Representativeness and statistical power of the New Zealand river monitoring network.

National Environmental Monitoring and Reporting: Network Design, Step 2

Prepared for Ministry for the Environment

July 2012

Authors/Contributors:
Scott Larned
Martin Unwin

For any information regarding this report please contact:

Scott Larned
Scientist
Freshwater Ecology
+64-3-343 8987
scott.larned@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
10 Kyle Street
Riccarton
Christchurch 8011
PO Box 8602, Riccarton
Christchurch 8440
New Zealand

Phone +64-3-348 8987
Fax +64-3-348 5548

NIWA Client Report No:	CHC2012-079
Report date:	July 2012
NIWA Project:	MFE12201

Contents

Tables

Figures

Reviewed by

Ton Snelder

Approved for release by

JS

Jochen Schmidt

Executive summary

In 2011, NIWA, GNS and Opus evaluated options for revising national-scale freshwater monitoring and reporting in New Zealand for the Ministry for the Environment (MfE) (Davies-Colley et al. 2011). These evaluations were the first steps in the MfE National Environmental Monitoring and Reporting (NEMaR) project. One of the primary aims of NEMaR is to generate information required to create a National Surface Water Monitoring Programme (NSWMP). Three broad issues in freshwater monitoring were considered: variables or analytes, indicators, and the spatial layout of monitoring networks. The reports and subsequent peer reviews identified issues that needed to be addressed in order to achieve the aims of NEMaR. To address these issues, an expert panel composed of regional council, central government, university, consultancy, and CRI staff was formed for each project area, and two workshops were held in 2011. The general aims of the workshops were to 1) seek consensus and provide recommendations on methods for improving national reporting on freshwater ecosystems; and 2) identify issues for which consensus was not reached and provide options to resolve those issues.

The expert panel for network design produced a number of consensus statements and recommendations, some of which provided directives for the current assessment of the SoE network. Those directives were:

1. The primary purposes of a national freshwater monitoring programme are to obtain 1) representative estimates of environmental states and trends, and 2) comparisons of state and trend among selected environmental classes, including reference classes. The first purpose has higher priority than the second.

2 Comparisons of reference classes or reference conditions versus impacted classes should be included. The panel considered that some comparisons between impacted classes are also important, but did not specify which classes should be included. The availability of unimpacted reference sites that generate reference-condition data is a high-priority for the monitoring network.

3. The current river monitoring network needs to be evaluated regarding its ability to meet the primary purposes of the NSWMP, before identifying prospective new sites or alterations to the network. For the first purpose (estimates of state and trend), analyses of the representativeness of the current river monitoring network, and precision and site number requirements are needed. For the second purpose (comparisons among selected classes), an analysis of statistical power and site number requirements is needed.

4. Use of both Freshwater Environments of New Zealand (FWENZ) and River Environment Classification (REC) for assessment of the environmental classes for the current ecological monitoring network was recommended. Updated site lists and datasets are required to carry out the assessments of the current networks.

This report provides the results and interpretations of the analyses of representativeness and statistical power analyses. The power analyses consisted of estimations of monitoring site-number requirements to achieve stated levels of precision, and to detect differences in mean river conditions in environmental classes. Representativeness refers to the comparative distributions of monitoring sites and rivers across the New Zealand environment. Precision refers to the variability associated with an estimate environmental conditions can be reported. Statistical power refers to the sensitivity of statistical comparisons (i.e., the power to detect statistical difference between groups or classes). Precision analyses produce estimates of the number of monitoring sites required to report environmental conditions with a given level of certainty; this is relevant for the first purpose of the NSWMP. Power analyses produce estimates of the number of sites required to detect between-group differences of a given size with a given level of confidence; this is relevant for the first purpose of the NSWMP.

To carry out the analyses of representativeness, precision and statistical power, we generated a dataset of river monitoring site locations and water quality and invertebrate data. The data were from the State of Environment (SoE) programmes run by all unitary authorities, and NIWA's National River Water Quality Network (NRWQN). Eight variables were used in the analyses: water clarity (CLAR), *E. coli* concentration (ECOLI), nitrate-nitrogen concentration (NO3N), total nitrogen concentration (TN), dissolved reactive phosphorus concentration (DRP), total phosphorus concentration (TP), invertebrate taxon richness (TAXA), and Macroinvertebrate Community Index score (MCI). These variables were selected because they are measured by most councils, and because they represent a wide across-sites gradient in variability. The original, raw dataset consisted of data from approximately 1500 SoE and NRWQN sites. This site list was filtered down to a list of 991 sites for which sampling duration and frequency were sufficient to calculate accurate medians for variables, and which regional councils appear to be committed to monitoring in the future. The sites retained for analyses met the following conditions.

1. Sites must be used for SoE monitoring, not point-source monitoring or short-term investigations;
2. Starting date must be no later than 1 January 2006, and the ending date must be no earlier than 1 January 2010;
3. Water quality sampling frequency must be quarterly or higher, and data must be available for at least 16 quarters within the study period;
4. Invertebrate sampling frequency must be annual or higher, and data must be available for at least 4 years within the study period;
5. Sites must be located on streams of order 2 or greater.

The 991 sites in the final dataset were assigned to FENZ classes at the 20-group level, and to REC classes at the Climate/Landcover level, and at the Climate/Source of Flow/Landcover level. The two REC levels were used to vary the environmental resolution and to vary the distribution of site numbers within environmental classes. An REC reference landcover class termed Natural (N) was created by pooling the landcover categories Bare, Indigenous Forests, Tussock, and Scrub. The Natural category was used for comparisons with the impacted landcover categories Pastoral (P), Urban (U) and Exotic Forest (EF).

The distribution of monitoring sites across FENZ classes is very uneven. Nine out of 20 FENZ classes are represented by one or more monitoring sites, and 90% these are in classes A and C. The distribution of monitoring sites across REC classes is also uneven:

over 70% of sites are in four Climate/Landcover classes: Cool Dry/Pastoral, Cool Wet/Pastoral, Cool Wet/Natural, and Warm Wet/Pastoral. The majority of monitoring sites are in the Pastoral landcover classes: 626 sites (63%) are categorised as Pastoral, 286 sites (29%) as Natural, 41 sites (4%) as Urban, and 36 sites (4%) as Exotic Forest.

Results of the representativeness analysis indicated that there are many gaps in the current river monitoring network, i.e., for many FENZ and REC classes with river reaches in New Zealand, there are no monitoring sites. However, these gaps do not create a serious problem for national reporting because they occur in classes that account for a very small proportion of the river length in New Zealand.

Shortages (under-represented environmental classes in which the proportion of monitoring sites is less than the proportion of river length) are likely to be more serious problems than gaps. In the FENZ framework, the largest relative shortage is in Class H (mid-elevation, dry climate streams), which only has 6% of the representative number of monitoring sites. The most common FENZ class (C, small, lowland, hill-country streams) is over-represented by about 30%. The most severely under-represented REC Climate/Landcover classes are in the Cool Extremely Wet/Natural, Cool Wet/Natural, and Cool Dry/Natural classes, with 22 to 119 too few monitoring sites. These classes include prospective reference sites. The most over-represented classes are Cool Wet/Pastoral (with a surplus of 80 sites), Cool Dry/Pastoral (with a surplus of 38 sites), and Warm Wet/Pastoral (with a surplus of 30 sites). The general pattern indicated by the analysis is that REC classes with natural landcover tend to be under-represented relative to river abundance, and classes with pastoral and urban landcover tend to be over-represented.

The results of the precision analyses indicated that current site numbers in a small number of environmental classes are sufficient to report mean values of water quality and invertebrate variables with moderately high precision. These environmental classes are common ones in terms of site numbers (e.g., FENZ classes A and C, REC classes CD/P and CW/P). In contrast, current site numbers for the majority of environmental classes are insufficient to maintain moderately high precision for some or all variables. In general, river conditions in pastoral landcover classes can be reported with greater precision than in natural, urban or exotic forest landcover classes, because there are more monitoring sites in pastoral landcover classes. For some natural landcover classes, > 100 additional sites would be required in order to report mean river conditions with a high level of precision (i.e., half of the standard deviation for all sites).

The results of the analyses of statistical power for comparisons indicated that that there are currently enough sites to detect differences in the means of most variables in the two most abundant FENZ classes, A and C. In all other cases, there is a shortage of sites corresponding to two or more variables. For 3 FENZ classes, there are too few sites to calculate statistical power.

Comparisons that involved one or two REC pastoral landcover classes had the highest frequency of sufficient site numbers. Comparisons that involved urban and exotic forest landcover classes, for which there are relatively few monitoring sites, had the highest frequencies of site-number shortages. For many comparisons among REC classes, the current SoE network has too few sites to detect between-class differences in the means of any variable. Comparisons involving the Natural landcover category are of interest because these include potential comparisons of impacted and reference conditions. Statistical power for comparisons with reference conditions is greatest when those comparisons include the CW/N or CD/N classes. There are sufficient sites to detect differences in variable means for about half of the comparisons involving CW/N versus CW/P, CW/U and CW/EF, and CD/N versus CD/P and CD/U.

The power analyses produced some very high site-number requirements (e.g., over 1000 classes required in both sites to detect some between site differences). Reasons for these high estimates include very small differences in the means being compared, and large pooled variances. It is unlikely that very large deficiencies will be addressed by adding many new sites to the SoE network. Small improvements in power can be gained by ensuring that more of the core variables are measured at existing sites. In many cases where a large number of new sites are required to detect small differences in mean values, there are enough existing sites to produce reasonably accurate and precise means. In these cases, it may be reasonable to conclude that the mean states of the classes being compared do not differ in an ecologically meaningful way.

One of the primary purposes of a national network identified by the expert panel is to monitor effects of landuse and human activities on river conditions. There are two important observations to be made about reference sites, based on the representativeness assessment. First, environmental classes where reference conditions are likely to prevail are poorly represented in the current SoE network. Most of the shortages identified at the REC Climate and the Source of Flow levels corresponded to classes with natural landcover, in which reference sites could be located. The under-representation of natural landcover classes in the network partly reflects the varied origin of monitoring sites; many sites originated as consent monitoring sites or sites used for investigations of human impacts, and few sites have been deliberately established as reference sites. Second, the primary value of reference sites and reference classes is not their contribution to representative networks, but their comparability to impacted sites and impacted classes. In other words, a reference environmental class may be scarce in terms of river length or numbers of reaches, but monitoring sites within that rare class may be very valuable for comparison with corresponding impacted classes. This situation is exemplified by the availability of monitoring sites in Cool Dry/Lowland and Warm Dry/Lowland environments. The CD/L/P class is one of the most abundant in New Zealand in terms of river length (about 12% of total river length), and is monitored at 145 SoE sites. Similarly, the WD/L/P class is abundant and is monitored at 34 sites. The appropriate reference classes for the CD/L/P and WD/L/P classes are the CD/L/N and WD/L/P classes. Rivers in these reference classes are rare due to historical conversion of low-elevation land to agriculture. There is only one SoE site in the CD/L/N class, and none in the WD/L/N class. Despite their rarity, the remaining river reaches classed as CD/L/N and WD/L/N should have high priority for establishing new reference sites.

1 Introduction

In 2011, the Ministry for the Environment (MfE) and its contractors evaluated options for revising national-scale freshwater monitoring and reporting in New Zealand (Davies-Colley et al. 2011). These evaluations were the first steps in the MfE National Environmental Monitoring and Reporting (NEMaR) project. One of the primary aims of NEMaR is to generate information required to create a National Surface Water Monitoring Programme (NSWMP). Three broad issues in freshwater monitoring were considered: variables or analytes, indicators, and the spatial layout of monitoring networks. The reports and subsequent peer reviews identified issues that needed to be addressed in order to achieve the aims of NEMaR. To address these issues, an expert panel composed of regional council, central government, university, consultancy, and CRI staff was formed for each project area, and two workshops were held in 2011. The general aims of the workshops were to 1) seek consensus and provide recommendations on methods for improving national reporting on freshwater ecosystems; and 2) identify issues for which consensus was not reached and provide options to resolve those issues. The expert panels were instructed to report points of consensus and points lacking consensus to a steering committee composed of regional and central government staff. This report provides analyses that were recommended by the expert panel for network design.

1.1 Background

In the last decade, over 1000 river sites in New Zealand have been monitored to provide water quality and ecological data, which are used for State of Environment (SoE) reporting. There is considerable variation among monitoring sites in terms of frequency and duration of sampling, range of variables measured, and field and laboratory procedures. Most of the SoE sites are or were monitored by regional councils and unitary authorities; NIWA monitors an additional 77 sites that comprise the National River Water Quality Network (NRWQN). Hereafter, the aggregate council and NIWA sites are referred to as the “SoE network”. The number of sites in the SoE network varies over time as sites are added or dropped. The number of sites that are suitable for national assessments of river conditions vary with the data requirements of each assessment, as discussed below.

Monitoring sites are initiated for a variety of reasons, including consent monitoring and site-specific investigations by councils. Only the NRWQN sites were originally intended for national-scale reporting. As a consequence, analysts undertaking national-scale assessments must recognise that the aggregated SoE sites are not optimally configured for their work. If an entirely new national network were designed without reference to sites in the current SoE network, it would likely have a different configuration than the current network. However, it is important to note that many different configurations would be suitable for a national network, depending on the primary purposes of national reporting. These purposes and corresponding network configurations are discussed in detail below.

Broadly speaking, there are three different approaches used in New Zealand for national-scale reporting on river water quality and ecology. The first consists of broad statements about nation-wide state and trends in river conditions (e.g., “The mean river nitrate concentration across New Zealand in 2007 was X mg L^{-1}”). The second approach consists of statements about state and trends that are national in extent, but refer to specific environmental classes[1]. Examples include water quality state in mountain streams across New Zealand, and comparisons between lowland streams in agricultural catchments versus lowland streams in native forest catchments. The third approach consists of statements about the state of individual rivers or a small set of rivers. Examples include the current OECD reporting format, which requires assessments of the water quality state at monitoring sites on several (currently six) large New Zealand rivers. All three approaches to national-scale reporting have the same extent (the entire country), but different levels of spatial resolution, from coarse-resolution, nation-wide assessments to fine-resolution assessments about multiple environmental classes. When assessing multiple environmental classes, the spatial resolution directly reflects the spatial scale of the underlying classes. For example, assessments of river conditions within large climate classes will have coarser resolution and greater within-class variability than assessments of smaller subdivisions of those climate classes. Note that environmental classes in the SoE network are composed of individual monitoring sites that have been grouped into multi-site classes using the classification frameworks discussed below.

Environmental classes of rivers have been defined using numerous systems. The River Environment Classification (REC) and Freshwater Ecosystems of New Zealand (FENZ) frameworks are the most frequently used in New Zealand (Snelder and Biggs 2002; Leathwick et al. 2010). The REC defines classes using a spatial hierarchy of environmental factors presumed to influence water quality, sediment, and flow regimes, with each factor (e.g., climate, topography, geology, landcover) associated with a characteristic spatial scale[2]. As a sampling location is assigned to progressively finer classification levels, the resolution or grain-size increases, and the specificity of the environmental description increases. The FENZ river classification is a multivariate alternative to the controlling factor-based REC. FENZ uses environmental factors in combination with datasets of native freshwater fish and invertebrate distributions, with the fish and invertebrate data used to select and weight the environmental factors (Leathwick et al. 2008, 2010). FENZ classes can also vary in scale; all river reaches in New Zealand can be classified into a few large and heterogeneous groups, and these can be subdivided into progressively smaller, more numerous, and less heterogeneous groups (Leathwick et al. 2010, 2011). The REC and FENZ frameworks have both been used to analyse data from water-quality and ecology monitoring programmes. In these analyses, monitoring sites are assigned to environmental classes within each framework, and the classes are used as the basis for data summaries, trend analyses, inter-class comparisons, and assessments of stream health (e.g., Snelder and Dey 2006, Ballantine et al. 2010, Clapcott et al. 2011).

1.2 Purposes of national river water quality and ecology monitoring

National river water quality and ecology datasets are used in several types of analysis, with different purposes (Table 1-1). The primary purposes listed in Table 1-1 fall into four broad categories: representative characterisation of river conditions across heterogeneous environments; detection of differences in conditions between rivers or environmental classes;

---

[1] Environmental classes are categorical descriptions of land areas or water bodies based on physical characteristics (e.g., mountain rivers in volcanic terrain). In contrast to an eco-region that occurs in a single geographic location (e.g., South Westland rainforest), an environmental class may occur in numerous locations.

[2] E.g., climate (10^3 – 10^5 km^2), geological terrain (10 – 100 km^2), landcover 1 – 10 (km^2).

a combination of characterisation and comparisons, and statements about the conditions of individual rivers.

Design criteria for the locations of monitoring sites depend on the monitoring purposes (Table 1-1). These criteria can be applied in two different ways: designing water quality monitoring networks, or selecting sites from existing networks to include in data analyses. In the following discussion, we focus on the two primary purposes identified by the expert panel: characterisation of river conditions across heterogeneous environments, and comparisons between rivers or environmental classes. A single monitoring network is unlikely to have an optimal configuration for both purposes. That is, the proposed NSWMP built on regional SoE monitoring will require compromises to be made in the capability of the network to achieve both purposes.

Table 1-1: Primary purposes of national-scale river water quality and ecology monitoring and their corresponding design criteria.

	Primary purpose	Primary design criterion
1	Nation-wide state or trend	Representativeness
2	National-scale state or trend within multiple environmental classes	Representativeness
3	National-scale comparisons of impacted environmental classes *versus* reference classes	Statistical power
4	National-scale comparisons among impacted environmental classes	Statistical power
5	National-scale comparisons across all environmental classes	Statistical power and representativeness
6	Water quality state or trend in individual rivers	Maximal sample size and period of record

1.3 Representativeness

The first two monitoring purposes in Table 1-1 concern accuracy in the estimation of water quality and ecological conditions. A representative estimate of river condition is one in which the influence or leverage of a sampling site or environmental class is proportional to its abundance or dominance. Data from a highly representative monitoring network will give accurate estimates of water quality and ecological state and trend. It is important to note that accurate estimates of may be imprecise, which can prelude identifying state and trends with a high level of certainty.

Representative estimates of river water quality state or trend across environmentally heterogeneous areas require two conditions to be met. First, data from rivers representing all environmental classes in the area should be included. If some environmental classes are excluded, the estimated state or trend will be over-influenced by classes that are included. For example, a regional water quality analysis including agricultural and urban landcover classes but excluding forested landcover classes in the same region may erroneously indicate that water quality is substantially worse than the true region-wide state.

The second condition is that the number of monitoring sites from each environmental class, and the abundance of each environment class in the assessment area, should have the same proportions. This condition can be applied before compiling data (by selecting sites in proportion to the size of their environmental class) or after compiling data (by assigning a weighting factor to the data from each environmental class). For example, if lake-fed rivers comprise 20% of the total number of rivers (or 20% of the total river length) in an assessment area, lake-fed rivers should comprise 20% of the monitoring sites used for assessment. If these conditions are not met, rare environmental classes may have too much influence on estimated water quality and common classes may have too little influence. Proportionality or weighting factors available for use include the number of rivers in each environmental class as a proportion of total river number, the length of river channel in each class as a proportion of total river length, the relative amount of river flow, or the catchment area of a river as a proportion of total area (Snelder et al. 2006).

In addition to the conditions listed above, statements about the condition of a class of rivers need to be made with a minimum level of precision (i.e., a maximum acceptable level of uncertainty). If a monitoring network is highly representative but the number of monitoring sites is low, estimated states and trends may be accurate but imprecise (i.e., there will be large uncertainties around medians and other statistics). Precision can be expressed in terms of the variation of values around a mean, median, or trend line (e.g., standard error, confidence intervals). In most cases, precision increases as the number of monitoring sites increases, and the number of sites required to achieve a minimum level of precision for a monitoring variable can be estimated using pilot datasets (Ward et al. 1990).

1.4 Inter-class comparisons and statistical power

Two general types of comparisons are common in water quality analyses: comparisons of modified or impacted landcover classes versus reference classes, and all possible comparisons among land-use or landcover classes. Modified landcover classes are those in which urbanisation, agriculture and other human activities have transformed landcover and may adversely affect water quality. Reference classes (see Section 1.5) are those in which the predominant landcover is native vegetation, and river water quality and biota are presumed to represent near-natural conditions. Comparisons of modified and reference classes are used to identify water quality degradation in modified classes, and trends indicating temporal water quality changes in modified classes. In the latter case, reference classes are used to identify temporal trends caused by non-anthropogenic factors such as natural climate variability, or by global anthropogenic factors such as atmospheric nutrient deposition. Trends in river conditions in modified classes can then be corrected by subtracting any trends detected in reference classes.

Detecting a statistically significant between-class difference in water quality or ecology depends on the number of sites in each class being compared, within-class variance, and the magnitude of the difference The power of statistical tests to detect differences generally increases directly with site numbers and with the magnitude of the differences. Here, statistical power refers to the probability of not rejecting the null hypothesis of no difference between means when a difference actually exists. For pair-wise comparisons (e.g., between modified and reference sites), the number of sites required in each class to detect a given difference with a given level of confidence can be estimated using pilot datasets (Ward et al. 1990). Sample-size requirement estimates can be used in two ways. The power of an existing SoE network to detect standardised differences can be assessed by comparing site-number requirements to the number of existing sites in each environmental class. Conversely, estimated site-number requirements can be used in the design of new SoE networks by ensuring that appropriate numbers of sites are established in each environmental class.

1.5 Reference sites and reference classes

Two types of reference sites are used in river water quality analyses: minimally disturbed, and least-disturbed (Stoddard et al. 2008). Minimally-disturbed sites are located in near-pristine catchments with minimal human activity. No locations on earth are entirely unaffected by human activities, which are often global in scale (e.g., atmospheric nutrient deposition, species invasions). In New Zealand, minimally-disturbed sites have been defined as those with intact indigenous landcover (e.g., native forest, ungrazed native grassland) in the upstream catchment, and with no major human impacts from roads, dams or other structures (Joy and Death 2003, Larned et al. 2004, Collier et al. 2007). Minimally disturbed sites have been used in New Zealand for comparison with impacted sites to identify the effects of land-use practices (e.g., Joy and Death 2003,), and as “targets” for stream restoration and rehabilitation (e.g., Parkyn et al. 2010). When used as rehabilitation targets, reference sites provide the water quality and biological conditions against which responses to rehabilitation are assessed, and the overall success of rehabilitation projects is judged.

Least-disturbed sites are located in non-pristine catchments, which have water quality conditions that represent the best achievable water quality associated with a predominant landuse. River assessments based on comparisons between impacted sites and least-disturbed sites have not been identified as a national objective for monitoring in New Zealand, and they are not considered further.

Water quality and ecological conditions at minimally-disturbed reference sites represent reference conditions. Relatively few long-term water quality monitoring sites in New Zealand have been established for the specific purpose of generating reference condition data (Collier 2008, Davies-Colley et al. 2011). Instead, reference environmental classes have been defined and mapped, and the existing SoE monitoring sites within those classes have been recommended as de facto reference sites (Snelder and Lessard 2009, Leathwick et al. 2010). Definitions of reference classes have been based on a predominance of indigenous landcover in the catchment upstream from monitoring sites. There are many additional anthropogenic factors that can affect water quality and ecological conditions (e.g., roads, non-native species, recreation activities). Including or excluding factors from the reference class definition will affect the resulting reference conditions, and the number of candidate reference sites. Water quality and biological conditions vary across reference sites in response to natural variability in climate, geology and other environmental factors. Aggregating reference sites into reference classes without regard to between-site differences in environmental factors will result in high within-class variability. The effect of this natural variability can be reduced by subdividing rivers using hierarchical frameworks such as the REC and FENZ (Collier et al. 2007).

In previous comparisons of impacted and reference river classes in New Zealand, reference classes were identified on the basis on their landcover classification in the REC or the Landcover Database-2 (Larned et al. 2003, Snelder and Lessard 2009). In these studies, several narrowly defined landcover classes (e.g., indigenous forest, flax-land, fern-land, matagouri, tussock) were pooled to form a general class of rivers in catchments where human land-use is presumed to be minimal. The number of river reaches that were potential reference sites varied widely across REC Climate, Source of Flow and Geology classes, with a notable scarcity of potential reference sites in low-elevation rivers (Larned et al. 2003).

1.6 Workshop recommendations

During the workshops described above, the expert panel for network design produced a number of consensus statements and recommendations, some of which provided directives for the current assessment of the SoE network. Those directives were:

1. The primary purposes of a national freshwater monitoring programme are to obtain 1) representative estimates of environmental states and trends, and 2) comparisons of state and trend among selected environmental classes, including reference classes. The first purpose has higher priority than the second.

2. Comparisons of reference classes or reference conditions versus impacted classes should be included. The panel considered that some comparisons between impacted classes are also important, but did not specify which classes should be included. The availability of unimpacted reference sites that generate reference-condition data is a high priority.

3. The current river monitoring network needs to be evaluated regarding its ability to meet the primary purposes of the NSWMP, before identifying prospective new sites or alterations to the network. For the first purpose (estimates of state and trend), analyses of the representativeness of the current river monitoring network, and statistical power and site number requirements are needed. For the second purpose (comparisons among selected classes), an analysis of statistical power and site number requirements is needed.

4. Use of both Freshwater Environments of New Zealand (FWENZ) and River Environment Classification (REC) for assessment of the environmental classes for the current ecological monitoring network was recommended. Updated site lists and datasets are required to carry out the assessments of the current networks.

The directives listed here provided us with a workflow consisting of assessment steps and datasets that needed to be compiled (Figure 1-1). In the remainder of this report, we describe the methods used to carry out the steps in the workflow, and the results of the assessments.

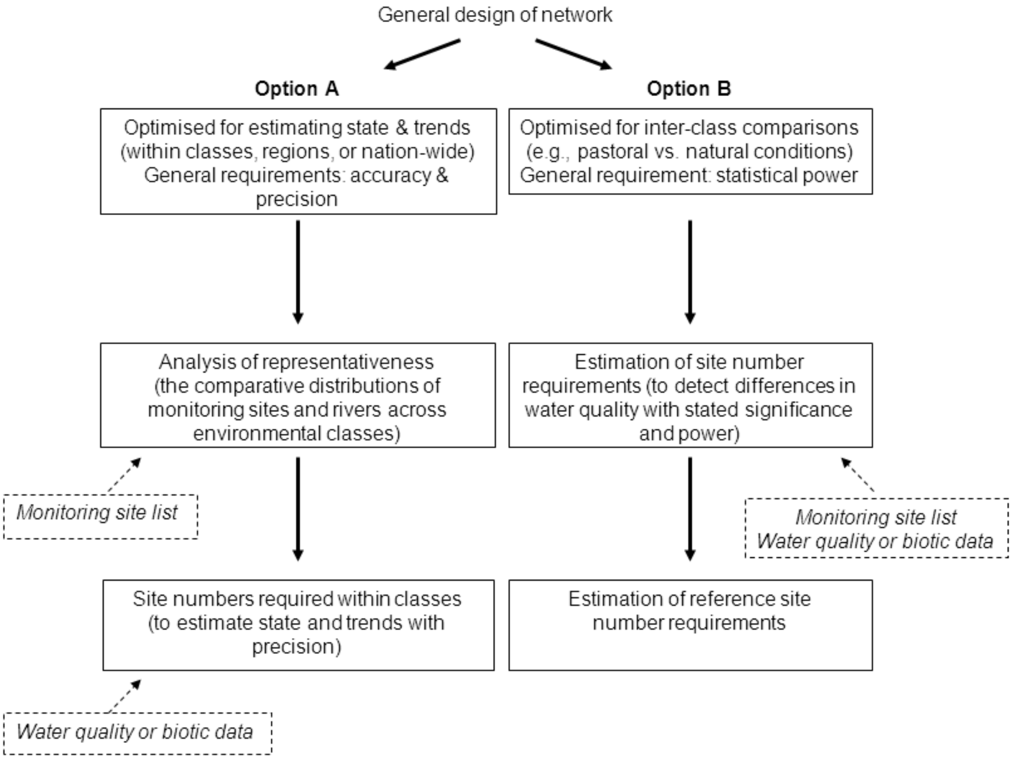

Figure 1-1: Flow chart indicating steps in the evaluation of the existing water monitoring network. The two pathways correspond to the primary purposes of the national network identified by the expert panel for network design: representative estimates of environmental states and trends, and comparisons of state and trend among selected environmental classes, including reference classes. Boxes with solid lines indicate evaluation steps; boxes with dashed lines indicate data requirements.

2 Methods

2.1 Compilation of sites, sampling dates, and water quality and invertebrate data

We requested river water quality and invertebrate monitoring data from all 16 unitary authorities (regional, district and city councils) and the NIWA National River Water Quality Network in 2011. We requested datasets that extended in time to late 2010 or later. For each monitoring site, we requested three types of location information: NZMG (NZMS260) or NZTM coordinates, council identification number, and location description (e.g., Opuha River at Skipton Bridge).

We requested data for eight variables: water clarity (CLAR), *E. coli* concentration (ECOLI), nitrate-nitrogen concentration (NO3N), total nitrogen concentration (TN), dissolved reactive phosphorus concentration (DRP), total phosphorus concentration (TP), invertebrate taxon richness (TAXA), and Macroinvertebrate Community Index score (MCI). These variables were selected because they are measured by most councils, and because they represent a wide across-sites gradient in variability. For example, median TN varies approximately 400-fold across monitoring sites, while median ECOLI varies 1000-fold across sites. This range of variability affects the results of statistical power analyses, and it is instructive to compare the number of sites required to achieve the same level of statistical power for different variables. To ensure consistency across sites, TAXA and MCI were recalculated from invertebrate datasets for each site, as discussed in the next section.

2.2 Data processing

The first data processing step was to assess methodological differences for each of the eight variables. For most of the variables, two or more measurement procedures were represented in our dataset. We grouped data by procedure, then pooled data for which different procedures gave comparable results. Data measured using the less-common and non-comparable methods were eliminated. Table 2-1 lists the most common procedures used for each variable, and the procedures corresponding to the data retained for analysis.

The data produced by multiple procedures used to measure ECOLI, NO3N, CLAR, TAXA and MCI were pooled, based on the assumption that the different procedures gave comparable results. In contrast, the different procedures used to measure TN and TP are unlikely to give comparable results. Most councils and the NRWQN use the persulfate digest method and unfiltered water samples. Hereafter, the data produced by this procedure are referred to as TNUNF and TPUNF. A smaller group of councils uses Kjeldahl digest procedures (TKN) and calculates TN as TKN + NNN. At least one council uses filtered samples, which corresponds to total dissolved nitrogen and phosphorus. These methods could generate substantial differences in TN and TP concentrations. Therefore, TNUNF and TPUNF data were retained for analysis, and data from other procedures were omitted.

The second data-processing step was to calculate the invertebrate variables TAXA and MCI. TAXA refers to the number of invertebrate taxa in a sample. Individual taxa are reported at a range of taxonomic levels from phylum to species, and it was necessary to standardise taxonomic levels to make sites comparable. For a standard taxa list, we used Table 1 of the User’s Guide to the Macroinvertebrate Community Index (Stark and Maxted 2007). For taxa identified to lower taxonomic levels than those in the standard list, we shifted the taxonomic level up (e.g., the genera *Alboglossiphonia*, *Barbronia*, and *Placobdelloides* were shifted to Subclass Hirudinea). Standardisation resulted in about 15% of our TAXA counts being lower than the counts provided by councils, but fewer than 2% of the counts differed by more than two taxa.

Table 2-1: Measurement procedures for water quality variables. TAXA and MCI procedures are from Stark et al. (2001). Procedures retained: data generated by the procedures in this column, and corresponding monitoring sites, were retained for analysis in this study.

Variable	Measurement procedures	Procedures retained
ECOLI	Colilert QuantiTray 2000 Membrane filtration	Both procedures (presumed to give comparable results)
NO_3N	Nitrate-N, filtered, Ion chromatography Nitrate-N + nitrite-N (or “NNN”), filtered, cadmium reduction Nitrate + Nitrite-N – Nitrite-N (filtered, Azo dye colourimetry)	All procedures (nitrite presumed to be negligible in unpolluted water)
TN	Unfiltered, persulfate digest Filtered, measured as dissolved inorganic+organic nitrogen Mixed, by Kjeldahl digest (TKN + NNN)	Unfiltered, persulfate digest
TP	Unfiltered, persulfate digest Filtered, measured as dissolved inorganic+organic phosphorus	Unfiltered, persulfate digest
DRP	Filtered, molybdenum blue colourimetry	molybdenum blue colourimetry
CLAR	Black-disks and clarity tubes	Both procedures (presumed to give comparable results)
TAXA and MCI	Collection procedures C1, C2, C3, C4 Processing procedures P1, P2, P3	All procedures (presumed to give comparable presence/absence data for calculating non-quantitative TAXA and MCI scores

MCI refers to the non-quantitative Macroinvertebrate Community Index. We used the non-quantitative MCI in lieu of the quantitative (qMCI) or semi-quantitative (sqMCI) forms of MCI because some council datasets did not include abundance data. MCI scores are based on presence/absence data, not abundance data. Two versions of MCI are available, one for hard-bottomed streams and one for soft-bottomed stream (Stark and Maxted 2007). We did not separate monitoring sites into hard- and soft-bottomed sites for MCI calculations, for two reasons. First, splitting the sites into two groups based on substrate would have required two parallel analyses with fewer sites in each, which would have affected the representativeness and power analyses. Second, there was insufficient information about site substrate to consistently assign sites to hard- and soft-bottomed groups. Instead, we calculated hard-bottom MCI scores for all sites. Several councils provided MCI scores as part of their datasets. However, we used raw invertebrate data for each site and sampling date to calculate new MCI scores. This ensured that the calculations were all made using the same suite of taxa and tolerance values. Our MCI scores differed from those provided by councils for about 15% of sites. For most of these sites, either our recalculated TAXA count differed from the original, the original MCI score was derived using the soft-bottomed version, or tolerance values used in the original score differed those listed in Stark and Maxted (2007). Most of the differences between the original scores and recalculated scores were very small.

Councils use a variety of formats for reporting data below analytical detection limits (BDL), including flagging the data (e.g., “BDL”, “<0.01”), replacing the data with fabricated values (e.g., 0.5×DL), and reporting the measured value regardless of the DL. We treated BDL issues in two steps. First, we flagged all data that were reported as or appeared to be BDL. If ≥ 15% of the data for a variable from a single site were BDL, the variable was eliminated from the site in subsequent analyses. This step ensured that no statistical analyses were carried out with large proportions of fabricated data. Second, for those variables with < 15% BDL data, the BDL data were replaced with values equal to 0.5×DL.

2.3 Rules for inclusion of monitoring sites in analyses

1. To determine which sites qualified for analyses of representativeness and statistical power, we applied five selection rules to the site lists provided by councils and NRWQN. Our aim was to identify sites for which sampling duration and frequency were sufficient to calculate accurate medians for variables, and which regional councils appear to be committed to monitoring in the future.

2. Sites must be used for SoE monitoring, not point-source monitoring or short-term investigations;

3. Starting date must be no later than 1 January 2006, and the ending date must be no earlier than 1 January 2010;

4. Water quality sampling frequency must be quarterly or higher, and data must be available for at least 16 quarters within the study period;

5. Invertebrate sampling frequency must be annual or higher, and data must be available for at least 4 years within the study period;

6. Sites must be located on streams of order 2 or greater.

The rationale for each of these rules is as follows:

Rule 1 (State-of-Environment). We requested separate lists of SoE and non-SoE sites from several councils, but in practice the definition of SoE monitoring varied widely among councils. We therefore used each council’s definitions wherever possible, but used details such as site location and description to identify and reclassify sites used for short-term investigations and consent monitoring, or to monitor point-source inputs (such as sites adjacent to sewage treatment and industrial outfalls). After removing these sites, and sites for which none of the eight variables used in this analysis were measured, we retained a list of **1472** long-term monitoring sites where at least one variable was measured. Invertebrates are or were monitored at **1105** of these sites.

Rules 2-4 (Duration and frequency). We set the starting date to no later than 2006 to ensure that recently established sites were included, as well as long-established sites. Setting the ending date to 2010 or later ensured that most or all sites were still in operation, without excluding sites for which the most recent finalised data were not yet available. Our sampling frequency rules (quarterly or higher for water quality, annually or higher for invertebrates) were chosen to ensure that sufficient data were available to calculate accurate means and standard deviations. Applying these rules reduced the number of monitoring sites available for analysis to **1069**, including **326** (31%) used for both invertebrate and water quality monitoring, **335** (31%) for invertebrate monitoring only, and **408** (38%) for water quality monitoring only. Sites used for water quality monitoring are those where at least one water quality variable was measured over the study period.

Rule 5 (Stream order). We omitted monitoring sites on order 1 streams for two reasons. First, over half of the reaches in New Zealand are first-order, both by number and by total stream length. Since fewer than 10% of monitoring sites are on order 1 streams, a representativeness analysis including these streams would inevitably indicate large mismatches in most environmental classes. Second, the majority of sites on order 1 streams are used only for invertebrate monitoring. To avoid these problems, first-order streams were excluded from both the monitoring site dataset and the river network before comparison. Of the **1069** SoE sites in the previous step, **78** (7%) are on order 1 streams, and **64** of those are used only for invertebrate monitoring, i.e., 10% of invertebrate monitoring sites are on first-order streams. Removing these reduced the total to **991** sites on second- through eighth-order streams, including **601** invertebrate monitoring sites.

2.4 Environmental classification

Monitoring sites were positioned using NZTM coordinates on a 1:50,000 scale digital map of the New Zealand river network; this map is part of the River Environment Classification GIS database. Accuracy in positioning was initially checked by determining the Euclidian distance between the site location indicated by NZTM coordinates and the centroid of the nearest NZReach. NZReaches are georeferenced river reaches defined by upstream and downstream confluences. Each of the 576,273 river reaches in New Zealand has been assigned a unique NZReach number. In most cases, the site location indicated by NZTM coordinates was within 200 m of a reach centroid. In cases where the distance was greater than 200 m, location information provided by councils (e.g., site name, stream order) was used to identify the correct NZReach. Over 600 of the monitoring sites used in this study had been assigned NZReaches in previous studies; less than 100 sites in the present study required manual checking. As part of the manual checking process, we ensured that no council and NRWQN sites coincided, and that there were no cases of multiple sites mapped to the same NZReach.

Each of the NZReaches has been classified in the REC system, and all but 9000 have been classified in the FENZ system. This small (1.5%) discrepancy is probably due to poor fits of some reaches to the FENZ spatial model. Those NZReaches lacking FENZ classification, plus reaches in a poorly defined REC class “M” (primarily small lake-margin and dune streams) were dropped from the representativeness analysis, leaving a set of 565,029 reaches. The REC and FENZ classes corresponding to the monitoring sites in our dataset were identified using the NZReach for each site.

We used the FENZ classification at the 20-group level. Brief descriptions of these classes are listed in Table 2-2. The 20-group level provides relatively course environmental resolution, and there is substantial within-group heterogeneity at this level. However, higher levels with more groups resulted in an absence of monitoring sites in most groups, which hindered the power analyses.

For classifications based on the REC we used three environmental factors: Climate, Source of Flow, and Landcover. We used two different levels of the REC, Climate/Source of

Flow/Landcover, and Climate/Landcover (Table 2-3). We began by defining a reference environmental landcover class, termed Natural (N), by pooling the categories Bare, Indigenous Forest, Tussock, and Scrub. We established three groups of impacted classes corresponding to the existing Landcover categories Pastoral (P), Exotic Forest (EF), and Urban (U). An eighth category, Wetland (W), which accounts for less than 0.1% of the REC network and 0.2% of monitoring sites, was included in our calculations for accuracy, but was not used in power analyses. The REC assigns landcover classes to NZReaches based on the predominant landcover in the upstream catchment, subject to several rules (e.g., if landcover is > 25% pastoral the reach is classed as Pastoral). As a result, reaches classed as Natural may have less than 100% natural landcover in their catchments.

The Climate/Source of Flow/Landcover classes provide relatively high environmental resolution, with 88 classes represented in New Zealand (excluding first-order reaches) after pooling landcover classes in the Natural category. However, many of these classes have few or no monitoring sites, which limits the number of power analyses that can be made. The Climate/Landcover classes (36 classes represented, excluding first order reaches) have lower environmental resolution, but a higher proportion of these classes have monitoring sites. We therefore focused our power analysis on landcover within climate as a primary determinant of stream condition, and a basis for inter-class comparison. We did not consider the effects of geology (the third level of the REC) on stream condition, in order to limit the complexity of the analyses. We did not consider stream order (the fifth level of the REC), based on results of a previous analysis in which effects of stream order on water quality were not detectable (Larned et al. 2004). Details of REC classes and categories are given in Snelder and Biggs (2002).

2.5 Analysis of representativeness

The analysis of representativeness was used to assess the degree to which the SoE monitoring sites in current use are distributed in environmental space in proportion to the abundance of river reaches. A close match in these distributions allows us to extrapolate water quality and ecological conditions from a small number of monitoring sites to a large number of river reaches.

Total river abundance, and the abundance within individual environmental classes, can be measured in terms of reach numbers, lengths of river, quantity of flow, or other metrics. We used river length rather than number of reaches as a measure of abundance in order to reduce the influence of very short reaches. We then calculated the representative number of monitoring sites in each class as the proportion of river length in each class (excluding first-order streams), multiplied by the total number of monitoring sites. For example, FENZ class A accounts for 86,420 river km, representing 22% of the total. In a representative monitoring network, 22% of monitoring sites would be on FENZ Class A river reaches.

Representativeness was calculated as the difference between the existing number of sites in each class and the representative number of sites. Positive values indicate that the class is over-represented (i.e., there are more sites than needed), and negative values indicate that the class is under-represented (i.e., there are fewer sites than needed).

Total river length in 44 of the 88 REC Climate/Source of Flow/Landcover classes is less than 100 km (i.e., less than 0.05% of the New Zealand total), and the representative number of sites was zero. To simplify the representativeness analysis, we excluded these minor classes unless there was a monitoring site in the class. In each of the latter cases, the class was over-represented.

To facilitate interpretation of the representativeness analysis, we distinguished two types of under-representation: “gaps” where there are no monitoring sites in a river class, and “shortages” where there are too few sites to achieve a minimum statistical power for a given variable. It is important to note that, if a single variable is measured at a site, that site is considered to be established and available for other variable measurements. Gaps are therefore defined in terms of presence or absence of sites, not in terms of the range of variables measured.

Table 2-2: FENZ classes at the 20-group level and their environmental attributes. Values in parentheses are proportions of total river length in New Zealand Source: Leathwick et al. (2008).

FENZ class	Description
A (21.5%)	Very small streams, mild temperatures, lowland, low gradient, sand, very unstable flow
B (0.4%)	Very small streams, mild temperatures, lowland, low gradient, sand, peaty, very unstable flow
C (43.9%)	Small streams, mild temperatures, lowland to hill country, coarse gravel, unstable flow
D (4.6%)	Very small streams, mild temperatures, lowland, inland, gravel, unstable flow, dry climate
E (0.3%)	Large rivers, mild temperatures, lowland, inland, coarse gravel, stable flow, dry climate
F (0.5%)	Very small streams, mild temperatures, lowland, inland, sand, dry climate
G (10.4%)	Small streams, cool temperatures, mid-elevation, coarse gravel, unstable flow, dry climate
H (7.2%)	Small streams, cool temperatures, mid-elevation, inland, cobbles, dry climate
I (0.4%)	Rivers, mild temperatures, mid-elevation, coastal, cobbles, stable flows, wet climate
J (4.2%)	Small streams, cool temperatures, mid-elevation, inland, cobbles, stable flows, wet climate
K (0.1%)	Small rivers, cool temperatures, mid-elevation, inland, cobbles, stable flows, glacial influence, wet climate
L (0.2%)	Streams, cold temperatures, mid-elevation, inland, cobbles, stable flows, glacial influence, wet climate
M (0.03%)	Rivers, mild temperatures, coarse gravel, strong glacial influence, wet climate
N (3.5%)	Very small streams, cold temperatures, high elevation, inland, cobbles, stable flow
O (1.1%)	Small streams, cold temperatures, high elevation, inland, wet climate, cobble, stable flow
P (0.8%)	Very small streams, very cold, high elevation, inland, cobbles, stable flow
Q (0.4%)	Very small streams, very cold, high elevation, inland, boulders, stable flow
R (0.01%)	Very small streams, very cold, high elevation, boulders, wet climate
S (0.4%)	Small streams, very cold, high elevation, inland, boulders, glacial influence, wet climate
T (0.2%)	Small streams, very cold, high elevation, inland, boulders, glacial influence, wet climate

Table 2-3: Summary of REC classes and notation for classifying stream reaches. Classes shown are for classifying river reaches at the Climate/Source of Flow/Landcover level and the Climate/Landcover level. For complete list of REC classes, see Snelder and Biggs (2002).

Classification level	Classes	Notation
Climate	Cool Extremely Wet	CX
	Cool Wet	CW
	Cool Dry	CD
	Warm Extremely Wet	WX
	Warm Wet	WW
	Warm Dry	WD
Source of Flow	Mountain	M
	Hill	H
	Lowland	L
	Lake	Lk
Landcover	Pastoral	P
	Urban	U
	Natural (Bare + Indigenous Forest + Tussock + Scrub)	N
	Exotic Forest	EF
	Wetland	W

2.6 Analyses of precision and statistical power

The purpose of our analyses was to determine the minimum number of monitoring sites needed in each environmental class to estimate water quality and invertebrate community variables for that class with specified levels of precision and statistical power. This approach is prospective rather than retrospective, seeking to inform future monitoring network design rather than characterise current environmental parameters based on existing data. This type of analysis, under the collective heading of *power analysis*, is often undertaken (or at least recommended) when designing studies to determine the level of sampling effort required to detect differences between treatments or sites (Cohen 1988, Zar 1999). It is also used in water quality network design to estimate numbers of sites required to detect trends or differences between areas at a stated level of certainty (Ward et al. 1990, Dixon and Chiswell 1996). In this study, we extend the power analysis approach to estimate site number requirements for within-class precision and between-class comparisons.

Power analysis requires sufficient a priori knowledge of the monitoring data to characterise variation within and between classes, and hence to quantify the minimum site numbers needed for each class. In particular, preliminary estimates are needed of the class means and variances, which determine the between-class differences we wish to detect, and the levels of precision we wish to achieve. Both of these objectives are referred to as “effect sizes”. For this study, we used estimated REC and FENZ class means and variances for each variable in the 991-site dataset. These estimates are the best a priori information available to for the existing network but we do not assume they are precise and accurate, particularly for classes that are poorly represented.

Power analysis also requires a priori specification of the desired statistical significance level (denoted α), and statistical power (denoted by $1–\beta$). In the language of statistical hypothesis testing, α is the probability of falsely rejecting the null hypothesis of no difference (i.e., a false-positive error), and β is the probability of failing to reject the null hypothesis when a significant difference exists (i.e., a false negative error). Therefore, statistical power, or “sensitivity” can be estimated as $1–\beta$. In this study we used $\alpha = 0.1$, equivalent to accepting a 10% probability of false positives, and $\beta = 0.2$, giving an 80% probability of avoiding false negatives. These values are relatively lenient, but are consistent with recommendations for power analyses for water quality testing (Ward et al. 1990, Snelder et al. 2006).

Data processing. We consolidated water quality and benthic invertebrate data for all data sources into a single data structure, with each record representing one measurement of one variable at one site on one date. After applying the data filtering rules described in Section 2.3, our working dataset comprised 143,757 records.

We calculated median values for each variable at each site, and used histograms and normal plots to assess the suitability of the resulting data for parametric analysis. Non-normal distributions confirmed that log transformation was appropriate for the six water quality variables, but that TAXA and MCI did not require transformation. Departures from normality for the resulting distributions were minimal (Figure 2-1), and were assumed to be small enough to ignore for all subsequent analyses. The resulting dataset contained 4,237 records, each representing a median for all available variable-by-site combinations.

We used this dataset to calculate means and variances of site medians for each variable by environmental class, for all variable-by-class combinations containing at least one median. We note that variances of site medians describe the spatial variability in a class, not the temporal variability. The class means and variances comprised a third data structure, representing a second level of condensation from the raw data, with a total of 501 records. These comprised 64 variable-by-class means for 9 FENZ classes; 144 means for 22 REC Climate/Landcover classes; and 293 means for 46 REC Climate /Source of Flow/Landcover classes. The number of variable-by-site combinations available within each class ranged from 1 to 454 for the FENZ classes; from 1 to 154 for the REC Climate/Landcover classes; and from 1 to 102 for the REC Climate /Source of Flow/Landcover classes. These data were the basis for all analyses of precision and power.

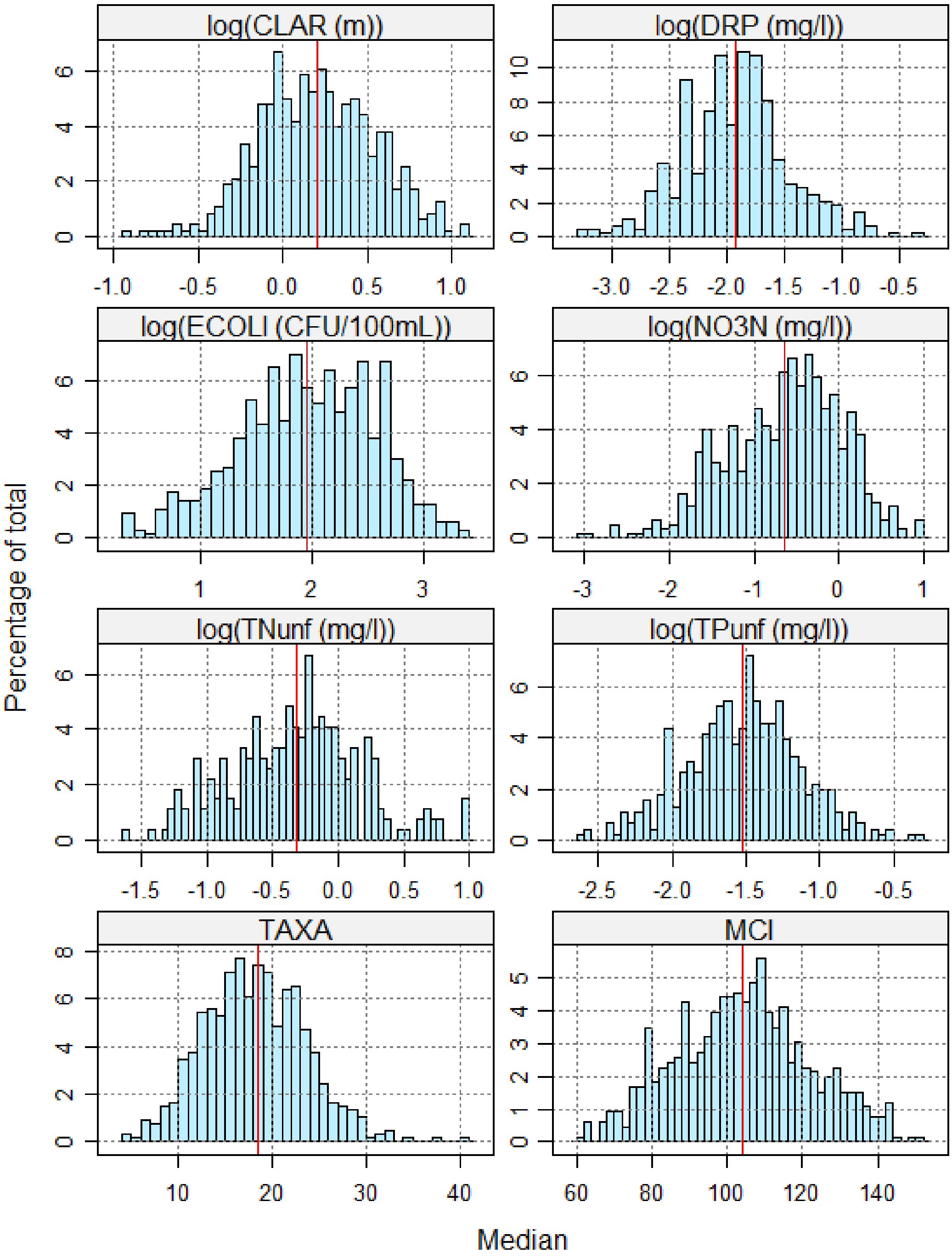

Figure 2-1: Frequency distributions of the values of water quality and invertebrate variables. Water quality variables were log10-transformed. Vertical red line indicates mean.

Site number estimations for precision. Having specified a desired significance level and power ($\alpha = 0.1$, $1–\beta = 0.8$), we must also specify the effect size in order to estimate site number requirements. In the case of our precision analysis, the effect size is the desired level of precision (Zar 1999). Due to the inherent differences in variation among the water quality and ecological variables, there was no single level of precision applicable to all variables. We therefore chose an effect size for each variable equal to the standard deviation for that variable (SD_{var}) over all classes of a given type (i.e., FENZ or REC). This definition can be applied objectively to all variables and classes. We ran these analyses using effect sizes of 1 (effect = $1SD_{var}$) and 0.5 (effect = 0.5 SD_{var}) to assess required sample sizes for different levels of precision. Further details and a worked example are given in Appendix 1.

Site number estimations for comparisons. We obtained a second measure of required site numbers in each environmental class by considering the power of the current network to detect differences in variable means between classes. The dataset used for these analyses consisted of all pairs of site means, for a given variable and class type (FENZ or REC), with at least three sites in each member of the pair (so as to avoid using means or SDs based on only one or two site medians). For these analyses, the relevant data were the means for each member of the pair being compared and their pooled standard deviation (Zar 1999), with the effect size defined as the absolute difference between the means. Further details and a worked example are given in Appendix 1.

The estimated site-number requirements for between-class comparisons assume equal site numbers in both classes. This approach was based on the fact that site number requirements are generally minimised when classes are equal-sized. When comparing the estimated site-number requirements with existing site numbers, there are three possible outcomes: both classes have sufficient sites; the larger class has sufficient sites, but the smaller class does not; neither class has sufficient sites. In some cases, there were too few sites in one or both classes to run the power analysis. In these cases, there are too few sites in at least one class to detect statistical differences in means, by definition.

Database and computations. Collated regional council and NRWQN datasets were stored in one of two MS-Access ™ databases, for water quality and stream invertebrates respectively. All subsequent data processing and analysis was performed using R 2.12.1 (R Development Core Team, 2010), with power analyses implemented using either the R *power.t.test* function, or a custom function based on Example 8.4 of Zar (1999).

3 Results

3.1 Distribution of sites and variables

As noted in Section 2.3, there were 991 sites in our dataset that satisfied the five rules for inclusion in the analyses. The distribution of sites across New Zealand is shown in Figure 3-1. Approximately 30% of the sites are used for both water quality and invertebrate monitoring; the remaining sites are used for water quality or invertebrate monitoring exclusively.

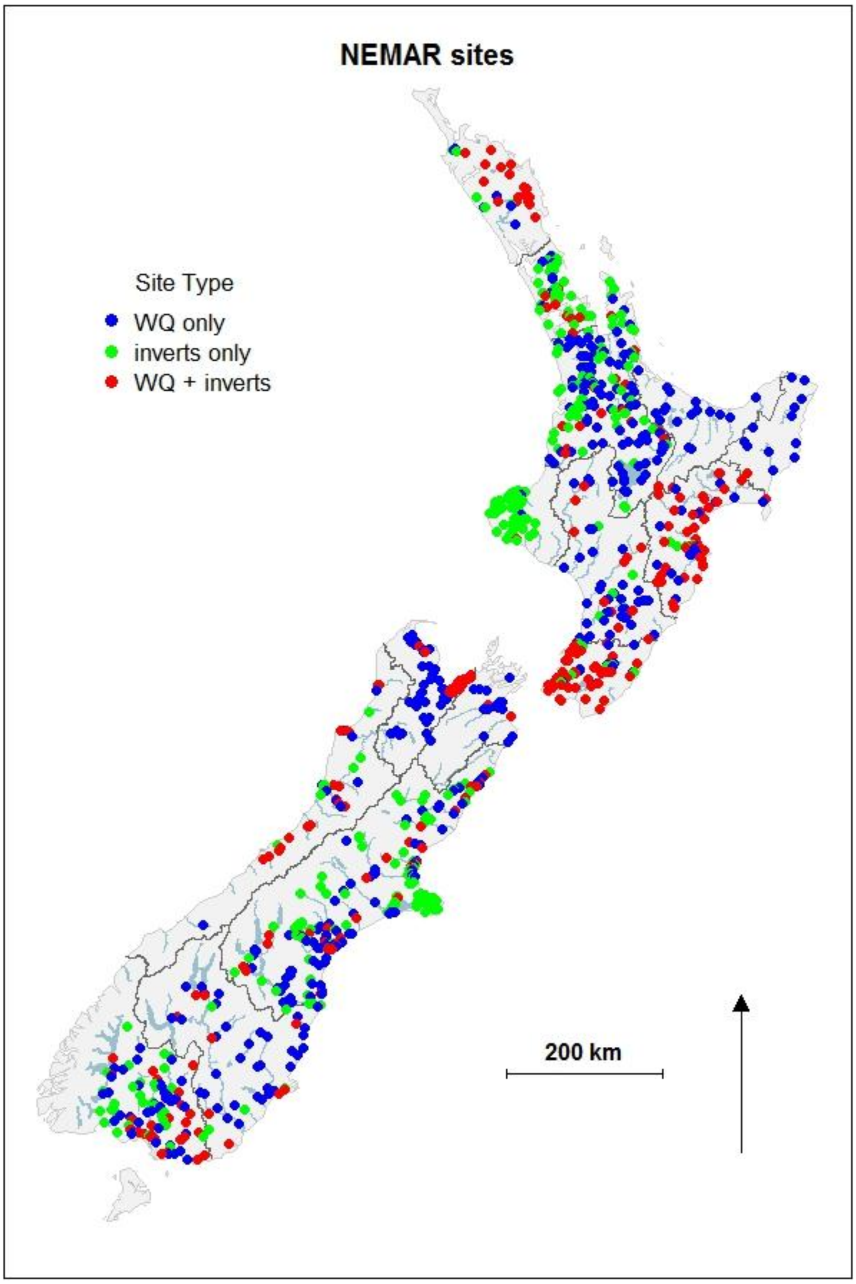

Figure 3-1: Distribution of 991 water quality (WQ) and invertebrate monitoring sites used in analyses. At least one of six water quality variables was measured at each WA site.

Nine out of 20 FENZ classes are represented by one or more monitoring sites, and 90% of the sites are in Classes A and C (Table 3-1). Class A is characterised by small, sand-bed, lowland streams in dry climate areas of the North Island and eastern South Island. Class C is characterised by small, gravel-bed streams in dry climate areas in lowland and hill country throughout New Zealand. FENZ classes are based on multivariate classifications, so each class includes some reaches that do not closely match the characterisations in Table 2.2.

Table 3-1: Monitoring sites in FENZ classes at the 20-group level. Values under each variable are the number of sites within a class at which the variable is measured. Values in the "Total" column are the number of sites in each class for which at least one variable is measured. Class descriptions are in Table 2-2.

FENZ class	Variable								
	CLAR	**DRP**	**ECOLI**	**NO3N**	**TNUNF**	**TPUNF**	**TAXA**	**MCI**	**Total**
A	65	128	130	132	61	115	105	105	**202**
B	3	2	2	3		3	1	1	**3**
C	351	273	399	377	159	258	432	432	**670**
D		15	19	14	14	16	11	11	**22**
E	7	9	11	9	9	9	7	7	**12**
G	35	40	50	51	18	42	50	50	**76**
H	1	1	1	1	1	1	2	2	**4**
I	1	1	1	1	1	1	1	1	**1**
J	0	1	0	1	0	0	0	0	**1**
Total	**463**	**470**	**613**	**589**	**263**	**445**	**609**	**609**	**991**

Twenty of the 34 REC Climate/Landcover classes present in New Zealand (excluding those composed only of first-order reaches) are represented by at least one monitoring site (Table 3-2). The distribution of monitoring sites is uneven across Climate/Landcover classes: over 70% of sites are in four classes, Cool Dry/Pastoral, Cool Wet/Pastoral, Cool Wet/Natural, Warm Wet/Pastoral. The distribution of monitoring sites is also uneven across Landcover classes: 626 sites (63%) are in the Pastoral class, 286 sites (29%) are in the Natural class, 41 sites (4%) are in the Urban class, and 36 sites (4%) are in the Exotic Forest class.

Forty-four of the 88 REC Climate/Source of Flow/Landcover classes present in New Zealand (excluding those composed only of first-order reaches) are represented by one or more monitoring sites (Table 3-3). The distribution of monitoring sites is very uneven across Climate/Source of Flow/Landcover classes.

The variable-class combinations in Tables 3-1, 3-2. and 3-3 indicate the differences among sites in the number of quality and invertebrate variables measured. For several variables, no measurements are made at sites in some FENZ and REC classes. These omissions include invertebrate monitoring in FENZ class J (the sixth most abundant FENZ class), and in REC classes WX/N, CD/L/N, and CD/Lk/N; all of these REC classes include potential reference sites. For several classes (e.g., FENZ classes B, H and I, and REC classes CD/M/N and WW/H/N), invertebrates are monitored at one or two sites, which is probably inadequate for deriving reference conditions, as discussed below.

Table 3-2: Monitoring sites in REC Climate/Landcover classes. Values under each variable are the number of sites within a class at which the variable is measured. Values in the "Total" column are the number of sites in each class for which at least one variable is measured. Classes and notation are listed in Table 2-3.

REC Class	Variable								Total
	CLAR	DRP	ECOLI	NO3N	TNUNF	TPUNF	TAXA	MCI	
CD/EF	0	0	0	0	0	0	1	1	**1**
CD/N	0	15	19	14	10	9	15	15	**29**
CD/P	51	129	152	143	95	112	109	109	**208**
CD/U	1	7	7	7	6	7	7	7	**13**
CD/W	2	1	1	2	0	2	1	1	**2**
CW/EF	20	16	21	18	3	8	15	15	**25**
CW/N	91	60	91	87	28	49	101	101	**147**
CW/P	126	85	108	111	61	105	122	122	**192**
CW/U	6	7	7	7	1	4	7	7	**7**
CX/N	40	23	45	39	18	21	49	49	**77**
CX/P	13	0	13	0	1	1	28	28	**31**
WD/P	14	20	23	22	4	11	14	14	**34**
WD/U	1	6	5	6	0	4	7	7	**11**
WW/EF	3	3	6	6	3	4	7	7	**9**
WW/N	8	7	15	11	4	10	22	22	**32**
WW/P	78	80	88	102	27	89	88	88	**149**
WW/U	4	6	6	6	1	6	9	9	**10**
WX/EF	1	0	1	1	0	0	1	1	**1**
WX/N	0	1	0	1	0	0	0	0	**1**
WX/P	4	4	5	6	1	3	6	6	**12**
Total	**463**	**470**	**613**	**589**	**263**	**445**	**609**	**609**	**991**

Table 3-3: Monitoring sites in REC Climate/Source of Flow/Landcover classes. Values under each variable are the number of sites at which the variable is measured. Values in the "Total" column are the number of sites in each class for which at least one variable is measured. Classes and notation are listed in Table 2-3.

REC Class	Variable								
	CLAR	DRP	ECOLI	NO3N	TNUNF	TPUNF	TAXA	MCI	Total
CD/H/N	0	12	16	12	9	8	13	13	**24**
CD/H/P	12	38	51	45	38	32	24	24	**62**
CD/L/EF	0	0	0	0	0	0	1	1	**1**
CD/L/N	0	1	1	1	0	0	0	0	**1**
CD/L/P	39	90	100	97	56	79	84	84	**145**
CD/L/U	1	7	7	7	6	7	7	7	**13**
CD/L/W	2	1	1	2	0	2	1	1	**2**
CD/Lk/N	0	1	1	1	1	1	0	0	**1**
CD/M/N	0	1	1	0	0	0	2	2	**3**
CD/M/P	0	1	1	1	1	1	1	1	**1**
CW/GM/N	0	0	0	0	0	0	1	1	**1**
CW/H/EF	12	8	11	9	2	6	6	6	**13**
CW/H/N	61	33	58	55	14	26	57	57	**96**
CW/H/P	61	38	53	59	31	45	50	50	**93**
CW/L/EF	6	7	8	7	0	0	9	9	**10**
CW/L/N	16	12	17	17	1	10	23	23	**27**
CW/L/P	54	37	44	41	23	48	68	68	**87**
CW/L/U	6	7	7	7	1	4	7	7	**7**
CW/Lk/EF	2	1	2	2	1	2	0	0	**2**
CW/Lk/N	8	7	7	7	7	7	6	6	**8**
CW/Lk/P	9	9	10	10	5	10	1	1	**10**
CW/M/N	6	8	9	8	6	6	14	14	**15**
CW/M/P	2	1	1	1	2	2	3	3	**2**
CX/GM/N	1	3	7	8	3	2	1	1	**8**
CX/H/N	20	10	19	16	7	10	28	28	**37**
CX/H/P	1	0	0	0	1	1	10	10	**11**
CX/L/N	8	3	8	3	1	2	7	7	**12**
CX/L/P	12	0	13		0	0	18	18	**20**
CX/Lk/N	8	5	8	9	5	5	5	5	**11**
CX/M/N	3	2	3	3	2	2	8	8	**9**
WD/L/P	14	20	23	22	4	11	14	14	**34**
WD/L/U	1	6	5	6	0	4	7	7	**11**
WW/H/N	0	0	1	1	1	0	1	1	**2**
WW/H/P	0	1	1	1	0	1	1	1	**2**
WW/L/EF	2	2	5	5	2	3	6	6	**8**
WW/L/N	8	7	14	10	3	10	21	21	**30**
WW/L/P	74	75	83	96	26	83	87	87	**141**
WW/L/U	4	6	6	6	1	6	9	9	**10**
WW/Lk/EF	1	1	1	1	1	1	1	1	**1**
WW/Lk/P	4	4	4	5	1	5	0	0	**6**
WX/H/N	0	1	0	1	0	0	0	0	**1**
WX/H/P	0	1	0	1	0	0	0	0	**1**
WX/L/EF	1	0	1	1	0	0	1	1	**1**
WX/L/P	4	3	5	5	1	3	6	6	**11**
Total	**463**	**470**	**613**	**589**	**263**	**445**	**609**	**609**	**991**

For the remaining water quality variables, differences among the number of sites at which measurements are made partly reflects differences among councils in the suite of variables used for SOE monitoring. For example, CLAR data were only available for a small proportion of the monitoring sites in Gisbourne and Marlborough Districts, and Northland and Otago Regions, which contributed to a shortage of CLAR data at the national level. The relative scarcity of sites that include TNUNF (27% of sites) and TPUNF (45% of sites) reflect the multiple procedures used to measure TN and TP (see Section 2.2).

3.2 Representativeness

In this section we compare the abundance of rivers (as proportion of total river length) in FENZ and REC classes with the proportion of monitoring sites in each class. The absolute and proportional lengths and numbers of river segments in each class are listed in Appendix B, Tables A3-A5.

When monitoring sites are grouped into FENZ classes, there are 11 gaps corresponding to classes with no monitoring sites (Table 3-4). The 11 FENZ classes lacking monitoring sites are rare classes; collectively, these classes account for 5% of the total length of order 2-8 rivers. The largest shortages in under-represented FENZ classes are in Classes H, J, G, D, and N. These classes are characterised by middle- and high-elevation streams without glacial influence, plus low-elevation streams in dry climate areas. The largest proportional shortage is in Class H (mid-elevation, dry climate streams), which only has 6% of the representative number of monitoring sites. The most common FENZ class (C, small, lowland, hill-country streams) is over-represented by about 30%, with a surplus of 193 sites. The second-most common FENZ class (A, very small, lowland, low-gradient streams) has a near-proportional number of sites, as do several small classes (e.g., Classes K, L, P, S, T)

When monitoring sites are grouped into REC classes at the Climate/Landcover level, there are 14 gaps (Table 3-5). The REC classes lacking monitoring sites are rare classes; collectively they account for less than 1% of the total length of order 2-8 rivers. The largest shortages in under-represented REC classes are in the Cool Extremely Wet/Natural, Cool Wet/Natural, and Cool Dry/Natural classes, with 22 to 119 too few monitoring sites. These classes include prospective reference sites. The largest proportional shortages are in the Warm Extremely Wet/Natural and Cool Dry/Exotic Forest classes, both of which have only one monitoring site. The most over-represented classes are Cool Wet/Pastoral (with a surplus of 80 sites), Cool Dry/Pastoral (with a surplus of 38 sites), and Warm Wet/Pastoral (with a surplus of 30 sites). The general pattern indicated in Table 3-5 is that REC classes with natural landcover tend to be under-represented relative to the abundance of rivers, and classes with pastoral and urban landcover tend to be over-represented. Exotic forest landcover is intermediate; Cool Wet/Exotic Forest is the largest exotic forest class by river length and has a representative number of monitoring sites, but the smaller exotic forest classes are under-represented.

For the assessment at the REC Climate/Source of Flow/Landcover level, minor classes (accounting for less that 0.05% of the total length of order 2-8 rivers) were excluded unless they monitoring sites were presented, as noted in Section 2.5. For the remaining classes, six are gaps with no monitoring sites: CD/H/EF, CX/H/EF, CX/L/EF, WD/L/EF, WD/L/N, WX/L/N (Table 3-6). The two unmonitored classes with natural landcover (WD/L/N and WX/L/N) include prospective reference sites. However, all six unmonitored classes are rare in New Zealand; collectively they account for less than 1% of the total length of order 2-8 rivers

The largest shortages in under-represented REC classes are in the CW/M/N, CX/M/N, and CX/H/N classes, with 41 to 61 too few monitoring sites. Each of these classes include prospective reference sites. The largest proportional shortages are in the CW/M/N and CX/M/N classes, which have less than 20% of the representative number of sites. The most over-represented classes are CW/L/P, CW/H/P, CD/L/P, and WW/L/P, which have 24 to 40 surplus sites. As in the case of Climate/Landcover classes, the general pattern illustrated in Table 3-6 is that classes with natural landcover tend to be under-represented and classes with pastoral and urban landcover tend to be over-represented. As expected, by subdividing groups of sites more finely, the degree of under- and over-representation indicated by the Climate/Source of Flow/Landcover level is greater that for the Climate/Landcover level.

Table 3-4: Distribution of monitoring sites among FENZ classes relative to distribution of river lengths. Blue: class is over-represented by monitoring sites. Yellow: class is under-represented. Representative numbers of sites are rounded to 1. Classes are ordered from the most to least abundant by river length.

FENZ class	% total length	Monitoring sites	% monitoring sites	Representative number of sites	Site difference
C	48.2	670	67.6	477	193
A	19.9	202	20.4	198	4
G	10.7	76	7.7	106	-30
H	6.2	4	0.4	62	-58
D	4.9	22	2.2	49	-27
J	4.0	1	0.1	39	-38
N	2.4	0	0	23	-23
I	0.8	1	0.1	8	-7
O	0.5	0	0	5	-5
P	0.5	0	0	4	-4
S	0.4	0	0	4	-4
B	0.3	3	0.3	3	0
E	0.3	12	1.2	3	9
L	0.3	0	0	3	-3
K	0.2	0	0	2	-2
T	0.2	0	0	2	-2
F	0.1	0	0	1	-1
M	0.1	0	0	1	-1
Q	0.1	0	0	1	-1
R	0	0	0	0	0
Total	**100.0**	**991**	**100.0**	**991**	

Table 3-5: Distribution of monitoring sites among REC Climate/Landcover classes relative to distribution of river lengths. Blue: class is over-represented by monitoring sites. Yellow: class is under-represented. Representative numbers of sites are rounded to 1. Classes are ordered from the most to least abundant by river mouth.

REC class	% total length	Monitoring sites	% monitoring sites	Representative number of sites	Site difference
CX/N	19.8	77	7.8	196	-119
CW/N	19.4	147	14.8	193	-46
CD/P	17.2	208	21	170	38
WW/P	12	149	15	119	30
CW/P	11.3	192	19.4	112	80
CD/N	5.2	29	2.9	51	-22
WD/P	4.6	34	3.4	46	-12
WW/N	2.7	32	3.2	27	5
CW/EF	2.5	25	2.5	24	1
CX/P	1.4	31	3.1	14	17
WW/EF	1.2	9	0.9	12	-3
CD/EF	0.6	1	0.1	6	-5
WX/P	0.6	12	1.2	5	7
WX/N	0.3	1	0.1	3	-2
CD/U	0.2	13	1.3	2	11
CX/EF	0.2	0	0	2	-2
WD/EF	0.2	0	0	2	-2
WD/U	0.2	11	1.1	2	9
WW/U	0.2	10	1	2	8
WD/N	0.08	0	0	1	-1
WX/EF	0.07	1	0.1	1	0
CW/U	0.05	7	0.7	0	7
WD/W	0.03	0	0	0	0
CD/W	0.02	2	0.2	0	2
CX/W	0.02	0	0	0	0
WW/M	0.008	0	0	0	0
CX/U	0.005	0	0	0	0
WD/M	0.004	0	0	0	0
WX/U	0.004	0	0	0	0
CW/W	0.003	0	0	0	0
CD/M	0.002	0	0	0	0
CW/M	0	0	0	0	0
CX/M	0	0	0	0	0
WW/W	0	0	0	0	0
Total	**100.0**	**991**	**100.0**	**991**	

Table 3-6: Distribution of monitoring sites among REC Climate/Source of Flow/Landcover classes relative to distribution of river lengths. Blue: class is over-represented by monitoring sites. Yellow: class is under-represented. Representative numbers of sites are rounded to 1. REC classes that account for less than 100 km in New Zealand (<0.05% of the river length) are excluded, unless there is a monitoring site in the class. In these classes, the representative number of sites is zero. Classes are ordered from the most to least abundant by river length.

REC class	% total length	Monitoring sites	% monitoring sites	Representative number of sites	Site difference
CD/L/P	11.9	145	14.6	118	27
WW/L/P	11.8	141	14.2	117	24
CW/H/N	8.7	96	9.7	86	10
CX/H/N	7.9	37	3.7	78	-41
CW/M/N	7.7	15	1.5	76	-61
CX/M/N	6.4	9	0.9	63	-54
CW/H/P	6.2	93	9.4	61	32
CD/H/P	5.2	62	6.3	52	10
CW/L/P	4.8	87	8.8	47	40
WD/L/P	4.6	34	3.4	45	-11
CD/H/N	3.7	24	2.4	37	-13
CX/GM/N	2.8	8	0.8	28	-20
CW/L/N	2.7	27	2.7	27	0
WW/L/N	2.5	30	3	25	5
CX/L/N	2.1	12	1.2	21	-9
CW/H/EF	1.9	13	1.3	19	-6
CD/M/N	1.1	3	0.3	11	-8
WW/L/EF	1.1	8	0.8	11	-3
CX/L/P	0.7	20	2	7	13
CX/H/P	0.6	11	1.1	6	5
CW/L/EF	0.5	10	1	5	5
CX/Lk/N	0.5	11	1.1	5	6
WX/L/P	0.5	11	1.1	5	6
CD/L/EF	0.4	1	0.1	4	-3
CD/L/N	0.3	1	0.1	3	-2

REC class	% total length	Monitoring sites	% monitoring sites	Representative number of sites	Site difference
CW/Lk/N	0.3	8	0.8	3	5
CD/H/EF	0.2	0	0	2	-2
CD/L/U	0.2	13	1.3	2	11
CW/Lk/P	0.2	10	1	2	8
WD/L/EF	0.2	0	0	2	-2
WD/L/U	0.2	11	1.1	2	9
WW/H/N	0.2	2	0.2	2	0
WW/L/U	0.2	10	1	2	8
WX/L/N	0.2	0	0	2	-2
CD/Lk/N	0.1	1	0.1	1	0
CD/M/P	0.1	1	0.1	1	0
CW/M/P	0.1	2	0.2	1	1
CX/H/EF	0.1	0	0	1	-1
CX/L/EF	0.1	0	0	1	-1
WD/L/N	0.1	0	0	1	-1
WW/Lk/P	0.1	6	0.6	1	5
WX/H/N	0.1	1	0.1	1	0
WX/H/P	0.1	1	0.1	1	0
WX/L/EF	0.1	1	0.1	1	0
CW/GM/N	0.04	1	0.1	0	1
CW/L/U	0.04	7	0.7	0	7
CW/Lk/EF	0.04	2	0.2	0	2
WW/H/P	0.04	2	0.2	0	2
CD/L/W	0.02	2	0.2	0	2
WW/Lk/EF	0.01	1	0.1	0	1
Total	**100.0**	**991**	**100.0**	**991**	

3.3 Site number requirement for precision

The estimated site-number requirements for maintaining specified levels of precision within FENZ and REC classes are shown in Tables 3-7, 3-8 and 3-9. For each variable in each class, site-number requirements were estimated for maintaining precision levels equal to or half of the standard deviation of all classes combined. There are three possible outcomes of this analysis for each class: sufficient site numbers to maintain both levels of precision (i.e., SD for the class ≤ 0.5SD for all classes), sufficient site numbers to maintain only the coarser level of precision (i.e., SD for the class ≤ SD for all classes), and too few classes to maintain the coarser level of precision (i.e., SD for the class > SD for all classes). These outcomes are colour-coded in the tables. In some cases, there were too few sites to calculate site-number requirements. In most of these cases, there were 0, 1 or 2 sites in one or both classes; these cases can be interpreted as site shortages.

For the two FENZ classes with most monitoring sites, A and C, there are currently sufficient site numbers to maintain a precision level of ≤ 0.5SD for each water quality and invertebrate variable (Table 3-7). For FENZ classes B, H, I and J, there were too few sites to calculate site number requirements for any variable. For the remaining FENZ classes, the precision level varied among the different variables.

Table 3-7: Estimated number of sites required to maintain a standard deviation (SD) equal to or half of the SD for FENZ classes. The two values in each cell correspond to the two effect sizes: the first (smaller) number corresponds to an effect size of 1SD, and the second (larger) number to an effect size of 0.5SD. Cell shading indicates the power of the current network. Blue: current site numbers are sufficient to maintain an effect size < 0.5SD. Yellow: current site numbers are sufficient to maintain an effect size < 1SD. Red: current site numbers are insufficient to maintain an effect size <1. Blank cells indicate that the existing network contains too few monitoring sites to calculate statistical power.

FENZ Class	CLAR	DRP	ECOLI	MCI	NO3N	TAXA	TNUNF	TPUNF
A	12 / 44	8 / 29	6 / 20	24 / 92	11 / 39	16 / 59	10 / 35	10 / 36
B			12 / 44		11 / 41			14 / 53
C	11 / 42	10 / 35	11 / 41	33 / 129	12 / 44	20 / 76	11 / 39	13 / 47
D		4 / 13	5 / 15	8 / 26	11 / 41	5 / 17	6 / 21	6 / 21
E	10 / 36	4 / 13	9 / 31	9 / 31	6 / 18	6 / 19	4 / 11	10 / 38
G	6 / 21	4 / 12	11 / 38	22 / 86	11 / 40	14 / 53	7 / 22	7 / 25

For the three REC Climate/Landcover classes with most monitoring sites, CD/P, CW/P, and WW/P, there are currently sufficient site numbers to maintain precision level of ≤ 0.5SD for each variable (Table 3.8). For three classes, CD/W, WW/EF and WX/P, there are insufficient sites to maintain a precision level of ≤ 1SD for any variable. There were too few sites in Classes WX/EF, WX/N, and CD/W to calculate site number requirements for any variable. . In the remaining classes, there are sufficient sites to maintain precision levels of ≤ 1SD or ≤ 0.5SD for some variables, but not others.

Subdividing the monitoring sites into REC Climate/Source of Flow/Landcover classes leads to an increase in the number of class-site combinations with too few sites to maintain the precision levels we tested (Table 3.9). No classes had enough sites to maintain precision level of ≤ 1SD for all variables. For six classes, there are insufficient sites to maintain a precision level of ≤ 1SD for any variable. For an additional nine classes, there were too few sites in the network to run the precision analysis. In the remaining classes, there are sufficient sites to maintain precision levels of ≤ 1SD or ≤ 0.5SD for some variables, but not others.

Table 3-8: Estimated number of sites required to maintain a standard deviation (SD) equal to or half of the SD for REC Climate/Landcover classes. Arrangements of values in cells and colour-coding are the same as Table 3-7. Brackets indicate that the analysis failed because the predicted number of sites was less than two.

	CLAR	DRP	ECOLI	NO3N	TNUNF	TPUNF	TAXA	MCI
CD/N		6 / 20	8 / 30	45 / 178	43 / 168	4 / 13	8 / 30	8 / 27
CD/P	9 / 32	13 / 49	16 / 59	28 / 107	25 / 96	16 / 62	8 / 30	12 / 46
CD/U		15 / 57	10 / 34	13 / 48	5 / 17	28 / 109	7 / 26	6 / 21
CD/W						17 / 62		
CW/EF	8 / 27	19 / 72	12 / 45	10 / 34	6 / 21	11 / 42	16 / 60	16 / 62
CW/N	15 / 54	20 / 75	15 / 55	15 / 58	9 / 30	12 / 46	21 / 78	12 / 46
CW/P	8 / 28	22 / 83	17 / 63	16 / 59	10 / 35	16 / 59	17 / 63	10 / 35
CW/U	5 / 15	7 / 24	7 / 26	6 / 18		3 / 6	28 / 109	12 / 43
CX/N	11 / 42	15 / 58	17 / 65	11 / 41	11 / 39	12 / 46	25 / 98	21 / 80
CX/P	13 / 48		5 / 17				12 / 44	16 / 61
WD/P	16 / 61	24 / 92	17 / 65	19 / 74	5 / 16	16 / 60	13 / 48	13 / 46
WD/U		29 / 112	4 / 11	8 / 28		3 / 7	7 / 23	9 / 31
WW/EF	26 / 102	13 / 50	10 / 36	21 / 80	12 / 45	41 / 159	20 / 75	15 / 55
WW/N	8 / 26	17 / 66	9 / 31	16 / 60	18 / 67	19 / 70	20 / 76	14 / 52
WW/P	14 / 53	16 / 61	11 / 40	14 / 53	6 / 19	12 / 43	21 / 81	16 / 60
WW/U	44 / 172	24 / 91	7 / 26	5 / 15	9 / 32	24 / 93	15 / 55	6 / 20
WX/P	30 / 117	6 / 19	6 / 20	(2) / 4		22 / 83	8 / 28	9 / 32

In general, the analyses of precision using both the REC Climate/Landcover classes and Climate/Source of Flow/Landcover classes indicate that statements about river condition in pastoral landcover classes can be made with greater precision than for natural, urban or exotic forest landcover classes. There are sufficient sites in some of the larger natural landcover classes to report some variable means with moderate precision. The majority of these cases are for the cool climate classes (CD, CW, CX). There are far fewer warm-climate, natural-landcover classes with sufficient site numbers. In several cases, over 100 monitoring sites are needed in some natural landcover classes to achieve the higher levels of precision.

Table 3-9: Estimated number of sites required to maintain a standard deviation (SD) equal to or half of the SD for REC Climate/Source of Flow Landcover classes. The two values in each cell correspond to the two effect sizes: the first (smaller) number corresponds to an effect size of 1SD, and the second (larger) number to an effect size of 0.5SD. Cell colour codes as in Table 3-7. Brackets indicate that the analysis failed because the predicted number of sites was less than two.

	CLAR	DRP	ECOLI	NO3N	TNunf	TPunf	TAXA	MCI
CD/H/N		4 / 11	6 / 20	49 / 192	38 / 148	4 / 11	8 / 29	9 / 33
CD/H/P	14 / 53	6 / 19	9 / 32	24 / 91	9 / 31	13 / 46	14 / 51	10 / 37
CD/L/U		10 / 38	8 / 28	12 / 45	5 / 14	22 / 84	7 / 25	7 / 22
CD/L/W						13 / 48		
CD/M/N							3 / 5	5 / 14
CW/H/EF	8 / 29	15 / 57	11 / 39	7 / 26	4 / 12	8 / 26	18 / 67	10 / 37
CW/H/N	14 / 52	11 / 39	8 / 27	9 / 32	5 / 15	8 / 28	19 / 73	11 / 42
CW/H/P	8 / 29	20 / 76	12 / 45	16 / 59	6 / 20	14 / 53	14 / 51	7 / 23
CW/L/EF	5 / 14	4 / 11	6 / 20	6 / 21			12 / 46	21 / 81
CW/L/N	7 / 26	3 / 7	12 / 43	7 / 25		4 / 11	16 / 60	13 / 48
CW/L/P	8 / 26	10 / 38	11 / 39	12 / 45	4 / 13	12 / 42	17 / 65	12 / 43
CW/L/U	5 / 15	5 / 16	6 / 21	5 / 17		3 / 5	27 / 106	12 / 45
CW/Lk/EF	18 / 67		11 / 39	23 / 90		21 / 81		
CW/Lk/N	14 / 52	6 / 20	12 / 43	28 / 108	3 / 7	8 / 27	12 / 45	4 / 13
CW/Lk/P	5 / 17	13 / 47	12 / 45	10 / 38	6 / 21	8 / 27		
CW/M/N	42 / 163	8 / 30	12 / 43	16 / 60	6 / 21	14 / 53	20 / 78	8 / 29
CW/M/P	8 / 30				4 / 11	3 / 8	12 / 43	5 / 14
CX/GM/N		5 / 14	23 / 90	8 / 28	18 / 69	12 / 45		
CX/H/N	10 / 37	9 / 31	9 / 32	7 / 25	3 / 8	9 / 32	24 / 91	30 / 117
CX/H/P							8 / 26	32 / 124
CX/L/N	11 / 42	12 / 43	15 / 57	12 / 43		6 / 19	16 / 62	5 / 15
CX/L/P	14 / 52		5 / 14				14 / 53	9 / 33
CX/Lk/N	6 / 21	9 / 33	19 / 74	14 / 54	18 / 69	14 / 52	26 / 100	22 / 85
CX/M/N	39 / 151	25 / 98	10 / 38	16 / 61	3 / 7	3 / 8	37 / 146	10 / 37
WD/L/P	16 / 61	16 / 61	14 / 53	18 / 69	4 / 13	12 / 46	13 / 47	13 / 49
WD/L/U		19 / 74	3 / 9	8 / 26		3 / 6	7 / 23	9 / 33
WW/L/EF	22 / 86	14 / 52	10 / 36	18 / 68		28 / 109	19 / 73	15 / 58
WW/L/N	8 / 27	12 / 44	7 / 24	16 / 60	19 / 73	14 / 54	20 / 76	15 / 56
WW/L/P	13 / 48	12 / 43	9 / 32	14 / 51	5 / 16	9 / 33	21 / 80	17 / 64
WW/L/U	44 / 173	16 / 60	6 / 21	5 / 14	7 / 26	19 / 71	14 / 54	6 / 21
WW/Lk/P	24 / 93	3 / 8	(2) / 3	12 / 44		5 / 17		
WX/L/P	30 / 118	5 / 16	5 / 16	(2) / 4		17 / 64	8 / 27	9 / 34

3.4 Site number requirements for comparisons

The estimated site-number requirements for between-class comparisons of FENZ classes, REC Climate/Landcover classes, and REC Climate/Source of Flow/Landcover classes are shown in Tables 3-10, 3-11, and 3-12, respectively. There are three possible outcomes with regard to detecting between-class differences in means: a surplus of sites in both classes, a shortage of sites in both classes, or a surplus in the larger class and a shortage in the smaller class. These outcomes are colour-coded in the tables. In some cases, there were too few sites to calculate site-number requirements. In most of these cases, there were 0, 1 or 2 sites in one or both classes; these cases can be interpreted as site shortages.

Table 3-10: Estimated site numbers required to detect differences in water quality and invertebrate variables between pairs of FENZ classes. Values are site numbers required in each class, assuming equal-sized classes. Blue: both classes have enough sites to detect the between-class difference. Yellow: the larger class has enough sites, but the smaller class does not. Red: both classes have too few sites. Blank cells: the existing network contains too few monitoring sites to calculate the pooled variance.

Class 1	Class 2	CLAR	DRP	ECOLI	NO3N	TNUNF	TPUNF	TAXA	MCI
A	B	21		33	4		548		
	C	16	27	10	15	10	22	41	10
	D		11	4	4	4	8	501243	22
	E	12			4		3	127	26
	G	35	45	69	29	14	36	40	12
B	C	5		30	13		18		
	D			7	71352		5		
	E	4		4	56		3		
	G	4		1195	8		18		
C	D		96	143	12	26	78	43	47
	E	433	4	7	8	5	7	20	39
	G	87	236	23	140	141	205	3507	190
D	E			6	86	6	7	38	843
	G		13	10	8	8	16	34	104
E	G	24		4	5		4	15	75

The table of site numbers required for between-class comparisons of FENZ classes indicates that there are currently enough sites to detect differences in the means of each variable in the two most abundant classes, A and C (Table 3-10). In all other cases, there is a shortage of sites corresponding to two or more variables. Classes H, I, and J are not shown as there were too few sites to site-number requirements for any variable. There are 4 SoE sites in Class H, and a single site in both Class I and J.

Table 3-11: Estimated site numbers required to detect differences in water quality and invertebrate variables between pairs of REC Climate/Landcover classes. . Values and cell shading conventions are the same as for Table 3-10.

Class 1	Class 2	CLAR	DRP	ECOLI	NO3N	TNUNF	TPUNF	TAXA	MCI
CD/N	CD/P		14	8	13	17	11	1764	14
	CD/U		4		14	18	6	7	
	CW/N		1829	198	86	9	234	47	20
	CX/N		34	267	218	12	17	33	13
	WW/N		5	5	50	41	26	5	18
CD/P	CD/U		28	27	3205	31788	78	8	5
	CW/P	317	155	102	141	34	419	16	11
	CX/P	4		94953				8	12
	WD/P	30	7	81	1696	461956	10	35	35
	WW/P	74	31	46	160	47	34	10	83
	WX/P	30	18	154	17		37	4	49
CD/U	CW/U		1142	247	747		4889	29	12
	WD/U		62	115	93		144	4	9
	WW/U		88	14	89		52	6	6
CW/EF	CW/N	4997	17	39	11		5	9303	84
	CW/P	6	231	27	76	631	1367	324	99
	CW/U	21	28	4	9		149	24	5
	WW/EF	48	7	17	31	13	851	6	82
CW/N	CW/P	8	12	9	7	4	7	577	22
	CW/U	38	7	3	4		4	23	3
	CX/N	438	78	20882	465	1200	40	705	77
	WW/N	37	9	6	129	30	17	10	4206
CW/P	CW/U	18	101	10	42		132	14	4
	CX/P	4		104				272	2288
	WD/P	16	14	25	213	16	12	102	5
	WW/P	33	148	17	18254	1077	65	133	31
	WX/P	45	16	33	19		22	26	33
CW/U	WD/U		71	1298	30		16	24	1259
	WW/U	16	110	20	104		39	40	551
CX/N	CX/P	188		6				359	18
	WW/N	18	5	6	46	52	8	13	120
CX/P	WD/P	4		63				31	6
	WW/P	4		32				1021	33
	WX/P	11		42				35	41
WD/P	WD/U		183	48	1282		44	10108	755
	WW/P	365	20	1044	245	11	22	35	16
	WX/P	15	5	943	13		5	9	11
WD/U	WW/U		1886	19	14		141	2410	6901
WW/EF	WW/N	606	10	97	21	26	49	144	19
	WW/P	9	12	14	13	9	49	12	60
	WW/U	22	39	4	5		29	5	5
WW/N	WW/P	10	635	29	5	4	9	21	9
	WW/U	14	53	4			8	7	3
WW/P	WW/U	7344	76	12	23		42	33	14
	WX/P	17	8	187	16		8	113	1745

Table 3-12: Estimated site numbers required to detect differences in water quality and invertebrate variables between pairs of REC Climate/Source of Flow/Landcover classes. Values and cell shading conventions are the same as for Table 3-10.

REC Class	Land-cover1	Land-cover2	CLAR	DRP	ECOLI	NO3N	TNUNF	TPUNF	TAXA	MCI
CD/H	N	P		100	64	726	1556	54	92636	298
CD/L	P	U		160	180	102	62	632	14	12
CW/H	EF	N	364	44	1028	24		20	200	1032
	EF	P	52	1616	92	16344		1632	316	168
	N	P	36	84	52	32	16	36	1672	100
CW/L	EF	N	52	96	640	84			52	84
	EF	P	12	44	72	48			176	484
	EF	U	10	8	8	8			102	12
	N	P	28	80	48	24		32	292	36
	N	U	190	8	10				16	
	P	U	20	260	44	232		322	26	10
CW/Lk	N	P	20		20	16				
CX/H	N	P							2744	148
CX/L	N	P	132		32				1568	28
WD/L	P	U		183	48	1282		44	10108	755
WW/L	EF	N			85	49		1464	137	20
	EF	P			28	18		24	24	122
	EF	U			4	4		12	5	5
	N	P	22	1422	62	10	10	20	44	18
	N	U	14	53	5			8	7	3
	P	U	2302	154	26	50		74	68	28

The table of site number requirements for REC Climate/Landcover classes (Table 3-11). shows two types of between-class comparisons. In the first, both classes being compared are in the same climate category, and the landcover categories are contrasted (e.g., CD/N versus CD/P). In the second, both classes being compared are in the same landcover category, and the climate categories are contrasted (e.g., CD/N versus CW/N). Comparisons that involved one or two pastoral landcover classes had the highest frequency of sufficient site numbers. Comparisons that involved urban and exotic forest landcover classes, for which there are relatively few monitoring sites, had the highest frequencies of site-number shortages. For most of the 45 comparisons in Table 3-11, the current SoE network has too few sites to detect between-class differences in the means of any variable. The deficiencies in sites vary widely by variable; there are sufficient sites for detecting differences in mean MCI scores in 19 out of 45 comparisons (42%), but this number drops to 4 out of 45 comparisons (9%) for detecting differences in mean TNUNF concentration.

Comparisons involving the Natural landcover category are of interest because these include potential comparisons of impacted and reference conditions. Statistical power in the current network is best for comparisons involving the CW/N and CD/N class. There are sufficient sites to detect differences in means for about half of the comparisons involving the CW/N versus CW/P, CW/U and CW/EF, and CD/N versus CD/P and CD/U. There is only one monitoring site in the CD/EF class, so this class is always deficient for making comparisons.

There are hundreds of possible between-class comparisons among the 44 REC Climate/Source of Flow Landcover classes for which there are monitoring sites. For brevity, we focused on comparisons among landcover categories while holding the climate and Source of Flow categories constant (Table 3-12). The 991 monitoring sites in our SoE dataset are subdivided into relatively small groups at the Climate/Source of Flow Landcover; there are an average of 23 sites per class at this level (Table 3-3). Therefore, there are relatively more classes for which there are too few sites to detect between-class differences in mean values, compared with the Climate/Landcover level. Of the 21 paired comparisons of landcover categories shown in Table 3-12, there are too few sites in one or both classes in about 75% of cases. The deficiencies in sites vary widely by variable; there are currently enough sites for detecting differences in TNUNF concentrations in only one case (CW/H/N versus CW/H/P), but there are enough sites for detecting differences in MCI scores in nine cases. For comparisons involving the natural landcover classes, statistical power in the current network is best for comparisons with the classes CW/H/N and CW/L/N, and worst for comparisons with the classes CX/H/N, CD/H/N, and WD/L/N. Note that there are no monitoring sites in the WD/L/N class.

4 Discussion

4.1 Representativeness

Water quality and invertebrates are currently (or have been recently) monitored by unitary authorities and NIWA at approximately 1400 SoE sites on order 2-8 rivers. We used a subset of 991 of these sites for analyses. We can assume that the 991 sites are an unbiased subset in terms of environmental representation, because the rules used for inclusion in analyses were based on sampling duration and frequency, not location or environmental class. Therefore, the distribution of the 991 sites in environmental space should be an accurate representation of the national SoE monitoring network. Some general observations can be then made about the degree to which the distribution of sites in the network matches the distribution of rivers in the New Zealand environment.

In the FENZ framework, there are numerous classes of small to very-small streams that are under-represented by monitoring sites (Classes D, G, H, J and N), and the common class C is over-represented by about 30%. Class A is the only common class with site numbers in proportion to river length. In the REC framework, the most under-represented river classes are in the CD/N, CW/N and CX/N Climate/Landcover classes. The same pattern holds when using the REC Climate/Source of Flow/Landcover classes; most shortages are in environmental classes with natural landcover. Environmental classes with pastoral landcover are over-represented in most climate and source of flow classes. The most over-represented classes are CW/P, CD/P and WW/P. It is important to note that representativeness was assessed relative to catchment conditions that we presume to be current. If this assumption is not valid, then there will be errors in the representativeness analysis. While climatic and topographic (i.e., source of flow) conditions can be considered invariable for the present study, landcover conditions could vary over short time-scales. Monitoring sites were assigned to REC landcover classes on the basis of GIS data from 2000-2001 satellite imagery. Our assessment of sites in operation in 2010 presumes that changes in landcover between 2000-2001 and 2010 have been minor. However, there may have been changes in the landcover associated with some monitoring sites since the photography date (Brockerhoff et al. 2008). More importantly, landcover associated with the monitoring sites may change substantially in the future, and the representation of environmental classes by monitoring sites will change.

As part of the representativeness analysis, we distinguished between gaps (environmental classes for which there are no monitoring sites) and shortages (environmental classes for which the proportion of monitoring sites is less than the proportion of river kilometres). Gaps are a potentially serious problem, as there is no monitoring information being generated about the corresponding environments. However, the analysis indicated that while there are many gaps, they all correspond to small environmental classes. Classes that lack monitoring sites collectively account for between <1 and 5% of the total length of rivers greater than order 1. These results indicate that gaps in environmental space are not currently a serious problem for the network.

In contrast to gaps, there are severe monitoring-site shortages in the network. It is possible that some shortages could be alleviated by shifting monitoring sites from over-represented classes to under-represented classes. However, monitoring sites are used for purposes other than national reporting (e.g., consent monitoring, targeted investigations). In addition, long-term sites have inherent values associated with their length of record. The multiple potential values of sites in over-represented environmental classes suggests that few of these sites will be closed to free up resources for new sites in under-represented classes.

4.2 Precision

The results of the precision analyses indicated that current site numbers in a small number of environmental classes are sufficient to report river conditions with moderately high precision. These environmental classes are common ones in terms of site numbers (e.g., FENZ classes A and C, and REC classes CD/P and CW/P). In contrast, current site numbers for the majority of environmental classes are insufficient to maintain moderately high precision for some or all variables. In general, river conditions in pastoral landcover classes can be reported with greater precision than in natural, urban or exotic forest landcover classes, because there are more monitoring sites in pastoral landcover classes. For some natural landcover classes, > 100 additional sites would be required in order to report mean river conditions with a high level of precision (i.e., half of the standard deviation for all sites). This raises the issue of trade-offs between precision and monitoring network costs.

Precision is an important issue when evaluating and forecasting changes (or lack of changes) in river conditions, and in understanding the effects of environmental heterogeneity. It is also important for selecting appropriate levels of environmental classification for reporting. In multi-scaled classifications such as FENZ and REC, increasing environmental specificity means shifting to lower classification levels, which increases the total number of classes and decreases the number of sites per class. This has two opposing effects on precision; reduced environmental heterogeneity within classes tends to increase precision, and the reduction in site numbers within each class tends to decrease precision. Ideally, there would be sufficient sites in the network to employ a low classification level with many classes, and have many sites per class. This is not the case with the current SoE monitoring network, as indicated by the precision analysis using the REC Climate/Source of Flow/Landcover level (Table 3-9). That classification level gave worse results (i.e., more class-variable combinations with insufficient site numbers) than the coarser FENZ and REC Climate/ Landcover classifications. The negative effect of reducing site numbers per class outweighed the positive effect of increased specificity. However, establishing new sites for the sole purpose of increasing the precision of river condition reports may be difficult to justify. The need to establish new reference sites in unrepresented or under-represented environments may have a higher priority. In this case, we note that moderate increases in precision could be made by ensuring that more variables are measured at existing sites. This point is discussed in more detail in the following section.

4.3 Power to detect between-class differences

Two general observations can be made about the numbers of monitoring sites required for sensitive comparisons between environmental classes. The first is that some estimated site number requirements are extremely high (e.g., over 1000 classes required in both sites to detect some between-site differences). Reasons for these high site number requirements include very small differences in the means being compared, and large pooled variances. For example, the difference in mean CLAR for many between-class comparisons is less than 10%, and for some comparisons, the pooled standard deviation was over 10 times larger than the mean of either class. It is unlikely that very large deficiencies will be addressed by adding many new sites to the SoE network. Small improvements in power can be gained by ensuring that more variables are measured at existing sites. However, in many cases where a very large number of sites is required to detect small differences in mean values, there are enough existing sites to produce reasonably accurate and precise means. In these cases, it is reasonable to conclude that the mean states of the classes being compared do not differ in an ecologically meaningful way.

The second observation is that there is little consistency across variables within pairs of FENZ or REC classes in terms of site-number sufficiency. For example, in comparisons between FENZ classes A and B, we estimated that 4 sites are required in each class to detect differences in mean NO3N, but 548 sites are required to detect differences in mean TPUNF (Table 3-10). This observation suggests that moderate improvements in statistical power could be made by either increasing site numbers to reduce shortages, or by ensuring that most or all variables are measured at most or all existing sites. For example, in comparisons of the WW/N versus WW/P classes, there are sufficient site numbers in both classes for five variables, but there are shortages for CLAR, ECOLI, and DRP (Table 3-12). CLAR and ECOLI are only measured at 25% and 50% of the 32 WW/N sites, respectively. By measuring both variables at all existing sites in the class, the site number requirements would be met, or nearly met. In the case of DRP, there is a very large shortage compared with current site numbers, and the only way to substantially improve statistical power is through a large increase in site numbers.

Examination of the tables of site number requirements for between-class comparisons (Tables 3-10, 3-11, and 3-12.) revealed few patterns in surpluses or shortages of sites required, either by variable (columns) or by REC classes (rows). The large differences in site number requirements among variables has three possible causes: 1) differences in the number of sites where individual variables are measured (see Section 3.1 above); 2) inherent differences in variation among the water quality and ecological variables; and 3) differences among variables in the magnitude of between-class differences. As an example of inherent variability, the coefficient of variation (CV, the ratio of standard deviation to mean) for CLAR across all Climate/Landcover classes is 2.2, which is 3-20 times higher than the CVs for the other variables. The variable TAXA provides an example of the effect of small between-class differences: the mean value of TAXA ranges from 10 and 20 in most classes, so there is a high probability that two classes will have similar values. More sites are required to detect small differences in means than large differences. This is reflected in the numerous between-class comparisons where more than 1000 sites may be required to detect differences in TAXA.

The effect-size used in a power analysis affects the estimated site-number requirements. In our analyses, we used the absolute difference in means for each variable as the effect-size to detect. If we had used larger effect sizes (e.g., two-times the absolute difference in means), the estimated site-numbers required would have been smaller. In a previous analysis of New Zealand river monitoring sites, Snelder et al. (2006) used one pooled standard deviation as the effect size to detect. In general, the site-number requirements estimated in the 2006 report were smaller than in the present study, because one pooled standard deviation is larger than the difference in means for most comparisons. Our approach was more conservative because, if there are enough sites in two classes to detect differences between mean values, there will likely be more than enough sites to detect differences of one standard deviation. We used the conservative approach in recognition of the fact that we rarely know the sizes of “ecologically meaningful” differences. For example, a small difference between reference sites and impacted sites in NO3N or DRP concentrations might be associated with excessive periphyton growth in the latter. In future analyses, a range of effect sizes can be tested, and different effect sizes can be used for each variable.

The power analyses presented in this report represent only one of many possible outputs from the combined water quality and invertebrate data set. For example, our results for the FENZ classes focus solely on the 20-class level, but could easily be extended to the 100-, 200-, or 300-class levels. At the 100 class level, for example, 25 FENZ classes are represented by at least 5000 NZReaches, and account for 81.8% of New Zealand waterways. These figures are very similar to those for REC climate / landcover classes with at least 5000 NZReaches (24 classes, 83.4% of waterways), suggesting that the two classification methods may provide comparable levels of resolution. Alternatively, the data set could be used to run complementary sets of power analyses where the effect size is specified in absolute terms, yielding estimates of site numbers needed within each class to estimate a specified variable (e.g., MCI) with a precision of ±5.

4.4 Reference site requirements

One of the primary purposes of a national network identified by the expert panel is to monitor effects of landuse and human activities on river conditions. These effects are assessed relative to the condition of rivers in the absence of human impacts, i.e., reference conditions, as measured at reference sites. There are two important observations to be made about reference sites, based on the representativeness assessment. First, environmental classes where reference conditions are likely to prevail are poorly represented in the current SoE network. Most of the shortages identified at the REC Climate and the Source of Flow levels corresponded to classes with natural landcover, in which reference sites could be located (Tables 3-5 and 3-6). The under-representation of natural landcover classes in the network partly reflects the varied origin of monitoring sites; many sites originated as consent monitoring sites or sites used for investigations of human impacts, and few sites have been deliberately established as reference sites (but see Collier et al. 2007). Second, the primary value of reference sites and reference classes is not their contribution to representative networks, but their comparability to impacted sites and impacted classes. In other words, a reference environmental class may be scarce in terms of river length or numbers of reaches, but monitoring sites within that rare class may be very valuable for comparison with corresponding impacted classes. The following examples, based on the REC Cool Dry/Lowland and Warm Dry/Lowland classes, illustrate this point.

The CD/L/P class is one of the most abundant in New Zealand in terms of river length (about 12% of total river length), and it contains the largest number of monitoring sites in our dataset (145 sites). This class accounts for a substantial amount of the agricultural land in southeast New Zealand. The appropriate reference class for use in evaluating water quality and ecological conditions in the CD/L/P class is the CD/L/N class. The CD/L/N class is currently rare (0.3% of total river length), and contains a single monitoring site. Similarly, the WD/L/P class accounts for a substantial amount of agricultural land in the Manawatu, Hawke’s Bay, Waikato and Northland regions, and is monitored at 34 sites. The appropriate reference class for WD/L/P class is the WD/L/N class, which is also rare (0.1% of total river length), and for which there are currently no monitoring sites. The primary reason that rivers in the CD/L/N and WD/L/N classes are scarce is historical conversion of natural, low-elevation landcover from forest, woodland, and wetland to agricultural land. Despite their current rarity, the remaining river reaches classed as CD/L/N and WD/L/N should have high priority for establishing reference sites (Larned et al. 2004).

The REC has been used in this and previous studies as a tool for classifying monitoring sites into reference classes (e.g., the “natural” landcover category), and for estimating reference conditions (Larned et al. 2004, Snelder and Lessard 2009). These classification and estimation steps give coarse-resolution results because the natural landcover category encompasses a wide range of catchment conditions, including the proportion of catchment composed of native vegetation, and the specific types of vegetation present. Despite the low resolution, a classification that distinguishes landcover categories associated with the presence and absence of human activity is an efficient screening tool. In contrast to the REC, the FENZ framework does not distinguish reaches using attributes that reflect human activities (Leathwick et al. 2010). Human impact variables were removed from the FENZ classification during its development, and the current FENZ classes are based on natural environmental factors that influence biological variation in rivers.

As noted above, the natural landcover category used in this study (or the original REC native forest, scrub and tussock categories) provides only a coarse screen for identifying reference site needs, and locating potential new reference sites. When selecting candidate reference sites, multiple criteria should be used, in addition to the broad environmental classifications. Two general classes of criteria should be considered, environmental and logistical criteria. The environmental criteria that we refer to include environment attributes that are related to “pristineness” (e.g., absence of mines, time since a catchment was logged), and attributes that help match reference sites to corresponding impacted sites (discussed below). Logistical criteria include accessibility and site operating costs.

We recommend identifying potential reference sites in approximately four stages: First, use the REC to define the general classes of reference sites needed and to map river reaches corresponding to those classes. For example, if reference conditions are needed for comparison with conditions in lowland pastoral and urban streams in cool-dry and warm-dry climates (CD/L/P, CD/L/U, WD/L/P, WD/L/U), the REC should be used to map river reaches in CD/L/N and WD/L/N classes. These reaches are potential reference sites, at a coarse screening level.

Second, increase the specificity of REC classification to improve matching between the impacted and reference classes. Matching impacted and reference classes in terms of natural environmental factors (e.g., climate, geology, stream order) reduces the between-class differences in water quality and biology that are independent of landcover effects. Specificity can be increased in this step by including the REC geology and network position levels when classifying impacted and reference classes. For example, a suitable reference class for comparisons with the Cool Dry/Lowland/Alluvial/Pastoral/Low Order class is the Cool Dry/Lowland/Alluvial/Natural/Low Order class. Matching can also incorporate environmental factors that are not part of the REC (e.g., flow regime).

Third, increase the specificity of landcover for impacted and reference classes. The Land Cover Database version 3 can be used for this step. For example, the pastoral landcover category for the impacted CD/L/AL/P class can be subdivided into high-intensity and low-intensity agriculture classes The identification of landcover categories for reference sites that correspond with each agricultural class is complicated by the fact that the indigenous landcover (prior to conversion by Polynesian or European populations) is unknown for many locations (Leathwick 2001). There is also a lack of information about differences in water quality and biology in rivers from different native vegetation classes (e.g., between catchments dominated by podocarp and beech forests). It is possible that these differences are negligible in comparison to differences between impacted and natural landcover.

Fourth, use logistical criteria to reduce the number of potential reference sites to a short-list of high-priority sites. GIS road layers can be used to apply road-accessibility criteria and to calculate driving times between sites. Establishment and operating costs can be calculated from driving times, sampling time and expenses, and installation and maintenance of flow recorders.

The steps listed above should be viewed as options for progressively screening potential reference sites until a list of high priority sites remains. There is no obligatory order of steps, and they should be used in an iterative way, by mapping the river reaches that meet the criteria after each step is applied. It is clear from the preceding discussion that the benefits of matching impacted and reference classes need to be balanced against the number of potential reference sites; the number of potential sites will decrease as the degree of matching and environmental specificity increases. If there are too few potential reference sites, unforeseen limitations (e.g., land-owner permission) may reduce the number of potential sites to zero.

5 Acknowledgements

We thank Lucy Baker and Lian Potter (Ministry for the Environment) for support of the NEMaR project. We thank Ton Snelder and Jochen Schmidt (NIWA) for report reviews. We thank regional, district, and Nelson City council staff and Graham Breyer (NIWA) for providing data for council SoE and NRWQN sites. And we thank the members of the expert panel for reviews, recommendations and participation in workshops. The expert panel members were:

Kevin Collier	Waikato Regional Council
David Kelly	Environment Canterbury
Summer Greenfield	Greater Wellington Regional Council
Adam Uytendaal	Hawke’s Bay Regional Council
Chris Daughney	Geological and Nuclear Science
Dave Kelly	Department of Conservation
Chris Arbuckle	Ministry of Agriculture and Forestry
Joanne Clapcott	Cawthron Institute
Roger Young	Cawthron Institute
Brenda Baillie	Scion
Dave West	Department of Conservation
Jon Roygard	Horizons Regional Council

6 References

Ballantine, D.; Booker, D.; Unwin, M.; Snelder, T. (2010). Analysis of national river water quality data for the period 1998–2007. NIWA Client Report CHC2010-038. Prepared for the Ministry for the Environment.

Boothroyd, I.; Harding, J.; Death, R. (2002). Guide to the selection of reference sites for environmental monitoring of New Zealand rivers and streams. New Zealand Macroinvertebrate Working Group Report No. 2.

Brockerhoff, E.G.; Shaw, WB.; Hock, B.; Kimberley, M.; Paul, T.; Quinn, J.; Pawson, S. (2008). Re-examination of recent loss of indigenous cover in New Zealand and the relative contributions of different land uses. *New Zealand Journal of Ecology 32*: 115–126.

Clapcott, J.; Young, R.; Goodwin, E.; Leathwick, J.; Kelly, D. (2011). Relationships between multiple land-use pressures and individual and combined indicators of stream ecological integrity. Department of Conservation Research and Development Series 326. Department of Conservation, Wellington. 57 p.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences, second edition. Lawrence Erlbaum Associates, New Jersey, USA.

Collier, K.J. (2008). Temporal patterns in the stability, persistence and condition of stream macroinvertebrate communities: relationships with catchment land-use and regional climate. *Freshwater Biology 53*: 603-616.

Collier, K.J.; Haigh, A.; Kelly, J. (2007). Coupling GIS and multivariate approaches to reference site selection for wadeable stream monitoring. *Environmental Monitoring and Assessment 127*: 29-45.

Collier, K.J.; Clapcott, J.E.; David, B.O.; Death, R.G.; Kelly, D.; Leathwick, J.R.; Young, R.G. (2012). Macroinvertebrate-pressure relationships in boatable New Zealand rivers: influences of underlying environment and sampling substrate. *River Research and Applications* doi:10.1002/rra.2564.

Davies-Colley, R.; Larned, S.; Unwin, M.; Verburg, P.; Hughes, A.; McBride, G.; Ballantine, D.; Hudson, N.; Daughney, C.; Hamill, K. (2011). Dependable monitoring of freshwaters for national-scale environmental reporting. NIWA Client Report HAM2011-055. Prepared for the Ministry for the Environment.

Dixon, W.; Chiswell, B. (1996). Review of aquatic monitoring program design. *Water Research 30*: 1935-1948.

Joy, M.K.; Death, R.G. (2003). Biological assessment of rivers in the Manawatu-Wanganui region of New Zealand using a predictive macroinvertebrate model. *New Zealand Journal of Marine and Freshwater Research 37*: 367-379.

Larned, S.T.; Scarsbrook, M.R.; Snelder, T., Norton, H.; Biggs, B.J.F. (2004). Water quality in low-elevation streams and rivers of New Zealand: recent state and trends in contrasting landcover classes. *New Zealand Journal of Marine and Freshwater Research 38*: 347-366.

Leathwick, J.R. (2001). New Zealand's potential forest pattern as predicted from current species-environment relationships. *New Zealand Journal of Botany 39*: 447-464.

Leathwick, J.R.; Julian, K.; Elith, J.; Chadderton, L.; Ferrier, S.; Snelder, T.H. (2008). A biologically-optimised environmental classification of New Zealand rivers and streams: reanalysis excluding human impact variables. NIWA Client Report HAM2008-027. Prepared for the Department of Conservation.

Leathwick J.R.; West, D.; Chadderton, L.; Gerbeaux, P.; Kelly, D.; Robertson, H.; Brown, D.; Ausseil, A.-G. (2010). Freshwater Ecosystems of New Zealand (FENZ) Geodatabase. User guide. Version one.

Leathwick, J.; Snelder, T.; Chadderton, W.; Elith, J.; Julian, K.; Ferrier, S. (2011). Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications. *Freshwater Biology 56*: 21-38.

Parkyn, S.; Collier, K.; Clapcott, J.; David, B.; Davies-Colley, R.J.; Matheson, F.; Quinn, J.; Shaw, W.; Storey, R. (2010) The restoration indicator toolkit. Envirolink Tools Programme, Foundation for Research, Science and Technology.

Snelder, T.H.; Biggs, B.J.F. (2002). Multi-Scale River Environment Classification for Water Resources Management. *Journal of the American Water Resources Association 38:* 1225–1240.

Snelder, T.; Dey, K. (2006). Test of the performance of two alternative river environmental classification systems, River Environment Classification (REC) and Freshwater Environments of New Zealand (FWENZ). NIWA Client Report CHC2006-083. Prepared for Department of Conservation.

Snelder, T.; Lessard, J. (2009). Assessment of the suitability of river water quality monitoring sites for development of water quality guidelines. NIWA Client Report CHC2009-190. Prepared for Ministry for the Environment.

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 1. Prepared for the Ministry for the Environment.

Stark, J.D.; Maxted, J.R. (2007). A user guide for the Macroinvertebrate Community Index. Prepared for the Ministry for the Environment. Cawthron Report No.1166. 58 p.

Stoddard, J.; Larsen, D.; Hawkins, C.; Johnson, R.; Norris, R. (2008). Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications 16*: 1267–1276.

Ward, R.C.; Loftis, J.C.; McBride, G.B. (1990). Design of water quality monitoring systems. Van Nostrand Reinhold, New York, USA..

Zar, J.H. (1999). Biostatistical analysis, Fourth edition. Prentice Hall, New Jersey, USA.

Appendix A Analyses of precision and statistical power

Power analysis is concerned with the relations between five parameters: sample size (N), standard deviation (SD), effect size (ES), significance level (α), and power (1–β). Specifying any four of these parameters uniquely determines the fifth. Given α, 1–β, and SD, we can therefore choose to specify N and ask what effect size we can be confident of detecting, or specify an effect size and ask what N is needed to detect this with our specified level of confidence (Zar 1999). This study is concerned with the second of these options. Note that the term effect size is used in this study to refer a minimum precision level, and to a minimum difference between pairs of environmental classes.

Site number estimations for precision. As an illustration, noting that we have already specified our desired significance level ($\alpha = 0.1$) and power ($1-\beta = 0.8$), we give an explicit calculation of an analysis of precision for the variable MCI at the REC Climate/Landcover level. The pooled council and NRWQN datasets include 21 classes which are represented by at least one site, with the number of sites per class ranging from 1 to 119 (Table A1). We average over all classes to estimate a mean of 101.5, with a standard deviation (SD) of 13.9.

Table A1: Analysis of precision worksheet for MCI score at REC climate/landcover class level

REC class	Number of sites	MCI		Required no. of sites	
		Mean	SD (= ES)	ES = 1	ES = 0.5
CD/EF	1	112.5			
CD/M	1	111.8			
CD/N	15	106.3	10.1	8	27
CD/P	118	93.3	13.2	12	46
CD/U	7	70.7	8.8	6	21
CD/W	1	94.9			
CW/EF	17	111.6	15.3	16	62
CW/N	103	116.8	13.2	12	46
CW/P	119	107.3	11.5	10	35
CW/U	7	82.6	12.7	12	43
CX/N	50	122.7	17.5	21	80
CX/P	29	108.2	15.2	16	61
WD/N	1	112.9			
WD/P	20	85.3	13.2	13	46
WD/U	12	83.7	10.8	9	31
WW/EF	8	105.7	14.4	15	55
WW/N	35	117.5	14.0	14	52
WW/P	99	98.8	15.0	16	60
WW/U	10	84.1	8.5	6	20
WX/EF	1	104.5			
WX/P	6	100.0	10.9	9	32
All classes		**101.5**	**13.9**		

We then estimate the site requirements for each class based on the SD for all classes (which provides a consistent measure of variability among classes), and the SD within each class (which is the ES we wish to detect). Using MCI in the REC CD/N class as an example, the third row in Table A1 indicates that a minimum of 8 sites are required to detect an ES of 1SD, or ±10.1, and 27 sites are required to detect an ES of 0.5SD or ±5.1. The existing number of sites (15 in this case) is therefore sufficient to detect an ES of 10.1, but insufficient to detect an ES of 5.1.

Similar calculations apply to all other class x variable combinations, after back transforming any log-transformed variables. For a given variable, the required number of sites for a given ES depends solely on the class SD, with higher SDs leading to higher site numbers. This accords with common-sense: we would expect to require more sites within a class where site to site variation is high (e.g., CX/N) than within a class where there is less variability (e.g., WW/U).

Site number estimations for comparisons. Our second set of analyses was designed to gauge the power of the current network to detect differences in variable means between pairs of classes. We illustrate these calculations by considering variation in MCI scores among FENZ classes, for all possible class pairs with sufficient data (Table A2). The difference between the two class means for each pair provides a natural measure of effect size. Thus, for the first row in the table (FENZ class A vs. FENZ class C), the effect size is 19.4 (= 109.2 – 89.8) and the pooled standard deviation 16.2.

Assuming normality and applying the standard method for a two sample t-test indicates that at least 10 sites are needed in each class to achieve the desired power (Zar 1999).. This calculation assumes equal site numbers in both classes, consistent with use of the pooled SD for the underlying test, but is not necessarily optimal if within-class SD varies between the two classes. A more efficient allocation can be made by apportioning the total number of sites (20 in this case) among the two classes in proportion to their variance, as in the final two columns of Table A2. Note that all site number estimates are rounded up to the nearest integer.

Minimum site numbers for each pair are strongly related to the difference between the two class means. For example, mean MCI for classes A and C differs by a large amount, so that relatively few sites are required to achieve our desired significance level and power . By contrast, mean MCI for classes D and E differs by very little, so that at least 843 sites would be required in each class.

Table A2: Worksheet for estimating site numbers for comparing MCI scores among pairs of FENZ classes.

										Power Analysis		
Class 1				Class 2				Effect size		Equal variance	Unequal variance	
Class	**N**	**Mean**	**SD**	**Class**	**N**	**Mean**	**SD**	**Δ means**	**Pooled SD**	**N**	**N1**	**N2**
A	135	89.8	14.2	C	454	109.2	16.8	19.4	16.2	10	9	12
A	135	89.8	14.2	D	11	100.5	7.4	10.7	13.8	22	34	10
A	135	89.8	14.2	E	7	99.6	8.1	9.8	13.9	26	40	13
A	135	89.8	14.2	G	49	104.9	13.6	15.1	14.0	12	12	12
C	454	109.2	16.8	D	11	100.5	7.4	8.7	16.6	47	78	16
C	454	109.2	16.8	E	7	99.6	8.1	9.6	16.7	39	63	15
C	454	109.2	16.8	G	49	104.9	13.6	4.2	16.5	190	229	151
D	11	100.5	7.4	E	7	99.6	8.1	0.9	7.7	843	771	916
D	11	100.5	7.4	G	49	104.9	13.6	4.4	12.8	104	48	160
E	7	99.6	8.1	G	49	104.9	13.6	5.4	13.1	75	39	112

Appendix B Absolute and proportional lengths and abundance of river reaches

Table A3: Absolute and proportional lengths and abundances of river reaches, by FENZ class. First-order streams are excluded.

FENZ class	Length (km)	% total length	No. reaches	% total reaches
A	39190.7	19.9	47580	17.3
B	672.2	0.3	711	0.3
C	94712.9	48.2	131106	47.7
D	9651.4	4.9	11296	4.1
E	641.5	0.3	939	0.3
F	136.7	0.1	85	0.0
G	21090.2	10.7	29769	10.8
H	12275.9	6.2	20702	7.5
I	1562.3	0.8	3026	1.1
J	7797.3	4.0	13724	5.0
K	304.4	0.2	567	0.2
L	545.3	0.3	1124	0.4
M	125.7	0.1	140	0.1
N	4661.7	2.4	8070	2.9
O	1006.4	0.5	1861	0.7
P	885.3	0.5	1572	0.6
Q	282.9	0.1	486	0.2
R	8.6	0.0	15	0.0
S	759.5	0.4	1537	0.6
T	315.5	0.2	613	0.2
Totals	**196626.4**	**100.0**	**274923**	**100.0**

Table A4. Absolute and proportional lengths and abundances of river reaches, by REC Climate/Landcover class. First-order streams are excluded.

REC class	Length (km)	% total length	No. reaches	% total reaches
CD/EF	1181.8	0.6	1607	0.6
CD/M	4.1	0.0	5	0.0
CD/N	10134.9	5.2	13313	4.8
CD/P	33809.2	17.2	39053	14.2
CD/U	382.1	0.2	418	0.2
CD/W	46.8	0.0	51	0.0
CW/EF	4830.7	2.5	5980	2.2
CW/M	0.0	0.0	1	0.0
CW/N	38212.0	19.4	58198	21.2
CW/P	22156.8	11.3	29743	10.8
CW/U	98.7	0.1	137	0.0
CW/W	5.6	0.0	9	0.0
CX/EF	463.8	0.2	616	0.2
CX/M	0.9	0.0	2	0.0
CX/N	38949.0	19.8	63471	23.1
CX/P	2708.7	1.4	2887	1.1
CX/U	9.7	0.0	12	0.0
CX/W	40.7	0.0	53	0.0
WD/EF	403.2	0.2	577	0.2
WD/M	6.9	0.0	9	0.0
WD/N	151.2	0.1	242	0.1
WD/P	9035.0	4.6	11059	4.0
WD/U	407.9	0.2	505	0.2
WD/W	61.6	0.0	58	0.0
WW/EF	2320.0	1.2	3313	1.2
WW/M	15.2	0.0	24	0.0
WW/N	5372.8	2.7	8109	2.9
WW/P	23518.5	12.0	32386	11.8
WW/U	384.2	0.2	509	0.2
WW/W	0.3	0.0	1	0.0
WX/EF	136.2	0.1	217	0.1
WX/N	681.9	0.3	986	0.4
WX/P	1088.2	0.6	1358	0.5
WX/U	7.9	0.0	14	0.0
Total	**196626.4**	**100.0**	**274923**	**100.0**

Table A5. Absolute and proportional lengths and abundances of river reaches, by REC Climate/Source of Flow/Landcover class. First-order streams are excluded.

REC class	Length (km)	% total length	No. reaches	% reaches	REC class	Length (km)	% total length	No. reaches	% reaches
CD/H/EF	426.6	0.2	584	0.2	CX/L/U	7.8	0.0	10	0.0
CD/H/N	7281.7	3.7	9478	3.4	CX/L/W	38.9	0.0	47	0.0
CD/H/P	10293.5	5.2	13444	4.9	CX/Lk/EF	1.6	0.0	1	0.0
CD/H/U	16.6	0.0	15	0.0	CX/Lk/M	0.9	0.0	2	0.0
CD/H/W	0.9	0.0	2	0.0	CX/Lk/N	1060.2	0.5	1723	0.6
CD/L/EF	755.1	0.4	1022	0.4	CX/Lk/P	4.9	0.0	7	0.0
CD/L/M	4.1	0.0	5	0.0	CX/Lk/W	1.7	0.0	5	0.0
CD/L/N	538.5	0.3	715	0.3	CX/M/EF	1.8	0.0	4	0.0
CD/L/P	23317.6	11.9	25384	9.2	CX/M/N	12546.8	6.4	22137	8.1
CD/L/U	365.1	0.2	402	0.1	CX/M/P	2.4	0.0	3	0.0
CD/L/W	45.9	0.0	49	0.0	WD/L/EF	386.9	0.2	546	0.2
CD/Lk/N	151.5	0.1	176	0.1	WD/L/M	6.9	0.0	9	0.0
CD/Lk/P	49.5	0.0	55	0.0	WD/L/N	148.6	0.1	237	0.1
CD/Lk/U	0.4	0.0	1	0.0	WD/L/P	8967.5	4.6	10969	4.0
CD/M/EF	0.0	0.0	1	0.0	WD/L/U	405.8	0.2	502	0.2
CD/M/N	2163.3	1.1	2944	1.1	WD/L/W	61.6	0.0	58	0.0
CD/M/P	148.6	0.1	170	0.1	WD/Lk/EF	16.3	0.0	31	0.0
CW/GM/N	86.6	0.0	143	0.1	WD/Lk/N	2.6	0.0	5	0.0
CW/H/EF	3671.5	1.9	4356	1.6	WD/Lk/P	67.5	0.0	90	0.0
CW/H/N	17148.4	8.7	25232	9.2	WD/Lk/U	2.1	0.0	3	0.0
CW/H/P	12183.3	6.2	16124	5.9	WW/H/EF	92.3	0.0	140	0.1
CW/H/U	14.7	0.0	22	0.0	WW/H/M	1.6	0.0	2	0.0
CW/H/W	0.2	0.0	1	0.0	WW/H/N	439.4	0.2	699	0.3
CW/L/EF	1063.5	0.5	1514	0.6	WW/H/P	86.2	0.0	120	0.0
CW/L/N	5346.8	2.7	8023	2.9	WW/L/EF	2199.8	1.1	3147	1.1
CW/L/P	9405.1	4.8	12848	4.7	WW/L/M	13.6	0.0	22	0.0
CW/L/U	84.0	0.0	115	0.0	WW/L/N	4919.3	2.5	7380	2.7
CW/L/W	5.4	0.0	8	0.0	WW/L/P	23217.5	11.8	31971	11.6
CW/Lk/EF	77.3	0.0	94	0.0	WW/L/U	383.8	0.2	508	0.2
CW/Lk/M	0.0	0.0	1	0.0	WW/L/W	0.3	0.0	1	0.0
CW/Lk/N	525.5	0.3	902	0.3	WW/Lk/EF	27.9	0.0	26	0.0
CW/Lk/P	325.0	0.2	420	0.2	WW/Lk/N	14.1	0.0	30	0.0
CW/M/EF	18.4	0.0	16	0.0	WW/Lk/P	214.8	0.1	295	0.1
CW/M/N	15104.7	7.7	23898	8.7	WW/Lk/U	0.4	0.0	1	0.0
CW/M/P	243.3	0.1	351	0.1	WX/H/EF	25.1	0.0	44	0.0
CX/GM/N	5593.6	2.8	10770	3.9	WX/H/N	262.4	0.1	388	0.1
CX/H/EF	258.8	0.1	375	0.1	WX/H/P	129.6	0.1	216	0.1
CX/H/N	15540.1	7.9	23730	8.6	WX/H/U	1.1	0.0	2	0.0
CX/H/P	1270.3	0.6	1419	0.5	WX/L/EF	111.1	0.1	173	0.1
CX/H/U	1.9	0.0	2	0.0	WX/L/N	418.3	0.2	596	0.2
CX/H/W	0.1	0.0	1	0.0	WX/L/P	957.0	0.5	1141	0.4
CX/L/EF	201.6	0.1	236	0.1	WX/L/U	6.8	0.0	12	0.0
CX/L/N	4208.2	2.1	5111	1.9	WX/Lk/N	1.1	0.0	2	0.0
CX/L/P	1431.0	0.7	1458	0.5	WX/Lk/P	1.6	0.0	1	0.0
Total						**196626.4**	**100.0**	**274923**	**100.0**