combined into groups. The results of the Saprobic Index are transferred into 'Saprobic Quality Classes' by applying stream type-specific boundaries. The boundaries were defined based on deviation from reference conditions and are as follows:

- 'high' for ≤ 5%
- 'good' for > 5% to ≤ 25%
- 'moderate' for >25% to ≤ 50%
- 'poor' for > 50% to ≤ 75%
- 'bad' for > 75% deviation from reference condition.

4.2.4. German Fauna Index

The German Fauna Index can be classified as a 'tolerance metric'. The Fauna Index (Lorenz et al. 2004) assesses the impacts of hydro-morphological degradation based on stream-type specific indicator taxa lists taking into account the relative abundances of each taxon. The indicator values have been assigned based on the association of invertebrate taxa to their habitat. Habitat here is defined by its substrate as well as current velocity and they can be characteristic of the natural condition or hydro-morphological degradation depending on the stream type. Hence, the Fauna Index is calculated from stream-type specific indicator taxa lists.

Among the list of applied tolerance metrics for assessment of general degradation (Table 2), the Fauna Index has shown to be specifically useful according to a large-scale practical test and hence will be described in more detail.

An indicator value can be either positive (value of 1 or 2) or negative (value of -1 or -2), respectively, representing the natural or degraded condition of a specific hydro-morphological habitat or structural element. Indicator values were assigned based on statistically proven associations, habitat preferences noted in the literature or based on expert knowledge, or rarity as an indication for high sensitivity. The values of 1/ -1 or 2/ -2 represent the strength of association with the habitat or element (statistically determined or expert knowledge), or the degree of rarity.

A high value indicates a high proportion of taxa requiring high quality habitat specific to a stream type. Due to the development of the indicator values according to habitat quality, the Fauna Index negatively responds to hydro-morphological degradation. However, since most of the taxa requiring high quality habitat are also sensitive to other perturbations (in particular organic pollution), the Fauna Index also negatively responds to these.

In contrast to the Saprobic Index, the Fauna Index can be affected by many factors related to habitat quality and their individual effects can often not be disentangled. The factors can originate from different spatial scales, ranging from the habitat scale (e.g. lack of microhabitats) to catchment scale (e.g. sedimentation from intensive land use).

More specifically, hydro-morphological degradation was defined as:

- Lack or rare occurrence of habitats characteristic of the natural condition (e.g. woody debris, gravel in lowland rivers)
- Presence or unusual occurrence of habitats characteristic of hydro-morphological degradation (e.g. stop banking, moving sands)
- Change in hydraulics (e.g. backed-up water levels, residual water)
- Hydraulic homogenisation (e.g. lack of riffles or littoral low-flow habitat)
- Lack of riparian vegetation and hence stronger exposure to sun radiation
- Increased sedimentation.

Development of the Fauna Index (and all metrics for the second module) were based on a large data set comprising 3,000 invertebrate sampling occasions for all of which supplementary abiotic data on habitat quality (determined using standardised habitat assessment protocols) and/or land use was also available. Certain other criteria for inclusion had to be met, such as sampling of all habitats (no Surber sampling), spring/summer sampling, knowledge of relative or coded abundances of the taxa, identification to an appropriate taxonomic level, and the absence of confounding effects from organic pollution (i.e. saprobic quality classification as ‘high’ or ‘good’).

The theoretical values of the Fauna Index range between -2 and 2. In order to make the Fauna Index comparable to other metrics used in the multi-metric index for general degradation, the scales of all metrics were standardised to range between 0 (lowest value) to 1 (highest value), called upper and lower anchor points, for each group of stream type. Due to the high performance of the Fauna Index to indicate hydro-morphological degradation, its index value is weighted more highly (50%) for calculation of the multi-metric index value than the values of the other core metrics. For each of the metrics as well as the multi-metric index, the boundaries for classification into a quality classes were defined as follows:

- ‘high’ for > 0.8
- ‘good’ for index values > 0.6 to 0.8
- ‘moderate’ for > 0.4 to 0.6
- ‘poor’ for > 0.2 to 0.4
- ‘bad’ for ≤ 0.2.

4.2.5. EPT and related metrics

EPT metrics have been described in Section 3.3.2. The German assessment system uses %EPT abundance and classified it as a ‘composition/ abundance’ metric. EPT richness is classified as a ‘richness/diversity’ metric and has been noted to be one of the most reliable metrics and to be very sensitive to water quality changes (in particular to toxic impacts) as well as sensitive to habitat degradation (decreasing).

EPTCBO richness (number of taxa belonging to EPT, Coleoptera, Bivalvia or Odonata), as with EPT richness, is more sensitive to perturbations and less variable with respect to different flow magnitudes across years than taxon richness. EPTCBO taxa generally require very good habitat quality (aquatic and terrestrial), hence high values indicate undisturbed, structurally complex streams with a high diversity of taxa and microhabitats.

Trichoptera richness can be informative as many Trichoptera taxa require relatively good habitat, in particular with respect to terrestrial and bankside structures, and some require woody debris as a food resource or coarse particular organic matter for construction of their cases. Hence, high Trichoptera richness indicates undisturbed, structurally complex streams with a high diversity of taxa and microhabitats. As all other EPT metrics, Trichoptera richness decreases with increasing perturbation. Note that any richness index unless expressed as a percentage of total richness will also be highly dependent on sample size/sampling effort.

4.2.6. Lake Outlet Typology Index, quantitative (LTI_{quan})

The LTI_{quan} is a ‘tolerance’ metric based on a list of indicator taxa typical of invertebrate communities at lake outlets. Indicator values of the taxa were assigned based on lake outlet preferences and characteristics and taxa also have different weighting factors. The quantitative version takes into account the relative abundances of the taxa. The local factors (flow, substrate etc.) seem to be more important than the physical characteristics or trophic state of the lake.

4.2.7. % Oligosaprobe

Percent Oligosaprobe is a ‘tolerance’ metric describing the percentage of individuals that prefer oligosaprobe conditions. These organisms predominantly require high oxygen concentrations and can live with few food resources but also require high quality habitat. Hence, high values indicate low organic pollution, good oxygen levels and high-quality habitat. The metric generally responds negatively to increasing perturbation with the exceptions of acidification and some toxic impacts.

4.2.8. Potamon-Typie-Index (PTI)

The PTI is a ‘tolerance’ metric assessing the naturalness of invertebrate communities of large rivers based on an indicator taxa list.

4.2.9. Rheoindex after Banning

The Rheoindex after Banning describes the ratio of taxa preferring flowing water to ubiquitous stream taxa or still-water taxa, taking account of their relative abundances. A value of close to 1 indicates a community composed of taxa with preferences for flowing water whereas a value of close to 0 indicates a community composed of still water and ubiquitous taxa. Most of the taxa preferring flowing water require high water quality and oxygen levels and are associated with course substrates, hence the Rheoindex responds negatively to pollution and changes to the substrate. However, channelising may positively affect the index values.

4.2.10. Functional metrics

Various functional metrics describe the percentage of individuals that have preferences for a specific zone along the longitudinal gradient of the river (i.e. Metapotamal = connected backwaters, Epirhithral = upper trout region, Hyporhithral = grayling region). The theoretical background for a strong longitudinal abiotic and biotic gradient is described by the 'river continuum concept' (Vannote et al. 1980). Typically only one of these metrics is applied for each stream type for assessment purposes. The taxa have been assigned based on the autecological information, in particular with respect to their feeding preferences.

The percentage of individuals with preferences for littoral habitats is used as a functional metric for lowland rivers while the percentage of individuals with preferences for macrophytes (Phytal) is used for lake-fed systems. Finally, the percentage of individuals with preferences for fine sediment habitat (Pelal) is used for stream types with naturally high percentages of fines in the substrate.

Generally, these functional metrics respond to perturbations such as organic pollution, changes to the trophic state or habitat but can sometimes be confounded by acidification.

4.3. Metrics used by states and tribes in the United States

4.3.1. Overview of biological assessment in the United States

Bioassessment using benthic macroinvertebrates is done for various reasons in the United States (Barbour et al. 1999; USEPA 2011). The most important purpose listed by each state agency, however, relates to specific requirements of the Clean Water Act (CWA) (Carter & Resh 2013), with the overall goal to restore and maintain the biological integrity of the nation's waters. For example, the CWA requires state agencies (and tribes to a certain extent) to identify water quality problems and evaluate the effectiveness of point and nonpoint source water quality controls. Bioassessment plays an important part in these activities as biological information is the most effective means of evaluating cumulative impacts from nonpoint sources and provides an understandable endpoint of relevance to society (Barbour et al. 1999). Bioassessment can therefore be used to determine the status of the water resource with respect to whether 'designated aquatic life uses' are being met ('attainment' or 'non-attainment' = impairment), to evaluate the causes of degraded water resources and relative contributions of pollution sources, and to determine the effectiveness of control and mitigation programs (Barbour et al. 1999).

Bioassessment is, as required by the European Union WFD, based on reference condition specific to the stream type, and is typically done using a multi-metric approach (i.e. with a suite of invertebrate metrics). The multi-metric approach to bioassessment in the United States involves selection of core metrics that are sensitive to pollution and most informative of the ecological relationships of the assemblage to specific stressors or cumulative impacts within the state or region (Barbour et al. 1999). The core metrics can be aggregated into a multi-metric index (after transformation to unitless metrics) or retained as individual metrics. The multi-metric index simplifies management in that a single index value is used to determine whether action is needed. However, analysis of the individual metrics and the raw data along with other ecological data is needed for impairment diagnostics and determination of the specific nature of action needed.

Also similar to the requirement by the European WFD, the core metrics should be representative of the following four categories:

1. richness measures for diversity or variety of the assemblage
2. composition measures for identity and dominance
3. tolerance measures representative of the sensitivity to perturbation
4. trophic or habit measures for information on feeding strategies and guilds (Barbour et al. 1999).

The most widely-used metrics across state programs are summarised in Table 3, however, there were many other metrics included by some programs when considered to be among the top three most informative metrics (Carter & Resh 2013).

Most monitoring programs in the U.S. calculate a few up to more than 100 different metrics (20 on average) but only a small set of metrics was considered to be useful across all States, most of these being based on richness or composition (Carter & Resh 2013). For example, total taxon richness and EPT richness were the metrics that featured most often within the top three metrics listed by each State in terms of their usefulness followed by the tolerance-type HBI (Carter & Resh 2013). Below we briefly describe those metrics not previously discussed.

Table 3. Most widely used metrics across state bioassessment programs in the USA. Only those metrics are listed that were included in the core metrics of at least five (out of 47) State programs (Carter & Resh 2013). Each metric is assigned to one of four metric types. All metrics expressed as percentages were calculated from abundances. # = number of; E=Ephemeroptera, P=Plecoptera, T=Trichoptera.

Richness	Composition	Tolerance	Trophic/habit
# Taxa	% EPT	Hilsenhoff Biotic Index (HBI)	# Scraper taxa
# EPT taxa	% Ephemeroptera		# Clinger taxa
# Ephemeroptera	% Chironomidae	% Dominant Taxon	% Scrapers
# Plecoptera	% Non-insects		% Clingers
# Trichoptera			
Shannon-Wiener Index of diversity			

4.3.2. Hilsenhoff Biotic Index (HBI)

The HBI uses tolerance values of taxa to weight abundance. While the tolerance values were assigned to evaluate organic pollution, the HBI appears to be responsive to overall pollution. The HBI is an equivalent of the MCI (Death & Collier 2010).

4.3.3. % Dominant taxon

The percentage of the dominant taxon is a simple measure of redundancy. A high level of redundancy equates to the dominance of a pollution-tolerant organism and to low diversity (Barbour et al*.* 1999).

4.3.4. Trophic/habit metrics

Trophic metrics based on functional feeding groups are surrogates of complex biological processes such as species interactions and production, and also reflect the availability of food sources. Generally, specialised feeders such as scrapers are more sensitive to pollution than generalists such as filterers but the usefulness of functional feeding groups has not been well demonstrated (Barbour et al*.* 1999). Habit metrics describe the prevalence of a movement or positioning mechanism, such as the number of clinger taxa or the percentage of individuals belonging to a clinger taxon.

4.4. The Australian assessment system

4.4.1. The Australian River Assessment System (AUSRIVAS)

Australian water management authorities primarily use the Australian River Assessment System (AUSRIVAS), which falls under multivariate approaches to stream health assessment in contrast to the multi-metric approaches used by Germany and the United States. AUSRIVAS and other multivariate approaches will be described in Section 5.1. However, some simple metrics are also used for rapid health assessment in Australia (see below).

4.4.2. Stream Invertebrate Grade Number – Average Level (SIGNAL) and taxon richness

SIGNAL stands for Stream Invertebrate Grade Number—Average Level and is a simple scoring system for benthic invertebrates in Australian rivers, SIGNAL 2 being the most recent scoring system (Chessman 2003). Most taxa have been assigned a sensitivity grade number (sensitivity value) based on the sensitivity to common forms of water pollution, akin to the tolerance values for the New Zealand specific MCI. Hence, SIGNAL 2 scores respond to a variety of stressors including salinity, turbidity, nitrogen and phosphorus. When interpreted together with taxon richness, SIGNAL 2 scores can provide indications of the broad types of pollution or degradation. SIGNAL 2 scores can be calculated with or without coded abundance weighting and a score can be calculated using the order-class-phylum or the family version.

5. CLASSES OF INVERTEBRATE METRICS

5.1. Multivariate approaches to assessment

Similar to a bioassessment using one or several metrics, multivariate approaches assess human impacts by comparing patterns observed at a site with that expected in the absence of human impact, i.e. the reference condition (Bonada et al. 2006). In contrast to the metric approaches reviewed above, the reference condition (expected biological pattern) is predicted using statistical analyses instead of using the range of patterns observed at reference sites. Furthermore, multivariate approaches assess changes in overall community composition based on taxon identity, from which metrics can be calculated. In essence, expected taxa occurrence in the absence of human impact is predicted from environmental conditions using a model developed using reference sites. Comparison of observed (O) versus expected (E) can be expressed as an O/E value of a certain metric. Typically O/E taxon richness is calculated but potentially any metric could be used. Development of the first multivariate approach of this kind began in October 1977 (Wright 1994) in the United Kingdom, and was called the River Invertebrate Prediction and Classification System (RIVPACS) (Wright et al. 2000). Equivalent approaches are now used widely in Australia, Canada, and by some agencies in the United States (Carter & Resh 2013). A few New Zealand studies have also trialled a multivariate approach.

Here, we provide a brief description of the Australian River Assessment System (AUSRIVAS) and a trial of adopting such a system for stream health assessment in New Zealand.

5.1.1. AUSRIVAS

The Australian River Assessment System (AusRivAs) is a predictive 'reference condition approach' model developed to assess the biological health of Australian rivers (eWater CRC 2012). AusRivAs models have been developed for each state and territory (e.g. Smith et al. 1999; Turak et al. 1999), for the main habitat types that can be found in Australian river systems. These habitats include the edge/backwater, main channel, riffle, pool and macrophyte stands. AusRivAs models have been used to report on the national health of Australian waterbodies (Harrison et al. 2011).

Like its overseas counterparts (e.g. RIVPACS), AusRivAs models match test sites to an existing set of reference sites using habitat and environmental features of the sites. The models can be constructed for a single season or using data from several seasons to provide more robust predictions. Taxa likely to occur at test sites are predicted using the reference data ("E" for "expected") and compared to those collected ("O" for "observed"), the ratio providing an assessment of degradation (O/E score) (Table 4). This provides an independent method of comparison, avoiding problems of upstream/downstream comparisons and comparisons to predetermined thresholds that may be unrealistic to expect for many sites. The AusRivAs model performance is poor where reference sites are problematic or lacking (Chessman et al. 2010). However, the approach may be less sensitive to climate change effects than other indices (Nichols et al. 2010).

Table 4 AusRivAs scores and interpretation (adapted from Gray 2004).

O/E score	Band Description	Aquatic Macroinvertebrate Biodiversity Status	Interpretation guide
>1.12	More biologically diverse than reference sites (needs detailed investigation)	Over 112% more biodiversity than reference sites	More macroinvertebrates found than expected Potential biodiversity “hot spot” Possible mild nutrient enrichment
Approx.1.0	Reference Site	Natural or near natural levels of biodiversity	Reference site used in the construction of the AusRivAs model Undisturbed / natural / pristine or least disturbed / impacted
0.85–1.15	Reference condition	Similar levels of biodiversity to reference sites	Most or all of the expected families were found at site Water quality and / or habitat condition roughly equivalent to reference sites Impact on water quality and habitat condition does not result in loss of biodiversity
0.55–0.84	Significantly impaired	Approx. 16% to 45% of macroinvertebrate biodiversity has been lost	Several expected families found Water quality and / or habitat condition significantly impaired Significant loss of biodiversity
0.25–0.54	Severely impaired	Approx. 46% and 75% of macroinvertebrate biodiversity has been lost	Many expected families not found Water quality and / or habitat condition severely impaired Sever loss of biodiversity
0–0.24	Extremely impaired	Approximately 75% to 100% of macroinvertebrate biodiversity has been lost	Extremely few of the expected macroinvertebrate families found Extremely poor water and / habitat quality Extreme loss of macroinvertebrate biodiversity Highly degraded

5.1.2. Trials of AusRivAs in New Zealand

AusRivAs was trialled in the Waikato region of New Zealand in response to a recommendation from the New Zealand Macroinvertebrate Working Group appointed by the Ministry for the Environment in 1997 (Coysh & Norris 1999). Two predictive models were constructed, one at the family level, the other at the generic level. The genus model outperformed the family model in most model diagnostics, including a very low error rate (22%). Habitat predictor variables used in the model to group reference sites were altitude, slope, latitude, longitude, catchment area, and distance from source, stream width and hydraulic indicators of macrohabitat diversity. Coysh & Norris (1999) demonstrated that the AusRivAs predictive modelling approach could work well in New Zealand to complement existing approaches through:

- Incorporating MCI in the model to validate reference sites and to provide O/E MCI scores that are directly equivalent and comparable across all sites
- Providing site-specific reference predictions, alleviating the need for future reference pairing when assessing impacts
- Providing an objective way of calculating bands of impairment to inform management decisions
- Identifying missing taxa, which with a knowledge of their biology aid in deciding on likely causes of impairment
- Rapid and standardised assessment; once the models have been created, site assessments are rapid.

In an independent review of Waikato AusRivAs trial results, Stark (1999) suggested AusRivAs has a combination of features that made it attractive in a New Zealand context, including a strictly defined methodology, internet-based support system, predictive ability (O/E ratios), embodiment of at least two (e.g. taxonomic richness and MCI) indices, ability to cover different ecotypes, and outputs that can be easily understood by water managers and the public. However, Stark (1999) also warned that the scientific defensibility and value for a water manager of AusRivAs models depends upon strict adherence to protocols for every aspect of model development and use, including sample collection, processing, model application and output interpretation.

Stark (1999) argued that the cost and time involved to adapt AusRivAs to the New Zealand context and then to train users should first be considered against the benefits of improved guidance in the use of current water management tools. He suggested that the AusRivAs approach is no more immune to problems related to substandard input data or inexperienced users than existing tools in use in New Zealand at the time. Subsequent effort in New Zealand was spent on development of standardised invertebrate sampling and processing protocols (Stark et al*.* 2001).

Joy & Death (2003) trialled an AusRivAs model in the Manawatu-Wanganui region. Bentic macroinvertebrates and related physico-chemical data were collected at 127 reference sites and the resulting model was used to assess 27 test sites. All test sites were assessed as impacted based on the 10th percentile of the reference data. The model predicted O/E taxon richness did not include a biotic index (e.g. MCI) and had a low error rate (28%) similar to that observed in previous Australian AusRivAs models.

Joy & Death (2003) suggested that a major advantage of the reference site predictive model is a lack of subjectivity associated with assigning taxa scores. However an assumption of the predictive model approach is, that stressors alter invertebrate communities leading to the extinction of taxa that would otherwise be present. It is probably a fair assumption if suitable reference sites have been sampled.

Death & Collier (2010) developed an RIVPACS/ AusRivAs model for the Waikato region. They used only 25 reference sites to build the model but found it performed similarly well compared to a model built from 87 sites in an adjoining region, i.e. Manawatu-Wanganui model. In a comparison of different metrics, they found that of the eight metrics considered, the MCI and taxon richness O/E had the strongest links with percent native forest cover in the catchment (Death & Collier 2010).

5.2. Pressure-specific invertebrate metrics

5.2.1. Deposited fine sediment

Various pressure-specific invertebrate indices have been developed overseas as a tool for measuring sedimentation impacts at the reach scale and determining ecologically relevant site-specific targets for reducing sedimentation in streams. For example, Extence et al. (2013) developed the Proportion of Sediment-sensitive Invertebrates (PSI), which is based on fine sediment sensitivity ratings (four categories) of British benthic invertebrate species and families. The taxa have been assigned to one of four categories based on information available in the literature and an assessment of anatomical, physiological and behavioural traits exhibited by individual taxa of relevance to fine sediment cover on the streambed. Calculation of the PSI score involves fine sediment ratings as well as abundance weighted scores (relative abundance data). The PSI has shown to be more strongly related to percent fine bed sediment cover than a pressure-specific flow invertebrate index or % EPT abundance, suggesting that the PSI can discriminate between multiple stressors (Glendell et al*.* 2013).

Zweig and Rabeni (2001) developed the Deposited Sediment Biotic Index (DSBI) using sediment tolerance values for 30 taxa occurring in Missouri, United States. The tolerance values (three categories) were assigned according to the deposited-sediment (< 2 mm diameter) level at which the taxon reached 50% cumulative abundance across the data set using quantitative invertebrate data. Sensitive taxa occur at low fine sediment levels. The fine sediment tolerance values did not correlate with tolerance values with respect to organic enrichment, suggesting that the DSBI also has potential to discriminate between multiple stressors (Zweig & Rabeni 2001).

Bryce et al. (2010) assigned deposited sediment tolerance values to taxa using the weighted-averaging technique based on relative abundances determined from a mountain stream data set in the western United States. The weighted average, or optimum tolerance value of each taxon, identified the sediment values (streambed cover of % fines (< 0.06 µm) or % fines and sand (< 2mm)) at which the highest relative abundances occurred. These values were used to calculate an IBI (index of biotic integrity) score.

5.2.2. Nutrient enrichment

Smith et al. (2007) developed two nutrient biotic indices for benthic invertebrate communities, one for total phosphorus (NBI-P), and one for nitrate (NBI-N) using a data set from New York State, United States, comprising invertebrate relative abundance data. The weighted-averaging technique was used to establish nutrient optima and subsequent nutrient tolerance values of 164 invertebrate taxa across total phosphorus and nitrate gradients. A three-tiered scale of eutrophication for these nutrients (nutrient concentration thresholds) was established through cluster analysis of invertebrate communities using Bray-Curtis (quantitative) similarity. Testing of the NBI-P and NBI-N showed that they were able to discriminate between stream trophic states as indicated by classification according to mean nutrient concentrations. The authors suggested that NBI-P and NBI-N are robust measures of stream nutrient status and can serve as a useful tool in enforcing regional nutrient criteria (Smith et al. 2007).

Haase and Nolte (2008) also developed a nutrient biotic index; the invertebrate species index (ISI), to assess stream health in southeast Queensland, Australia. This eutrophication index is calculated as the weighted average of species' sensitivity scores, i.e. using species-specific indicator weights (describing whether the species has a broad or narrow ecological amplitude) and abundance data (counts of individuals per unit area) for each species. Development of the species' sensitivity scores and indicator weights, as well as calculation of site-specific ICI, are based on quantitative invertebrate data. Reference conditions were calculated for both upland and lowland sites to be able to establish O/E scores and to define ecological quality classes.

5.2.3. Flow

Extence et al. (1999) developed the Lotic-invertebrate Index for Flow Evaluation (LIFE), a flow index primarily based on recognised flow associations of British invertebrate taxa. Taxa were assigned into one of six flow groups based on information found in the literature as well as on expert knowledge. The site-specific LIFE score is calculated for a semi-quantitative sample using the flow category as well as abundance weighted scores of the taxa. Higher flows should result in higher LIFE scores. This flow index should be useful for monitoring and assessing the effects of low flows as well as abstraction and could provide a basis for setting benchmark flows suitable for protecting and maintaining ecological integrity (Extence et al. 1999).

Recently, Greenwood et al. (2016) developed two flow indices for New Zealand (LIFENZ and a weighted variant: LIFENZ_W) for the purpose of determining impacts of river flow alterations on aquatic ecosystems and in setting environmental flows that sustain certain ecological values. Similarly to the UK-based LIFE (Extence et al. 1999), indicator values represent preference for a water velocity category which were assigned to aquatic invertebrate taxa using professional judgement, and the site-specific index scores are calculated using velocity category and relative abundance (New Zealand's coded abundance categories) of each scoring taxon. The LIFENZ_W variant down-weights the scores if the taxon has a general compared to a more specific velocity preference. Testing the two indices showed that they were correlated with depth-averaged water velocity. Furthermore, changes in LIFENZ and LIFENZ_W scores were associated, both between sites and temporally within sites, with changes in hydrological attributes, such as the magnitude and length of time since a recent high flow event, and to a lesser degree with other physico-chemical parameters. In comparison, the MCI was unrelated to local flow conditions but also correlated with antecedent flow conditions (likely influenced by effects of flow stability on algal growth). The flow indices appear to be useful tools for understanding and managing the effects of hydrological alteration on aquatic invertebrate communities in New Zealand (Greenwood et al. 2016).

5.2.4. Acid mine drainage

Gray & Harding (2012) developed a benthic invertebrate biotic index for assessing coal mining impacts in New Zealand streams. The acid mine drainage index (AMDI) was developed by associating water chemistry and benthic invertebrate community data collected from 91 sites on the West Coast of the South Island. Indicator values for 57 taxa were calculated using weighted averaging. Site scores can range from 0 (severely impacted) to 100 (unimpacted) and sites can be categorised as 'severely impacted', 'impacted' or 'unimpacted'. Comparisons between. Use of the AMDI and traditional indices indicated that the AMDI is more accurate at detecting mine drainage. The AMDI was shown to be a sensitive indicator of mining impacts across a broader spatial scale on the West Coast of the South Island (Clapcott et al. 2015).

5.3. Traits metrics

Invertebrate taxa differ in their feeding behaviour, morphology, attachment to the substrate, and modes of respiration and reproduction. Trait information typically has been compiled from the literature for many taxa covering a large geographical area and is held in a trait data base. This information can be used to calculate from a site-specific taxa list the relative abundances of the different trait expressions (e.g. proportion of the filter-feeding or grazing trait) or the richness of taxa expressing a certain trait category (e.g. number of filter-feeding or grazing taxa). Human perturbation often alters invertebrate food sources and habitats, hence invertebrate trait metrics should respond to perturbation and may even be able to discriminate between different types of stressors.

For example, a New Zealand study identified individual invertebrate traits that responded to water abstraction but not the percentage of farming intensity in the catchment (Lange et al. 2014). Furthermore, in this study common community metrics (EPT metrics, taxon richness, MCI) either failed to detect an effect of percentage water abstraction or showed non-additive, antagonistic response patterns for the two landscape-scale stressors, suggesting that trait metrics are particularly valuable for management of the effects of water abstraction (Lange et al. 2014). An experimental study approach also found that some traits responded to a gradient of deposited fine sediment but not to a nutrient concentration gradient (Wagenhoff et al. 2012).

Invertebrate trait metrics can add significant value to stream assessment because they can detect and potentially discriminate between different stressors. Trait metrics also offer a mechanistic alternative to traditional taxonomy-based descriptors, i.e. give clues about the underlying mechanisms of stressor effects, and they are also indirect ecosystem functional measures (Menezes et al. 2010). For example, changes in the proportion of the shredding trait or changes in the number of shredding taxa in a community may indicate changes to organic matter processing. Furthermore, trait diversity metrics as surrogates for functional diversity have shown to be potentially useful for stream health assessment (Peru & Dolédec 2010; Gallardo et al. 2011; Lange et al. 2014), in particular because of their stability in response to natural environmental variability (Peru & Dolédec 2010). Finally, the biological trait approach has large-scale applicability (even across continents) as all invertebrate taxa can be described and compared on the same 'trait' scale (Bonada et al. 2006). The performance of multiple biological traits approach has been rated highly by (Bonada et al. 2006) which met 10 of their 12 criteria for an 'ideal' biomonitoring tool.

Despite the large range of different types of invertebrate traits, few of them appear to feature in the set of core metrics applied in the reviewed overseas assessment systems (section 4). Mainly functional feeding groups (e.g. grazers) or attachment to the substrate (e.g. clingers) have been considered because human impacts often alter the food source or habitat features. However, the scientific literature has shown that more traits are promising for metric development, either to detect perturbation at a landscape scale (e.g. pastoral land-use effects, Dolédec et al. 2006; Feld & Hering 2007; Dolédec et al. 2011; Lange et al. 2014) or at the instream reach scale (e.g. fine sediment and nutrients, Wagenhoff et al. 2012; Lange et al. 2014). Perhaps, as for pressure-specific invertebrate metrics, the reason for the paucity of a wide range of trait metrics featuring in the European and United States assessment systems is related to the fact that the scientific literature had only just started to investigate invertebrate traits as tools for stream health assessment.

New Zealand has its own New Zealand Macroinvertebrate Trait Database held by NIWA as part of the Freshwater Biodata Information System (FBIS). This database contains information on 16 traits with 59 trait categories (i.e. the different trait expressions) for invertebrate life history (eight traits for body size, reproduction and development), resistance and resilience (five traits for morphology, locomotion and dissemination) and general biology (three traits for feeding and respiration) which is based on genus or higher-level identifications. Use of invertebrate traits for stream biomonitoring in New Zealand has since been encouraged (Phillips 2004) and has raised interest in managers.

For example, Waikato Regional Council and Auckland Council investigated invertebrate traits for use in their stream health monitoring programmes as traits reflect changes in ecosystem function (rather than just changes in macroinvertebrate composition) and provide a mechanistic tool (Phillips & Reid 2012b; Phillips & Reid 2012a). A wide range of traits responded to increasing pastoral development in both hard-bottomed and soft-bottomed streams. This suggests that trait metrics were as equally powerful at detecting different levels of pastoral land-use impacts as the traditional metrics (e.g. EPT, MCI). Overall, these reports concluded that trait-based biomonitoring would fit readily into existing biomonitoring frameworks employed by councils, as the necessary information (taxon list) is already collected, and recommendations were made for the development of trait-based tools (e.g. for diagnostics of instream stressors) (Phillips & Reid 2012a, 2012b). Furthermore, at a national scale, the percentage of taxa that reproduce only once (Cycle 1) showed a strong predictable response to catchment land use (Clapcott et al. 2012). Finally, traits can also be used to explore variation in biotic indices. Stark and Phillips (2009), for example, used species traits to help explain seasonal variation in observed MCI scores.

5.4. Trout prey indices

Benthic invertebrate information provides an insight into the food resources available for fish. Trout prey indices (TPI) are specifically designed to assess benthic (or drift) invertebrate data in relation to food quality for adult trout. Three separate indices have been developed for the different trout foraging strategies: benthic-browsing, drift-feeding and cruise-feeding (the latter mode of feeding being relevant in a backwater or lake environment) (Shearer, unpublished data). The indices are based on the premise that some invertebrates will be more suitable as prey for the fish than others. Invertebrate traits for size, activity and availability were weighted for vulnerability to predation. Invertebrate taxa weightings were multiplied over a sample taxa list to calculate trout prey indices according to a specific trout foraging behaviour.

The TPI scores are calculated based on the presence of taxa using a modified version of the equation used by Stark (1985) to calculate the MCI. A quantitative version of a

TPI (the QTPI) can also be calculated based on taxa abundances (requiring quantitative sampling) and weights the overall index scores towards the dominant taxa.

The TPI have only been recently developed and still requires further testing and validation. The robustness of these indices require testing, i.e. how well does the TPI score reflects trout abundance in a river (e.g. using Jowett's 1990 100-rivers dataset); and, how well does the TPI correlate with the actual diet of trout. Provided there is a strong relationship between the TPI/QTPI and trout abundance (or biomass), boundaries ('bands') can be defined for assessment of invertebrate food quality for adult trout within the NPS-FM framework.

In summary, the TPI will allow managers to use invertebrate data to infer potential effects for highly valued recreational trout fisheries. Ultimately this approach should be adaptable to other predatory fish such as eels. Currently, no such tools exist for New Zealand water managers and we are unaware of any other attempts to develop such indices. At present the TPIs provide a theoretical basis to evaluate if changes in an invertebrate community result in a change to the quality of invertebrate food for adult trout.

6. SUMMARY AND RECOMMENDATIONS

Benthic macroinvertebrates are important organisms in streams and used worldwide as indicators of stream ecosystem health as they respond to human pressures and are taxonomically diverse and easy to sample. In New Zealand, Hirsch (1958) published the first biological evaluation of organic pollution in streams and since the 1980s benthic invertebrate monitoring has been widely used to assess the 'ecosystem health' of streams (Pridmore & Cooper 1985). Subsequently, New Zealand holds a large amount of invertebrate information collected annually for SoE monitoring and reporting from over 1000 wadeable sites across the country.

New Zealand's benthic invertebrate information has a diverse basis. Most councils use Stark et al.'s (2001) standard collection and processing protocols, thereby providing reasonable consistency across the country. Most councils sample semi-quantitatively. Half of the councils have been using coded abundance methods while the rest have been using fixed or total counts of all individuals in a sample. While sampling in hard-bottomed streams is generally done in riffle/run habitat, there is less consistency across the methods used for soft-bottomed streams.

6.1. Recommendation 1: Adopt nationally consistent approaches to macroinvertebrate sampling and processing

While there is some consistency in the methods used to collect and process macroinvertebrates in New Zealand, there is a need to improve this consistency for development of new metrics and NOF-relevant attributes. Currently, there remains debate as to what protocols are best to use for SoE reporting (i.e. current practice vs. NEMaR recommendations), which needs to be resolved for NOF attribute development also.

International assessment systems include standardised methods for sample collection and processing and the analysis of data, including robust data quality assurance. This allows for the calculation of reliable metrics. While our literature review shows that there is a large number and types of invertebrate metrics that could be calculated using existing invertebrate information in New Zealand[2], the application of those metrics at both a regional and national level must be considered. For example, the recent report updating MCI tolerance scores attempted to address the variability in MCI methods applied throughout New Zealand (Greenwood et al. 2015); in addition to collecting samples differently, councils were using different TVs to calculate the MCI site scores. A logical next step on from the Greenwood report is to ensure each

[2] The metrics used in the EU and US assessment systems are based on relative abundance data, either coded or quantitative. Most of the reviewed pressure-specific metrics are also calculated using relative abundance data and tolerance values of the taxa used for calculation of these metrics have also often been developed based on relative abundances.

council calculates MCI in a consistent way. MCI scores could then be normalised to a regional/stream type reference condition (e.g. O/E, see Section 6.2).

The logical vehicle for developing consistent methods in New Zealand is the National Environmental Monitoring Standards (NEMS). We understand that the EMaR invertebrate group headed by Dean Olsen is moving ahead on this already. In addition to providing guidance on quality assurance for data collection and processing, NEMS could provide tools for standard metric calculation. What those metrics could be is discussed next.

6.2. Recommendation 2: Making invertebrate metrics fit for purpose for working within the NPS-FM framework

6.2.1. Recognise and test the strength of existing invertebrate metrics

In New Zealand, the invertebrate metrics that are most commonly used in environmental reporting include variants of the Macroinvertebrate Community Index (MCI) and of the three insect orders Ephemeroptera, Plecoptera and Trichoptera (EPT) taxa. These two core metric types are sensitive to a wide range of human impacts on streams including stressors that are considered to be of most concern in agricultural (and also in urban) catchments, namely nutrients and fine sediment. Hence, they can provide an indication of overall ecosystem health in catchments where these two stressors are of main concern. A clear strength of these metrics is that they automatically take into account multiple-stressor effects.

However, these metrics are possibly not sensitive, or not sensitive enough, to all potential stressors. To our knowledge, there are several studies that have investigated the response patterns of EPT and MCI-type metrics to stressors other than those mentioned (e.g. water abstraction, temperature rise, heavy metals etc). As a first step, we recommend reviewing this information to gain more knowledge on which other stressors (than organic/nutrient enrichment and sedimentation) they respond to. This knowledge would give more confidence in that using these tools would adequately protect species from all or most stressors of concern. If these metrics were shown to be unresponsive to an important stressor, additional tools (e.g. other metrics or ecotoxicological approaches) would be required to manage for the effects of that particular stressor.

6.2.2. Develop stressor-specific invertebrate metrics

Generally EPT taxa are sensitive to eutrophication but also tend to be sensitive to sedimentation and vice versa, which is why EPT and MCI-type metrics respond to both. However, research has shown that some taxa are sensitive to one stressor but less to another, hence it should be possible to develop stressor-specific metrics based on stressor-specific tolerance values, which can be assigned using multiple possible methods. In fact, most of our reviewed studies on stressor-specific metrics showed that their metrics could successfully discriminate between two different stressors. We recommend development of stressor-specific metrics for the most important stressors, in particular those which do not lend themselves for being tested using ecotoxicological assays. A stressor-specific metric should provide a better link between the metric and the single stressor and hence meet the current requirements for inclusion in the NOF as an attribute. Stressor-specific metrics would further be useful as diagnostic tools, i.e. help with identification of the stressor(s) constraining ecosystem health.

One caveat of stressor-specific metrics is that they do not capture multiple-stressor effects. Even if such indicators were developed for all potential stressors generic limits (i.e. set on a national or regional scale) defined from the stressor-response relationships would not guarantee that ecosystem health is appropriately protected as additive multiple-stressor effects, and let alone interactive effects, would not be accounted for.

The second caveat is related to the fact that a stressor-specific invertebrate metric probably does not link directly to values such as fisheries quality or biodiversity, hence it would be difficult to decide where to define the attribute boundaries for these metrics. Definition of boundaries for a metric with a direct link to a human value, on the other hand, will help stakeholders to decide what level of protection (A, B or C) is needed to fulfil their specific goals for a given stream or river. Therefore, during the development and selection of metrics it will be important to consider the quantitative link to stressors but equally the representativeness of human values.

There are numerous methods that can be used to determine stressor-specific metrics and to test their independence from other stressors. Community turnover and sensitivity species analysis have previously been used to inform stressor-specific responses, but only in a single stressor framework (e.g. Threshold Indicator Taxa Analysis; Baker & King 2010). Whereas species response in the presence of multiple stressors can be explored using Gradient Forest analysis (e.g. Wagenhoff et al. in review). Stressor-specific metrics will have an inherent quantitative link to stressors.

6.2.3. Single metrics

One option for determining NOF-suitable attributes is to develop metrics following the Europe and the United States examples, beginning with the identification and development of core metrics. These would be grouped into four categories covering the different aspects of invertebrate communities (i.e. richness/diversity, composition, tolerance, functional aspects). They could include metrics that have been developed in recent years or even metrics currently in development (Table 5). However, metrics would need to be able to be quantitatively linked to stressors to be included as attributes within the NPS-FM framework. Metric scores could be combined to report on overall stream health (see Section 6.2.2).

Table 5. Potential metrics to explore for a macroinvertebrate attribute(s) of Ecosystem Health, Mahinga kai (Fishing) value. See sections 3.3. and 5 for description of metrics. Also, Cycle 1 = the proportion of taxa reproducing only once; % >365days is the proportion of taxa living longer than 365 days.

Ecosystem health component or value	Metric	Stressor specific	Comment
Richness/diversity	# Taxa	No	Widely used
	# EPT taxa	No	Widely used
	O/E taxa richness	No	Requires reference model approach
Composition/abundance	% EPT taxa	No	Widely used
	% EPT abundance	No	Widely used
Tolerance	MCI	eutrophication, sediment	Widely used
	Nutrient index	eutrophication	Requires NZ development
	Fine sediment index	sediment	Requires NZ development
	LIFENZ	hydrological alteration	Requires national testing
	AMDI	coal mine effects	Requires national testing
Functional aspects	% Cycle 1	No	Requires national testing
	% >365days	hydrological alteration	Requires national testing
	% predators	hydrological alteration	Requires national testing
	% burrowers	sediment	Requires national testing
Mahinga kai	% large	No	Requires national testing
	TPI	No	Requires national testing

Richness is the number of taxa, either of all taxa or of only those belonging to the EPT insect orders, known to be generally pollution-sensitive. Richness metrics, especially percent total richness, are used worldwide as measures of biodiversity, which is known to be linked to ecosystem resilience. Biodiversity and resilience are important aspects of healthy ecosystems and specifically mentioned in the description of the national compulsory value Te Hauora o te Wai.

Composition metrics are measures of the relative abundance of known pollution-sensitive/tolerant taxa and have also been successfully implemented worldwide in stream health assessment. For example, % EPT abundance responds to various types of perturbation therefore performing well in detecting overall human impact such as freshwater chemistry, excessive nutrients, algal blooms, high sediment levels, high temperatures, low oxygen, invasive species, and changes in flow regime.

Tolerance metrics are designed to respond to specific types of human impacts or stressors. This is achieved via assignment of indicator values to taxa. Metric calculation may involve weighting of the species according to their relative abundance and/or according to the specificity of their indicator value. Tolerance metrics do not directly reflect the status of biodiversity, resilience or ecosystem functioning. However, they should be able to indicate adverse effects of important stressors on invertebrate communities which have to be taken into account when managing for Te Hauora o te Wai (see Section 1.2). These metrics sometimes respond to stressors other than the one they were designed for, which means that at a given site or group of sites they may not be able to discriminate between these. As long as different stressors do not confound the response (i.e. having opposing effect directions), response to multiple stressors at worst sacrifices their discriminatory power.

Functional metrics are measures of taxa traits, such as life history, habitat preference or feeding modes. These metrics respond to perturbations such as organic pollution, changes to the trophic state or habitat. Hence traits may also be used for managing aspects that affect Te Hauora o te Wai, such as excessive nutrients, algal blooms, high sediment levels, high temperatures, low oxygen, invasive species, and changes in flow regime. Traits metrics can also inform Mahinga kai (fishing) in terms of the food availability for secondary consumers.

6.2.4. Multi-metrics

The representativeness of the ecosystem health value is probably best addressed through the use of a multi-metric index. A multi-metric index would encompass metrics that provide measures of ecosystem structure and function as well as the impacts of major stressors. One approach to calculating a multi-metric index would be to use an Average Score Per Metric. Which metrics were included in the ASPM could be determined by the need for an holistic assessment of ecosystem health components (e.g. Clapcott et al*.* 2014) or based on which single metrics best discriminated between reference and impacted sites (Collier 2008). The flexibility of a multi-metric index is the ability to incorporate a few or many single metrics. The shortcoming of an ASPM is that alone it is unlikely to be suitable as a NOF attribute because it will not link to management actions, although component metrics might. The application of an ASPM in the NOF would need to be considered carefully. For example, would thresholds need to be set on an ASPM or component metrics?

Moving beyond macroinvertebrates, there is a need to include other component qualities of ecosystem health in the NOF. For example, the condition of freshwaters in South East Queensland, Australia, is assessed by measuring ecosystem processes, fish, invertebrates, physical chemical, pollutant loads, and riparian indicators. Managing for Te Hauora o te Wai will need to account for more than just benthic invertebrates. How to incorporate multiple components requires further consideration. One possible approach is to normalise metrics by a reference or expected reference value, e.g. O/E. In this way the percentage deviation from a reference condition becomes a common scale by which to determine management bands and report ecosystem health.

6.3. Recommendation 3: Future data and research needs

There are some clear knowledge gaps that require addressing, specifically the collection of paired invertebrate-stressor data and the collection of reference data to develop and/or validate reference condition approaches.

6.3.1. Paired invertebrate-site data

To develop stressor-specific metrics requires the collection of benthic macroinvertebrate data in concert with direct measures of stressors, such as nutrients, sediment, or water abstraction. Many biological monitoring sites in New Zealand are not paired with water quality sites. Furthermore, not all stressors are measured at biological sites. For example, benthic sediment measurements are a recent addition to a few regional monitoring networks. Recent surveys have identified a strong link between macroinvertebrate metrics and catchment drivers of these stressors (e.g. land use), but to inform management decisions, quantitative links to proximate (e.g. nutrient and sediment loads) drivers are needed. Analysis of independent measures of proximate stressors and ecological responses, using causal pathway analysis or similar, would address the requirements of an attribute within the NPS-FM framework. This has not been possible to date at a national scale and instead stressor-response relationships have been inferred from the response of invertebrate metrics to catchment land use or modelled stressors (e.g. CLUES N).

In addition to paired invertebrate-stressor data, sufficient description of stream character is required. For example, the classification of geomorphological character such as eroding or depositional zones or dominant substrate classes. The habitat template is a primary driver of natural variation in invertebrate communities. Currently in New Zealand we are dependent on broad-scale tools such as REC and FENZ to provide estimates of character at the stream segment scale. Despite the development of standardised habitat assessment methods (Harding et al*.* 2009) and a rapid scoring method for stream habitat quality (Clapcott 2015), few benthic invertebrate monitoring sites have accompanying site-scale habitat data. Additional habitat data would ensure that habitat variation was accounted for when quantifying stressor-response relationships. It would also help identify stream ‘types’ should a classification system be needed when developing NOF thresholds (e.g. differing MCI bands for soft versus hard bottomed streams).

6.3.2. Defining reference condition

Defining the reference condition is necessary to determine deviation from reference and inform management thresholds. Currently in New Zealand less than 10 percent of regularly monitored sites are reference sites. In the past this has been used as the reason to preclude development of an O/E community composition metric (e.g. AusRivAs) approach at a national scale. However, prediction of reference condition can be achieved using a data set across a land-use intensity gradient rather than just using reference sites (which may not exist in some regions). This has been done for the prediction of reference MCI scores for all stream segments in the country (Clapcott et al*.* 2013) as well as for the probability of occurrence of many macroinvertebrate species (Leathwick et al*.* 2009). Similarly, an offset approach was used to predict reference conditions for large rivers (Collier & Hamer 2013). These approaches could also be used to predict reference condition for other/new metrics. Despite these promising modelling approaches, further reference data is needed to validate models and improve model predictive performance especially for environment types where currently little data is available. New macroinvertebrate data from reference sites could also be used to inform any frequency component of a macroinvertebrate attribute within the NPS-FM framework. For example, an understanding of the temporal variability at reference sites can be used to how often samples need to be collected (Collier et al. 2014)

6.4. Conclusion

The proposal to include MCI in the NPS-FM as a mandatory measure of water quality (MfE 2016) requires further consideration, especially the possibility of including MCI, or other macroinvertebrate metrics, in a non-attribute role in the NPS-FM. A possible consequence of the requirement to fit metrics into the NPS-FM framework could be the abandonment or reduction in the use of indices like the MCI that have a long (> 30 years) history of use in New Zealand. The MCI responds to multiple stressors and is useful for SoE reporting and for rapid screening in order to detect stream health issues that may warrant further investigation. It is not our recommendation that the MCI be abandoned, rather we suggest that the relationship between MCI, along with other potential metrics[3], and proximate stressors needs to be quantified to determine their suitability as NOF attributes. To do this requires the collection of paired site/stressor-invertebrate data.

NOF thresholds need to be determined for macroinvertebrate metrics. Metric values could be expressed as O/E ratios allowing for a common scale to set thresholds and also potentially the combination of metrics in a multi-metric index. To do so requires a robust estimate of reference condition. More reference data is needed in New Zealand, either to inform stream type or regionally specific reference state or to

[3] Stressor-specific metrics and traits included, but especially those metrics already in wide use (Table 4).

validate reference model predictions. An improved reference data set would also allow for the development of community composition O/E models (e.g. AusRivAs) that could be used to inform the diversity and composition components of ecosystem health assessments.

Finally, to ensure the robust calculation of metrics there needs to be national consistency in sample collection, processing and analysis. The National Environmental Monitoring Standards provides a vehicle to achieve national consistency. Together with site-specific stressor information and knowledge of reference state, consistent and quality assured data would allow for the exploration of the suitability of multiple potential benthic invertebrates metrics as NOF attributes to assess Ecosystem Health and Mahinga kai (fishery) values in New Zealand freshwaters.

7. REFERENCES

Armitage PD, Moss D, Wright JF, Furse MT 1983. The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. Water Research 17: 333-347.

Baker ME, King RS 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. Methods in Ecology & Evolution 1: 25-37.

Barbour MT, Gerritsen J, Snyder BD, Stribling JB 1999. Rapid bioassessment protocols for use in streams and wadeable rivers: periphyton, benthic macroinvertebrates and fish, 2nd Edition. EPA 841-B-99-002. Environmental Protection Agency, Washington D.C. 339 p.

Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund W, Zampoukas N, Hering D 2012. Three hundred ways to assess Europe's surface waters: An almost complete overview of biological methods to implement the Water Framework Directive. Ecological Indicators 18: 31-41.

BMWP 1978. Final report: assessment and presentation of the biological quality of rivers in Great Britain. Unpublished report. Department of the Environment, London, Biological Monitoring Working Party. 37 p.

Bonada N, Prat N, Resh VH, Statzner B 2006. Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. Annual Review of Entomology 51: 495-523.

Booker DJ, Snelder TH, Greenwood MJ, Crow SK 2014. Relationships between invertebrate communities and both hydrological regime and other environmental factors across New Zealand's rivers. Ecohydrology 8: 13-32.

Boothroyd IKG, Stark JD 2000. Use of invertebrates in monitoring. In: Collier KJ, Winterbourn MJ eds. New Zealand Stream Invertebrates: Ecology and Implications for Management. Christchurch, New Zealand, New Zealand Limnological Society. Pp. 344-373.

Bryce SA, Lomnicky GA, Kaufmann PR 2010. Protecting sediment-sensitive aquatic species in mountain streams through the application of biologically based streambed sediment criteria. Journal of the North American Benthological Society 29(2): 657-672.

Carter JL, Resh VH 2013. Analytical approaches used in stream benthic macroinvertebrate biomonitoring programs of state agencies in the United States. U.S. Geological Survey Open-File Report 2013-1129. 50 p.

Chessman BC 2003. New sensitivity grades for Australian river macroinvertebrates. Marine and Freshwater Research 54: 95-103.

Chessman BC, Jones HA, Searle NK, Growns IO, Pearson MR 2010. Assessing effects of flow alteration on macroinvertebrate assemblages in Australian dryland rivers. Freshwater Biology 55(8): 1780-1800.

Clapcott J 2015. National rapid habitat assessment protocol development for streams and rivers. Prepared for Northland Regional Council. Cawthron Report No. 2649. 29 p.

Clapcott J, Goodwin E 2014. Relationships between MCI and environmental drivers. Prepared for Ministry for the Environment. Cawthron Report No. 2507 21 p.

Clapcott J, Goodwin E, Snelder T 2013. Predictive models of benthic macroinvertebrate metrics. Prepared for Ministry for the Environment. Cawthron Report No. 2301. 32 p.

Clapcott J, Goodwin E, Harding J 2015. Identifying catchment-scale predictors of coal mining impacts on New Zealand stream communities. Environmental Management 57(3): 711-721.

Clapcott JE, Goodwin EO, Young RG, Kelly DJ 2014. A multimetric approach for predicting the ecological integrity of New Zealand streams. Knowledge and Management of Aquatic Ecosystems 415 (3):10.

Clapcott JE, Collier KJ, Death RG, Goodwin EO, Harding JS, Kelly D, Leathwick JR, Young RG 2012. Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. Freshwater Biology 57(1): 74-90.

Collier K, Clapcott J, Neale M 2014. A macroinvertebrate attribute to assess ecosystem health for New Zealand waterways for the national objectives framework - Issues and options. Environmental Research Institute report 36, University of Waikato, Hamilton. 31 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(4): 453-465.

Collier KJ 2008. Average score per metric: an alternative metric aggregation method for assessing wadeable stream health. New Zealand Journal of Marine and Freshwater Research 42(4): 367-378.

Collier KJ, Hamer MP 2013. Ecological response differentials: an alternative benchmark to inform stream and river bioassessment. Freshwater Biology 58(7): 1471-1483.

Coysh J, Norris R 1999. Waikato Region AUSRIVAS / RIVPACS trial. Report prepared for the Ministry for the Environment. SMF Contract No. 5091. 89 p.

Davies-Colley R, Hughes A, Verburg P, Storey R 2012. Freshwater monitoring protocols and quality assurance (QA). National environmental monitoring and reporting (NEMaR) variables step 2. Prepared for the Ministry for the Environment. NIWA Client Report HAM2012-092. 92 p.

Death RG, Collier KJ 2010. Measuring stream macroinvertebrate responses to gradients of vegetation cover: when is enough enough? Freshwater Biology 55(7): 1447-1464.

Dolédec S, Phillips N, Townsend CR 2011. Invertebrate community responses to land use at a broad spatial scale: trait and taxonomic measures compared in New Zealand rivers. Freshwater Biology 56(8): 1670-1688.

Dolédec S, Phillips N, Scarsbrook M, Riley RH, Townsend CR 2006. Comparison of structural and functional approaches to determining landuse effects on grassland stream invertebrate communities. Journal of the North American Benthological Society 25(1): 44-60.

eWater CRC (eWater Cooperative Research Centre) 2012. AUSRIVAS: Australian River Assessment System. eWater, Canberra, Australia. (Available from: http://AusRivAs.ewater.com.au/).

Extence CA, Balbi DM, Chadd RP 1999. River flow indexing using British benthic macroinvertebrates: A framework for setting hydroecological objectives. Regulated Rivers-Research and Management 15(6): 543-574.

Extence CA, Chadd RP, England J, Dunbar MJ, Wood PJ, Taylor ED 2013. The assessment of fine sediment accumulation in rivers using macro-invertebrate community response. River Research and Applications 29(1): 17-55.

Feld CK, Hering D 2007. Community structure or function: Effects of environmental stress on benthic macroinvertebrates at different spatial scales. Freshwater Biology 52(7): 1380-1399.

Gallardo B, Gascón S, Quintana X, Comín FA 2011. How to choose a biodiversity indicator – redundancy and complementarity of biodiversity metrics in a freshwater ecosystem. Ecological Indicators 11(5): 1177-1184.

Glendell M, Extence C, Chadd R, Brazier RE 2013. Testing the pressure-specific invertebrate index (PSI) as a tool for determining ecologically relevant targets for reducing sedimentation in streams. Freshwater Biology 59(2): 353–367.

Gray B 2004. Monitoring River Health Initiative Report Number 38. Australian Government Department of the Environment and Heritage.

Gray DP, Harding JS 2012. Acid Mine Drainage Index (AMDI): a benthic invertebrate biotic index for assessing coal mining impacts in New Zealand streams. New Zealand Journal of Marine and Freshwater Research 46(3): 335-352.

Greenwood M, Booker D, Stark J, Suren A, Clapcott J 2015. Updating MCI tolerance values for freshwater invertebrate taxa. Prepared for Environment Southland and Hawkes Bay Regional Council. NIWA client report no. CHC2015-008.

Greenwood MJ, Booker DJ, Smith BJ, Winterbourn MJ 2016. A hydrologically sensitive invertebrate community index for New Zealand rivers. Ecological Indicators 61: 1000-1010.

Haase R, Nolte U 2008. The invertebrate species index (ISI) for streams in southeast Queensland, Australia. Ecological Indicators 8(5): 599-613.

Harding JS, Clapcott JE, Quinn JM, Hayes JW, Joy MK, Storey RG, Greig HS, Hay J, James T, Beech MA, Ozane R, Meredith AS, Boothroyd IKD 2009. Stream habitat assessment protocols for wadeable rivers and streams of New Zealand.

Harmsworth GR, Young RG, Walker D, Clapcott JE, James T 2011. Linkages between cultural and scientific indicators of river and stream health. New Zealand Journal of Marine and Freshwater Research 45(3): 423-436.

Harrison ET, Nichols S, Gruber B, Dyer F, Tschierschke A, Norris R 2011. AUSRIVAS: Australia's in-stream biological health 2003–2010. 2011 State of the Environment report. Report prepared for The Australian Government, Department of Sustainability, Environment, Water, Population and Communities, Canberra, Australia. (Available from: http://www.environment.gov.au/system/files/pages/ba3942af-f815-43d9-a0f3-dd26c19d83cd/files/soe2011-supplementary-water-AusRivAs.pdf).

Hawkes HA 1997. Origin and Development of the Biological Monitoring Working Party (BMWP) Score System. Water Research 32(3): 964-968.

Hayes JW 2014. Evidence on potential effects of changes to water quality conditions, periphyton and benthic macroinvertebrates on trout and native fish associated with Palmerston North City Council's Wastewater Treatment Plant discharge to the Manawatu River. Presented on behalf of Palmerston North City Council to the Environment Court in the matter of a review under section 128(1)(a)(iii) of the RM Act of Palmerston North City Council's resource consent to discharge treated wastewater to the Manawatu River. Cawthron Institute, Nelson.

Heck KL, van Belle G, Simberloff D 1975. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. Ecology 56(6): 1459-1461.

Hickey CW, Clements WH 1998. Effects of heavy metals on benthic macroinvertebrate communities in New Zealand streams. Environmental Toxicology and Chemistry 17(11): 2338-2346.

Hirsch A 1958. Biological evaluation of organic pollution of New Zealand streams. New Zealand Journal of Science 1(4): 500-551.

Jowett IG 1990. Factors related to the distribution and abundance of brown and rainbow trout in New Zealand clear-water rivers. New Zealand Journal of Marine and Freshwater Research 24(3): 429-440.

Joy MK, Death RG 2003. Biological assessment of rivers in the Manawatu-Wanganui region of New Zealand using a predictive macroinvertebrate model. New Zealand Journal of Marine and Freshwater Research 37: 367-379.

Lange K, Townsend CR, Matthaei CD 2014. Can biological traits of stream invertebrates help disentangle the effects of multiple stressors in an agricultural catchment? Freshwater Biology 59:2431-2446

Leathwick JR, Julian K, Smith B 2009. Predicted national-scale distributions of freshwater macro-invertebrates in all New Zealand's rivers and streams. NIWA Client Report HAM2009-042.

Lorenz A, Hering D, Feld CK, Rolauffs P 2004. A new method for assessing the impact of hydromorphological degradation on the macroinvertebrate fauna of five German stream types. Hydrobiologia 516(1): 107-127.

Matthaei CD, Piggott JJ, Townsend CR 2010. Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. Journal of Applied Ecology 47(3): 639-649.

Matthaei CD, Weller F, Kelly DW, Townsend CR 2006. Impacts of fine sediment addition to tussock, pasture, dairy and deer farming streams in New Zealand. Freshwater Biology 51(11): 2154-2172.

Meier C, Böhmer J, Biss R, Feld CK, Haase P, Lorenz A, Rawer-Jost C, Rolauffs P, Schindehütte K, Schöll F, Sundermann A, Zenker A, Hering D 2006. Extension and adaptation of the national assessment system for benthic invertebrates to international requirements. UFOPLAN-No. 202 24 223. 198 p.

Menezes S, Baird DJ, Soares AMVM 2010. Beyond taxonomy: a review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. Journal of Applied Ecology 47(4): 711-719.

MfE 2014. National Policy Statement for Freshwater Management 2014. Wellington: Ministry for the Environment. 34 p.

MfE 2016. Next steps for fresh water: Consultation document. Wellington: Ministry for the Environment. 45 p.

Nichols SJ, Robinson WA, Norris RH 2010. Using the reference condition maintains the integrity of a bioassessment program in a changing climate. Journal of the North American Benthological Society 29(4): 1459-1471.

Niyogi DK, Koren M, Arbuckle CJ, Townsend CR 2007. Stream communities along a catchment land-use gradient: subsidy-stress responses to pastoral development. Environmental Management 39(2): 213-225.

Peru N, Dolédec S 2010. From compositional to functional biodiversity metrics in bioassessment: A case study using stream macroinvertebrate communities. Ecological Indicators 10(5): 1025-1036.

Phillips N 2004. Stream biomonitoring using species traits. Water & Atmosphere 12(4): 8-9.

Phillips N, Reid D 2012a. Biological traits: Application to the Auckland Council River Ecology Monitoring Programme. Auckland Council Technical Report 2012/001. 79 p.

Phillips N, Reid D 2012b. Biological trait analysis: Application to the Waikato Regional Council Monitoring Programme. Technical Report for Waikato Regional Council. NIWA Client Report No. HAM2011-071. 76 p.

Piggott JJ, Townsend CR, Matthaei CD 2015. Climate warming and agricultural stressors interact to determine stream macroinvertebrate community dynamics. Global Change Biology 21(5): 1887-1906.

Plafkin JL, Barbour MT, Porter KD, Gross SK, Hughes RM 1989. Rapid bioassessment protocols for use in streams and rivers: Benthic macroinvertebrates and fish. U.S. Environmental Protection Agency, Office of Water Regulations and Standards. Washington, D.C.

Pridmore RD, Cooper AB (Eds) 1985. Biological monitoring in freshwaters: proceedings of a seminar, Hamilton, 21-23 November 1984. Published for the National Water and Soil Conservation Authority, by the Water and Soil Directorate, Ministry of Works and Development. Water and soil miscellaneous publication, no. 82-83.

Quinn JM, Hickey CW 1990. Characterisation and classification of benthic invertebrate communities in 88 New Zealand rivers in relation to environmental factors. New Zealand Journal of Marine and Freshwater Research 24(3): 387-409.

Quinn JM, Cooper AB, Davies-Colley RJ, Rutherford JC, Williamson RB 1997. Land use effects on habitat, water quality, periphyton, and benthic invertebrates in Waikato, New Zealand, hill-country streams. New Zealand Journal of Marine and Freshwater Research 31(5): 579-597.

Shearer KA, Hayes JW 2010. Invertebrate drift and trout growth potential in the Taharua and Upper Mohaka rivers: an investigation of effects of dairy farming across three seasons. Cawthron Report No. 1832. 54 p.

Shearer KA, Hay J, Hayes JW 2007. Invertebrate drift and trout growth potential in Didymo (*Didymosphenia geminata*) affected reaches of the Mararoa and Oreti Rivers: April and August 2006. Cawthron Report No. 1214. 73 p.

Smith AJ, Bode RW, Kleppel GS 2007. A nutrient biotic index (NBI) for use with benthic macroinvertebrate communities. Ecological Indicators 7(2): 371-386.

Smith MJ, Kay WR, Edward DHD, Papas PJ, Richardson KSJ, Simpson JC, Pinder AM, Cale DJ, Horwitz PHJ, Davis JA, Yung FH, Norris RH, Halse SA 1999. AusRivAs: using macroinvertebrates to assess ecological condition of rivers in Western Australia. Freshwater Biology 41: 269-282.

Stark J 1999. Review of the AUSRIVAS Trial in the Waikato. Cawthron Report no 506. 18 p.

Stark J, Boothroyd I, Harding J, Maxted J, Scarsbrook M 2001. Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate Working Group Report No. 1. Prepared for the Ministry for the Environment, Sustainable Management Fund Project No. 5103. 57 p.

Stark JD 1985. A macroinvertebrate community index of water quality for stony streams Water & Soil miscellaneous publication 87: 1-52.

Stark JD 1993. Performance of the Macroinvertebrate Community Index: effects of sampling method, sample replication, water depth, current velocity, and substratum on index values. New Zealand Journal of Marine and Freshwater Research 27(4): 463-478.

Stark JD 1998. SQMCI: A biotic index for freshwater macroinvertebrate coded-abundance data. New Zealand Journal of Marine and Freshwater Research 32(1): 55-66.

Stark JD, Maxted JR 2004. Macroinvertebrate community indices for Auckland's softbottomed streams and applications to SOE reporting. Cawthron Report No. 970.

Stark JD, Maxted JR 2007a. A biotic index for New Zealand's soft-bottomed streams. New Zealand Journal of Marine and Freshwater Research 41(1): 43-61.

Stark JD, Maxted JR 2007b. A user guide for the Macroinvertebrate Community Index. Prepared for the Ministry for the Environment. Cawthron Report No.1166. 58 p.

Stark JD, Phillips N 2009. Seasonal variability in the Macroinvertebrate Community Index: are seasonal correction factors required? New Zealand Journal of Marine and Freshwater Research 43(4): 867-882.

Suren AM, Stark JD, Wech J, Lambert P 2010. Development of a South Island Wetland Macroinvertebrate Community Index Score. NIWA Client Report: CHC2010-0. 49 p.

Towers DJ, Henderson IM, Veltman CJ 1994. Predicting dry weight of New Zealand aquatic macroinvertebrates from linear dimensions. New Zealand Journal of Marine and Freshwater Research 28(2): 159-166.

Townsend CR, Uhlmann SS, Matthaei CD 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45(6): 1810-1819.

Townsend CR, Tipa G, Tierney LD, Niyogi DK 2004. Development of a tool to facilitate participation of Maori in the management of stream and river health. EcoHealth 1: 184-195.

Turak E, Flack LK, Norris RH, Simpson J, Waddell N 1999. Assessment of river condition at a large spatial scale using predictive models. Freshwater Biology 41: 283-298.

USEPA 2011. A Primer on Using Biological Assessments to Support Water Quality Management. EPA 810-R-11-01. Washington, D.C., U.S. Environmental Protection Agency, Office of Science and Technology, Office of Water. Pp. 108.

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE 1980. The River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences 37(1): 130-137.

Wagenhoff A, Townsend CR, Matthaei CD 2012. Macroinvertebrate responses along broad stressor gradients of deposited fine sediment and dissolved nutrients: a stream mesocosm experiment. Journal of Applied Ecology 49(4): 892-902.

Wagenhoff A, Townsend CR, Phillips N, Matthaei CD 2011. Subsidy-stress and multiple-stressor effects along gradients of deposited fine sediment and dissolved nutrients in a regional set of streams and rivers. Freshwater Biology 56(9): 1916-1936.

Wagenhoff A, Clapcott J, Lau KEM, Lewis GD, Young RG in review. Identifying congruence in stream assemblage thresholds in response to nutrient and sediment gradients for limit setting. Ecological Applications.

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(2): 461-470.

Wright JF 1994. Development of RIVPACS in the UK and the value of the underlying data-base. Limnetica 10(1): 15-31.

Wright JF, Sutcliffe DW, Furse MT 2000. Assessing the biological quality of fresh water: RIVPACS and other techniques. Ambleside, UK, Freshwater Biological Association.

Zweig LD, Rabeni CF 2001. Biomonitoring for deposited sediment using benthic invertebrates: a test on 4 Missouri streams. Journal of the North American Benthological Society 20(4): 643-657.

8. APPENDIX

Appendix 1. Tolerance values for MCI-based biotic indices in HB (Stark et al. 2001) and SB streams; '-' taxa not yet assigned SB tolerance values; from Stark & Maxted (2007a).

Taxon	HB	SB	Taxon	HB	SB	Taxon	HB	SB
COELENTERATA			**Odonata (continued)**			**Diptera** (continued)		
Hydra	3	1.6	*Procordulia*	6	3.8	Sciomyzidae	3	3.0
PLATYHELMINTHES	3	0.9	*Uropetala*	5	0.4	Stratiomyidae	5	4.2
RHABDOCOELA	-	0.9	*Xanthocnemis*	5	1.2	Syrphidae	1	1.6
BRYOZOA	-	4.0	**Hemiptera**			Tabanidae	3	6.8
NEMATODA	3	3.1	*Anisops*	5	2.2	Tanypodinae	5	6.5
NEMATOMORPHA	3	4.3	*Diaprepocoris*	5	4.7	Tanytarsini	3	4.5
NEMERTEA	3	1.8	*Microvelia*	5	4.6	*Tanytarsus*	3	-
OLIGOCHAETA	1	3.8	Saldidae	5	3.9	Thaumaleidae	9	8.8
POLYCHAETA	-	6.7	*Sigara*	5	2.4	Tipulidae	5	3.4
HIRUDINEA	3	1.2	**Coleoptera**			*Zelandotipula*	6	3.6
TARDIGRADA	-	4.5*	*Antiporus*	5	3.5	**Trichoptera**		
CRUSTACEA			*Berosus*	5	-	*Alloecentrella*	9	-
Amphipoda	5	5.5	*Copelatus*	5	3.7	*Aoteapsyche*	4	6.0
Cladocera	5	0.7	Dytiscidae	5	0.4	*Beraeoptera*	8	7.0
Copepoda	5	2.4	Elmidae	6	7.2	*Confluens*	5	7.2*
Halicarcinus	-	5.1	*Enochrus*	5	2.6	*Conuxia*	8	-
Helice	-	6.6	Hydraenidae	8	6.7	*Costachorema*	7	7.2*
Isopoda	5	4.5	Hydrophilidae	5	8.0	*Cryptobiosella*	9	-
Mysidae	-	6.4*	*Liodessus*	5	4.9	*Diplectrona*	9	-
Ostracoda	3	1.9	*Onychohydrus*	5	-	*Ecnomina*	8	9.6
Paracalliope	5	-	*Podaena*	8	-	*Edpercivalia*	9	6.3
Paraleptamphopus	5	-	Ptilodactylidae	8	7.1	Ecnominidae	8	-
Paranephrops	5	8.4	*Rhantus*	5	1.0	*Helicopsyche*	10	8.6
Paranthura	-	4.9	Scirtidae	8	6.4	*Hudsonema*	6	6.5
Paratya	5	3.6	Staphylinidae	5	6.2	*Hydrobiosella*	9	7.6
Tanaidacea	4	6.8	**Neuroptera**			*Hydrobiosis*	5	6.7
INSECTA			*Kempynus*	5	-	*Hydrochorema*	9	-
Ephemeroptera			**Diptera**			*Kokiria*	9	-
Acanthophlebia	7	9.6	Anthomyiidae	3	6.0	*Neurochorema*	6	6.0
Ameletopsis	10	10.0	*Aphrophila*	5	5.6	*Oecetis*	6	6.8
Arachnocolus	8	8.1	*Austrosimulium*	3	3.9	Oeconesidae	9	6.4
Atalophlebioides	9	4.4	*Calopsectra*	4	-	*Olinga*	9	7.9
Austroclima	9	6.5	Ceratopogonidae	3	6.2	*Orthopsyche*	9	7.5
Austronella	7	4.7	Chironomidae	2	3.8	*Oxyethira*	2	1.2
Coloburiscus	9	8.1	*Chironomus*	1	3.4	*Paroxyethira*	2	3.7
Deleatidium	8	5.6	*Corynoneura*	2	1.7	*Philorheithrus*	8	5.3
Ichthybotus	8	9.2	*Cryptochironomus*	3	-	*Plectrocnemia*	8	6.6
Isothraulus	8	7.1	*Culex*	3	-	*Polyplectropus*	8	8.1
Mauiulus	5	4.1	Culicidae	3	1.2	*Psilochorema*	8	7.8
Neozephlebia	7	7.6	Diptera indet.	3	2.9	*Pycnocentrella*	9	-
Nesameletus	9	8.6	Dixidae	4	7.1	*Pycnocentria*	7	6.8
Oniscigaster	10	5.1	Dolichopodidae	3	8.6	*Pycnocentrodes*	5	3.8
Rallidens	9	3.9	Empididae	3	5.4	*Rakiura*	10	-
Siphlaenigma	9	-	Ephydridae	4	1.4	*Synchorema*	9	-
Tepakia	8	7.6	Eriopterini	9	7.5	*Tiphobiosis*	6	9.3
Zephlebia	7	8.8	*Harrisius*	6	4.7	*Triplectides*	5	5.7
Plecoptera			Hexatomini	5	6.7	*Triplectidina*	5	-
Acroperla	5	5.1	*Limnophora*	3	4.5	*Zelandoptila*	8	7.0
Austroperla	9	8.4	*Limonia*	6	6.3	*Zelolessica*	10	6.5
Cristaperla	8	-	*Lobodiamesa*	5	7.7	**Lepidoptera**		
Halticoperla	8	-	*Maoridiamesa*	3	4.9	*Hygraula*	4	1.3
Megaleptoperla	9	7.3	*Mischoderus*	4	5.9	**COLLEMBOLA**	6	5.3
Nesoperla	5	5.7	*Molophilus*	5	6.3	**ACARINA**	5	5.2
Spaniocerca	8	8.8	Muscidae	3	1.6	**ARACHNIDA**		
Spaniocercoides	8	-	*Nannochorista*	7	-	*Dolomedes*	5	6.2
Stenoperla	10	9.1	*Neocurupira*	7	-	**MOLLUSCA**		
Taraperla	7	8.3*	*Neolimnia*	3	5.1	*Gundlachia = Ferrissia*	3	2.4
Zelandobius	5	7.4	*Nothodixa*	4	9.3	*Glyptophysa = Physastra*	5	0.3
Zelandoperla	10	8.9	Orthocladiinae	2	3.2	*Gyraulus*	3	1.7
Megaloptera			*Parochlus*	8	-	*Hyridella*	3	6.7
Archichauliodes	7	7.3	*Paradixa*	4	8.5	*Latia*	3	6.1
Odonata			*Paralimnophila*	6	7.4	Lymnaeidae	3	1.2
Aeshna	5	1.4	*Paucispinigera*	6	7.7	*Melanopsis*	3	1.9
Anisoptera	5	6.0	Pelecorhyncidae	9	-	*Physa = Physella*	3	0.1
Antipodochlora	6	6.3	*Peritheates*	7	-	*Potamopyrgus*	4	2.1
Austrolestes	6	0.7	Podonominae	8	6.4	Sphaeriidae	3	2.9
Hemianax	-	1.1	*Polypedilum*	3	8.0			
Hemicordulia	5	0.4	Psychodidae	1	6.1			
Ischnura	-	3.1	*Scatella*	7	-			