NIWA
Taihoro Nukurangi

Implications for regional council SOE monitoring from adopting the NEMS macroinvertebrates protocols

Analysis of the macroinvertebrate metric scores from paired sampling

Prepared for Northland Regional Council

November 2023

Climate, Freshwater & Ocean Science

Prepared by:
Michelle Greenwood
Elizabeth Graham
Alex Rose

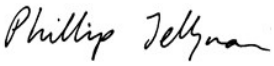
For any information regarding this report please contact:
Michelle Greenwood
Freshwater Ecologist
Freshwater Ecological Modelling
+64 3 343 8061
michelle.greenwood@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 8602
Riccarton
Christchurch 8440

Phone +64 3 348 8987

NIWA CLIENT REPORT No: 2023309CH
Report date: November 2023
NIWA Project: ELF23203

Revision	Description	Date
Version 1.0	Final version sent to client	6 November 2023

Quality Assurance Statement		
	Reviewed by:	Anika Kuczynski
	Formatting checked by:	Rachel Wright
	Approved for release by:	Phillip Jellyman

Contents

Executive summary **6**

1 Background **10**

Report structure 11

2 Methods **13**

2.1 Sample and site metadata collection 13

2.2 Feedback from sample processors 13

2.3 Macroinvertebrate metric calculation 13

3 Data availability **15**

4 How do the NEMS protocols differ from council in-house protocols? **18**

4.1 Sample collection 18

5 Sample processing and metric calculation methods **23**

6 Macroinvertebrate community and metric differences between NEMS and in-house samples **30**

6.1 Unique taxa 30

6.2 Community composition 32

6.3 Metric differences 37

6.4 Potential causes of differences in metric values between protocols 43

6.5 Summary 48

7 Precision of metric values from the protocols **49**

8 Overall summary and conclusions **58**

8.1 Recommendations 59

8.2 Future monitoring considerations 59

9 Acknowledgements **61**

10 References **62**

Appendix A Key comments from laboratory processors **64**

Appendix B Requirements for macroinvertebrate monitoring under the NPS-FM. **65**

Appendix C Plots of NEMS vs in-house metric values by council **66**

Tables

Table 3-1: The minimum, maximum and median percentage of the catchment upstream of each sampling site (n = 129) in indigenous, pastoral or urban land cover. 16

Table 3-2: The number and percentage of samples (n = 258) with macroinvertebrate metric values within the range of MCI and QMCI attribute Bands of the NPS-FM. 17

Table 4-1: Summary of methods used by councils for collection and processing of the paired samples. 22

Table 5-1: Changes in macroinvertebrate metric scores when taxa from the missed taxa scan were included and excluded from metric calculations. 29

Table 6-1: The total taxonomic richness and taxa unique to each protocol for samples collected using NEMS and in-house protocols for each council. 31

Table 6-2: Differences and correlations between macroinvertebrate metrics scores calculated on 129 paired samples collected using NEMS and council in-house protocols. 39

Table 6-3: Shift in provisional NPS-FM bands for QMCI and MCI when samples were collected using in-house or NEMS protocols by council. 40

Table 6-4: Summary results (p values) from separate generalised linear models (GLMs) predicting the difference in metric values between NEMS and in-house samples (NEMS minus in-house) for five macroinvertebrate metrics. 45

Figures

Figure 3-1: Sampling sites where paired macroinvertebrates samples were collected. 15

Figure 4-1: Number of habitat types (e.g., riffle, run, pool, woody debris) sampled within individual sites using in-house and NEMS collection methods. 19

Figure 4-2: The proportion of sample that was processed for macroinvertebrates when collected with in-house and NEMS sampling methods. 20

Figure 5-1: Macroinvertebrate metrics as provided by councils and recalculated using NEMS tolerance values and taxa list. 26

Figure 5-2: Macroinvertebrate metrics calculated for 201 samples processed using the 200+ Count with Scan for Missed Taxa using data with and without taxa from the missed taxa scan. 28

Figure 6-1: Non-metric multidimension scaling (NMDS) ordination on 258 paired macroinvertebrate samples collected using NEMS (red circle) and council in-house (blue triangle) protocols at 129 sites by ten councils using relative abundance (left) and presence/absence (right) data. 35

Figure 6-2: Non-metric multidimensional scaling (NMDS) ordinations of paired macroinvertebrate samples collected at the same sites using NEMs (red circle) and in-house (blue triangle) protocols by Environment Canterbury (ECAN) analysed using relative abundance data (left) and presence/absence data (right). 36

Figure 6-3: Plots of macroinvertebrate metrics calculated for paired samples collected using NEMS and council in-house protocols. 41

Figure 6-4: Bland Altman plots for macroinvertebrate metrics calculated from samples collected using NEMS or in-house protocols. 42

Figure 6-5: Plots of between-council differences in metric scores between NEMs and in-house samples for metrics in which council was identified as a significant predictor in GLM analyses. 46
Figure 6-6: Difference between paired NEMS and in-house sample taxonomic richness by the percentage of riffle/run habitat sampled during the NEMS habitat proportional sampling. 47
Figure 7-1: Long-term MCI values from ECAN SOE monitoring sites included in this report. 51
Figure 7-2: Long-term MCI values provided by councils at SOE monitoring sites included in this report. 52
Figure 7-3: Long-term QMCI values from ECAN SOE monitoring sites included in this report. 53
Figure 7-4: QMCI values calculated for 13 Surber samples collected within three reaches (of nine permanently marked) at two sites on the Selwyn River (Scotts Road and Whitecliffs) over multiple visits between 2020 and 2022. 55
Figure 7-5: MCI values calculated for 13 Surber samples collected within three reaches (of nine permanently marked) at two sites on the Selywn River (Scotts Road and Whitecliffs) over multiple visits between 2020 and 2022. 56
Figure C-1: Proportion EPT abundance calculated from paired samples using NEMS and council in-house protocols for 10 regional councils. 66
Figure C-2: Taxonomic richness calculated from paired samples using NEMS and council in-house protocols for 10 regional councils. 66

Executive summary

The development of the National Environmental Monitoring Standard for macroinvertebrates (NEMS macroinvertebrates) is aimed at improving national consistency in collection, processing and reporting of macroinvertebrate metrics for State of the Environment (SOE) monitoring. Regional councils currently use slightly different collection and processing protocols, but commonly focus collection on riffle or run habitat. The NEMS requires a habitat-proportional sampling approach. To maintain the continuity of long-term data sets, it is critical to assess whether macroinvertebrate data and calculated metrics using the NEMS protocol are directly comparable to a council's historical methods.

Ten regional councils provided paired macroinvertebrate samples during the 2021/22 summer, one collected using a riffle/run-focussed protocol and one collected using the NEMS protocol, from a total of 129 sites. Paired samples were collected from Northland, Auckland, Bay of Plenty, Waikato, Hawke's Bay, Wellington, Taranaki, Manawatū-Whanganui, Canterbury and Southland regions. Councils provided riffle/run focussed samples for their 'in-house' samples, although this is not the methodology currently employed by all councils for their SOE monitoring. The 129 sampling sites were predominantly hard-bottomed (12 soft-bottomed sites) and represented a range of climatic, topographic and land-use conditions.

The goal of this report was to investigate patterns in this dataset to assess: 1) the magnitude of differences in macroinvertebrate metrics between samples collected using the different protocols and 2) if we can identify conditions under which differences in metric values from the protocols can be predicted. The overall aim was to provide councils with information that can be used to make a more informed choice when considering whether and how to switch to the NEMS protocol. An online workshop was also held with the laboratories that processed the macroinvertebrate samples to collate feedback on sample size and processing differences between the protocols.

We assessed differences in sample collection and processing methods between the NEMS and in-house protocols, and compared the resulting differences in taxa identity, community composition and metric values. Key macroinvertebrate metrics assessed were taxonomic richness, EPT (Ephemeroptera, Plecoptera and Trichoptera) richness, EPT percentage taxa, EPT percentage abundance and the National Policy Statement for Freshwater Management (NPS-FM) compulsory macroinvertebrate attributes: Macroinvertebrate Community Index (MCI), its quantitative variant (QMCI) and Average Score Per Metric (ASPM). We also discuss the importance of estimates of precision for each protocol and provide recommendations for councils potentially switching to the NEMS.

The main differences in sample collection between the in-house council protocols and the NEMS protocols were:

- The NEMS protocols were well followed by the councils for sample collection and processing, with some differences in the size of the sampling device used.
- Council in-house collection and processing protocols varied somewhat between councils, which made conclusions about the impact of specific methodological differences difficult.
- NEMS protocols resulted in up to six additional habitats being sampled within a site compared to in-house riffle/run-focused protocols.

- No obvious difference in the proportion of the sample that was processed between NEMS and in-house samples was observed or reported by laboratory processors.

We considered the impacts of processing methods on calculated metric values. Specifically, these processing differences were: 1) variation from the standard NEMS list in the taxonomic resolution and MCI tolerance values assigned to taxa before metric calculations and 2) the inclusion/exclusion of the scan for missed taxa in the NEMS recommended 200+ count processing method. The missed taxa scan is a scan of the full sample for the presence of taxa not already identified and is conducted after the sub-sample with at least 200 individuals has been processed. Key findings from the consideration of processing methods were:

- The processing laboratories reported no consistent differences in sample size and processing time between the protocols, but that large samples of either type were more difficult to process.
- The small between-council differences in taxonomic resolution and tolerance values used in metric calculation had relatively small effects on metric values.
- While on average, few (0.7 per sample) additional taxa were reported by councils that were not on the standard NEMS taxonomic list, up to 10 additional taxa were identified in one sample.
- Switching between the MCI tolerance values for hard-bottomed (HB) and soft-bottomed (SB) streams for taxa had an impact on metric values for some, but not all, samples from SB sites. This difference was not related to council methods, but rather an artefact due to the unavailability of council-provided SB metric scores for all SB sites.
- In just over half the samples, the missed taxa scan resulted in no additional total or EPT taxa being reported.
- The inclusion of taxa from the missed taxa scan resulted in up to 12 additional taxa and 6 additional EPT taxa reported in one sample.
- MCI, %EPT taxa richness and ASPM were almost always higher in samples when taxa from the missed taxa scan were included.
- Abundance-based metrics (QMCI, %EPT abundance) were largely unimpacted by the inclusion or exclusion of the taxa from the missed taxa scan.

Differences in macroinvertebrate communities and metric values between the NEMS and in-house protocols were assessed as: 1) the unique taxa collected by each protocol, 2) community compositional differences between the protocols using multivariate ordinations, and 3) metric differences between the protocols using correlation and Bland-Altman plots. We also assessed whether differences in metric values between the protocols were related to environmental differences between sites, the diversity of habitat sampled, or regional differences between councils. Key findings were:

- For eight out of ten councils, more taxa and more unique taxa were found in NEMS protocol samples than in the in-house protocol samples, although unique taxa were collected by both protocols.

- The composition of the community collected by the two protocols did not differ significantly overall using data from all councils or for most councils individually (except for ECAN).
- The metric scores from the two protocols were highly correlated and, on average, differences between the protocols were low. However, for individual sites, the differences in metric values between the protocols could be quite large (large enough to shift sites 1 or 2 NPS-FM attribute bands).
- Metric values from one protocol were not consistently higher or lower than the other. At some sites, in-house metric values were higher and at other sites the NEMS protocol metric values were higher.
- The difference between the metrics from the two protocols was not explained by many of the variables for which data were available, particularly environmental differences between sites.
- Taxonomic richness was generally higher in NEMS samples when a higher proportion of non-riffle/run habitat was sampled, but this pattern was not observed for all councils.
- The identity of the collecting council was the only predictive variable that explained differences in metric values between the protocols for more than one metric, likely a reflection of regional environmental differences (e.g., climate, geology, flow regime, etc.) or perhaps variation in the in-house methods between councils.
- Using the available predictors, we cannot identify between-site or between-protocol differences that explain the difference in macroinvertebrate metrics calculated from samples collected using in-house or NEMS samples.

The precision, or repeatability, of metric values from samples collected using each method is an important consideration when discussing how different the values are from the two protocols. Samples collected using the same protocol, at the same time, at the same site will show some variability due to spatial variation in community composition and sampling effort. The key question for councils is whether the variation in metric scores associated with differences between NEMS and in-house protocols is within these expected ranges of natural variation. For context, plots of 1) long-term temporal variability in in-house metric scores from a subset of sites in this report and 2) spatial variation in metric values within a site using data from another study are provided. We did not have replicate samples from each site available to calculate precision for the protocols. However, Stark (1993) and Stark (1998) provide precision estimates of approximately 10 MCI units for a single sample using riffle-focussed kicknet methodology. At 79% of sites the difference in MCI values between the NEMs and in-house protocols was less than this value.

Recommendations and future monitoring considerations

The NEMS macroinvertebrate protocol provides benefits for nationally consistent monitoring and reporting.

Our recommendations are that if councils do switch to the NEMS protocol:

1. Co-ordination between councils on the timing of the switch is considered so that national consistency is attained over a short period of time.

2. Dual monitoring for some time period is considered, particularly in sites where the risk of getting the metric value wrong is high, such as sites with metric values close to critical management thresholds or those where metric values may be challenged (e.g., locations of potential future development). An appropriate length of time for dual monitoring could be five years, to match the time period over which NPS-FM macroinvertebrate attribute scores are calculated. Samples could be collected and stored for later processing, if required, to minimise processing costs.

Continued paired monitoring would also provide further data to investigate the implications of switching protocols.

It is also worth noting that under Clause 3.10 (4) in the NPS-FM, "attribute states and baseline states may be expressed in a way that accounts for natural variability and sampling error." An increased knowledge of the magnitude of natural variability in macroinvertebrate communities and thus metrics, and in the precision of in-house and NEMS protocols under different environmental conditions, could assist councils in this process.

Several options to investigate these sources of variation are:

- Further investigation of expected variation between sampling methods at a site during different environmental conditions (i.e., longer-term paired monitoring) to understand if recent environmental conditions affect the difference in metrics scores between protocols.

- Replicate samples of both sampling protocols under the same environmental conditions would help identify the scale of spatial variation expected within a site.

1 Background

Aquatic macroinvertebrate communities are monitored in rivers and streams across Aotearoa New Zealand (NZ) for many purposes, including assessment of environmental impacts, characterising biodiversity and monitoring changes in ecological health over time. Macroinvertebrate community data are commonly used as a fundamental measurement of aquatic ecosystem health by regional councils and other agencies. Monitoring benthic macroinvertebrates for this purpose is now mandatory under the National Policy Statement for Freshwater Management 2020 (NPS-FM; New Zealand Government 2023), where the macroinvertebrate metrics Macroinvertebrate Community Index (MCI), its quantitative variant (QMCI) and the Average Score Per Metric (ASPM) are included as attributes. However, many councils have slightly different approaches to macroinvertebrate sample collection, processing and/or metric calculation.

The different approaches mean that sample collection and processing methods can be adjusted to environmental conditions specific to a region or site and to the goals of the sampling programme. However, these different methods introduce sources of uncertainty when data are compared between regions and compiled to make national assessments, such as required when calculating macroinvertebrate metrics and assigning attribute bands under the NPS-FM.

To improve national consistency, particularly in State of the Environment (SOE) reporting, a new National Environmental Monitoring Standard (NEMS) for macroinvertebrates has been developed (NEMS 2022). This introduces a standard approach to macroinvertebrate sample collection and processing for councils to follow in their long-term SOE monitoring programmes. The aim of the NEMS is to reduce uncertainty caused by different sample collection and processing methods when comparing macroinvertebrate metrics scores nationally, particularly for SOE reporting under the NPS-FM (see Appendix B for a summary of requirements of the NPS-FM and the macroinvertebrate metrics).

Most councils' existing macroinvertebrate collection methods are based on the Stark et al. (2001) national protocols. The NEMS is intended to supersede the Stark et al. (2001) protocols for SOE monitoring and differs in several ways. For example, many councils currently collect samples in hard-bottomed streams from riffles, in line with Stark et al. (2001), which promotes the sampling of riffles to minimise variability between sites. In contrast, the NEMS requires that samples are taken proportionally from all habitat types present within the sample reach, i.e., runs, pools, wood debris, etc. The NEMS recommends a shift to habitat proportional sampling for SOE monitoring for the following reasons:

1. A sample collected from all habitats present was thought to likely represent overall ecological health of a reach better than a sample collected from a targeted habitat type.

2. Riffles are not always abundant in hard-bottomed streams (particularly in many lowland stream systems). The NEMS suggests that sampling across the most common mesohabitats in proportion to their abundance provides a more representative assessment of stream condition at a site.

3. Large geomorphological change may occur at a site, potentially removing the previously sampled habitat type. For example, Cyclone Gabrielle has caused immense geomorphological changes in many waterways in the Hawke's Bay region. Previous riffle habitat may have been replaced by a different habitat type in some sites.

Habitat proportional sampling will represent the macroinvertebrate community at a site, regardless of the habitats that are present.

Councils are considering switching to the NEMS protocol for macroinvertebrate samples collected for SOE reporting to improve national consistency. However, to maintain the continuity of long-term data sets at these sites, it is critical to assess whether macroinvertebrate data and calculated metrics from samples collected using the NEMS protocols are directly comparable to a council's historical methods of sample collection and processing. Otherwise, a switch in protocols could create a disjunct in long-term macroinvertebrate community data and calculated macroinvertebrate metrics.

To investigate and quantify the potential impact of a shift from current council macroinvertebrate sample collection and processing protocols to the NEMS protocol, ten regional councils provided data from paired macroinvertebrate samples, one collected using the habitat proportional NEMS protocol and one collected targeting riffles or runs. Riffle/run focussed sample collections are commonly the councils' current in-house sampling method. Paired samples were collected from 129 sites across ten regions: Northland, Auckland, Bay of Plenty, Waikato, Hawke's Bay, Wellington, Taranaki, Manawatū-Whanganui, Canterbury and Southland.

Previous investigations of the paired data from individual councils (23 paired samples by Greater Wellington Regional Council and 31 paired samples by Environment Canterbury) have reported differences in the macroinvertebrate communities and calculated metrics from samples using the different protocols (e.g., Hornblow 2022). The goal of this report is to investigate patterns in the combined dataset from the ten councils to assess: 1) whether differences in macroinvertebrate metrics between protocols is likely to be a widespread issue and 2) if we can identify conditions under which differences in metric values between the protocols can be predicted.

The overall aim is to provide councils with information that can be used to make an informed choice between the options available to them in response to the new proposed NEMS standards for SOE monitoring:

1. Continue using in-house protocols to maintain consistency of long-term record unless required to switch to NEMS.
2. Switch to the NEMS protocol to improve national consistency, without any further paired monitoring.
3. Collect additional data at all (or a subset of) sites for some time to further investigate the potential implications of switching methods, e.g., continued paired monitoring with both protocols at some sites.

1.1 Report structure

This report is structured in the following way:

Section 2 presents details of the general methods of metric calculation, collection of sample metadata and input from sample processors**.** More specific details of analyses are provided in subsequent sections.

Section 3 describes the available data from the paired sample collections by the ten councils**.**

Section 4 compares how the NEMS protocol differs from the council in-house protocols, particularly the sampling devices used, the number of habitats processed, sample processing techniques and the

proportion of the sample that was processed. Feedback about sample size and processing difficulty from the staff at laboratories that processed the samples is also briefly discussed (further details in **Appendix A**).

Section 5 considers the influence of differences in metric calculation (identifying taxa to different taxonomic levels and using different tolerance values) and the effect of including or excluding taxa in the missed taxa scan on metric values**.**

Section 6 presents recalculated macroinvertebrate metrics from raw macroinvertebrate data using the standard NEMS taxonomic resolution and tolerance values**.** Differences in taxonomic richness, community composition and metric values between the NEMS and in-house protocols are investigated using scatterplots, boxplots, ordinations, linear correlations, and Bland-Altman plots (a method to visualise biases in the difference between protocols). We also investigate whether the magnitude and direction of difference in metric values between the protocols can be explained by between-site differences in physical conditions.

Section 7 considers the precision of metric scores calculated for an individual sampling event and provides some context for between-protocol metric differences relative to temporal and spatial variability in metric scores.

Section 8 provides an overall summary with the main conclusions**,** including options for potential further monitoring.

2 Methods

The term 'in-house' is used in this report to refer to the riffle/run focussed sample collection protocol used by the councils to compare to the habitat proportional NEMS protocol. Note that Bay of Plenty Regional Council already used habitat proportional sampling protocols for SOE monitoring and collected samples using a riffle-focussed methodology to add to the dataset. The other nine councils collected samples using their standard riffle/run focussed in-house SOE protocols.

2.1 Sample and site metadata collection

All participating councils filled out a provided spreadsheet, so that we could collect details about their sample collection and sample processing methods under both protocols. This spreadsheet collated details, where provided by the council, of the type, opening size and mesh size of the sampling device used, sampling effort (area sampled, number of sub-samples compiled or time spent sampling), sample size (number or size of pottle), number of habitat types sampled and the taxonomic list used during metric calculation.

Information on Climate, Source of Flow and Land cover categories for each site was extracted from the River Environment Classification (REC v2[1]), a database of catchment spatial characteristics summarised for every segment in New Zealand's river network based on the site's location and associated river segment number. For each site, the upstream catchment area was generated from the REC and the proportion of the upstream catchment in different land cover categories was extracted from the Land Cover Database (LCDB5[2]), which assigns 33 different categories (e.g., indigenous forest and built-up area).

2.2 Feedback from sample processors

A short online workshop was held on 23 May 2023 with the laboratories that processed the collected samples to gather feedback on sample size and processing time for samples collected using both protocols.

2.3 Macroinvertebrate metric calculation

Macroinvertebrate metrics (Macroinvertebrate Community Index – MCI; its quantitative variant – QMCI; taxa richness; Ephemeroptera, Plecoptera and Trichoptera (EPT) richness; percent EPT taxa; percent EPT abundance; and Average Score Per Metric – ASPM) were collated or calculated for all 258 samples in the following ways:

- As provided by the councils, where available
- Recalculated using standardised taxonomic lists and tolerances scores from Annexes D and H from the macroinvertebrate NEMS:
  - Including taxa that were picked up in the missed taxa scan
  - Excluding taxa from the missed taxa scan, where possible.

[1] Available here: https://data-niwa.opendata.arcgis.com/maps/3a4b6cc2c1c74fbb8ddbe25df28e410c/about
[2] Available here: https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/

The missed taxa scan is conducted during the 200+ Count and Scan for Missed Taxa sample processing method recommended by the NEMS. After a sub-sample containing at least 200 individuals is processed (individuals identified and counted) a scan of the rest of the sample is undertaken to note the presence of any taxa not yet identified.

Several points about metric calculations should be noted. Firstly, all EPT metrics excluded hydroptilidae caddisflies. Secondly, for abundance-based metrics, such as the QMCI, the NEMS recommends that taxa from the missed scan should be given an abundance of 0.5. We used a value of 1 to standardise between councils as it wasn't always possible to identify taxa identified in the missed taxa scan or determine the value assigned to all taxa from the scan. Finally, the taxonomic list in the NEMS is the minimum list; councils may include additional taxa for metric calculations. However, to be consistent across councils, we excluded taxa not listed or without tolerance values provided in Annexes D and H of the NEMS when recalculating metric scores.

The area from which samples were collected was relatively consistent (commonly 0.7 m^2 to 1 m^2, although a minimum of 0.3 m^2), when provided. However, abundances could not always be standardised to a per unit area value (e.g., per square metre) because the area sampled was not always available. This is not an issue for metrics calculated using presence/absence data (MCI, taxa richness) or those standardised by the total abundances of individuals (e.g., % EPT abundance, QMCI). However, the data used in ordinations of community composition need to be standardised because the abundances of individual taxa influence the results. For metric calculations, we scaled sub-samples to a full sample abundance using information on the proportion of the same processed. Community composition data for ordinations were transformed to relative abundances for samples where sub-sampling methods meant that abundances could not be scaled to the full sample (e.g., when including ECAN in-house samples which are processed with a fixed 100 count of individuals).

Soft-bottomed (SB) streams, i.e., those with high (generally classified as >50%) deposited sediment cover, were self-identified by the councils (12 of 129 sites were soft-bottomed). Following the NPS-FM, we used MCI soft-bottomed (SB) variant tolerance scores (Stark and Maxted 2007, updated in Clapcott et al. 2017) to calculate metrics for sites with soft bed substrate and the original hard-bottomed (HB) tolerance values (Stark 1985, updated in Clapcott et al. 2017) to calculate metrics for streams with hard bed substrate.

The assignment of macroinvertebrate metric values from the samples to NPS-FM attribute bands (MCI, QMCI, ASPM) are provisional as they are based on individual samples and not the median of 5 years, as required under the NPS-FM.

All analyses were completed in R. More detailed methods relevant to each analysis are provided in the following sections.

3 Data availability

Paired samples collected using NEMS and in-house protocols were provided for 129 sites by ten regional councils: Auckland Council, AC - 8 sites; Bay of Plenty Regional Council, BOP - 6 sites; Environment Canterbury, ECAN - 31 sites; Environment Southland, ES -15 sites; Greater Wellington Regional Council, GWRC - 23 sites; Hawke's Bay Regional Council, HBRC - 14 sites; Horizons Regional Council, HRC - 12 sites; Northland Regional Council, NRC - 5 sites; Taranaki Regional Council, TRC - 4 sites and Waikato Regional Council, WRC - 11 sites; Figure 3-1, Table 3-1).

Sites were sampled using both protocols on the same day, apart from one site sampled by HRC (Ohau at Haines Farm) where the samples were collected two days apart.

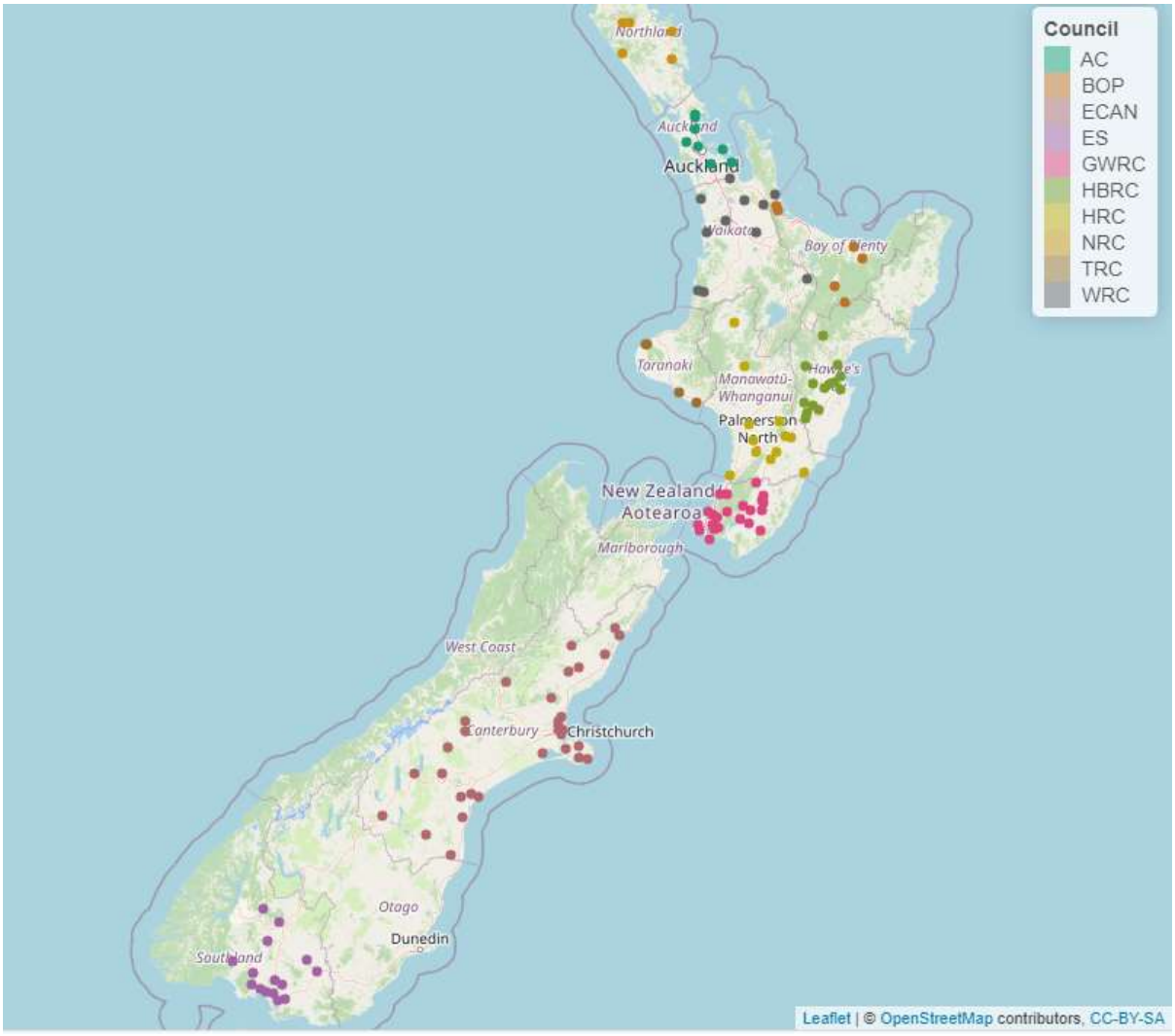


Figure 3-1: Sampling sites where paired macroinvertebrates samples were collected. Paired samples were collected at 129 locations by ten regional councils: Auckland Council; AC, Bay of Plenty Regional Council; BOP, Environment Canterbury; ECAN, Environment Southland; ES, Greater Wellington Regional Council; GWRC, Hawke's Bay Regional Council; HBRC, Horizons Regional Council; HRC, Northland Regional Council; NRC, Taranaki Regional Council; TRC and Waikato Regional Council; WRC.

Most sites were assigned as 'hard-bottomed' by the councils (90.7%) with twelve sites (9.3%) assigned as 'soft-bottomed'. Sites were located within a range of climatic conditions, topography and land cover types according to the REC. Most sites were Cool-Wet (36%), Cool-Dry (27%) and Warm-Wet (21%), with a few sites in the Warm-Dry (9%) or Cool-Extremely wet (5%) climate categories. Most sites were in the Lowland REC topographic category (71%), while some sites were in Hill (26%) and five sites in the Mountain category (2%). Pasture was the predominant land cover class across the sites (59%), with some sites in exotic forest (21%), indigenous forest, tussock or shrub (15 %) or urban landcover (9%), according to the REC categories.

The upstream catchment area for individual sites ranged from 0.1 to 10894 km^2, with a median of 77 km^2. The proportion of the catchment with indigenous land cover ranged from 0 to 96%, and urban landcover ranged from 0 to 44% (Table 3-1).

The health of macroinvertebrate communities, as indicated by the macroinvertebrate metrics MCI and QMCI, varied widely between sites, with scores ranging from indicating relatively unimpacted communities to severely degraded. MCI values (calculated using the standardised NEMS list, see Section 2.3) across all 258 paired samples ranged from 44 to 148. Under the NPS-FM, sites with five-year median values exceeding an MCI of 130 fall into Band A, indicative of communities in pristine condition with minimal organic pollution. Sites with median MCI values less than 90 fall within Band D, indicative of severe organic pollution and below the NPS-FM national bottom line. Ten percent of samples (26 samples) were in the range of MCI Band A (MCI >130) and 27% of samples had values below the national bottom line (MCI <90, Table 3-2).

Table 3-1: The minimum, maximum and median percentage of the catchment upstream of each sampling site (n = 129) in indigenous, pastoral or urban land cover. Land cover was calculated using LCDB5. Indigenous combines the LCDB categories Fernland, Sub-alpine shrubland, Matagouri or grey scrub, Manuka/Kanuka, Broadleaved indigenous hardwood and Indigenous forest. Pastoral/Crop combines the Short rotation cropland, Low producing grassland and High producing exotic grassland categories. Urban combines Transport infrastructure and Built-up area/Settlement categories.

	Indigenous	Pastoral/Crop	Urban
Minimum %	0	0	0
Maximum %	96 %	76 %	44 %
Median %	37 %	24 %	0.06 %

Table 3-2: The number and percentage of samples (n = 258) with macroinvertebrate metric values within the range of MCI and QMCI attribute Bands of the NPS-FM. Metric values were re-calculated for all samples using a consistent taxonomic resolution and tolerance value table (NEMS Annexes D and H). NPS-FM attribute bands were calculated from one sample, not the median of five annual samples, as recommended in the NPS-FM, and therefore are indicative only. Metric scores in Band D are below the national bottom line under the NPS-FM.

	Band A	Band B	Band C	Band D
MCI				
Number sites	26	60	103	69
Percentage sites (%)	10	23	40	27
QMCI				
Number sites	48	50	43	117
Percentage sites (%)	19	19	17	45
ASPM				
Number sites	45	81	66	66
Percentage sites (%)	17	31	26	26

4 How do the NEMS protocols differ from council in-house protocols?

In this section, we use information provided by the councils to compare the sample collection and processing protocols of the NEMS and in-house samples collected by the councils.

4.1 Sample collection

4.1.1 Collection device and sampling intensity

The sampling devices and sampling intensity recommended by the NEMS protocols were generally followed by the councils. Kicknets with the recommended mesh size (0.5 mm) were used by all councils to collect samples for the NEMS protocol samples (Table 4-1). The size of kicknets varied from 25 cm to 50 cm between councils (30 – 40 cm recommended under NEMS). The area of habitat sampled using the NEMs protocol, where reported by the councils, varied from 0.75 m^2 to 1.0 m^2, similar to NEMS guidelines (0.6 m^2 to 0.9 m^2). The councils reported that between one and 16 subsamples (kicks) were compiled into the composite sample (four to eight recommended, Table 4-1). Note that the NEMS states that additional sampling effort may be required in streams with low macroinvertebrate densities to obtain at least 200 individuals, which may have been the case in some of the sites sampled by councils.

Councils differed in the sampling device and sampling effort they used to collect in-house samples (Table 4-1). This also varied for some councils between the collection of the in-house and NEMs paired samples. HRC, WRC, BOP and NRC used the same sampling device and sampled the same area under both in-house methods and the paired NEMS samples (Table 4-1). The other councils had differences between the NEMS and in-house samples in either sampling device, e.g., ES and HBRC used Surber samplers for their in-house method, or in sampling intensity, e.g., ECAN spent 9 minutes collecting the in-house composite sample and 10 kicks for the NEMS samples, or in both device and sampling intensity (Table 4-1).

4.1.2 Habitats sampled

All in-house samples collected by councils prioritised riffle or run habitat, where available. Data on the number of additional habitat types (such as pool, woody debris) sampled for each sample using both methods were available for 109 of the 129 sites (Figure 4-1). Of these, at 13 sites (12%) the same number of habitats were sampled under both protocols. At the other 96 sites (78%) at least one more habitat was sampled under the NEMS protocol than the in-house protocol, up to a maximum of six additional habitats (at four sites).

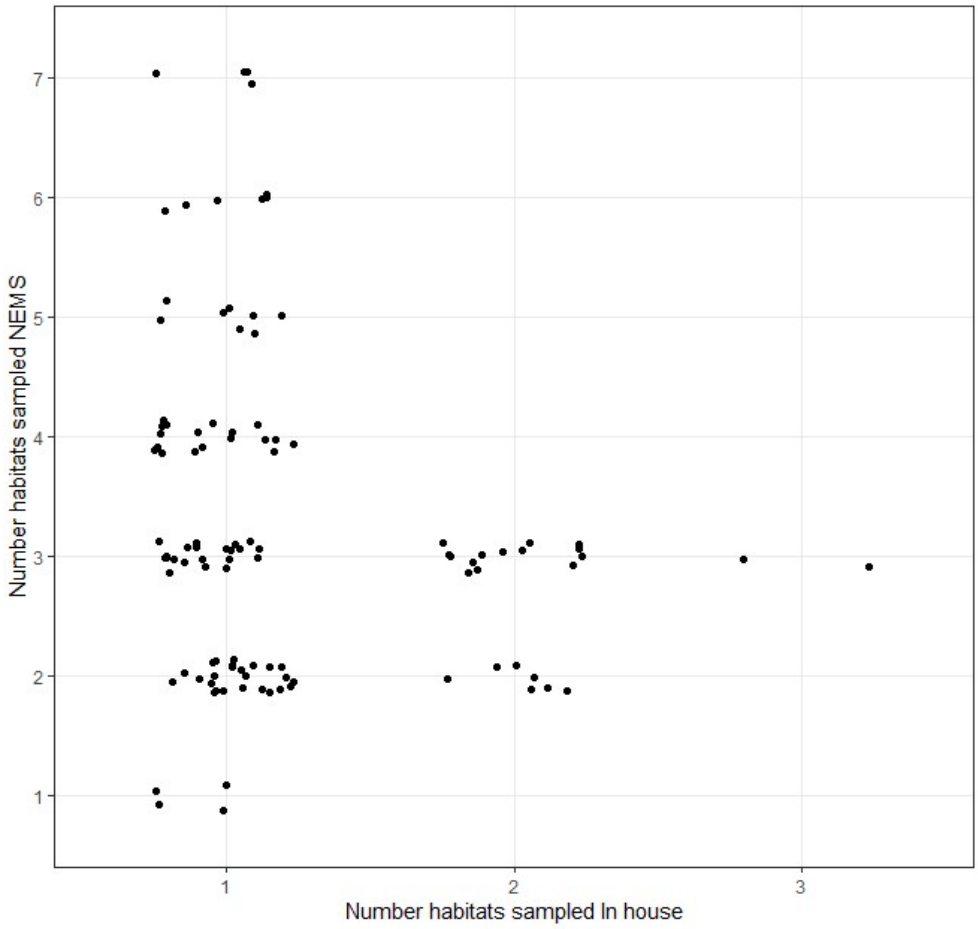


Figure 4-1: Number of habitat types (e.g., riffle, run, pool, woody debris) sampled within individual sites using in-house and NEMS collection methods. n = 109 sites with data. A small amount of random vertical and horizontal scatter has been added to points to aid visibility.

4.1.3 Sample processing

Processing method

The macroinvertebrate NEMS recommends that samples are processed using the protocol 200+ Count with Scan for Missed Taxa (Annex E of NEMs). In this method a random representative subsample is processed, which must contain at least 200 individuals. A visual scan by eye is made of the rest of the sample to find taxa missed in the subsample. Samples can be subsampled using a grid, cookie cutter or barrel splitter until at least 200 individuals are counted. All of the subsample must be processed after the 200 individuals have been counted and the fraction of sample sorted recorded. Macroinvertebrate samples may also be processed using the Full Count with subsampling option, modified from P3 of Stark et al. (2001).

All councils except ECAN processed both sets of samples using the NEMS protocols, i.e., either the 200 count with scan for missed taxa or the full count methodology (NRC used full count for both protocols and AC for the in-house protocol with 200 count for NEMS samples). To compare both collection and processing differences between their in-house and the NEMS protocols, ECAN processed their in-house samples using their standard method of subsample with a barrel splitter and fixed count of 100 individuals without complete processing of the sub-sample. Seven councils

processed both paired samples using the 200+ Count method, and one council used Full count with subsampling for both paired samples (Table 4-1).

Proportion of the sample processed

Sixty-seven paired samples had information about the proportion of sample processed for macroinvertebrates using both collection methods and were processed using the same methods. Differences in the proportion of the sample that was processed between the paired samples ranged from -0.69 less of the sample under NEMS to 0.2 more of the sample, with a median difference of 0 (Figure 4-2). Twenty-three samples (34%) had the same proportion of the sample processed. A higher proportion of sample was processed under NEMS for 18 sites (27%) and a lower proportion under NEMS for 26 sites (39%).

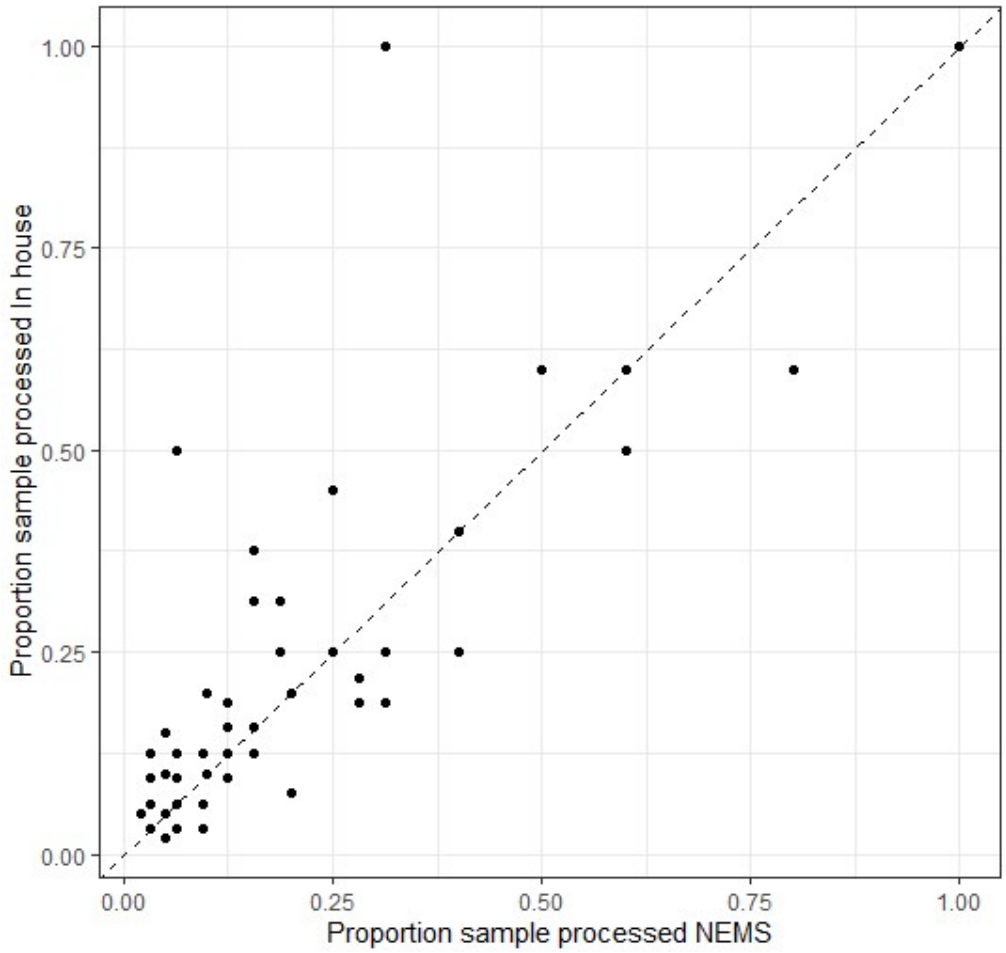


Figure 4-2: The proportion of sample that was processed for macroinvertebrates when collected with in-house and NEMS sampling methods. n = 67 paired samples with information available for both protocols.

4.1.4 Feedback from sample processors

A video conference was held on the 23 May 2023 with the laboratories who processed the paired samples to discuss implications of the switch to NEMS protocols for sample processing. The key findings from that meeting are in Appendix A. Overall, while there were some differences in the size of samples supplied, the proportion that was processed and the time taken to process between the samples collected using NEMS and in-house protocols, the differences were often site-specific and few generalisations between the protocols could be drawn. A common difficulty for laboratory processors was large samples (from both protocols) and further guidance on reducing sample volume in the field was suggested as a useful outcome.

4.1.5 Summary

- The NEMS protocols were well followed by the councils for sample collection and processing, with some differences in the size of the sampling device used.
- NEMS protocols resulted in up to six additional habitats being sampled within a site compared to in-house riffle-focused protocols. At 12% of sites the same number of habitats were sampled using both protocols.
- No obvious differences in the proportion of the sample that was processed was observed between NEMS and in-house samples. Processing laboratories identified that large samples were a problem impacting their processing ability, but that large samples seemed to be site-specific with no general conclusions able to be drawn about the relative size of samples collected using NEMS and in-house protocols.
- For some councils the differences in sampling device, effort and processing protocol between in-house and NEMS protocols were greater than for other councils. This between-protocol variation between councils makes conclusions about the specific impact of differences in sampling device, effort and processing methods between protocols more difficult to identify.

Table 4-1: Summary of methods used by councils for collection and processing of the paired samples. Note BOP collected 'in-house' samples from riffles to allow comparison with the other councils. Their current in-house method is habitat proportional. SB = soft-bottomed, K = kicknet, Clapcott = Clapcott et al. (2017). Stark 2001 = Stark et al. (2001).

Council	Method	No. sites	SB sites	Sampler	Habitats sampled	Sampling effort	Processing method	Tolerance scores	Missed taxa ided?
ECan	NEMS	31	0	K, 0.5mm, 25cm	Habitat proportional	0.75 m^2, 10 kicks	200+ fixed count	Clapcott	Y
ECan	In-house	31	0	K, 0.5mm, 25cm	100 % runs	Sampled for ~ 9 minutes	ECan method - 100 count +scan for rare taxa	In house list	N
TRC	NEMS	4	2	K, 0.5mm, 50cm	Habitat proportional	0.9m^2, 8 kicks	200+ fixed count	Stark 2001	Y
TRC	In-house	4	2	K, 0.5mm, 50cm	100% riffles	400 ml sample	200+ fixed count	Stark 2001	Y
HRC	NEMS	12	0	K, 0.5mm, 36cm	Habitat proportional	5 kicks over an area of 45 to 120 m	200+ fixed count	Stark 2001	Y
HRC	In-house	12	0	K, 0.5mm, 36cm	Riffle if present, otherwise run	5 kicks over an area of 45 to 120 m	200+ fixed count	Stark 2001	Y
WRC	NEMS	11	1	K, 0.5mm, 50cm	Habitat proportional	0.75m^2, 5 kicks, 30 sec each	200+ fixed count	Clapcott	Y
WRC	In-house	11	1	K, 0.5mm, 50cm	100% riffles, or 100% wood in SB	0.75m^2, 5 kicks, 30 sec each	200+ fixed count	Clapcott	Y
BOP	NEMS	6	0	K, 0.5mm, 30cm	Habitat proportional	30 sec per kick	200+ fixed count	Clapcott	Y
BOP	In-house	6	0	K, 0.5mm, 30cm	100% riffles	30 sec per kick, 10 to 12 kicks	200+ fixed count	Clapcott	Y
AC	NEMS	8	4	K, 0.5mm, 31cm	Habitat proportional	0.8 m^2, 8 kicks	200+ fixed count	Clapcott	Y
AC	In-house	8	4	K, 0.5mm, 31cm	HB: 100% riffle (C1) SB: Proportional macrophytes margin, woody debris (C2)	0.3m^2, 10 reps	Full count with subsample, P3	Clapcott	Y
NRC	NEMS	5	0	K, 0.5mm	Habitat proportional	1 m^2, 5 kicks	Full count with subsample, P3	Stark 2001	Y
NRC	In-house	5	0	K, 0.5mm	Riffle and runs	1 m^2, 5 kicks	Full count with subsample, P3	Stark 2001	Y
GWRC	NEMS	23	4	K, 0.5mm, 40cm	Habitat proportional	1 m^2, 1 to 16 kicks	200+ fixed count	Clapcott	Y
GWRC	In-house	23	4	K, 0.5mm, 40cm	HB: Riffle, otherwise run (C1) SB: Proportional macrophytes, margin, woody debris (C2)	1 m^2, 5 kicks for HB	200+ fixed count	Clapcott	Y
ES	NEMS	15	1	K, 0.5mm, 35cm	Habitat proportional	6 to 7 kicks	200+ fixed count	Stark 2001	Y
ES	In-house	15	1	Surber, 0.5mm, 0.1m^2	HB: Riffles, runs SB: Proportional	5 to 7 Surbers or kicks, mix at a site	200+ fixed count	Stark 2001	Y
HBRC	NEMS	14	0	K, 0.5mm, 40cm	Habitat proportional	5 kicks, ~ 0.8 m^2	200+ fixed count	In house list	Y
HBRC	In-house	14	0	Surber, 0.5mm, 32cm	Riffle if present, otherwise run	5 samples, ~ 0.5 m^2	200+ fixed count	In house list	Y

5 Sample processing and metric calculation methods

Macroinvertebrate metric scores calculated from collected samples can be influenced by how the samples are processed and how the metric is calculated, particularly by:

1. Differences in the taxonomic level of identification and associated tolerance values used to calculate metric scores, and
2. How sub-sampling is undertaken when the whole sample is not processed, which may affect the recorded taxa and their abundances, from which the metrics are calculated.

Differences in taxonomic resolution and tolerance values occur because not all taxa have assigned MCI tolerance values and because taxa are often identified to different taxonomic levels, e.g., genus or family level by different agencies or sample processors. There are currently published MCI HB tolerance values for 180 taxa (NEMS Annex D, from Clapcott et al. 2017, an update of Stark and Maxted 2007). A further recommended list of HB tolerance values for 20 taxa is provided in NEMS Annex H. Greenwood et al. (2015) calculated tolerance values for additional taxa, resulting in HB scores for a total of 234 taxa. However, pre-existing scores of some taxa were revised in this report and the updated values have not been widely adopted or included in the NEMS. Often, agencies add taxa that are commonly identified in their region to the standard list of 180 to increase the taxa pool from which the metrics are calculated. Due to regional differences in community composition and in perception of taxon tolerances to local stressors, the taxonomic list and tolerance values used by individual councils to generate MCI, QMCI and ASPM metrics are likely to differ slightly.

Sub-sampling increases efficiency of sample processing as the full sample does not need to be processed. However, the methods must balance efficiency with the sub-sample being a 'representative' measure of the taxa and their abundances in the full sample. The missed taxa scan increases sample processing time and influences resulting metric scores. Understanding the impact that the additional taxa from the missed taxa scan have on metric scores can help quantify the benefits of this processing step.

In this section, we compare differences in macroinvertebrate metrics scores between:

1. Council-provided metric values, where available, and recalculated values using a standardised taxonomic resolution and tolerance value list (Annexes D and H from NEMS).
2. Recalculated standardised metric values both including and excluding the taxa from the missed taxa scan, where possible.

5.1 Influence of taxonomic resolution and tolerance values

We compiled macroinvertebrate metrics as provided by the councils and recalculated metrics from the councils' raw invertebrate data for all samples using the standard list in the macroinvertebrate NEMS (Annex D) and the additional taxa in Annex H. The NEMS states that "Macroinvertebrate taxonomic resolution shall be at least to a level consistent with that outlined under the MCI, as provided in Annex D." Thus, taxa can be identified to lower taxonomic levels than those listed in NEMS Annexes D and H; however, we used this list to standardize taxonomic resolution across councils.

There were 258 samples in total, and councils provided metric values for some samples:

- taxonomic richness for 168 samples,
- MCI-HB for 188 samples,
- QMCI-HB for 170 samples,
- ASPM-HB for 161 samples,
- EPT taxa richness for 170 samples,
- EPT percent taxa for 102 samples, and
- EPT percentage abundance for 187 samples.

Not all councils provided MCI-SB metrics for the 24 soft-bottomed samples, so we used the council-provided MCI-HB values for samples from both hard-bottomed and soft-bottomed sites. For the NEMS samples, we re-calculated MCI-HB and MCI-SB metrics using HB tolerance values for samples from HB sites and SB tolerance values for samples from the 12 SB sites, respectively. We note that, for the SB sites, differences between council-provided and recalculated metrics were likely due to the different tolerance values used (SB vs HB), an artefact of the data we had available, rather than differences in taxonomic resolution and tolerance values between councils and NEMS protocols. However, any differences in metric values between those recalculated and provided by the council for HB sites are due to differences in the taxonomic resolution and tolerance values.

Graphs of provided and recalculated metrics were plotted with a 1:1 line. We identify SB and HB sites in the plots to allow visualisation of the differences caused largely by SB vs HB tolerance values (SB sites) and those caused by taxonomic resolution and tolerance values (HB sites).

5.2 Results

Across the 168 samples with taxa richness values provided by the councils, in just under half (46%) the samples the councils identified more taxa than were on the list of taxa and tolerance scores available in the NEMS (Annexes D and H; Figure 5-1A). The average and maximum additional taxa identified per sample by the councils compared to the NEMS taxa list were 0.7 and 10 taxa, respectively.

The differences in metric values provided by the councils (using their own taxonomic resolution and tolerance values) and those calculated using the standard NEMS list were relatively small, particularly for the NPS-FM attributes (MCI, QMCI and ASPM, Figure 5-1). For example, the difference in MCI was generally <5, where NPS-FM MCI attribute band boundaries are 20 units different.

The largest differences between metric values were when SB or HB tolerance values were used to generate metrics for the same sample. For example, samples with an MCI-HB value of approximately 75 had MCI-SB values ranging from <45 to >90 (Figure 5-1C). Similar, but not quite so extreme, differences were observed for QMCI and ASPM values calculated using SB or HB tolerance values. However, differences between SB and HB metric scores were sample-specific, with some samples showing little difference in metric scores regardless of whether SB or HB tolerance values were used.

Several metrics showed outliers that may be caused by data entry or calculation errors. For example, the provided ASPM value of 0.8 is perhaps a data entry or calculation error (Figure 5-1E). Likewise,

the one sample with a lower taxa richness provided by the council than calculated from the raw data is a data error in one of the datasets as no additional taxa could be identified using the NEMS taxonomic list. For a few samples, the EPT taxa richness values were higher in recalculated values than in those provided by the councils. We are unsure why this has occurred. Both metrics exclude hydroptilidae caddisflies.

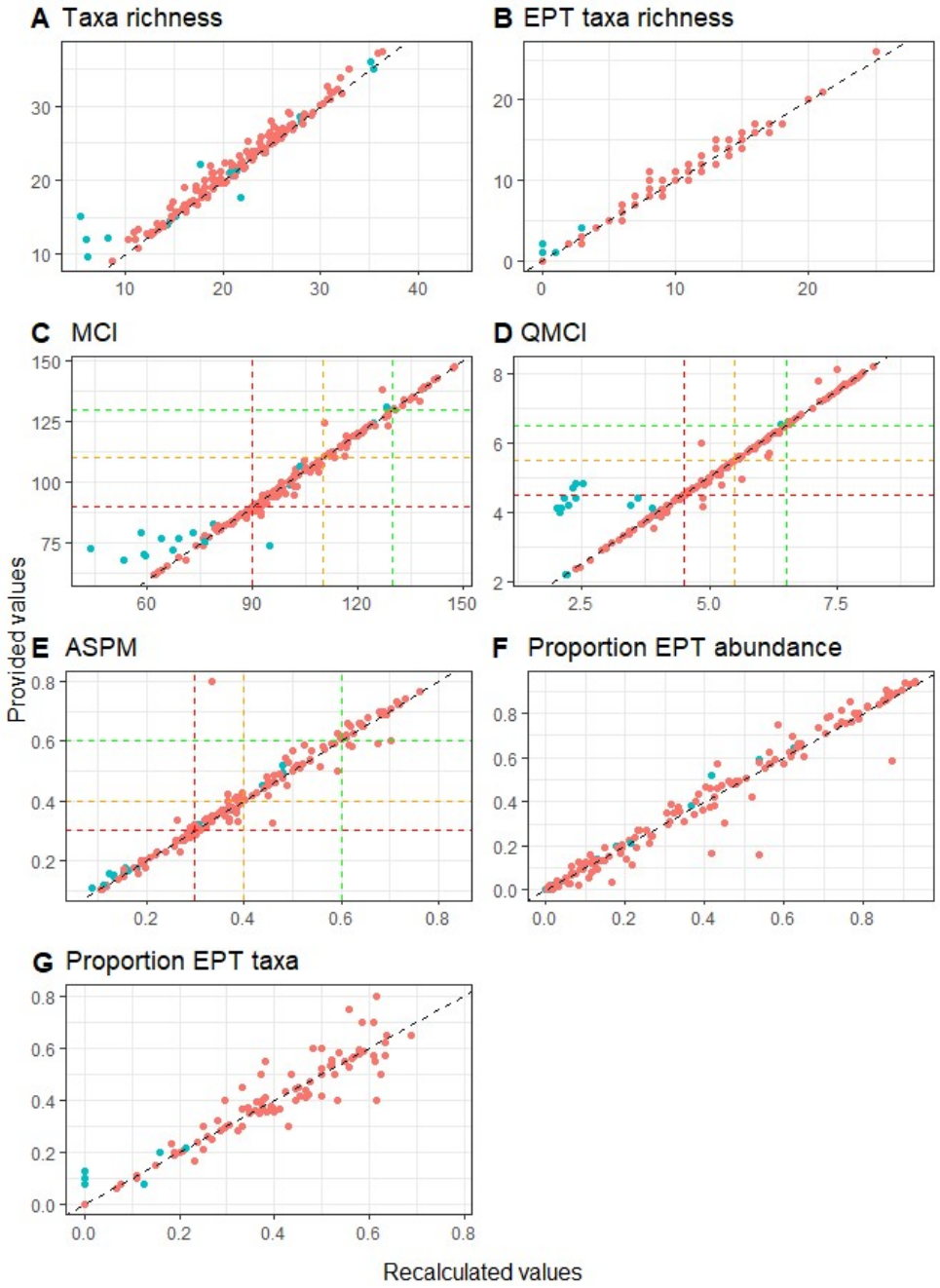


Figure 5-1: Macroinvertebrate metrics as provided by councils and recalculated using NEMS tolerance values and taxa list. Note that the MCI, QMCI and ASPM values provided by the councils used HB tolerance values because SB variants had not been calculated for all soft-bottomed sites. For the 12 sites identified as soft-bottomed by councils (blue dots) we calculated metrics using SB tolerance values. EPT taxa richness points have been jittered to aid visibility of individual points. For NPS-FM macroinvertebrate attributes, green dashed lines indicate the Band A/B boundary, orange the Band B/C boundary and red the band C/D boundary.

5.3 Influence of the missed taxa scan

The influence of the missed taxa scan on macroinvertebrate metrics was evaluated for 201 of the paired samples analysed using the 200+ Count Plus Missed Taxa Scan protocol and in which the taxa identified in the missed taxa scan could be detected. ECAN in-house (n = 31) and TRC (n = 8) samples could not be included as taxa found in the missed taxa scan could not be identified either due to the processing method or where not identified separately in the datasheet provided. AC in-house samples and NRC samples were processed using full count methods and thus excluded from this analysis.

Graphs of metrics with and without the missed taxa scan were plotted with 1:1 lines and summary statistics were generated of the difference between provided and recalculated values for each metric.

5.4 Results

Between zero and 12 additional total taxa and zero and six additional EPT taxa were found per sample in the missed taxa scan. In 52% and 54% of the samples, the missed taxa scan identified no additional total taxa or EPT taxa, respectively (Table 5-1, Figure 5-2).

When taxa from the missed scan were included, the percentage of EPT taxa was between 11% lower and 21% higher than when they were excluded. Fifty four percent of samples retained the same value (Table 5-1, Figure 5-2).

Abundance-based metrics such as the QMCI and proportion EPT individuals were only very slightly different with or without the missed taxa scan (Table 5-1), as would be expected given only the presence (recorded as a value of 1 or 0.5) of additional taxa in the missed scan is recorded (Table 5-1, Figure 5-2).

MCI was generally, but not always higher if the missed taxa scan was included, leading to a change in metric values between -14 units and +18 units (compared to NPS-FM band thresholds 20 units apart for MCI), with an average increase of 1.8 units when the missed taxa scan was included (Table 5-1). Fifty six percent of samples retained the same MCI value with or without the missed taxa scan. ASPM was always higher when the missed taxa scan was included, with 11% of samples retaining the same value and changes ranging between no difference and an increase of 0.1 (Table 5-1, Figure 5-2).

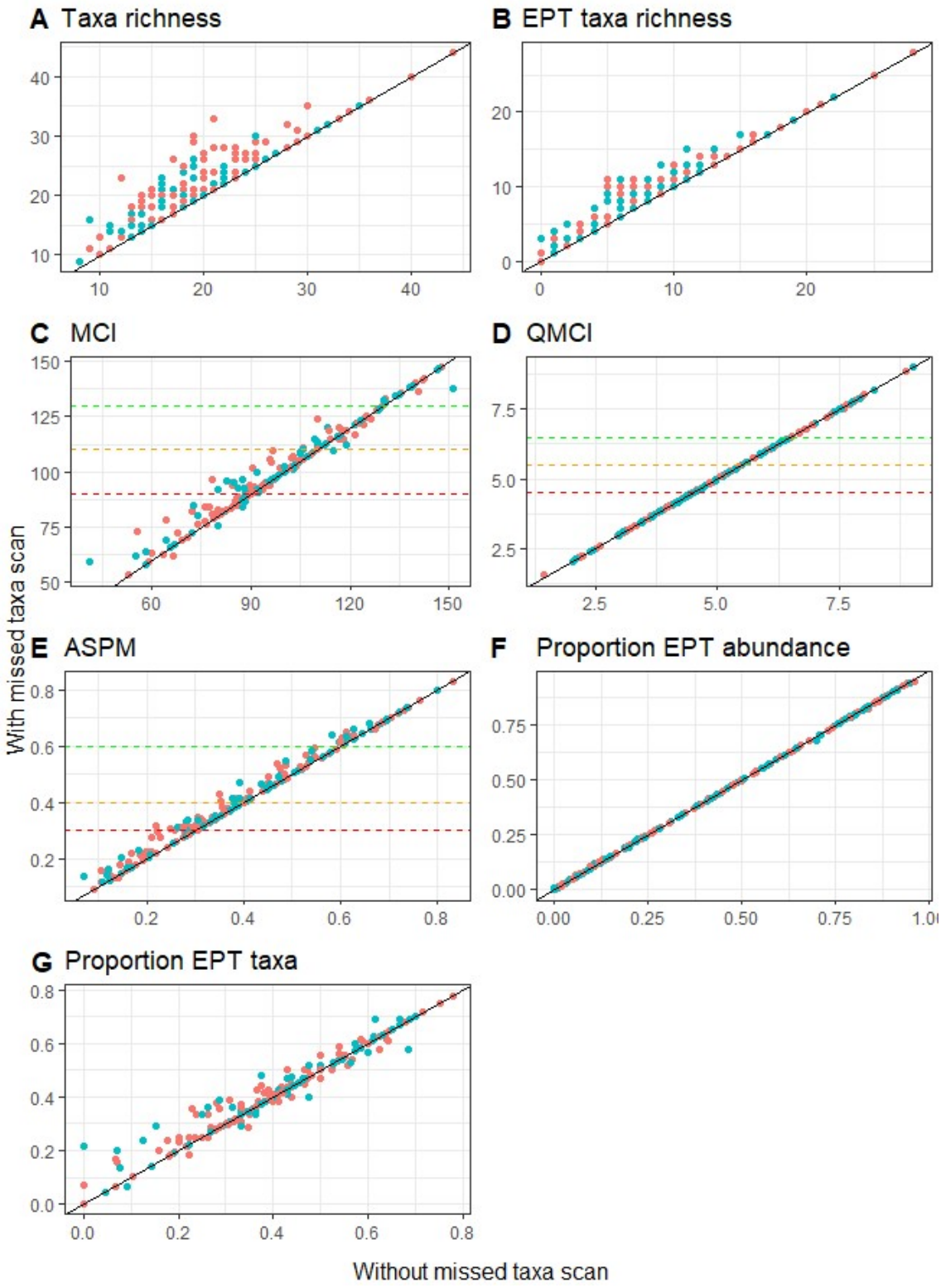


Figure 5-2: Macroinvertebrate metrics calculated for 201 samples processed using the 200+ Count with Scan for Missed Taxa using data with and without taxa from the missed taxa scan. Blue symbols are in-house samples and peach are NEMS protocol samples. Red, orange and green horizontal lines indicate the Band C/D, Band B/C and Band A/B boundaries for NPS-FM attributes.

Table 5-1: Changes in macroinvertebrate metric scores when taxa from the missed taxa scan were included and excluded from metric calculations. Differences are calculated from metrics inclusive of the missed taxa (MT) scan – metrics excluding the missed taxa scan. n = 201 samples.

	Max decrease with MT scan	Max increase with MT scan	Mean change	No. (%) samples same
Taxa Richness	0 taxa	12 taxa	1.7 taxa	105 (52%)
EPT taxa richness	0 taxa	6 taxa	0.9 taxa	108 (54%)
Percentage EPT taxa	-11 %	21 %	1.0%	108 (54%)
Percentage EPT individuals	-1.5 %	1.2 %	0.002%	22 (11%)
MCI	-13.6	18.6	1.6	105 (52%)
QMCI	-0.06	0.08	0.002	20 (10%)
ASPM	0.0	0.1	0.01	21 (10%)

5.5 Summary

- Councils are using similar but not identical taxonomic lists and tolerance values (see Table 4-1 for council reported lists). The small regional differences in taxonomic resolution and tolerance values from in-house lists and the NEMS recommended list had relatively small effects on most resulting metric values.
- While on average there were few (0.7 per sample) additional taxa reported by councils than present on the NEMS taxonomic list, up to 10 additional taxa were identified in one sample.
- Switching between HB and SB tolerance values had much greater impacts on metric values for some, but not all SB samples. This difference was not related to council methods and was an artefact due to the unavailability of provided SB metric scores for all SB sites. However, it does highlight that appropriate use and consistency in the use of SB or HB tolerance values at a site is required for robust temporal and spatial comparison of MCI and QMCI scores.
- The inclusion of taxa from the missed taxa scan resulted in up to 12 additional total taxa and 6 additional EPT taxa reported in a sample.
- In just over half the samples, the missed taxa scan resulted in no additional total or EPT taxa being reported.
- Abundance-based metrics (QMCI, %EPT abundance) were largely unimpacted by the inclusion or exclusion of the taxa from the missed taxa scan.
- MCI, %EPT taxa richness and APSM were almost always higher in samples when taxa from the missed taxa scan were included.

6 Macroinvertebrate community and metric differences between NEMS and in-house samples

In this section, the recalculated metrics using consistent taxonomic resolution and metric values were compared between samples collected using NEMS and in-house protocols. The goal was to investigate differences in macroinvertebrate metric values due to collection and processing protocols while excluding any effects of differences in taxonomic resolution and tolerance values used to calculate the metrics.

The samples collected and processed using council in-house and NEMS protocols were assessed by:

1. Identifying the unique taxa collected using each protocol both overall and for each council separately.
2. Using ordinations to investigate community compositional differences between NEMS and in-house samples overall and individually for each council. Any key taxa associated with significant differences in composition between the protocols were identified.
3. Assessing differences in metric scores (MCI, QMCI, ASPM, taxa richness, %EPT taxa, %EPT abundance) by:
    A. linear correlations, and
    B. assessments of the difference in metric scores across the range of mean metric values (Bland-Altman analyses) to visually assess differences in metric scores between the sample types. The magnitudes of the differences were assessed relative to differences in metric values between thresholds for NPS-FM attribute bands for MCI, QMCI and ASPM.
4. Using generalised linear models across the combined council data set, including a term for council, to investigate potential causes of differences in metrics scores between NEMS and in-house samples, such as the number of additional habitats sampled under NEMS.

More detailed methods are provided in each section below.

6.1 Unique taxa

Across all 258 samples, 138 different taxa were identified (to MCI level under Annex D and H of the NEMS). Across all regions unique taxa were collected using both in-house and NEMS protocols. Taxa unique to in-house samples were the dipterans *Zelandotipula*, Tipulidae and Sciomyzidae, the worms Rhabdocoela and Nematomorpha and the freshwater mussel Hyridella = Ecyhridella. The taxa collected only using the NEMS protocol were the caddisfly *Alloecentrella*, damselflies *Austrolestes* and *Ischnura*, the dipterans *Calopsectra*, *Nothodixa*, *Parochlus,* Pelecorhynichidae and the snail *Glptophysa* = *Physella*.

In all individual regions, unique taxa were also collected using both sampling protocols (Table 6-1). Across all paired samples collected by a council, all councils except HRC and NRC collected both more taxa overall and more unique taxa using the NEMS rather than in-house protocols (Table 6-1).

See Section 6.3.2 for more detailed discussion of taxonomic richness differences between the protocols.

Table 6-1: The total taxonomic richness and taxa unique to each protocol for samples collected using NEMS and in-house protocols for each council. The taxonomic list recommended by the macroinvertebrate NEMS (Annexes D and H) was used to allow comparisons between councils. TR = taxonomic richness.

Council	TR NEMS	TR In-house	Taxa unique to NEMS	Taxa unique to in-house
AC	72	69	*Ameletopsis, Beraeoptera,* Hydrophilidae, Ceratopogonidae*,* Chironomidae*, Gyraulus, Ichthybotus,* Nematoda, Psychodidae, Scirtidae, *Tepakia, Zelandobius, Zelandotipula*	*Aprophila, Dolomedes, Enochrus, Hudsonema,* Hyridella = Echyridella, *Megaleptoperla, Mischoderus,* Nemertea, Oeconesidae, *Oxyethira, Zelandoptila*
BOP	64	53	*Acanthophlebia*, *Ameletopsis*, Anisoptera, *Berosus,* Collembola, *Corynoneura,* Ephydridae, Hydrophilidae, *Ichthybotus, Microvelia,* Nematoda, *Paradixa, Paratya, Paroxyethira,* Psychodidae, Ptilodactylidae	*Lobodiamesa*, *Molophilus, Siphlaenigma*, *Stenoperla, Zelandobius*
ECAN	91	76	Amphipoda, Anisoptera, *Calopsectra*, *Chironomus*, Collembola, *Corynoneura*, *Gundlachia* = *Ferrissia*, *Harrisius*, Hexatomini, Hirudinea, *Hydrochorema, Ischnura, Liodessus, Lobodiamesa, Maoridiamesa,* Nemertea, *Oniscigaster, Parochlus, Polypedilum, Stictocladius, Tanytarsini, Xanthocnemis*	*Atalophlebioides, Beraeoptera,* Isopoda, *Mischoderus,* Tabanidae, Tipulidae, *Zephlebia*
ES	72	59	Amphipoda, *Austrolestes, Beraeoptera, Berosus, Corynoneura, Dolomedes,* Empididae, Ephydridae*, Hygraula, Microvelia, Neozephlebia, Paraleptamphopus, Paralimnophila, Paranephrops,* Psychodidae, *Sigara,* Tanypodinae, *Xanthocnemis, Zephlebia*	Collembola, Hirudinea, *Limonia,* Nematomorpha, *Neocurupira, Neurochorema*
GWRC	91	88	Anisoptera, *Atalophlebioides, Austrolestes, Enochrus,* Isopoda, *Lobodiamesa,* Lymnaeidae, *Neocurupira,* Stratiomyidae, *Zelandobius, Zelolessica*	Ceratopogonidae, *Megaleptoperla,* Oeconesidae, *Paralimnophila,* Psychodidae, Ptilodactylidae, Sciomyzidae, *Tepakia*
HBRC	66	65	*Acroperla*, Ceratopogonidae, *Corynoneura*, *Gundlachia* = *Ferrissia*, Isopoda, *Nothodixa, Paratya,* Pelecorhynchidae, *Rallidens, Triplectides*	Chironomidae, *Helicopsyche, Liodessus,* Lymnaeidae, *Molophilus, Neozephlebia, Paralimnophila, Sigara,* Sphaeriidae
HRC	51	54	*Ameletopsis*, Hirudinea, Nematoda, *Physa* = *Physella*, Platyhelminthes, Tabanidae, *Zelandobius*	*Corynoneura*, *Enochrus, Gundlachia* = *Ferrissia*, Hexatomini, *Ichthybotus*, *Megaleptoperla*, Muscidae, *Neozephlebia*, *Polyplectropus*, *Stenoperla*

Council	TR NEMS	TR In-house	Taxa unique to NEMS	Taxa unique to in-house
NRC	55	57	*Chironomus,* Copepoda, *Gundlachia* = *Ferrissia,* Hydrophilidae, *Microvelia*, *Paratya, Paroxyethira, Polyplectropus, Sigara, Xanthocnemis*	*Austroperla,* Ceratopogonidae, *Costachorema, Latia,* Lymnaeidae, Muscidae, *Neurochorema, Oniscigaster, Paracalliope, Plectrocnemia, Pycnocentria, Zelandobius*
TRC	33	32	Isopoda, *Neurochorema*, Oeconesidae, *Zelandobius*	*Austroperla*, Hydraenidae, *Paranephrops*
WRC	84	69	*Acroperla, Antipodochlora, Arachnocolus,* Ceratopogonidae, *Chironomus,* Copepoda, Empididae, *Glyptophysa* = *Physastra, Harrisius,* Hirudinea, *Hygraula, Lobodiamesa,* Lymnaeidae, *Paralimnophila, Paratya, Paroxyethira, Siphlaenigma,* Tabanidae, *Xanthocnemis, Zelandobius*	*Hydrochorema,* Hydrophilidae, Rhabdocoela, Scirtidae, *Stictocladius*, *Zelolessica*

6.2 Community composition

6.2.1 Methods

To identify differences in community composition between samples collected using council in-house or NEMS protocols, we used non-metric multidimensional scaling (NMDS) ordinations on macroinvertebrate community data for all councils together, and then separately for each individual council, where sufficient data were present.

Ordinations cluster sites in multidimensional space based on community composition with the first few dimensions explaining most of the variation. Generally, the first two or three dimensions (or axes) are visualised in bi-plots. Sites that have more similar community composition are clustered more closely together than sites with more different community composition. Ordinations based on taxa presence/absence data rely on the addition or removal of taxa to identify differences between samples. The use of abundance data provides additional information about changes in taxa abundances before a taxon is added or lost. However, for comparisons of taxa abundances between samples to be robust, they need to be in the same units of abundance. To reduce as much of the variation in abundances between methods, we scaled abundances per subsample to the full sample, for samples where sub-sampling was undertaken (e.g., using the 200+ Count or barrel splitter). This was possible for all samples except the ECAN in-house samples because the full subsample was not processed.

We were also not able to standardise invertebrate abundances in the samples to a per unit area value (e.g., abundance per square metre) for all councils because we could not calculate the area sampled for all samples (see Table 4-1). We assumed that abundances per sample approximated abundances over a similar area (reported areas sampled ranged from 0.3 m^2 to 1 m^2 per sample, although commonly was between 0.75 m^2 and 1 m^2).

Due to these differences in sample collection and processing methods within and between protocols and the inability to scale up to comparable areal abundances (e.g., individuals per square metre) for all samples, the abundance data for ordinations was processed in several different ways:

1. Relative abundance data were used for the all-council comparison and individually for councils that utilised a different processing method between NEMS and in-house samples (ECAN and AC).
2. Abundances per sample were used for all other councils.
3. Ordinations were also run using presence/absence data for all councils together and all councils individually.

Abundance data were square-root Wisconsin transformed before ordinations. Presence/absence and relative abundance data were not transformed. Bray-Curtis distances were calculated as the dissimilarity measure. Ordinations were considered appropriate if the stress value was <0.2 and the model converged. Bi-plots of the first three ordination axes were visualised to investigate community compositional differences between NEMS and in-house samples.

Permanova (permutational ANOVA, with 100 permutations) was used to determine whether any significant differences in community composition existed between the NEMS and in-house protocols for individual councils and between protocols, councils and their interaction for the combined dataset from all councils.

6.2.2 Results

For the combined dataset from all councils (258 samples from 129 sites), community composition (based on relative abundances) varied significantly between councils (permanova: $R^2 = 0.17$, $F_{9, 238} = 5.5$, $p = 0.001$), as would be expected due to regional differences in communities. Community composition was not significantly different between the NEMS and in-house samples ($R^2 = 0.005$, $F_{1, 238} = 1.6$, $p = 0.1$; Figure 6-1) and there was no interaction between council and sample type (permanova $R^2 = 0.01$, $F_{9, 238} = 0.5$, $p = 1.0$). The ordination based on presence/absence data ordination showed similar results for differences between councils and the interaction of council and sample type. The effect of sample type (NEMS vs in-house) was borderline significant at an alpha of 0.05 ($F_{1, 238} = 2.0$, $p = 0.05$), although very little of the variation in presence/absence of taxa was explained by the sample type ($R^2 = 0.006$, c.f. council $R^2 = 0.22$) and little evidence of separation in the protocols was visually evident in any 2-dimensional ordination plots involving the first three axes (for an example of axis 1 and axis 2 see Figure 6-1).

ECAN (64 samples from 32 sites) was the only council to show a significant difference in community composition between NEMS and in-house samples in either abundance (relative or raw abundance, depending on council) or presence/absence community data. The relative abundance of taxa in paired samples collected by ECAN differed significantly between protocols (permanova $R^2 = 0.06$, $F_{1, 60} = 3.6$, $p = 0.001$), although with relatively little of the variation between samples explained by the two protocols (7%, Figure 6-2). The main macroinvertebrate taxa that were significantly correlated with the shift in community composition between sample types were the mayfly *Neozephlebia,* the caddisfly *Helicopsyche,* the dipterans *Neocurupira* and *Mischoderus,* and the Hydraenidae beetle family, all of which were generally more common in in-house samples rather than in NEMS samples (Figure 6-2). The ordination based on presence/absence data also showed a significant difference between samples using the two protocols for ECAN samples (permanova $R^2 = 0.06$, $F_{1, 60} = 3.8$, $p =$

0.004). Differences in the taxa present or absent in NEMS and in-house samples were best explained by NMDS axis 3 (Figure 6-2).

There were no significant differences in the macroinvertebrate communities between NEMS or in-house samples collected by GWRC, NRC, HRC, WRC, AC, ES, HBRC or BOP based on either abundance (relative abundance for AC) or presence/absence data. Ordinations on abundance data for TRC did not converge, probably due to insufficient data (only four paired samples), and no conclusions could be drawn for this council.

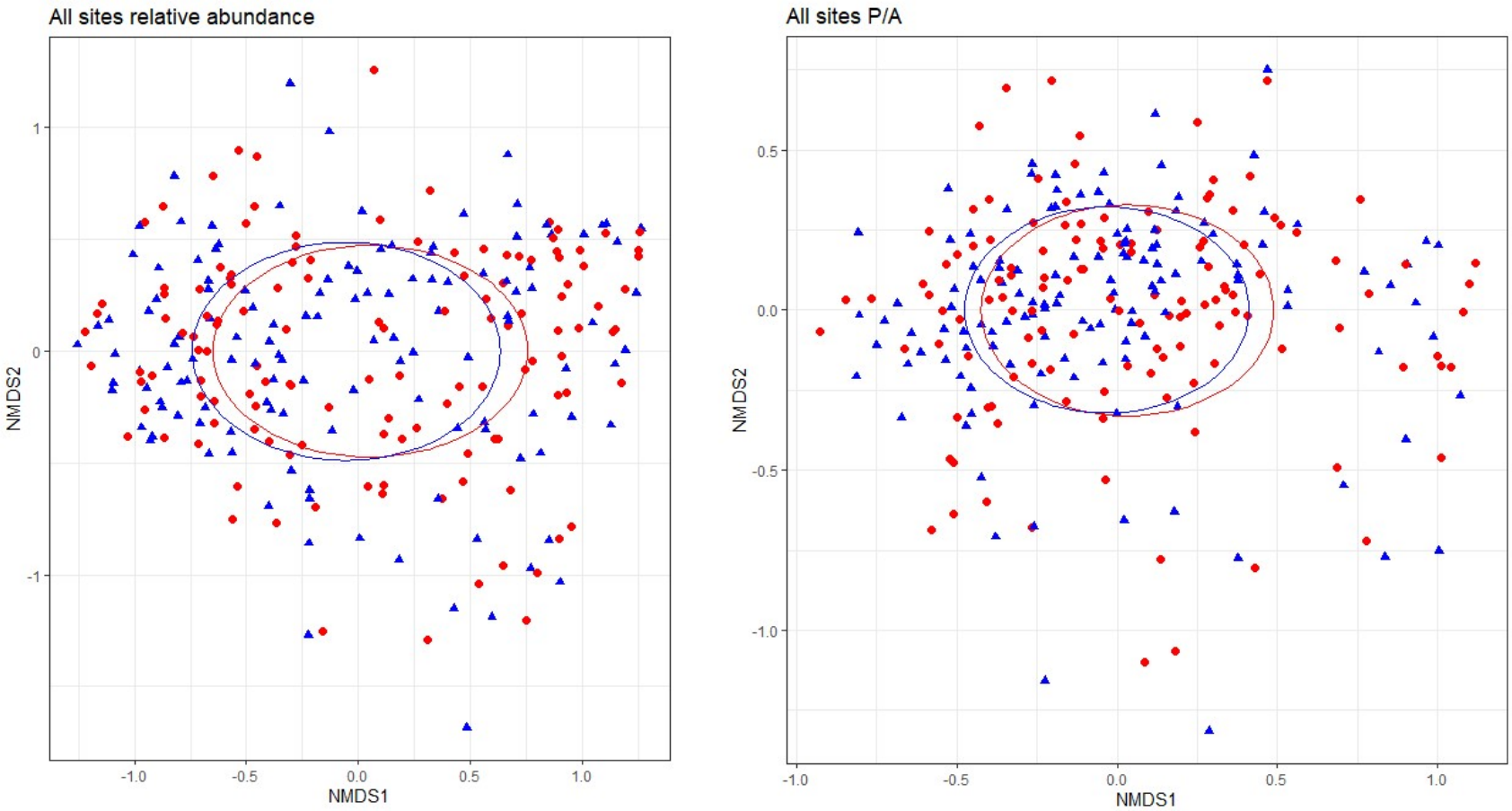


Figure 6-1: Non-metric multidimension scaling (NMDS) ordination on 258 paired macroinvertebrate samples collected using NEMS (red circle) and council in-house (blue triangle) protocols at 129 sites by ten councils using relative abundance (left) and presence/absence (right) data. Ellipses are centred on the average value for the NEMS and in-house protocols. The stress value was 0.15 for relative abundance and 0.16 for presence absence data.

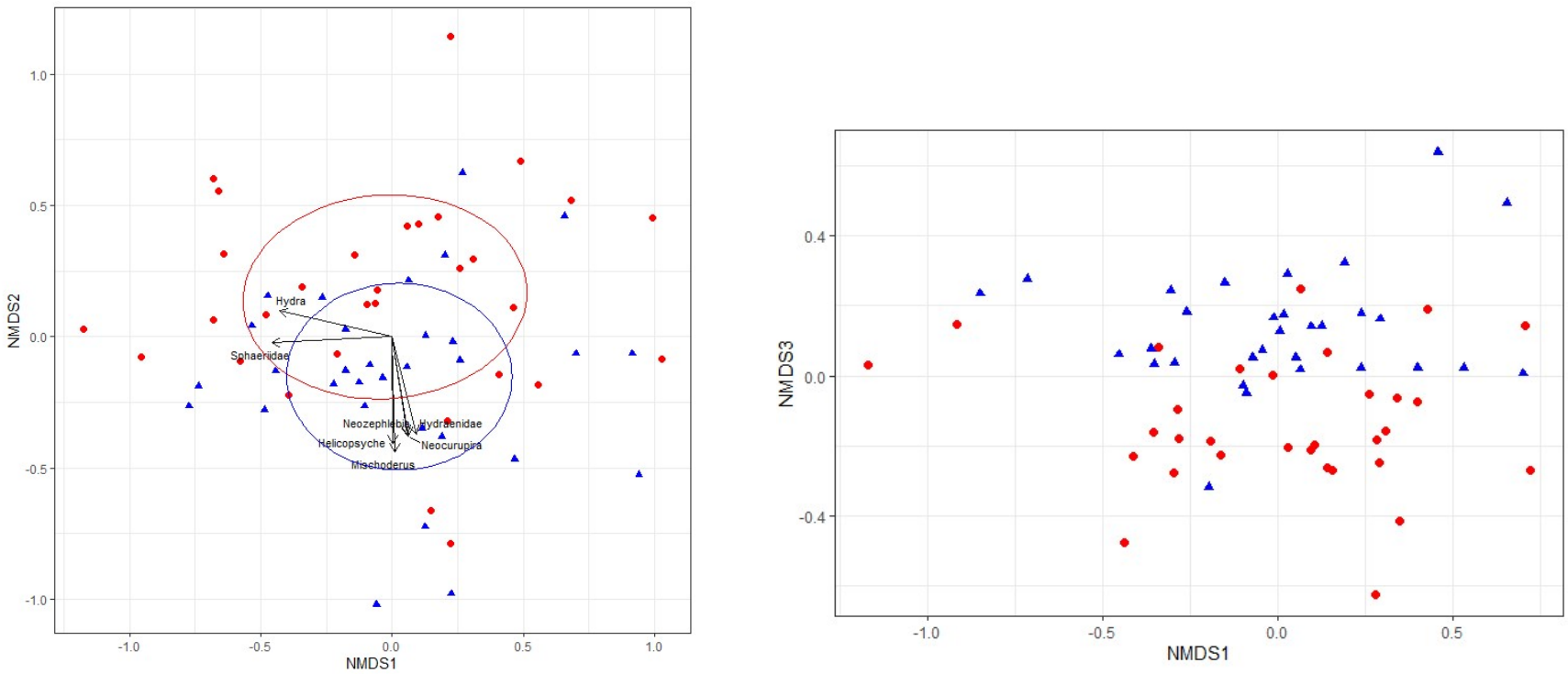


Figure 6-2: Non-metric multidimensional scaling (NMDS) ordinations of paired macroinvertebrate samples collected at the same sites using NEMs (red circle) and in-house (blue triangle) protocols by Environment Canterbury (ECAN) analysed using relative abundance data (left) and presence/absence data (right). The community composition of samples collected by ECAN differed significantly between the two sampling methods using both data types (see main text). Note the different NMDS axes plotted on the two graphs. Macroinvertebrate taxa that were significantly correlated with the NMDS axes ($p<0.01$) are indicated on the relative abundance plot. Ellipses are centred on the average value for the NEMS and in-house protocols. Stress values were 0.13 and 0.14 for relative abundance and presence absence data, respectively.

6.3 Metric differences

6.3.1 Methods

Pearson correlations were used to calculate the degree of correlation between metrics calculated from the NEMS samples and the in-house samples using all council data combined. See Appendix C for plots for individual councils for each metric.

Correlations can indicate whether the metric values from the two protocols co-vary, but further investigation is required to determine whether: 1) differences between the metrics from the two protocols are sufficiently small to allow them to be used interchangeably and 2) if the relationship between the metrics follows a 1:1 line (for example, if one protocol consistently gives higher numbers than the other, then intercept ≠ 0, or if there is bias between the methods across the metric values, then slope ≠ 1).

Bland-Altman plots (Altman and Bland 1983) are among the most highly cited methods for use in assessing concordance between two methods of measurement of continuous variables (i.e., >39,000 citations of Bland and Altman 1986). The Bland-Altman method involves creating plots with the difference in the measurements from the two methods on the y-axis and the average of the measurements from the two methods on the x-axis (as an indicator of the 'true value'). Inspection of the scatter of points on the plots allows assessment of whether:

1. There is a systematic bias in measurements from the two methods, i.e., does one method consistently record higher values than the other?
2. Measurement error is associated with the true value, i.e., does the difference between the two methods get consistently larger or smaller as the average of the methods increases or decreases?
3. There are any occasional large differences between methods or outliers.

Lines indicating two standard deviations of this mean value are also added to the plots to indicate the limits of agreement between the methods. The user then assesses whether the magnitude of the limits of agreement (i.e., the values within which 95% of the data points fall) are acceptable, depending on the level of accuracy required between the two methods.

We created Bland-Altman plots separately for each metric using the combined council data to visualise the degree of difference between scores from the two protocols and used them to:

1. Assess whether measurement error was associated with the average metric values e.g., were sites with higher average MCI values more likely to have a greater difference in MCI values between the two protocols. This was done visually and through separate linear models for each metric to test whether the difference in metric values from the two protocols was significantly related to the mean metric values between the protocols.
2. Determine overall mean bias and limits of agreement (two standard deviations or range of values within which 95% of difference values between the protocols occur).
3. Compare the magnitude of the differences between protocols to potentially acceptable differences in metric scores between protocols for use in SOE monitoring.

For NPS-FM attributes we compared the differences to metric values required to move between NPS-FM band thresholds.

The differences in metric scores on the y axis were plotted as NEMS minus in-house, with the average of the NEMS and in-house metric scores on the x axis. Mean bias was plotted as the overall mean of the differences in metric scores. To aid in interpreting the scale of the difference in metric values between the protocols, we added horizontal lines indicating limits of agreement that were either 2 standard deviations of the mean bias (for non-NPS-FM metrics) or the minimum difference between bands for MCI, QMCI and ASPM attributes under the NPS-FM.

6.3.2 Results

Total and EPT taxonomic richness results

Both total and EPT taxonomic richness were significantly positively correlated between the two protocols (total TR $r = 0.67$, EPT TR $r = 0.85$, $p < 0.001$ for both; Figure 6-3).

Bland-Altman plots and associated regressions of the difference between the protocol metric values indicated no bias in differences between the protocols associated with variation in the average metric value, a relatively low overall bias between the protocols (green lines; Figure 6-4A and B). However, some of the differences in overall richness and EPT richness between protocols were large (Figure 6-4A and B). For example, the difference in taxonomic richness between NEMS samples and in-house samples ranged from 13 fewer to 13 more taxa identified in NEMS than in in-house samples, with an average value of 1.8 additional taxa collected in NEMS samples than in in-house samples (Figure 6-4A and B). Taxonomic richness was the same in 13 (10%) of the 129 paired samples (Table 6-2).

The taxonomic richness of EPT in paired NEMS and in-house samples ranged between nine fewer to seven more taxa collected in the NEMS samples than in in-house samples, with an average of 0.2 more EPT taxa collected in NEMS samples (

Table 6-2, Figure 6-4A and B). EPT taxa richness was the same in 34 (26%) of the paired 129 samples (

Table 6-2).

For three councils (AC, ES, WRC), taxonomic richness between the two sample types was generally positively correlated but was more commonly higher in the NEMS samples than in the in-house samples (Appendix C). For the other councils, the relationship between taxa richness in both sample types was either generally positive with a variable degree of higher taxonomic richness occurring at a roughly similar frequency in both sample types (e.g., HBRC and HRC) or showed very little correlation in taxonomic richness between the sample types (e.g., GWRC, NRC; Appendix C).

Percentage EPT abundance

Generally, the percent abundance of EPT individuals in NEMS samples was positively correlated with a higher percentage EPT abundance in in-house samples (Table 6-2; Figure C-1). There was no significant bias between protocols across the mean values and the average difference was low (4.1% lower EPT abundance in NEMS samples). However, for some samples, the difference in metric scores between protocols was high, ranging between 44% lower and 1.6 % higher in NEMS samples compared to in-house samples (Table 6-2, Figure 6-4). Note also that the Bland-Altman plot for percent abundance is influenced by the upper and lower limits of percent abundance. For example,

samples with average values approaching either 0% or 100% are more constrained in how different they are likely to be, compared to samples with percent abundance values more in the middle of the possible values. See Appendix C for correlation plots for individual councils.

MCI, QMCI and ASPM

The three NPS-FM metrics MCI, QMCI and ASPM were positively and significantly correlated between protocols (Figure 6-3). See Appendix C for correlation plots for councils separately.

Bland-Altman plots and correlations indicated no significant bias between the protocols across the average metric values, and low average differences between the protocols for all metrics, with all metrics on average slightly lower in the NEMS samples (Figure 6-4, Table 6-2).

However, the three metrics all had at least one site where the difference between NEMS and in-house values was greater than the difference between NPS-FM band thresholds (Table 6-2).

Relative to the NPS-FM attribute bands, at 39 (30%) sites the MCI value was different by one provisional NPS-FM Band between sampling methods. At 30 of those sites the MCI value moved to a lower band with NEMS methods instead of in-house methods, and at nine sites the MCI value moved into a higher band. For QMCI, at 32 sites (25%) the difference in value between the two methods shifted the value between NPS-FM Bands. At 20 sites the NPS-FM Band was lower in the NEMS samples and at three of these sites this shift was across two bands. At twelve sites the QMCI value was in a higher band with NEMS methods than with in-house methods.

The proportion of sites that shifted MCI and QMCI bands differed by council (Table 6-3), with 50–100% of sites retaining the same provisional attribute band under the NPS-FM depending on the protocol.

Table 6-2: Differences and correlations between macroinvertebrate metrics scores calculated on 129 paired samples collected using NEMS and council in-house protocols. Values were calculated as NEMS – in-house. Min. = minimum, max. = maximum, med. = median. The NPS-FM bands for MCI, QMCI and ASPM are 20, 1 and either 0.1 or 0.2 units apart, respectively. Pearson correlation tests were run on the combined council data set and correlation coefficients and p values are shown. See Figure 6-3 for regression plots for metrics.

	Difference (NEMS– in-house)					Correlation test	
	Min.	Max.	Med.	Mean	No. samples same value	Correlation coef. r	P value
Taxonomic richness	-13	13	2	1.8 taxa	13 (10 %)	0.67	<0.001
EPT taxa richness	-9	7	0	0.2 taxa	34 (26 %)	0.85	<0.001
Percentage EPT abundance	-44	29	-1.6	-4.1 %	1	0.90	<0.001
MCI	-20.8	21.0	-3.3	-2.9	0	0.92	<0.001
QMCI	-2.2	1.8	-0.08	-0.12	0	0.93	<0.001
ASPM	-0.22	0.14	-0.007	-0.16	0	0.93	<0.001

Table 6-3: Shift in provisional NPS-FM bands for QMCI and MCI when samples were collected using in-house or NEMS protocols by council.

Council	Number sites sampled	% (number) sites QMCI band same	% (number) sites MCI band same
AC	8	88 % (7)	88 % (7)
BOP	6	50 % (3)	67 % (4)
ECAN	31	71 % (22)	68 % (21)
ES	15	87 % (13)	67 % (10)
GWRC	23	74% (17)	70 % (16)
HBRC	14	57 % (8)	93 % (13)
HRC	12	83 % (10)	58 % (7)
NRC	5	80 % (4)	60 % (3)
TRC	4	100 % (4)	50 % (2)
WRC	11	82 % (9)	64 % (7)

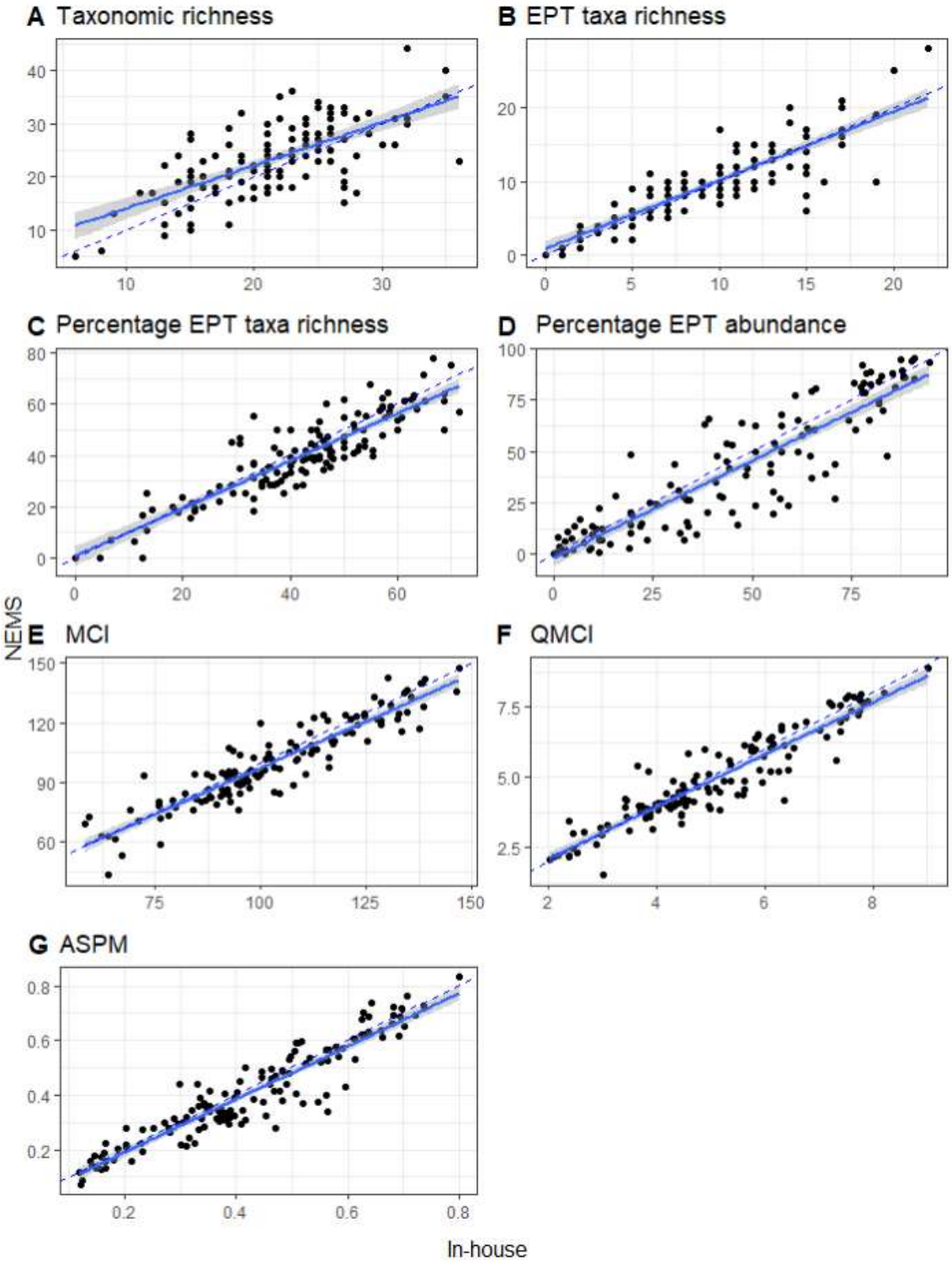


Figure 6-3: Plots of macroinvertebrate metrics calculated for paired samples collected using NEMS and council in-house protocols. Dashed lines are 1:1. Blue lines are from linear regression with 95% confidence intervals (grey shading). Taxonomic resolution and tolerance values were standardised using the NEMS list prior to metric calculation.

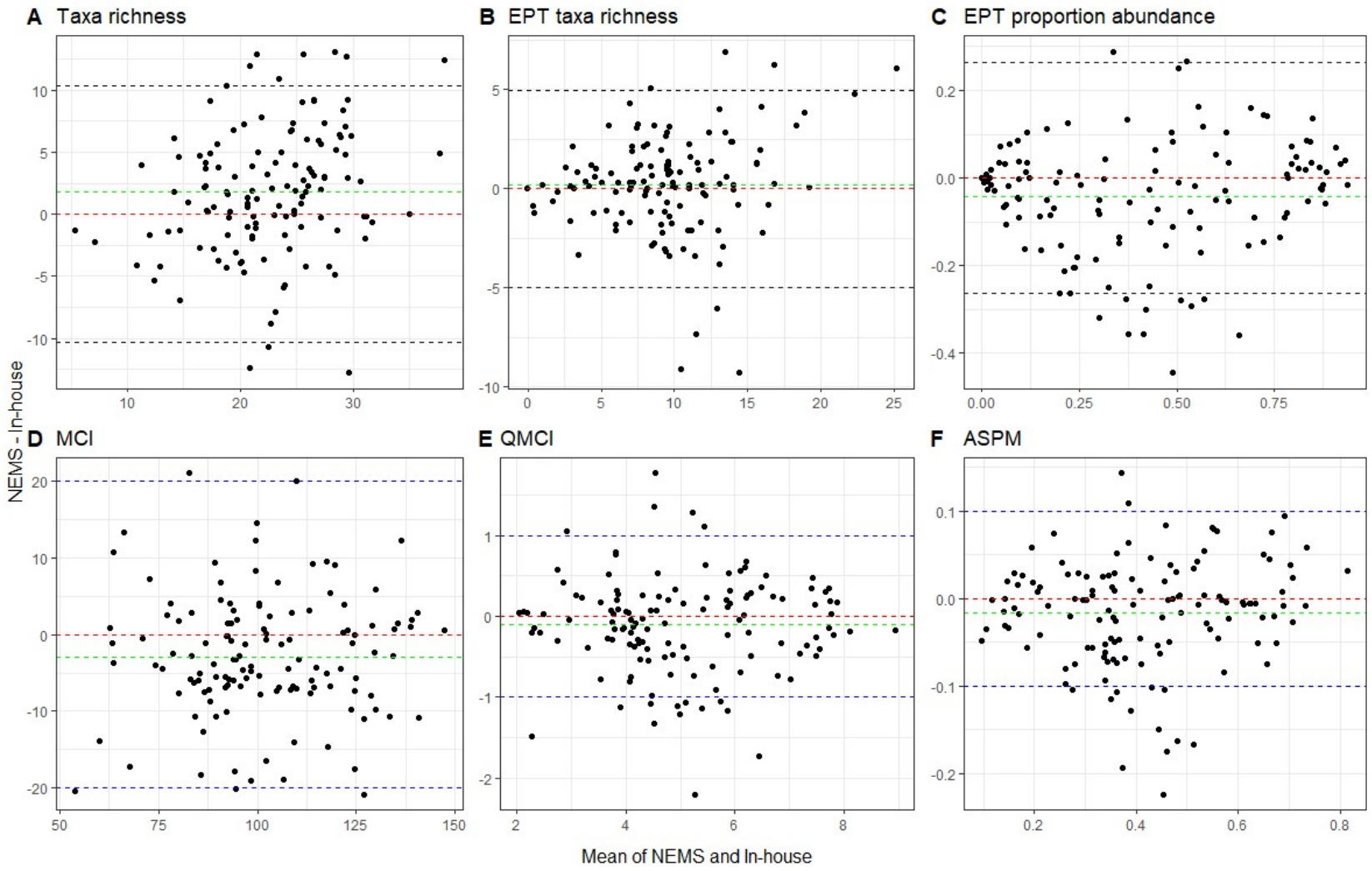


Figure 6-4: Bland Altman plots for macroinvertebrate metrics calculated from samples collected using NEMS or in-house protocols. The difference in metric values between the protocols is plotted against the mean for each sample. Red dashed lines indicate no difference between metric values and green lines the average difference (or mean bias) across all sites. Blue dashed lines indicate the minimum difference between NPS-FM bands for the three metrics in the NPS-FM. For metrics that are not NPS-FM attributes, black lines indicate 2 standard deviations of the overall difference between protocols.

6.4 Potential causes of differences in metric values between protocols

In this section, we investigated whether the difference in metric values between the protocols could be explained by differences in physical in-stream conditions, land-use, stream topography or differences between the protocols (such as the difference in number of habitats sampled under each protocol). The goal was to identify if environmental differences could be used to predict the magnitude or direction of differences in metric values between the two protocols.

6.4.1 Methods

Separate generalised linear models (GLM) were run for each macroinvertebrate metric using the combined council dataset with the following predictors:

1. REC land-use classes (e.g., pastoral, exotic forest, indigenous forest, urban).
2. Percent fine silt and mud cover from NEMS protocol. This was negatively correlated with macroinvertebrate metric scores (MCI, QMCI, ASPM, percentage EPT abundance and taxa richness).
3. Percent riffle/run habitat in the NEMS protocol. The rationale of including this was that NEMS samples that were collected predominantly from riffle/run habitat may be more similar to in-house samples, and result in more similar metric values, than those collected from a larger percentage of different habitats.
4. Number of additional habitats sampled under NEMS. The rationale of including this was that NEMS samples collected from a greater number of additional habitats than the in-house samples could be expected to be more different and result in more different metric values than those collected from a similar number of habitats as the in-house samples.
5. The collecting council.

GLMs were run twice: once including the council term and once excluding it to investigate the potential influence of other included predictors co-varying with the council identity on model results. Very limited differences in the significance of the other terms were observed when the council term was included or excluded. Below we report the results from GLMs including the council term.

We considered including other predictors but removed any that were highly correlated or covaried with the list above. For example, the percentage of sampled habitat that was riffles or runs for the NEMS protocol was significantly lower in sites in the Lowland REC topography class than in Hill or Mountain classes (binomial glm: z value = -2, p = 0.04). We included percentage of riffle/run habitat as a predictor and excluded topographic class. Likewise, REC climate categories covaried with council, as expected due to regional climatic differences. We included council rather than climate category in the models.

The number of additional habitats and percentage of riffle/run habitat during NEMS samples were not available for 20 sites. Due to the low number of SB sites (12 of 129), SB and HB sites were not analysed separately.

6.4.2 Results

Differences in metric values for all three NPS-FM macroinvertebrate attributes (MCI, QMCI, ASPM) between the NEMS and in-house protocols were not significantly related to any of the environmental or sampling protocol differences that were included in the GLMs (Table 6-4). The only significant relationship for these three metrics was that the magnitude of difference between ASPM values calculated from in-house vs NEMS protocol samples varied between councils (Table 6-4, Figure 6-5C). For most councils, the difference in ASPM values between the protocols was a mix of higher and lower under NEMS; however, for HRC, NRC,

TRC and to a lesser extent ES, ASPM values were more commonly higher from in-house samples than NEMS samples (Figure 6-5C).

The difference in percent EPT abundance between NEMS and in-house samples was only significantly related to differences between councils (Table 6-4, Figure 6-5B). Most councils observed no consistent difference in percentage EPT between the protocols; however, for ES and WRC, most samples had higher percent EPT abundance in the in-house samples than in the NEMS samples (Figure 6-5B).

The difference in taxonomic richness between NEMS and in-house samples was significantly greater in sites where a lower percentage of run/riffle habitat was sampled using the NEMS protocol (Table 6-4, Figure 6-6), which may relate to greater habitat diversity sampled, resulting in greater macroinvertebrate diversity. However, note that while this pattern was significant across all councils, when this relationship is plotted separately for the councils, the relationship does not appear to be consistent (Figure 6-6). Differences in taxonomic richness between the protocols was generally higher in in-house samples for NRC but more commonly higher in NEMS samples for AC, BOP, ES and WRC (Figure 6-5A).

Table 6-4: Summary results (p values) from separate generalised linear models (GLMs) predicting the difference in metric values between NEMS and in-house samples (NEMS minus in-house) for five macroinvertebrate metrics. TR = Taxa richness. NS = non-significant, bold values are p values when α <0.1. The predictors riffle/run percentage in NEMS survey and the number of additional habitats in NEMS were missing values for 20 sites. n = 129 sites. (-) indicates a negative relationship.

Predictor	QMCI	MCI	ASPM	TR	%EPT abundance
REC Landcover class	NS	NS	NS	NS	NS
Mud/silt bed cover	NS	NS	NS	NS	NS
No. additional habitats in NEMS	NS	NS	NS	NS	NS
Riffle/run % in NEMS	NS	NS	NS	**0.004 (-)**	NS
Council	NS	NS	**0.03**	**<0.001**	**0.05**

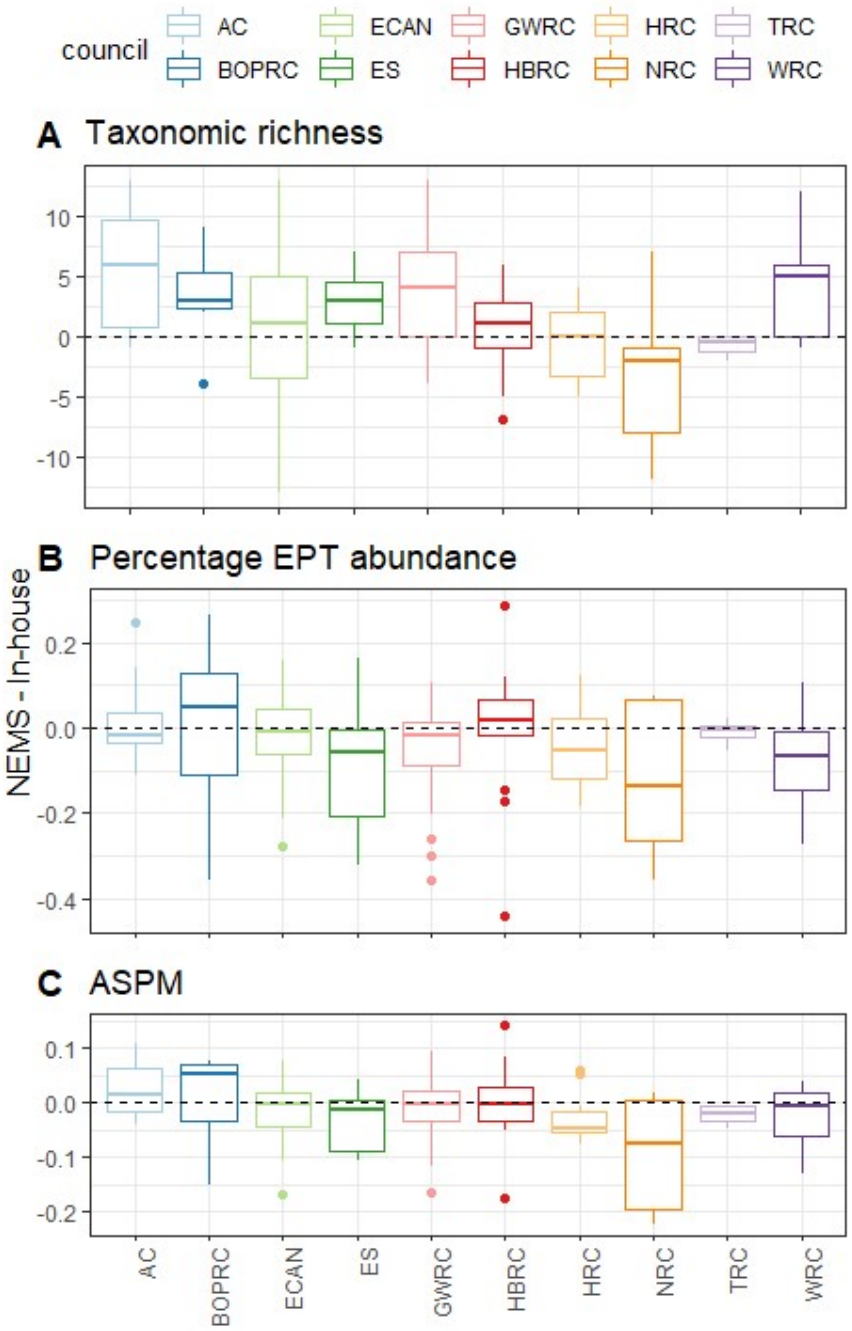


Figure 6-5: Plots of between-council differences in metric scores between NEMs and in-house samples for metrics in which council was identified as a significant predictor in GLM analyses. See Table 6-4 and main text for details of statistical analyses.

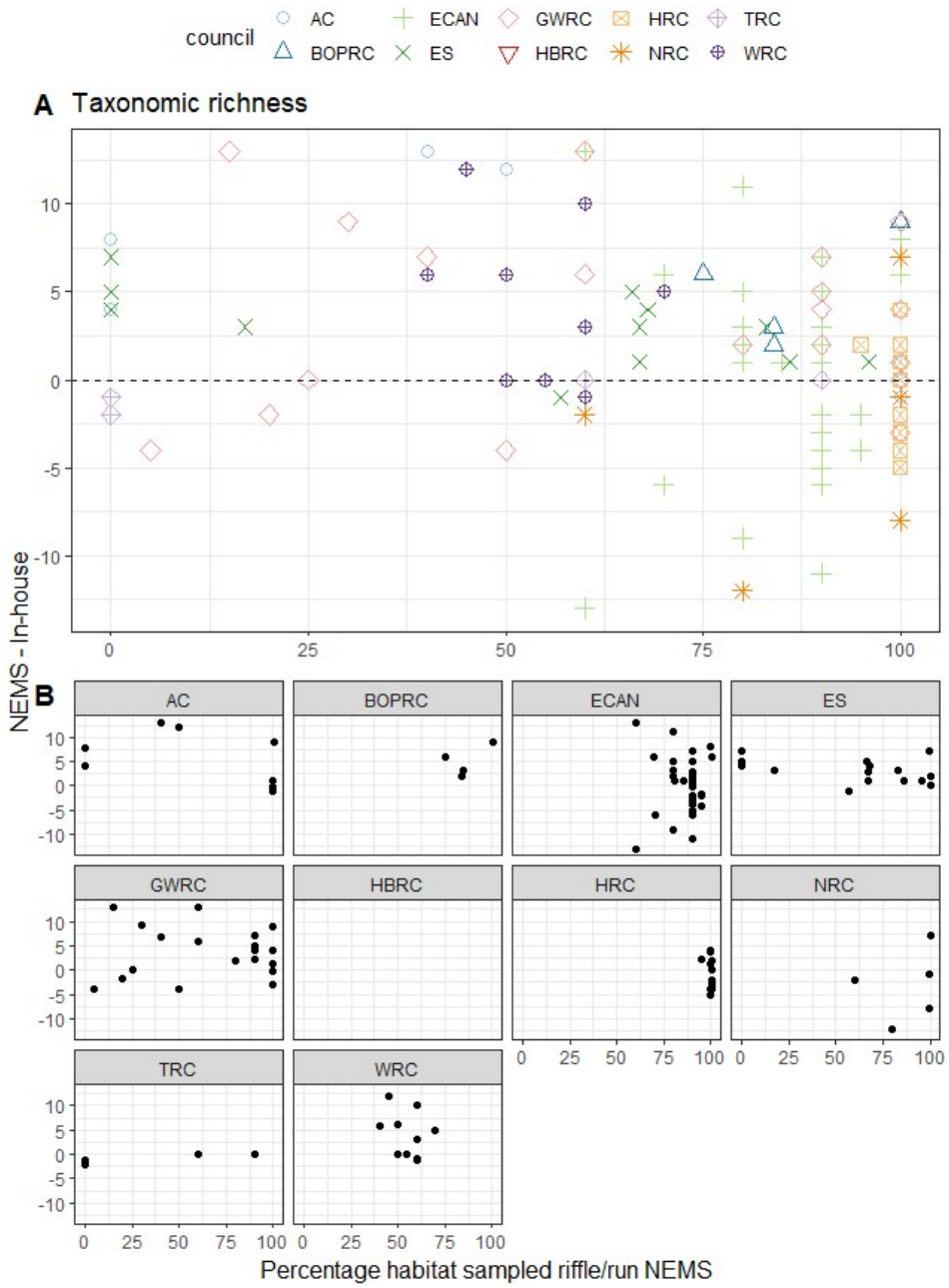


Figure 6-6: Difference between paired NEMS and in-house sample taxonomic richness by the percentage of riffle/run habitat sampled during the NEMS habitat proportional sampling. Plots are shown with all councils together (top) visualising the negative relationship identified by a GLM (see main text) and councils separately (bottom).

6.5 Summary

Key findings for taxonomic identity and richness were:

- Unique taxa were present in samples collected using both NEMS and in-house protocols for all councils together and separately for each council.
- For eight out of 10 councils, more taxa and more unique taxa were collected using NEMS protocols versus the in-house riffle/run-focussed sampling.
- On average, 1.8 more taxa were collected using the NEMS protocol than the in-house method (range -13 to 13).
- On average, 0.2 more EPT taxa were collected using the NEMS protocol than the in-house protocol (range -9 to 7).
- The relationship between taxa richness from NEMS protocols and in-house protocols was generally positive for many councils, although not all (e.g., NRC, Figure C-2).
- Differences in taxonomic richness between the protocols varied significantly by council. Taxonomic richness was generally higher in in-house samples for NRC but more commonly higher in NEMS samples for AC, BOP and WRC.
- Taxonomic richness was higher in NEMS samples than in-house samples when a low proportion of NEMS-sampled habitat was riffle or run. This relationship varied for individual councils, however.

Key findings for differences in community composition between the NEMS and in-house samples were:

- No differences in community composition, when assessed as relative abundance, were observed between NEMS and in-house samples across all councils. Differences in community composition based on taxa presence/absence were almost significant, but the explained variance was low and there was no obvious visual separation between the protocols on the first three NMDS axes.
- The community composition (both relative abundance and presence/absence) was significantly different between the NEMS and in-house samples collected by ECAN. This is similar to differences observed by Hornblow (2022) for the ECAN dataset.
- No other councils showed significant differences in community composition between sample type for either abundance or presence/absence data.

Key findings for the macroinvertebrate metrics MCI, QMCI, ASPM, percentage EPT individuals and percentage EPT taxa were:

- The metric scores from the two protocols were highly correlated and, on average, differences between the protocols were low. However, for individual sites, the differences in metric values between the protocols could be quite large (1 or 2 NPS-FM attribute band ranges).
- Metric values from one protocol were not consistently higher or lower than the other. At some sites, in-house metric values were higher and at other sites the NEMS protocol metric values were higher.

The difference between the metrics from the two protocols could not be explained by environmental differences. Taxonomic richness was generally higher in NEMS samples when a higher proportion of non-riffle/run habitat was sampled, but this pattern was not consistent between councils. The identity of the

collecting council was the only predictive variable that explained differences in metric values between the protocols for more than one metric, likely in part due to regional differences in climate and topography. There may be other environmental or methodological differences in the application of the two protocols, particularly differences between in-house protocols between councils, that could explain these differences better. However, currently we cannot identify why macroinvertebrate metrics calculated from samples collected using in-house or NEMS samples differ.

7 Precision of metric values from the protocols

Previous sections identified differences in the accuracy and bias of metrics calculated from samples collected and processed using council in-house and NEMS protocols. In addition, precision or repeatability of metric scores is also important when comparing different sampling methodologies (Diamond et al. 1996). Samples collected using the same protocol at the same site at the same time are likely to differ somewhat due to spatial variation in the macroinvertebrate community at a site, which influences how representative the collected sample is of the community present. Thus, replicate samples of each protocol at a site are required to assess the precision of each method. The minimum detectable difference in metric values can then be determined for individual protocols, i.e., the magnitude of difference that exceeds the expected natural variability in replicate samples within a site.

Stark (1993) assessed the precision of twelve replicate kicknet and Surber samples collected from riffle habitat (similar to the in-house council methods in this report) from four sites. Across the four sites, the range in MCI values from the 12 individual kicknets were 20.4, 20.4, 15.1, and 13.4 (c.f., 20 units between band thresholds for the MCI NPS-FM attribute). Stark (1993) calculated the minimum detectable difference as the smallest difference that can be detected statistically for a given sample size using the standard deviation in MCI values within a site and number of samples collected. This was reported as 10.3 MCI units for a single kicknet sample for MCI. This means that sites that had MCI scores that were >10.3 units apart were likely to have statistically different MCI values.

Stark (1998) recalculated the minimum detectable difference between sites using the same methodology but for a larger data set (827 kicknet samples with 2 to 12 replicates per site). Stark (1998) reported a very similar minimum detectable difference for MCI values from one kicknet sample of 10.8 (c.f. 10.3 reported by Stark 1993).

If we assume a minimum detectable difference of 10 MCI units is relevant for both NEMS and in-house protocols, we can consider at how many sites the difference in metric scores between the protocols is likely to be different, while accounting for potential variation in metric scores within a site. At 79% of sites (102 of 129 sites) the difference in MCI values between the two protocols was likely to be less than the minimum detectable difference calculated for riffle-habitat based kicknet sampling (using Stark 1993 and 1998 estimates).

The dataset analysed in this report does not contain replicate samples, and therefore we are unable to calculate precision or minimum detectable differences between council in-house and NEMS protocols. We also recognise that these data would be expensive to collect, and difficult at small sites where there may be insufficient in-stream habitat to collect replicate samples.

In the absence of replicate samples for each method, we drew upon additional existing datasets to visually compare the variation in metric scores:

1. Temporally within sites, using long-term data from a subset of the same sites collected using in-house protocols (data provided by the councils).

2. Spatially within individual sampling reaches, using a NIWA dataset where 13 replicate Surber samples were collected on the same day within several reaches of the same sites.

We provide these visualisations for consideration only, and recognise that long-term temporal variation in metric scores is largely caused by environmental differences between sampling years and is greater than the difference in metric values between samples collected on the same day using the same protocol.

We also recognise that in-house and NEMS monitoring samples combine replicate kicknets into one sample to reduce the spatial variation in community composition. The metric scores calculated for individual Surber samples in our spatial dataset will likely be more variable than a composite sample collected using NEMS or in-house protocols. However, given that minimum detectable differences for riffle-based kicknet MCI values have previously been estimated as ~10 (Stark 1993, 1998), and that for 79% of sites the difference in MCI values between in-house and NEMS protocols was less than this, we consider that a visualisation of the relative magnitude of metric score variation spatially within sites is beneficial.

7.1.1 Temporal variability within a site

Annual macroinvertebrate metric scores (MCI and QMCI) were provided by councils for a subset of the sites where the paired in-house and NEMS samples were collected. The data per site ranged between 3 years and 25 years of data. We plotted these data with the paired metric values from the in-house and NEMS protocol samples. Examples of long-term temporal variability in MCI and QMCI metrics from council in-house metrics are provided in Figures 7-1 (ECAN MCI), 7-2 (other council MCI) and 7-3 (ECAN QMCI). Often, but not always (see MCI values at TRC site WIU000700; Figure 7-2) the difference in metric scores between the two protocols is within the range of historical variation in metric values at a site. Most of the temporal variation in metric scores at a site is attributed to environmental differences between years affecting the macroinvertebrate community and calculated metrics, but it is informative to view the magnitude of the between protocol metric differences in the context of this temporal variation.

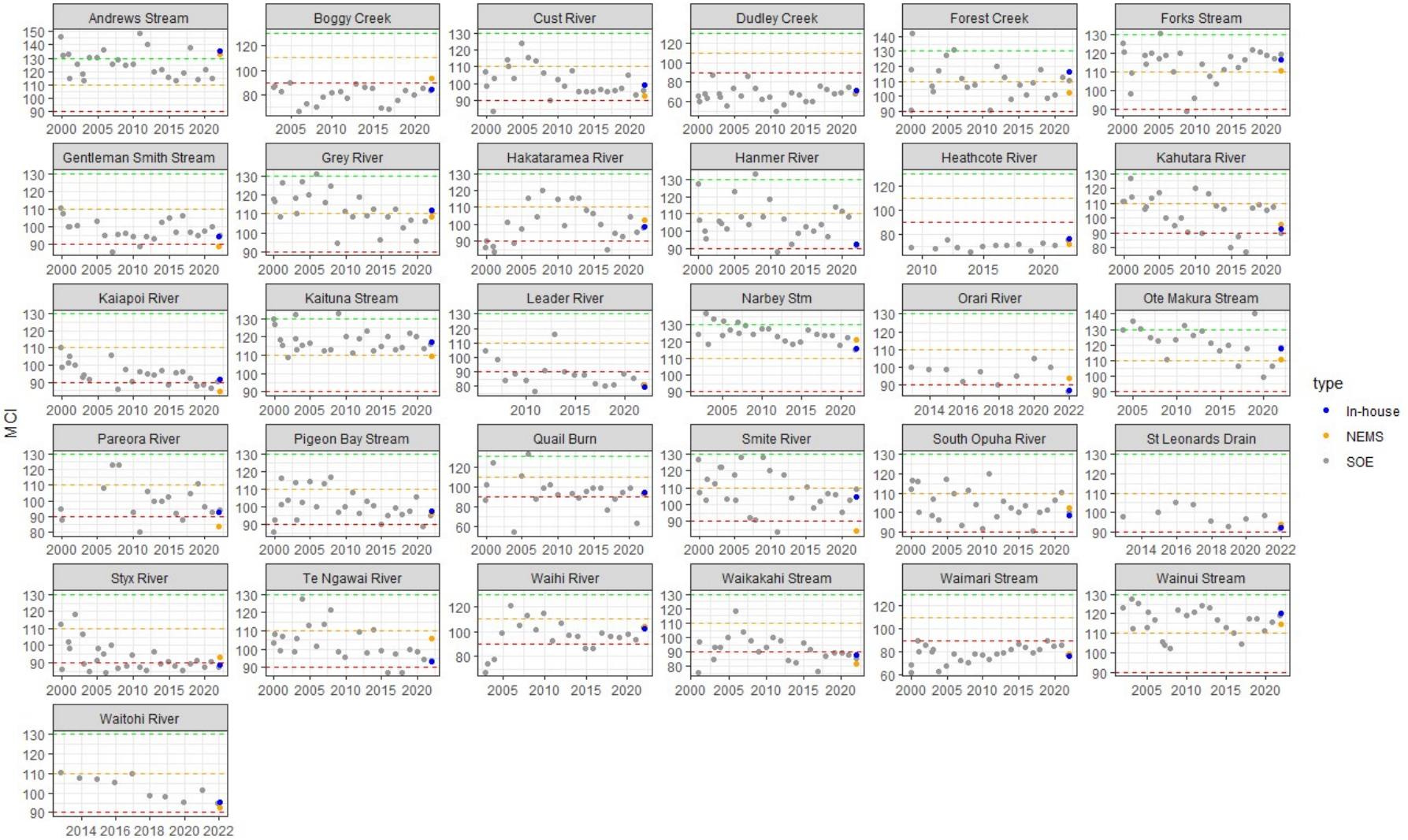


Figure 7-1: Long-term MCI values from ECAN SOE monitoring sites included in this report. Historical SOE values are in grey and recent paired sampling using NEMS and in-house collection methods in orange and blue, respectively. SOE metric values are as provided by the council.

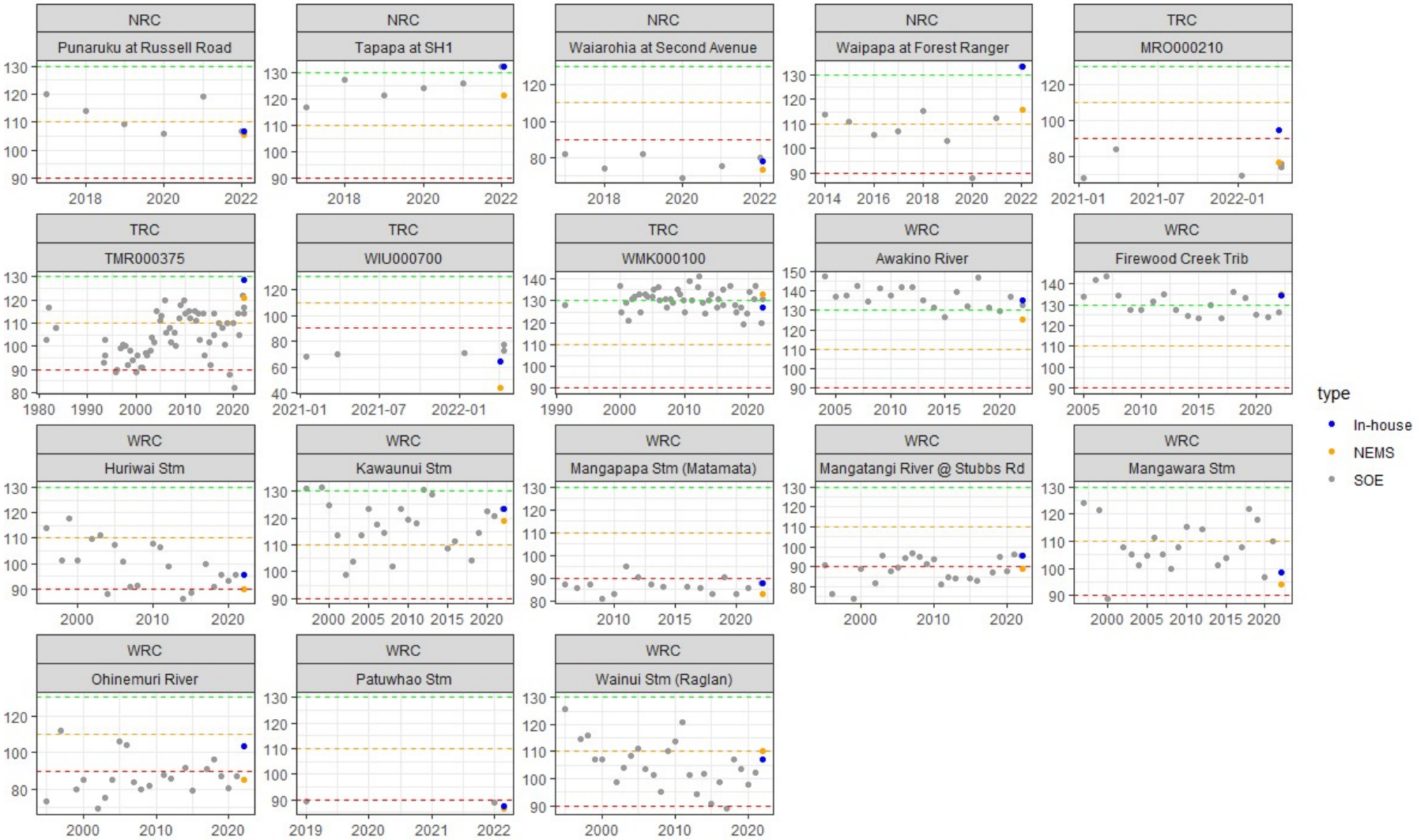


Figure 7-2: Long-term MCI values provided by councils at SOE monitoring sites included in this report. Historical SOE values are in grey and recent paired sampling using NEMS and in-house methods are in orange and blue, respectively. SOE metric values are as provided by the council.

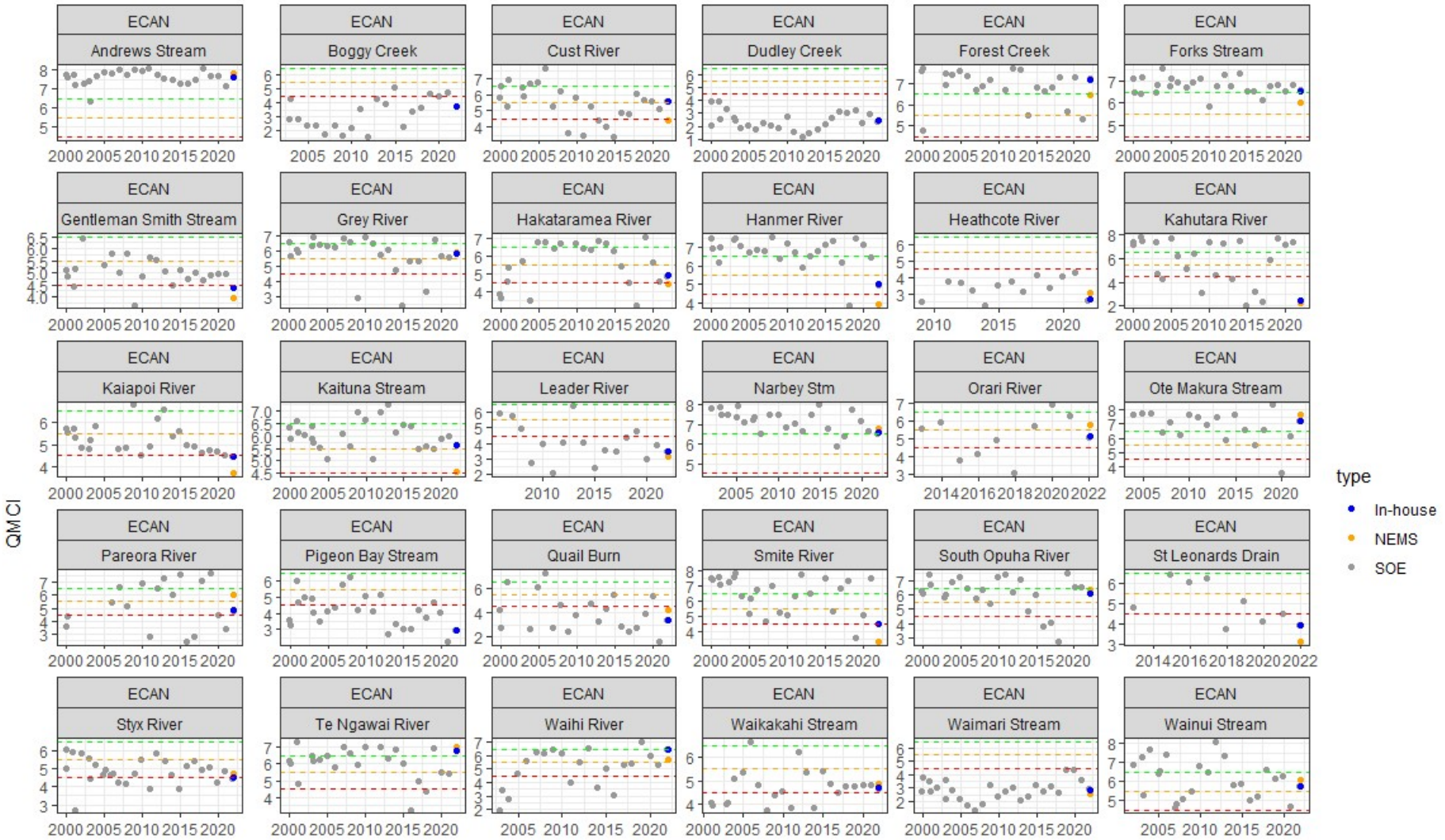


Figure 7-3: Long-term QMCI values from ECAN SOE monitoring sites included in this report. Historical SOE values are in grey and recent paired sampling using NEMS and in-house collection methods in orange and blue, respectively. SOE metric values are as provided by the council.

7.1.2 Spatial variability within a site

Data from a NIWA Strategic Science Investment Funding project were analysed to explore spatial variability within sites. Macroinvertebrate data were collected at two sites on the Selwyn River, Canterbury (Scotts Road and Whitecliffs) using Surber samplers multiple times between 2020 and 2022. At each site visit, three reaches (of nine permanently marked at the site) were surveyed. Within each reach on each visit, a facies map was drawn by visually classifying the in-stream bed substrate, depth and water velocity into similar categories. There were up to six facies (labelled A to F) identified at a reach. On each site visit at each reach, 13 Surber samples were collected from random locations distributed across the facies present in proportion to their abundance. We calculated MCI and QMCI metrics from each sample and plotted them in Figures 7-4 and 7-5.

Metric values from replicate Surber samples were sometimes highly variable within a reach, and even within a facie within a reach on the same date. Occasionally, metric values calculated from replicate samples ranged across two or three NPS-FM attribute bands.

However, we note that individual Surbers are collected from smaller areas than composite kicknets and often more Surber samples are required to attain the same level of precision as a single composite kicknet (Stark 1993). The variation in metric scores in the plots is thus likely greater than would occur for replicate kicknet samples. Stark (1998) found that, in riffle habitat, four Surber samples (0.1 m^2 area) provided estimates of MCI similar in precision (just over ± 10%) to a single kicknet sample collected from an area of 0.3–0.6 m^2. Three D-net samples or 8 Surber samples were required to achieve precision of around ± 10% for the SQMCI and QMCI variants respectively.

NEMS and council in-house samples are composites of multiple individual kicks (each individual kick would sample an area approximately the same as the individual Surber) to account for such spatial variability in macroinvertebrate communities within a site.

Figure 7-4: QMCI values calculated for 13 Surber samples collected within three reaches (of nine permanently marked) at two sites on the Selwyn River (Scotts Road and Whitecliffs) over multiple visits between 2020 and 2022. Facies are visually identified and mapped for each reach on each visit based on the substrate composition, water depth and velocity. Red, orange and green dashed lines indicate the NPS-FM attribute band thresholds for C/D, B/C and A/B, respectively.

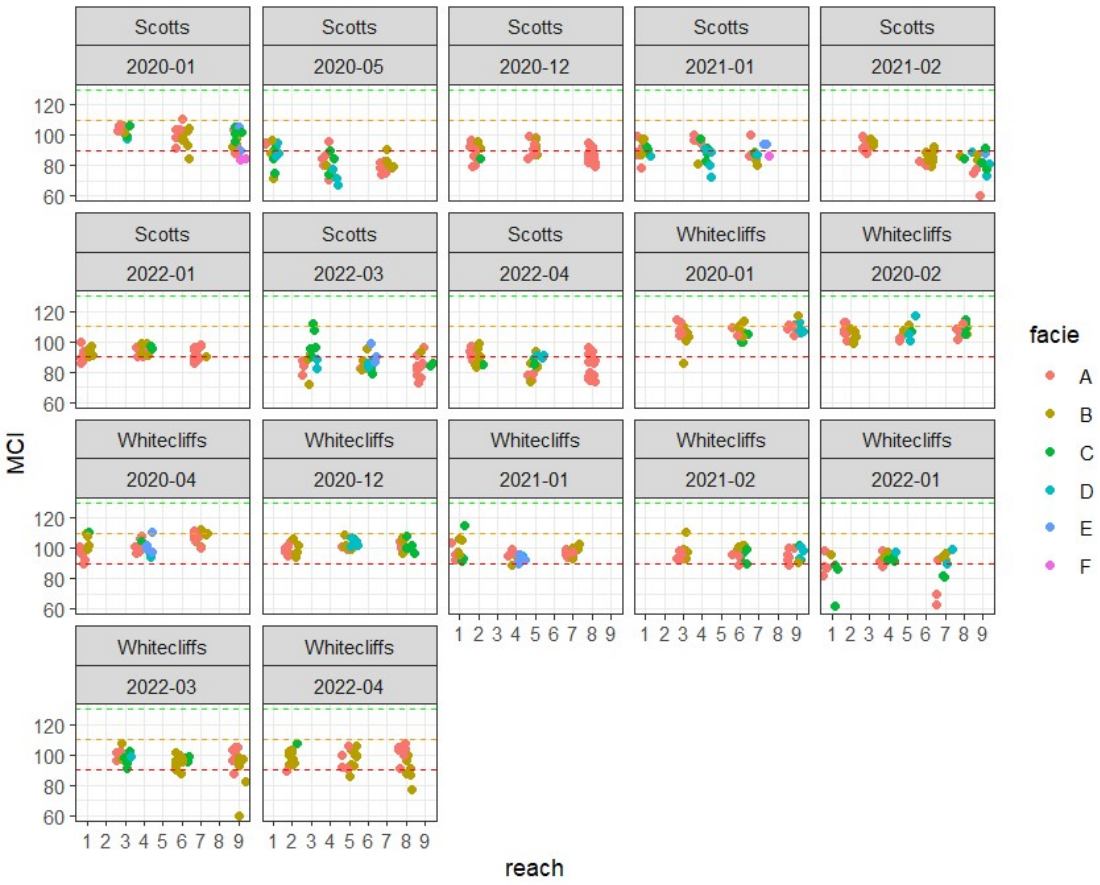


Figure 7-5: MCI values calculated for 13 Surber samples collected within three reaches (of nine permanently marked) at two sites on the Selywn River (Scotts Road and Whitecliffs) over multiple visits between 2020 and 2022. Facies are visually identified and mapped for each reach on each visit based on the substrate composition, water depth and velocity. Red, orange and green dashed lines indicate the NPS-FM attribute band thresholds for C/D, B/C and A/B, respectively.

7.1.3 Summary

While we cannot assess the precision of the in-house and NEMS protocols without replicate samples at a site, we can observe that in-house metrics vary both spatially and temporally. Much of this variation is due to temporal differences in the environment (e.g., antecedent flow conditions) or due to spatial variation in macroinvertebrate communities (which both protocols minimise by using composite samples). We cannot directly equate either the temporal or spatial variation in metric score that we observed to differences expected from replicate samples of the same protocol at a site.

However, Stark (1993) and (1998) generated minimum detectable differences for a method equivalent to most councils' in-house methodology that indicated values needed to be >10 MCI units different for values to exceed expected variation from replicate samples at a site. At 79% of sites, the difference in MCI values between the NEMS and in-house protocols was likely to be less than this value.

Overseas investigations also provide evidence suggesting that proportional habitat sampling may generate macroinvertebrate metrics that are similar enough to be interchangeable with riffle-focussed sampling in some areas, depending on the use of the metrics. For example, Rehn et al. (2007) compared targeted riffle and random proportional habitat sampling approaches and found

that the variability in replicate metric values calculated from the samples was generally larger than differences between the metric values calculated from the two methods at a site. They concluded that biological metrics generated from the two data sets may be generally interchangeable, for the metrics they tested. Ostermiller and Hawkins (2004) found that macroinvertebrate metrics generated from targeted-riffle and timed multihabitat samples collected from wadeable streams in western Oregon and Washington were approximately equally precise. They did also show that assessments using the two different protocols sometimes resulted in different site-specific inferences about ecological degradation, but that agreement improved as the fraction of the sample that was processed was increased.

8 Overall summary and conclusions

- Samples collected using NEMS protocols were not observed to be consistently larger than in-house samples by the processing laboratories, but large samples in general made processing difficult.
- The slight differences in taxonomic resolution and tolerance values used for some taxa by the councils generally only resulted in small differences in macroinvertebrate metric values (e.g., commonly less than 5 units for MCI, where 20 is the difference between NPS-FM bands).
- Including the missed taxa scan resulted in up to 12 additional taxa (when using the NEMS Annexes D and H lists) and generally resulted in slightly higher MCI and ASPM scores.
- Generally, more taxa and more unique taxa were found in NEMS protocol samples than in the in-house protocol samples, although unique taxa were collected by both protocols.
- The composition of the community collected by the two protocols did not differ significantly across all councils or for most councils individually, except ECAN.
- Unique taxa were collected for all councils and overall using both protocols; however, apart from ECAN samples, the community composition was similar between protocols. That is, the taxa collected differed between the protocols but not in a consistent way.
- The metric scores from the two protocols were highly correlated and, on average, differences between the protocols for both metrics were low. However, for some individual sites the metric differences between protocols could be very large, and occasionally large enough to move values from being in one NPS-FM band to another (or even two bands apart, occasionally for QMCI).
- Metric values are expected to differ between the protocols, just as there is some variability in replicate samples of the same method at a site. The challenge is determining what level of difference is acceptable between the methods. In addition, if the direction and/or magnitude of the difference is predictable, it may be possible to identify situations in which the protocols result in 'acceptably' similar results, and perhaps even a conversion factor that could be applied in some situations for metric values collected by the different protocols.
- Metric values from one protocol were not consistently higher or lower than the other. At some sites, in-house metric values were higher and at other sites the NEMS protocols metric values were higher.
- At 27 of the 129 sites (21%) the difference in metric values between the two protocols was greater than ±10 (the minimum detectable difference identified by Stark 1993 and Stark 1998) for riffle-focussed kicknet samples. At 78% of the sites, the difference in MCI values between the two protocols was likely to be less than the ±10 minimum detectable difference.

9 Recommendations

The options available to councils when considering adopting the NEMS protocols for macroinvertebrate monitoring are to:

1. Remain with in-house protocols, which has the benefits of long-term consistency in monitoring methods.
2. Switch to the NEMS macroinvertebrate protocol, which has benefits for national consistency and may better future-proof sampling in events that significantly alter in-stream habitat. However, switching protocols may result in a step-change in metric values at some sites.
3. Undertake a period of dual monitoring using both protocols at some sites, which allows for investigation of the implications of the protocol switch for metric scores but is associated with the costs of dual sampling.

We recommend that if councils do switch to the NEMS protocol, that co-ordination between councils is considered, so that national consistency is attained over a short period of time.

We also recommend dual monitoring, if possible, particularly at sites where the risk of getting the metric value wrong is high. For example, sites with metric values close to critical management thresholds or those where metric values may be challenged, i.e., locations of potential future development. Dual samples could be stored for later processing, if required, to save on processing costs.

Continued dual monitoring would also provide data to investigate the implications of switching protocols. Ideally, replicate samples of both council in-house and NEMS protocols at a subset of sites should be collected to quantify the expected variation in samples of the same protocol.

It is also worth noting that under Clause 3.10 (4) in the NPS-FM, attribute states and baseline states may be expressed in a way that accounts for natural variability and sampling error. Better knowledge of the magnitude of natural variability in macroinvertebrate communities and thus metrics, and in the precision of in-house and NEMS protocols under different environmental conditions could assist councils in this process.

9.1 Future monitoring considerations

Differences in metrics scores between the paired samples were identified, which in some cases were relatively large. Site characteristics did not explain the difference in metric scores between NEMS and in-house protocols, apart from some differences in taxa richness.

Differences in metrics scores between protocols at a site were often smaller than previous estimates of the within site variance associated with composite kicknet samples in riffles.

Samples collected using the same protocol at the same time at the same site will always show some variability due to spatial variation in community composition and sampling effort, and sites will always show natural temporal variability in macroinvertebrate community composition associated with environmental conditions. The key question for councils is whether the variation in metric scores associated with differences between NEMS and in-house protocols is within these expected ranges of natural variation.

Several potential options to investigate these sources of variation with further monitoring are:

- Further investigation of expected variation between sampling methods at a site during different environmental conditions (i.e., longer-term paired monitoring) would be useful to understand if recent environmental conditions affect differences in metrics scores between protocols.
- Replicate samples of both sampling protocols under the same environmental conditions would help identify the scale of spatial variation expected within a site.

10 Acknowledgements

Many thanks to the participating councils for collecting, processing and providing us with the macroinvertebrate data. We are grateful to the staff from the processing laboratories who provided feedback on the samples collected using both protocols. Juliet Milne was instrumental in getting this project initiated and in organising the workshop for laboratory staff. Rick Stoffels at NIWA provided the SSIF project data for the analysis of spatial patterns in macroinvertebrate metric values.

11 References

Altman, D.G., Bland, J.M. (1983) Measurement in medicine: the analysis of method comparison studies. Journal of the Royal Statistical Society Series D: The Statistician. 32: 307-317.

Bland, J.M., Altman, D.G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1: 307-310.

Diamond, J.M., Barbour, M.T., Stribling, J.B. (1996) Characterizing and comparing bioassessment methods and their results: a perspective. Freshwater Science. 15: 713-727.

Greenwood, M.J., Booker, D, Stark, J., Suren, A., Clapcott, J. (2015) Updating MCI tolerance values for freshwater invertebrate taxa. Prepared for Environment Southland and Hawke's Bay Regional Council. Envirolink report. NIWA client report ChC2015-008.

Hornblow, K. (2022) Macroinvertebrate monitoring: A comparison of sampling and lab processing methodologies. Report prepared by Boffa Miskell for Environment Canterbury.

NEMS (2022) National Environmental Monitoring Standards Macroinvertebrates: collection and processing of macroinvertebrate samples from rivers and streams. Version 1.0.0. https://www.nems.org.nz/documents/macroinvertebrates/

New Zealand Government (2023) National Policy Statement for Freshwater Management 2020. https://environment.govt.nz/assets/publications/National-Policy-Statement-for-Freshwater-Management-2020.pdf

Ostermiller, J.D., Hawkins, C.P. (2004) Effects of sampling error on bioassessments of stream ecosystems: application to RIVPACS-type models. Freshwater Science. 23: 363-382.

Rehn, A.C., Ode, P.R., Hawkins, C.P. (2007) Comparisons of targeted-riffle and reach-wide benthic macroinvertebrate samples: implications for data sharing in stream-condition assessments. Freshwater Science. 26: https://doi.org/10.1899/0887-3593(2007)26[332:COTARB]2.0.CO;2

Stark, J. D. 1985. A Macroinvertebrate Community Index of water quality for stony streams. Water & Soil miscellaneous publication 1985 vol. 87: 1-52

Stark, J.D. (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: 463-478

Stark, J. D. (1998) SQMCI: a biotic index for freshwater macroinvertebrate coded abundance data. New Zealand Journal of Marine and Freshwater Research 32: 55- 66.

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

Stark, J. D., Boothroyd, I. K. G., Harding, J. S., Maxted, J. R., Scarsbrook, M. R. (2001) Protocols for sampling macroinvertebrates in wadeable streams. New Zealand

Macroinvertebrate Working Group Report No. 1, Prepared for the Ministry for the Environment.

Appendix A Key comments from laboratory processors

On 23 May 2023 an online meeting was held with lab representatives involved with the development of the *NEMS Macroinvertebrates* to discuss their experiences in processing recent regional council State of the Environment (SOE) macroinvertebrate samples. Ten councils had carried out different forms of paired sampling or processing over the previous summer to help understand what impact adopting NEMS QC 600 requirements might have on long-term datasets obtained from their existing sampling/processing methods.

The meeting attendees were: Karen Shearer (Cawthron), Emily Demchick (EOS Ecology), Ben Ludgate (4Sight/SLR), Brian Smith (NIWA), Duncan Gray (ECan) and Juliet Milne (NIWA). Tanya Blakely (Boffa Miskell) was also invited to the meeting but was unable to make it. Juliet followed up with Tanya on 8 June.

Key comments made from the VC and follow-up call to Tanya are summarised below.

1. One processor found that the samples were large with a lot of material to work through. This was in part during the processing of the samples collected using in-house methods, which were composites of 7 pooled Surbers. In some cases, the NEMS QC time restriction of 30 minutes for the missed taxa scan was not applied.
2. In samples from urban environments with homogeneous material (lots of organic matter including silt and algae), there was not a noticeable increase in sample volume.
3. One processor processed samples from five councils and four had paired existing and NEMS sampling methods and requested sample processing using their existing (i.e., pre-NEMS) methods. The samples collected using the NEMS methods were generally bigger, so a slightly smaller sub-sample volume needed to be processed to obtain a fixed count of 200+ individuals. It probably took ¾ hour longer on average to process samples according to NEMS methods.
4. One processor reported that receiving multiple sample containers (2–3 but up to 5 x 1 L) was not uncommon and became logistically challenging (e.g., setting up and completing the processing of a large sample in one session). There is also a need for good sample pre-processing in the field.

General comments and questions:

- The impact of adopting the NEMS whole-of-habitat approach to sampling on sample volume is very much site-specific.
- NEMS has not created drastic changes to processing methods; more clarification and improvements to QC requirements.
- Cyclone Gabrielle aftermath suggests that the NEMS approach is future proofing SOE sampling; for example, riffles that were previously samples in some Hawke's Bay rivers simply do not exist now – instead they are silt-laden runs.
- The labs that process SOE-type samples would benefit from an annual sample processing workshop to support NEMS implementation and best practice, etc.

Appendix B Requirements for macroinvertebrate monitoring under the NPS-FM.

There are compulsory macroinvertebrates attributes that must be monitored as part of the Ecosystem Health value under the NPS-FM. Three macroinvertebrate metrics are included as attributes:

- The Macroinvertebrate Community Index (MCI) is a presence/absence-based metric calculated based on the average tolerance scores of taxa. Values range from 0 to 200, with a higher score indicating fewer impacts of organic pollution or nutrient enrichment.
- The quantitative variant of the MCI (QMCI) incorporates abundance data as well as taxa presence.
- The Average Score Per Metric (ASPM) is calculated from normalised scores of MCI, the taxa richness of sensitive Ephemeroptera, Plecoptera and Trichoptera (EPT TR) and the percentage abundance of sensitive EPT individuals (% EPT).

The current state for each attribute at a site is calculated as the median value of the last five years' annual samples collected between November and April.

Samples can be processed using either full counts of all individuals within a sample, or fixed counts of at least 200 individuals.

Each taxon is assigned a tolerance value to use when calculating the MCI. Tolerance values are provided for both soft-bottomed (SB, generally assigned when >50% of the bed substrate comprises fine material <2 mm in diameter) and hard-bottomed sites (HB, assigned when <50% of bed substrate comprises fines). Soft-bottomed sites are required to use the SB tolerance scores for taxa, while HB tolerance values are used for hard-bottomed sites.

The taxonomic resolution and tolerance values for taxa in HB sites are defined in Stark and Maxted (2007). Tolerance values for taxa in SB sites are in Clapcott et al. (2017). These are also the lists in Annex D of the NEMS.

Sites are required to have target attribute states, which cannot be lower than the national bottom line. If the current state at a site falls below the target attribute state, or a deteriorating trend is detected, then an action plan is required, which aims to improve attribute values (NPS-FM 2020).

Under Clause 3.10 (4) in the NPS-FM, target attribute states and baseline states may be expressed in a way that accounts for natural variability and sampling error.

Appendix C Plots of NEMS vs in-house metric values by council

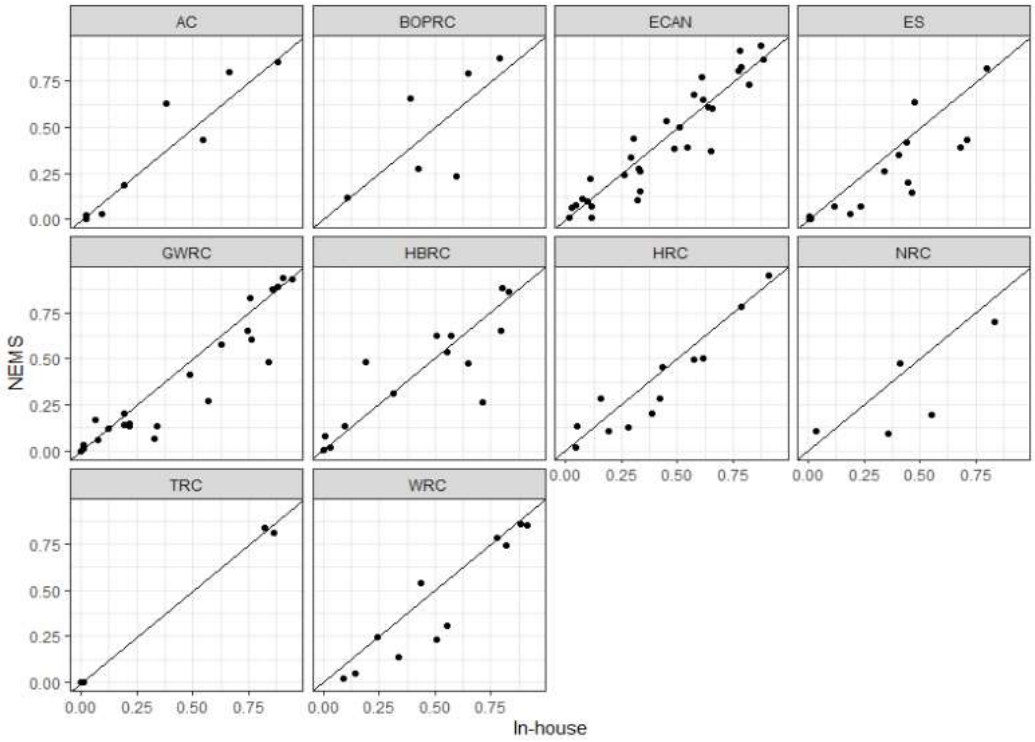


Figure C-1: Proportion EPT abundance calculated from paired samples using NEMS and council in-house protocols for 10 regional councils. Taxonomic resolution was standardised to follow NEMS macroinvertebrate before metric calculation.

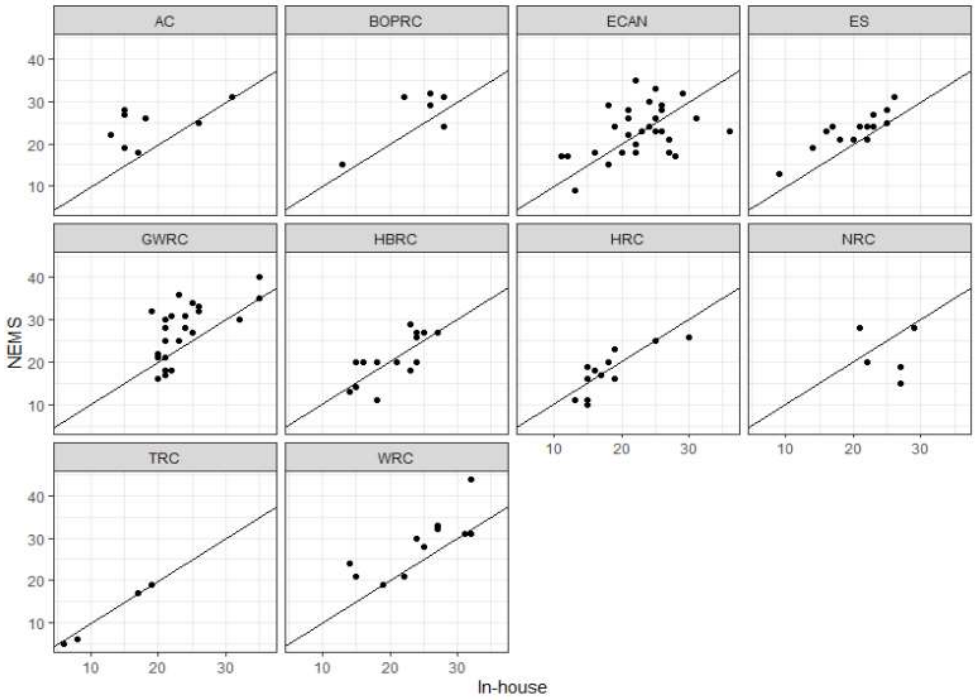


Figure C-2: Taxonomic richness calculated from paired samples using NEMS and council in-house protocols for 10 regional councils. Taxonomic resolution was standardised to follow the NEMS macroinvertebrate Annexes D and H.

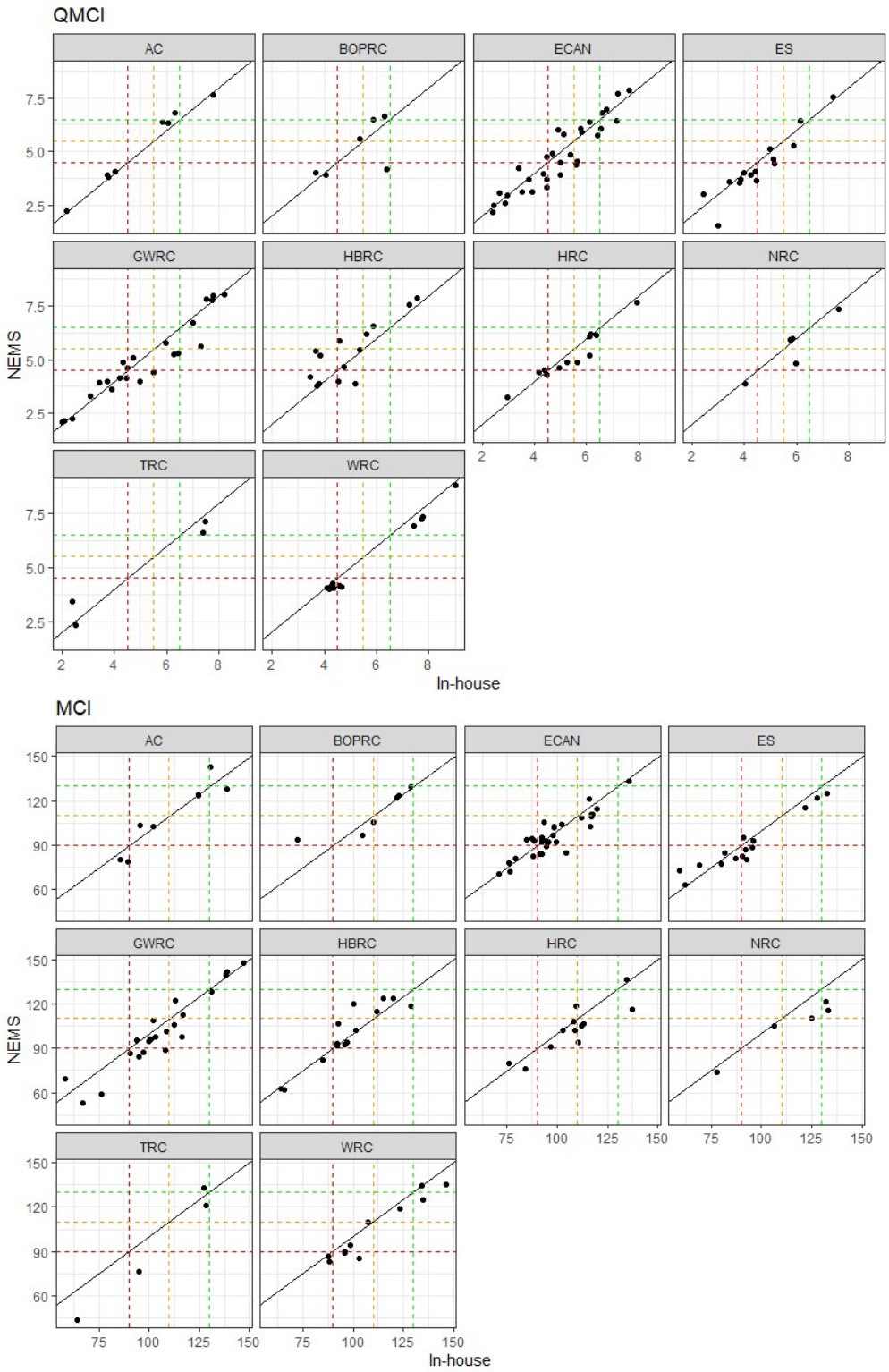


Figure C-3: QMCI (top) and MCI (bottom) metric scores calculated from paired samples using NEMS and council in-house protocols for 10 regional councils. Taxonomic resolution and tolerance values were standardised to follow NEMS macroinvertebrate before metric calculation. Red dashed lines indicate the NPS-FM attribute band C/D boundary, orange the B/C boundary and green the Band A/B boundary.

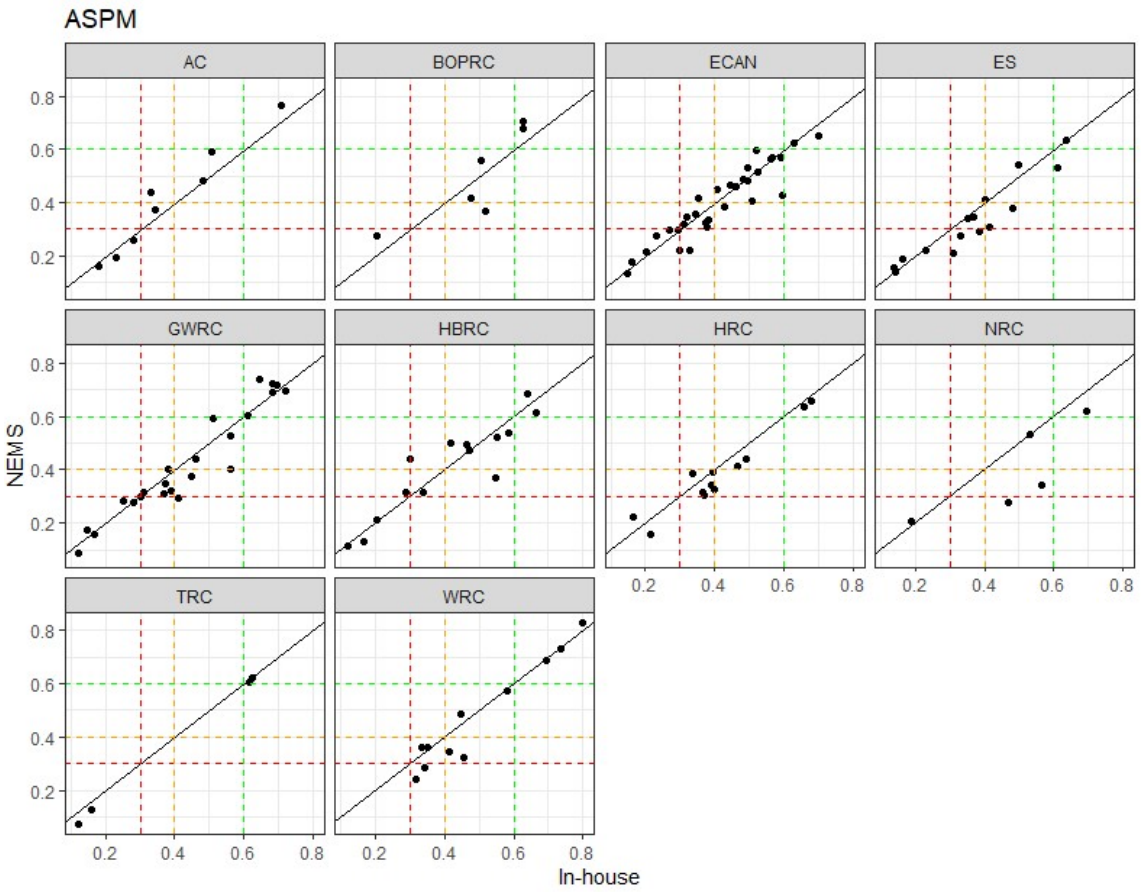


Figure C-4: ASPM macroinvertebrate metric scores calculated from paired samples using NEMS and council in-house protocols for 10 regional councils. Taxonomic resolution and tolerance values were standardised to follow NEMS macroinvertebrate before metric calculation. Red dashed lines indicate the NPS-FM attribute band C/D boundary, orange the B/C boundary and green the Band A/B boundary.