Freshwater Monitoring Protocols and Quality Assurance (QA)

National Environmental Monitoring and Reporting (NEMaR) Variables Step 2

Prepared for the Ministry for the Environment

August 2012

Authors/Contributors:
R. J. Davies-Colley
A. O. Hughes
P. Verburg
R. Storey

For any information regarding this report please contact:

Rob Davies-Colley
Principal Scientist
Aquatic Pollution
+64-7-856 1725
r.davies-colley@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
Gate 10, Silverdale Road
Hillcrest, Hamilton 3216
PO Box 11115, Hillcrest
Hamilton 3251
New Zealand

Phone +64-7-856 7026
Fax +64-7-856 0151

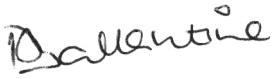
NIWA Client Report No:	HAM2012-092
Report date:	August 2012
NIWA Project:	MFE12201

Contents

Tables

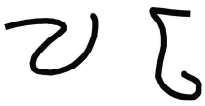
Figures

Reviewed by

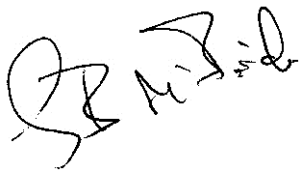


Dr Deborah Ballantine

Approved for release by



Dr Jochen Schmidt



Graham McBride

Formatting checked by

Executive summary

Background

The National Environmental Monitoring and Reporting (NEMaR) project has the overall goal of achieving 'dependable' regional monitoring of New Zealand's inland surface waters as a basis for national state-of-environment (SoE) reporting. In the first phase of NEMaR, NIWA staff prepared two reports for the Ministry for the Environment (MfE) regarding the assessment and reporting of the 'condition' of New Zealand's freshwaters. NIWA was subsequently engaged to lead a second phase of work in NEMaR with the objective of designing and implementing a National Surface Water Monitoring Programme (NSWMP) to support water SoE reporting in New Zealand. The (Phase 2) project was divided into three 'workstreams': Indicators (for national reporting), Variables (on measurements and timing plus protocols and quality assurance) and Networks (design of monitoring locations). Each workstream was based on an expert panel comprising regional council scientists and water resource officers supported by scientists from research institutes and universities.

The expert panels met in two series of workshops in late 2011. During these workshops, variables (and indicators) considered likely to fulfil national water quality assessment and reporting purposes were identified and assessed. The first Variables workshop, held on 26th October 2011, discussed variables identified in the Indicators workshop held the previous day, considered timing and frequency of sampling, and rated the variables for importance, cost and feasibility. The second Variables workshop, held on 30th November 2011, was more concerned with protocols and quality assurance (QA) – which, furthermore, are the subject of this report. The expert panel decided that a first step in defining 'national' protocols and QA measures to support national state-of-environment (SoE) reporting was to review current regional protocols.

Approach

Surface Water Information Management (SWIM) contacts within the sixteen regional councils and TLAs were asked to forward copies of their regional monitoring protocol documents and field sheets. It quickly became apparent that protocols were comprehensively documented by most councils, but with considerable diversity. Summarizing these sometimes very large and differently-structured documents would have been a massive effort. So the project team decided to use a questionnaire approach to obtain basic 'metadata' about different regional water monitoring protocols and QA measures. Simple questionnaires were designed to capture basic features of monitoring protocols and QA for each of river water quality, river bio-monitoring and lake monitoring. The regional council contacts were again approached and we achieved a 100% return.

Protocols

This report summarises the important characteristics ('metadata') of current regional water monitoring (with emphasis on variables selected as 'core' variables in previous NEMaR reports) as regards protocols and such QA as there is. An overview of current protocols is given and then protocols considered suitable for each of the three main surface freshwater 'sub-domains' of river water quality, river bio-monitoring and lake monitoring are recommended. The report then goes on to recommend 'national' protocols and QA. In general, the protocols we recommend for adoption nationally are those that are used by a

majority (or substantial minority) of the regional councils – *so long as these were judged 'scientifically sound' and of sufficient precision for national reporting purposes*. In some cases, refinements on common protocols are suggested however.

Note that it is imperative that protocols be *consistent over time* to be suitable for robust time-trend analysis – which is very demanding in terms of data quality. This principle has major implications for the *change* in protocols implied for certain variables and certain councils if all councils are to have consistent protocols. The transition to new protocols, implying a period of overlap of, say, 12 month, will need to be funded to achieve this where valuable (for time-trend analysis) historical datasets exist measured with previous protocols.

Water quality protocols are significantly diverse between councils for both lakes and rivers, i.e., there is a general lack of consistency across councils, at least in details. If greater consistency is to be achieved this implies a *change* in protocols for some regions and some water bodies. Such changes will need to be made very cautiously to avoid inducing a discontinuity in data that will confound time-trend analysis. A workshopping approach is recommended (separately for lakes and river water quality) to consider the recommended protocols and their ramifications for individual councils.

LakeSPI (Lake Submerged Plant Index) is recommended as a holistic index of lake condition to complement the TLI and other indicators based on lake water quality. However, LakeSPI protocols are not discussed because this index requires specialist expertise and resources and is probably not suitable for monitoring at the regional level (not least because of QA considerations).

Protocols are lacking or insufficiently developed for some bio-monitoring variables, notably for (invertebrates, fish and periphyton in) non-wadeable streams and for habitat. We recommend that a working group under the auspices of SWIM be conducted with priority (say early next financial year) to develop protocols suitable for national river bio-monitoring. QA in river biomonitoring could be usefully considered at the same time.

Quality Assurance (QA)

There is comparatively little current activity by councils on quality assurance (QA). Regional council scientists are aware of the need for QA, but the pressure of other work and on their monitoring budgets, seems to result in QA being side-lined. Several questionnaire returns mentioned international standards or training of field staff by way of response on QA. However, although such activities *contribute* to QA they do not *assure* data quality. Only replication of measurements can do that.

We propose that, as a guideline, about 10% of (regional) monitoring budgets should be devoted to QA activities with perhaps 5% of actual measurements being duplicated. QA activities could include: forwarding 'unknowns' (e.g., duplicates of 'known' samples) to laboratories and batch checking of data for internal consistency. Monitoring staff are encouraged to undertake **joint field exercises** with staff from neighbouring regions and NIWA field staff to promote good practice in water SoE monitoring.

More formally, we propose that an **audit of water monitoring** should be conducted with all councils, about every two years. The auditor or auditing team would accompany regional

field staff for one day on part of a routine monitoring 'run', and would (1) note field protocols and (2) take duplicate measurements and samples. The intention of such auditing would not be to 'police' monitoring activity, but to encourage good practice. A crucial output of these audits would be *provision of independent, duplicate data with which to assess monitoring accuracy*.

A valuable QA check (for rivers) would be **deliberate duplication of NRWQN sampling** (macro-invertebrates, and periphyton as well as water quality) by councils at one or more NRWQN sites in each region. (This is already done routinely in one region.) Good agreement would confer confidence in both datasets – and confidence in regional data from *other* SoE sites. In the absence of any national network of lake monitoring like the NRWQN for rivers, the field auditing of lakes is regarded as of particularly high priority.

Improved and consistent national protocols

Remaining issues with regards to nationally consistent protocols and QA need to be addressed, ideally by a workshop approach involving regional council staff and science advisors. In particular, protocols for some aspects of river biomonitoring need to be developed, and approaches to QA need to be discussed with a field auditing system instituted using 'national' field forms to promote consistency. We recommend that working groups are convened on sub-domain lines – separately for lakes, river water quality and various specialist aspects of river biomonitoring.

1 Introduction

1.1 Background

During 2011, NIWA investigated national monitoring and reporting on water 'condition' on behalf of the Ministry for the Environment (MfE). State-of-Environment (SoE) reporting would be built mainly on consistent and dependable water monitoring undertaken by regional councils. The outcomes of these investigations were summarised in two complementary reports "Investigation of single indicators for national water quality assessment and reporting" (Hudson et al. 2011), and "Dependable monitoring of freshwaters for national-scale environmental reporting" (Davies-Colley et al. 2011a).

To progress the development and implementation of monitoring for water quality assessment and reporting, MfE convened workshops of experts (comprising regional council scientists and water resources officers supported by research institute and university researchers) to further consider indicators, variables, and network design. These expert panel assessments became the basis for Phase 2 work, now referred to as the "National Environmental Monitoring and Reporting" project (NEMaR). Two separate workshop sessions were held for each of the "Indicators", "Variables" and "Networks" Workstreams of NEMaR during October and November 2011. Appendix A outlines the process towards implementing a national surface water monitoring programme, with the Variables workstream highlighted.

For the Variables workstream, the key objectives to be achieved in Phase 2 of the NEMaR project were defined as follows:

Step 1: Define and obtain agreement on a set of variables for national freshwater, and fresh and marine recreational water quality monitoring based on [the Phase I report of Davies-Colley et al. 2011a].

Step 2: Develop and document agreement on national protocols and quality control of these variables.

The current report deals with Step 2 of the Variables workstream, i.e., recommending national protocols and quality assurance in regional water monitoring.

1.2 The Brief

The agreed brief for this Step 2 Variables workstream (letter of April 2, 2012 from Tom Bowen, MfE to Jochen Schmidt, NIWA) was as follows:

1. Compile statements of protocols and quality assurance (QA) currently used by councils for water monitoring in NZ.
2. Review these protocols and recommend common ('national') protocols for core variables as defined in step 1. This is based on a clear framework, e.g., national protocols exist and are used; national protocols exist and are not well implemented; protocols are diverse and/or not up-to-date; no protocols existing.

3. Recommend QA measures to promote 'dependability' of regional water monitoring.

4. Summarize findings in a report by end June 2012.

1.3 The approach

Workshops of the NEMaR Variables expert panel were held on Wednesday 26 October 2011 and Wednesday, 30 November 2011. These workshops culminated in our recommendation in the NEMaR Variables Step 1 report (Davies-Colley et al. 2012) of a minimal set of core variables for water monitoring (Table 1-1; Table 1-2; Table 1-3). The Variables expert panel also agreed to standardisation on monthly monitoring of most variables for river water quality, irrespective of river state-of-flow. This has implications for protocols for river sampling during storm flows, as is considered below. Likewise, sampling of lakes should be monthly irrespective of lake level; although it was recognised that safety considerations during rough weather may call for a stand-down of lake field work to the first following safe day. (Protocols for stand-down in lake sampling are considered below).

At the second workshop, the philosophy of approach to recommending 'national' protocols and quality assurance was considered. The Variables Expert panel recognised that protocols must be strictly consistent over time for data to be suitable for long term time-trend analysis – to answer the question: "is the water body getting better or worse?" The expert panel recommended that a survey of protocols currently used by councils be undertaken as a basis for establishing national protocols. Accordingly, we requested water resources scientists in the 16 councils (see the Acknowledgements) to provide protocol documents for river and lake water quality and for river bio-monitoring, together with example field forms. It soon became apparent that this approach was unwieldy because of the length and diversity of the current council protocol documents. Instead of attempting to summarize these sometimes large documents we decided on a different approach. We composed questionnaires to elicit basic characteristics of ('metadata' on) existing monitoring programmes and sent these to our regional council contacts. This report is very largely built on the returns to these questionnaires (Appendix B-D).

1.4 Structure of this Report

The emphasis of this report, like its predecessor (Davies-Colley et al. 2012) **is on the variables that need to be measured *routinely and indefinitely* in the 'national' network** – either because they are needed for national reporting ('core' variables, for example total nitrogen, macro-invertebrates; Hudson et al. 2012) or because they are valuable 'supporting' variables that assist with interpretation of the 'core' reporting variables (notably turbidity, also pH and electrical conductivity). (Refer Table 1-1; Table 1-2; Table 1-3 for the 'core' variables).

In this report we consider first, certain general aspects of protocols, including the need for stability of protocols (as well as monitoring sites and variables) over time for long-term trend analysis. Laboratory protocols should be specified in contracts for analysis, including, notably, that *un*censored data should be provided by laboratories to maximize information yield.

We then consider, in turn, protocols for the three main sub-domains of freshwater: river water quality, river bio-monitoring and lake monitoring. Care was taken to define protocols that are suitable for *indefinite*, *routine* monitoring that will be stable over time – so that resulting data are suitable for time-trend analysis.

This is followed by a discussion of quality assurance (QA; designed to ensure data are unbiased and sufficiently precise), starting with the philosophy of QA, which, ultimately, requires replication of some measurements. A field auditing system incorporating duplication of council measurements is a key recommendation.

Finally recommendations are collected in a final chapter. This includes recommendations for a 'workshopping' approach to development of protocols for some bio-monitoring components, and for implementing change in protocols generally (without incurring data discontinuities with valuable long-term data records), together with consideration of QA measures notably a national field audit system.

Table 1-1: Minimal set of water quality variables ('Core' variables) for rivers.

All listed variables should be measured *monthly*, although flow, temperature and DO are ideally measured continuously. Several other 'supporting' variables would normally be measured too, such as conductivity as an indicator of total ionic content. (After Davies-Colley et al. 2012.)

River WQ Variable	Needed for?
Discharge (flow 'stamping')	State-of-flow for WQ interpretation and load calculation
Temperature	Thermal conditions (climate change)
Dissolved oxygen (DO)	Oxygen conditions for aquatic life
Black disc visibility	Measure of visual clarity (aquatic life and human use)
* Total suspended solids (TSS)	Relevant to sedimentation effects and sediment loads
Ammoniacal-nitrogen	Nutrient (immediately bio-available), toxic (as NH_3)
Oxidised-nitrogen (NO_x)	Nutrient (immediately bio-available), toxic
Total nitrogen (TN)	Nutrient
DRP	Nutrient (immediately bio-available)
Total phosphorus (TP)	Nutrient
E. coli	Indicator of faecal microbial pollution

*TSS and visibility are inversely correlated so there is a degree of redundancy in monitoring both (Davies-Colley and Smith 2001).

Table 1-2: Minimal set of variables ('Core' variables) for lakes.

All listed variables should be measured *monthly* except LakeSPI. Several other 'supporting' variables would normally be measured too, such as conductivity as an indicator of total ionic content. (After Davies-Colley et al. 2012).

Lake WQ Variable/index	Needed for?
Total nitrogen (TN)	Trophic level index (TLI), Nutrient
Total phosphorus (TP)	Trophic level index (TLI), Nutrient,
Chlorophyll *a*	Trophic level index (TLI), Index of phytoplankton biomass
*Secchi depth	TLI4*, Optical characterisation
Temperature profile	Thermal stratification (climate change)
DO profile	Hypolimnetic de-oxygenation
LakeSPI index (based on submerged aquatic plant survey)	'Integrating' index of lake condition

*Secchi depth should still be measured for general optical characterisation even if *not* used in TLI4 calculation (i.e., if TLI3 is calculated.

Table 1-3: Biological components ('Core' variables) for bio-monitoring of rivers.

Periphyton should be visually assessed monthly at water quality sites. Other biotic variables would normally be measured annually (probably in late summer) or less often – depending on resources. Macrophytes would be done (if at all) as part of physical habitat. Hydrological indices need to be defined to support bio-monitoring. (Ecosystem processes are not recommended for now because of staff time and instrumental limitations in most regions.) (After Davies-Colley et al. 2012.)

Biotic component	Needed for?
Macro-invertebrates (annually)	Well-established 'integrating' indicator of stream 'health'
Periphyton (monthly, warm months)	Basis of stream primary production, nuisance growths
Fish (2-5 yearly rotation?)	Valued component of biota, biodiversity issues
Macrophytes (annually?)	Flow and water quality effects, habitat
Physical habitat (annually?)	Important influence on stream ecology

2 General principles

2.1 Difficulties around monitoring "redesign"

2.1.1 Time-trend analysis and the need for stability in protocols

A key objective of national monitoring and reporting (as defined in the NEMAR project) is to document trends in freshwater condition (Davies-Colley et al. 2011; Table 3-3). To enable trend analysis, consistency in protocols is crucial to answer the question: "is the water 'condition' changing?" (e.g., Smith et al. 1996). Trend analysis can only be reliably conducted on environmental data that is collected at the *same sites* for the *same variables* (usually at the same time of day) by the *same protocols*. It is probably not possible to do meaningful trend analysis if sites or variables change, and even a subtle protocol change can cause a data 'discontinuity' (step change) that will confound trend analysis. Time-trend analysis is fundamental to choice of protocols and QA. Indeed, time-trend analysis is probably *the* most 'demanding' application of water monitoring data as regards protocols and data quality.

Consistency of monitoring variables seems 'absolute'. That is, if variables are changed there is no way to use historical data with the 'old' variables in trend analysis up to the present. Even change in time of day may have unfortunate ramifications for diurnally variable measures, such as DO, pH or temperature. Changes in monitoring site, however well-considered, are also likely to be very difficult to accommodate in trend analysis, although a closely-related site (say, fairly close by on the same main-stem river or in the same lake basin) might not preclude time-trend analysis (but the possibility of a discontinuity should be checked).

Stability in water monitoring protocols *between* councils is important to permit comparisons between regions and to facilitate national summaries. As reported below (Chapters 3-5) there is currently appreciable diversity between councils which makes comparisons between regions and national summaries very difficult to prepare (Ballantine et al. 2010; Verburg et al. 2010). Achieving common ('national') protocols is a major aim of this report. However we recognise this is problematic where valuable long-term datasets exist. If such historical data were obtained with 'old' protocols that are now being recommended for change there is the strong possibility of a step change or discontinuity that will confound time-trend analysis.

The need for stability over time in sites, variables, and protocols amounts to a major disincentive to "redesign" of monitoring programmes, however worthy and scientifically well-intended. Where valuable long term data exists then the default should always be to continue the programme that generated that data. Changes to improve the design may well risk destroying the continuity of time series needed for time-trend analysis, and must be avoided if at all possible.

2.1.2 Innovation in monitoring

Water science will continue to advance, and some advances will spin off new technologies and new understandings leading to new methods for water monitoring. However, innovation in monitoring is inevitably in tension with the need for stability of protocols over time (Davies-Colley et al. 2011b). Great caution is needed in implementing new monitoring methods and techniques that may not be entirely comparable with previous techniques so leading to a step

change in the data record and the potential for spurious trends over time. In fact, it would be fair to say there is considerable *inertia* in environmental monitoring for the very good reason that stability over time is essential to provide data suitable for detecting subtle changes in environmental condition. The longer a stable historical record, the greater the 'inertia' to change in monitoring protocols even if a new method arises that is demonstrably better.

Eventually a new method may become compellingly better than the original (e.g., much cheaper or more precise). (We are assuming that good practice including frequent calibration is followed such that bias is minimal.) If and when such a new method is to be introduced an overlap period with the original method is needed, say of 12 months so as to enable seasonal patterns to be considered and perhaps longer if high flows (with water quality contrasting with that at baseflow) need to be intercepted. Ideally a functional relationship relating the new to the old method can be developed from paired data during the overlap so that any step change can be quantified and data suitably adjusted for trend analysis. This approach is routinely applied in climate monitoring where a change in site or protocol is applied. Clearly, measuring a variable twice by two different methods during the period of overlap, is costly – which is itself a disincentive to change.

As an example of a protocol that has persisted for very good reasons despite innovation, consider visual clarity measurement in lakes. Arguably black disc visibility is a preferred method for optical characterisation of waters compared to the traditional Secchi disc depth (Davies-Colley and Smith 2001; Zanevald and Pegau 2003), and is routinely used in New Zealand rivers. However, many valuable long-term Secchi records exist on certain iconic lakes in NZ (permitting trend analysis) (also, the Trophic level Index (TLI) uses Secchi depth) so it is generally preferable to continue with Secchi measurements in lakes rather than changing to black disc. (Note that although Secchi depth is well correlated with black disc visibility, the former measure is around 25-30% greater than the latter (Davies-Colley et al. 2003)).

A change in laboratory bench turbidimeter brings with it a more subtle (and related) change in method because different bench turbidimeters can and do give different numerical response in NTU even though identically calibrated to formazin (e.g., Davies-Colley & Smith 2001). Again an overlap period of measurement with *both* turbidimeters would ideally be undertaken before the old turbidimeter is retired. Of course if the old turbidimeter simply fails without warning and is unrepairable there is no opportunity for an overlap with the replacement instrument. In that case the 'new' turbidity values cannot be directly compared with the 'old' and a step change in the data record cannot be adjusted for. Fortunately, this problem is not insurmountable with turbidity relegated to a 'supporting' variable (compared to visual clarity – which is a core variable, e.g., Table 1-1). Figure 2-1 shows the typical fairly close inverse relationship between turbidity (measured by a particular instrument) and visual clarity at a particular monitoring site. This type of relationship, which varies somewhat between different waters (i.e., there is no 'universal' relationship) can be used to 'anchor' turbidity if there is a change in instrumentation. Although time trend analysis on visibility data can be very useful, we recommend great caution with such analysis applied to turbidity data because of the possibility of (unrecorded) change in nephelometers. It follows that any and all reporting of turbidity data should be accompanied with the make and model of nephelometer.

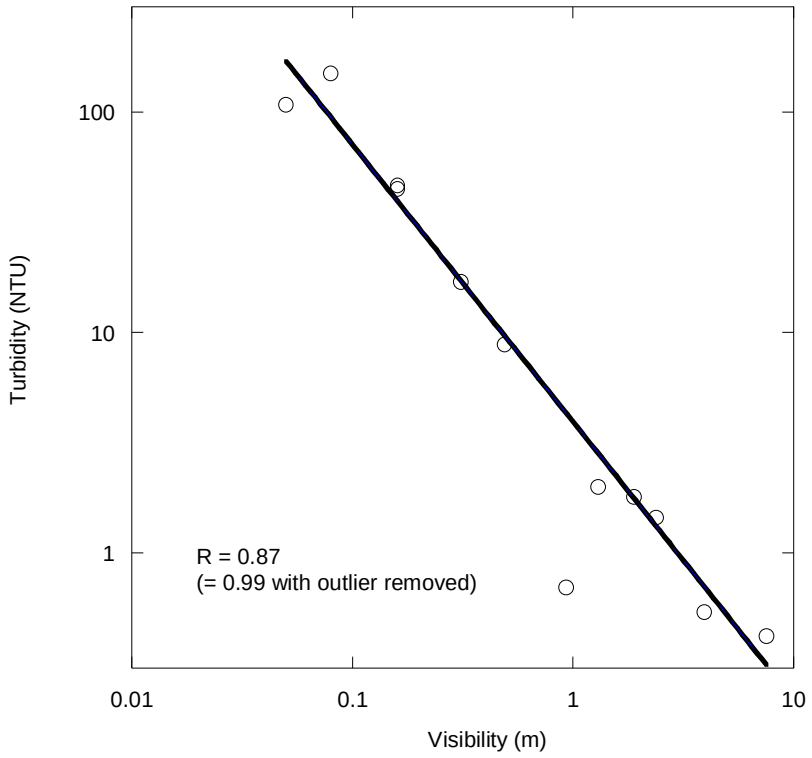


Figure 2-1: Laboratory turbidity versus visual clarity data for 12 monthly measurements in a river. There is good overall correlation, but one outlier point where the visibility and turbidity are not consistent. On this occasion the turbidity was consistent with the flow, but not the visibility, so the latter was judged to be erroneous.

2.1.3 Implementing change in monitoring protocols

As we have seen, the need for stability in monitoring programmes is in major tension with scientific innovation and the emergence of improved methods. The same considerations arise when recommending changed protocols to achieve national consistency. In this report we recommend protocols that are common to most councils, so long as the methods are judged scientifically sound. This should permit the maximum use of historical data without discontinuities. However, where a particular council needs to change its protocols to be nationally consistent with other councils, a conundrum arises – namely that historical data collected with the 'old' protocol may not be immediately comparable with new data. There seems no alternative with implementing protocol changes to running a period of overlap, say of 12 months with *both* protocols to provide cross-calibration data. The overlap paired data can be used to develop an algorithm allowing inter-conversion of 'old' and 'new' data so as to permit the full record to be analysed for time-trends without confounding step-changes.

Again we seem to be pushing against the 'inertia' of valuable historical records. If councils are going to implement national protocols, that in some cases will be new for them, a step-change should be expected and must be corrected to permit time-trend analysis. An overlap period, say of 12 monthly samplings, seems unavoidable. This transitional overlap implies duplicated measurements and analyses.

2.2 Accuracy, precision and bias

Accuracy has two main aspects: precision, referring to random error, and bias referring to systematic error. These two aspects are often confused, but are easily understood with reference to the schematic diagram in Figure 2-2.

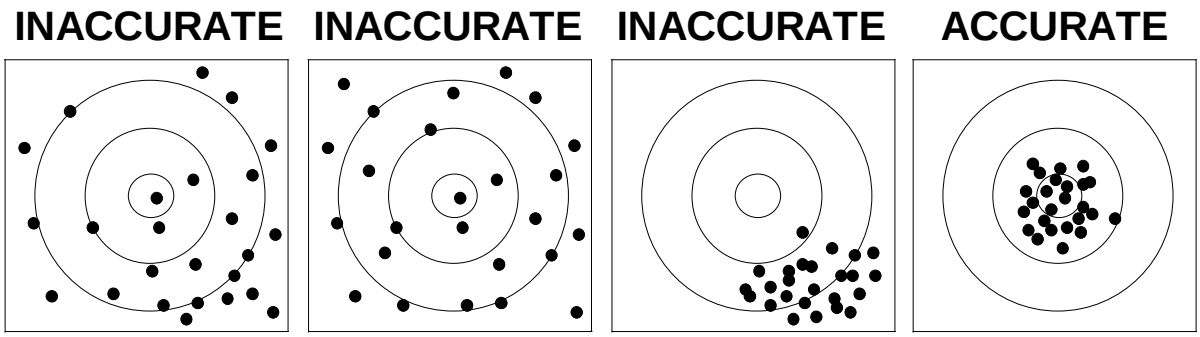


Figure 2-2: Accuracy, precision and bias in environmental measurements. These concepts are illustrated by the analogy of a dart-board in which the true environmental value of some variable is represented by the bulls eye (after McBride 2005). "Accurate" implies unbiased *and* precise.

Clearly water monitoring data needs to be sufficiently accurate for the purpose of properly characterising the water mass or its biota and for tracking water 'condition' over time (McBride 2005). Our aim in water monitoring must be to minimise bias – by frequent calibration of instruments and checking of standards.

Elimination of random error, in contrast to systematic error, is not possible. However random error must be small enough for the purposes of defining water 'condition' and tracking that condition over time. As a useful rule of thumb, 10% precision (i.e., a coefficient of variation of replicate measurements of 10%) should be sufficient for most water monitoring variables (Davies-Colley et al. 2011a). A lesser precision may be suitable for some variables that are intrinsically very variable. For example, 15-30% precision is typical, depending on method, for bacterial indicators such as *E. coli*, but that precision is usually sufficient because concentrations typically vary by four orders of magnitude in the one river with a strong contrast between baseflow and stormflow (e.g., McKergow and Davies-Colley 2010).

Precision of water quality variables depends on concentration. At relatively high concentration, 10% precision might be routinely attained. However at low concentrations the random error, expressed as a percentage, starts to rise as concentration continues to fall (Figure 2-3) and rises rapidly as concentration gets very low.

An alternative term to precision is often used in the laboratory context – "uncertainty of measurement" (UoM) (Standard ISO 17025), and usually defined in terms of 95% confidence interval (i.e., about ± 2 standard deviations).

A difficulty arises with maintaining precision with low level data. For example DRP in a phosphorus-limited lake may be easily measured to 10% precision through much of the year, but DRP concentrations may be driven very low during algal blooms – possibly to a level below the 'detection limit' into a range where the CoV of replicates is appreciably higher than 10%. Precision needs to be good enough for purpose over the concentration range of interest, but it may not always be possible to maintain better than 10% precision over the full range ever encountered.

Precision may be expected to change somewhat between batches of samples analysed in laboratories, and should normally be reported along with batch results. However if a particular batch is appreciably worse that the 'typical' precision it seems reasonable that analyses should be repeated (on frozen back-up subsamples, where this is feasible as for nutrients) with the cost borne by the laboratory. Regional councils, as laboratory clients, need to be alert to such issues of laboratory data quality.

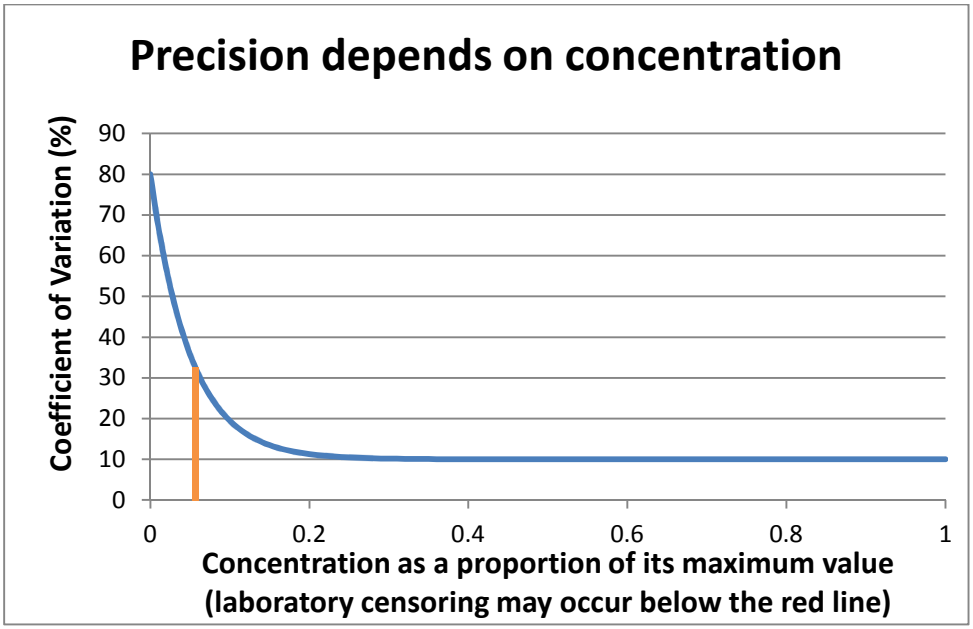


Figure 2-3: Precision of water quality data depends on concentration. Precision, expressed as coefficient of variation (CoV) of replicates, is typically fairly similar over much of the concentration range, but the CoV rises at low concentrations and gets very large (i.e., relative precision becomes increasingly poor) when concentration approaches zero.

2.3 Laboratory protocols

The emphasis of this report is on field protocols because laboratory protocols are (probably) a smaller source of imprecision and bias for most core variables than field protocols (Robinson 2010). Chapters 3-5 consider field protocols with different sub-domains. However some general issues are covered here regarding laboratory protocols relevant to both lakes and rivers.

2.3.1 Detection limits and data censoring

Common practice with laboratory data is to define a 'detection limit' in terms of the laboratory precision and 'censor' data below that level (which are reported merely as "<DL"). APHA (2005) defines the detection limit as 3 X the standard deviation of replicates at very low level (so that the Coefficient of Variation, by definition, is about 33% at the detection limit - Figure 2-1) and this definition is followed by Hill Laboratories among others in New Zealand. Although the practice of 'censoring' may have some rationale for materials that are not usually present and are being 'detected' in water the practice and the term itself seem to have little relevance to analytes in water such as phosphorus which we *know* are present (because its concentration was reported for all samples greater than DL) and for which we want to estimate concentration even when that may be low. Censoring hides information (Porter et al. 1988) and numerical fixes to replace censored data usually introduce bias. Furthermore, although one isolated analytical result should be treated with caution when the relative precision is low (say below the nominal 'detection limit' when CoV of replicates is high), SoE reporting, including time-trend analysis, uses *multiple* analyses of the same variable so that the low precision of any one measurement is not so critical.

The different perspectives of laboratory staff and their clients (environmental managers) is probably the source of much of the difficulty here (Porter et al. 1988): laboratory staff are concerned with the integrity of individual results taken in isolation, environmental managers are concerned with detecting change in environmental condition – for which the maximum information must be extracted from datasets taken as a whole.

The censoring of data may well make the difference between detecting incipient eutrophication of an oligotrophic lake and not detecting this trend. If the 'detection limit' changes over time, spurious trends may result from censoring. Censoring of data only seems to occur when laboratories do analyses, not in any other areas of science and technology. In particular, censoring is obviously not applied when field instruments are used to measure various water quality variables, so censoring of associated laboratory SoE data seems doubly inappropriate.

In a previous NEMaR report (Davies-Colley et al. 2011a) we recommended that data censoring be discontinued in laboratory provision of SoE data. Here we strengthen the recommendation: **SoE data provided by laboratories should not be censored.** Instead laboratories should be specifically contracted to provide their best estimate of concentrations together with a precision estimate. The precision estimate is nearly as important as the concentration estimate itself. The detection limit is sometimes used as a precision index, although intrinsically less fundamental than the standard deviation of replicates against which it is defined.

Although the above recommendation to avoid censoring will involve a change in practice it is not difficult in principle, and laboratories can cover themselves for perceptions that data may be misused by suitable caveats in laboratory reports of results to clients.

Precision may be expected to change somewhat between batches of samples analysed in laboratories, and should normally be reported along with batch results. However if a particular batch is appreciably worse that 'typical' (e.g., as indicated by detection limit) analyses should be repeated (on frozen back-up subsamples, where this is feasible as for

nutrients). Regional councils, as laboratory clients, need to be alert to such issues of data quality.

2.3.2 Nutrient analyses

We are aware that nutrient (N and P) analyses (and nomenclature) currently vary appreciably between councils. There is a need to standardize protocols nationally for N and P analysis. Some councils are currently estimating TN as TKN + NNN, where TKN = total Kjeldahl nitrogen (which includes ammoniacal-N), and NNN = nitrate plus nitrite-N. However, the TKN method is more suited to high level (e.g., vegetation or wastewater) than environmental sample analysis. Furthermore, adding components combines the errors (variances are additive) contributed by both measurements. Consequently TN calculated by adding components including TKN is noisy (low precision).

Total N is best measured *directly* (on unfiltered, digested samples) rather than indirectly by addition of components. Furthermore, unfiltered samples should be used for both TN and TP, not filtered samples (on which total *dissolved* N and total *dissolved* P would be measured, usually referred to as TDN and TDP).

Field filtering is sometimes suggested as a means for arresting change in water samples for nutrient analysis (Robinson 2010). We are not sure if field filtering is done routinely in any region, but would caution against this because of the high potential for contamination under field conditions – e.g., by lint from clothing or dust. Instead filtration is best left to the clean conditions of the analytical laboratory. But proper sample handling (containment in ice chests with ice chips) and prompt freight (usually by overnight courier) is important for stability of nutrient and some other analyses, as is considered in Chapters 3 and 5 below.

3 River Water Quality Protocols

3.1 Regional Council SOE water quality sampling protocols and quality assurance

3.1.1 Summary of current protocols and quality assurance procedures

All 16 councils completed a 20 question questionnaire (Appendix B) on water quality sampling protocols and QA procedures carried out during their SOE monitoring. Table 3-1 summarises the responses. The questionnaire responses indicate that there is considerable diversity in both the SOE river water quality sampling protocols and QA approaches taken by the councils. The diversity in responses is probably to be expected given that councils have developed their procedures and protocols autonomously, and mostly independently, over a number of years. More detail on the current SOE water quality monitoring and QA carried out the by 16 councils is outlined below.

3.1.2 SOE water quality sites

All councils visit the same core sites on every visit, however four councils have a combination of core and rotational sites. The majority (11) councils are currently (or will be by July 2012) carrying out monthly sampling at all their SOE sites. One council carries out a mix of monthly and quarterly sampling depending on the site. The remaining four councils sample less frequently, either bi-monthly or quarterly.

Ten councils visit the sampling sites at the same time (±1 hr) on every visit. Five other councils follow the same routes on sampling days but considered it unlikely that the sites were visited within ±1 hr on every visit. A further council also does not visit sampling sites at the same time but no information was provided as to whether the same routes were followed on sampling days – which would imply reasonable consistency of sampling time.

Thirteen councils have formal metadata on SoE sites which may include maps, GPS coordinates, photographs, and notes on access and health and safety. Only three councils do not have any, or have only very limited, site information available to guide field staff.

Only six councils are able to determine flow (at the time of sampling) for all of their SOE sites. For these councils, flow is determined by either using information from an associated hydrometric site (where applicable), readings from staff gauges or by gauging at the time or sampling.

3.1.3 Water quality sampling procedures

Twelve councils carry out sampling if the target date and time happens to intercept high flows/floods, although one of these does so for its monthly sampling but not its quarterly sampling. Of those councils that do not sample during high flows/floods, two specifically stated that high flow sampling was avoided so as to avoid unrepresentative data. Interestingly, although twelve councils stated they carry out sampling during high flows/floods, eight stated they would reschedule if there was bad weather on the sampling day.

Table 3-1: Summary of Regional Council responses to the freshwater quality field sampling questionnaire.

Questionnaire question	Summary of responses
1) What is the sampling frequency	11/16 carry out monthly sampling with remainder carrying out bi-monthly or quarterly sampling
2) Are sites visited at the same time (± 1hr) on every visit?	10/16 stated (in an unqualified way) that they visit the sampling sites at the same time (±1hr) on every visit
3) Are the same sites visited on every cycle?	16/16 visit the same core sampling sites on every visit
4) Do you have metadata on each site available to guide field staff?	13/16 have metadata
5) Do you estimate streamflow on each sampling visit?	6/16 are able to determine flow at all their sampling sites
6) Does sampling occur during high flows/floods	12/16 carry out sampling during high flows/floods
7) How do you cope with bad weather on the target day/time?	8/16 carry out sampling regardless of weather
8) Are samples collected in bottles that have been acid-washed?	8/16 use acid-washed bottles
9) Where in streams are samples collected?	All councils sample from channel centre or from the main flow by telescopic pole
10) Are bridges or other artificial structures used for access?	6/12 use bridges for access
11) Describe water sampling procedure	14/16 carry out sampling in a manner comparable to the NRWQN procedures
12) Are samples collected from pools or riffles?	13/16 carry out sampling in riffles or runs
13) Are samples placed into a chilled/light sealed storage bin?	All councils store samples in cooled storage bin
14) Are samples delivered to lab within 24 hours of collection?	All councils aim to deliver samples to lab within 24 hours
15) Is a field sheet completed for each site visited/sample taken?	All councils complete field sheets
16) How often are field instruments calibrated?	7/16 carry out calibrations on day of use; highly variable otherwise
17) Is visual clarity (e.g., black disc) measured?	11/16 measure black disc (BD) clarity; others do not measure
18) If yes, describe visual clarity protocol	All those councils that measure BD follow MfE (1994) BD protocols
19) Do you QA your river water quality field procedures?	10/16 carry out (some) QA
20) Describe QA approach	Highly diverse between the 10 councils that carry out QA

Only six councils stated that they used bridges or artificial structures for river access. One Council specifically stated they do not sample from bridges because of H&S issues (traffic). Based on the responses to the questionnaire question, there may have been a wording problem with the questionnaire in which we were trying to determine whether bridges or artificial structures are used for sampling in some situations (notably, high flow events). Some respondents might have thought we were referring to the general use of bridges to access sampling sites.

With regards to the actual task of collecting water samples from rivers there is a high degree of consistency between councils with consistent sampling from the middle of the channel or by telescopic bottle holder where wading is not possible (e.g., H&S issues, large rivers). Thirteen councils obtain their samples from riffles or runs or areas of 'flowing water' with three councils sampling from pools. Fourteen use a bottle filling technique which is similar to that described in the NRWQN field protocols (as described broadly in Davies-Colley et al. 2011b). One council stated that water is first collected in a bucket and then bottles filled from the bucket. One council did not describe their technique.

Eight councils stated they used acid-washed bottles where appropriate with many of these councils using bottles supplied by their contracted laboratory. The remaining councils either used rinsed (or new) bottles or did not state how their sample bottles are cleaned.

All councils use field forms (either paper or electronic) to record site information and sample details for each sampling site.

Eleven councils measure visual clarity at their SOE sites, with all of these following the black disc measurement protocols outlined in MfE 1994).

3.1.4 Water quality sample handling

There is a high degree of consistency in the handling of water samples once collected. All councils place sample bottles in cooled storage bins after collection. Most respondents did not state what was used to chill the samples. Of those that did, there was some variability with ice, flake salt ice and cooler pads all being used.

All the councils stated that it was their aim to send the samples to their lab services provider within 24 hours of sampling. Several councils stated, however, that there are occasions when this might not be possible. At least one council specifically avoids Friday sampling so there is no risk that samples remain unprocessed over weekends.

3.1.5 Quality assurance approaches

Six councils stated that they do not carry out any quality assurance of field procedures. Out of the ten that do carry out QA there was considerable diversity as to what is actually done. Most respondents stated that QA consists of training staff on procedures with an emphasis on senior staff training or inspecting junior staff in the field. One council stated that they share knowledge and improvements with a neighbouring council and that there is some visiting of each other's sites. Another council specifically stated that some duplicate sampling has taken place in the past and checking of visual clarity measurements between operators is periodically carried out. While no details were provided, another council stated they carried out QA in accordance with their ISO9001 accreditation.

Some variables are evaluated in some regions by the use of field instruments (e.g., pH, conductivity), although only for temperature, visibility and perhaps DO are field measurements mandatory. All councils stated that they carry out calibration (when possible) of their field instruments. Seven of these councils perform calibration tests on the day of sampling with the others calibrating their instruments at intervals ranging from weekly to 6-monthly.

3.2 Recommendations

The above summary indicates that, while most water quality sampling tasks are carried out in a consistent way within councils, there is considerable diversity in protocols between all 16 councils. Despite this, many of the councils are already carrying out sampling procedures that follow widely accepted procedures and this is reflected in those questionnaire answers that had a majority or 100% response. This is a reassuring finding which means that the following recommendations for changes to sampling protocols should not be too onerous for many councils (Table 3-2) and should cause least disruption overall in achieving consistency.

Table 3-2: Summary of recommended river water quality monitoring protocols.

Sampling Protocol	Recommendation
Sampling frequency	Monthly
Timing of site visits	Same sites visited at the same time (±1hr) on every visit
Core sites	All SOE sites should be sampled on each visit
Site metadata	Detailed metadata should be available on sites
Measurement of flow	All samples should be 'flow-stamped'
High flow/flood sampling	If high flows are intercepted on the target date and time sampling should be carried out as usual (assuming it is safe to do so).
Sample bottles	Only use new or acid washed (reused) sample bottles
Sampling location	Mid-stream or by telescopic pole
Use of bridges for sampling during high flows	When access to sampling site may be hazardous, bridge sampling (by bucket) should be considered (where possible)
Water sampling procedure	NRWQN water sampling procedure
River sampling location	Runs or riffles (avoid pools – where dead zones are common)
Sample handling and storage	Sample stored *unfiltered* in chilled storage bins. (Filtering is best done in laboratories – Section 2.3.2.) Ice chips are the preferred method of chilling.
Sample delivery to laboratory	Deliver samples to lab within 24 hours. Interim refrigeration may be appropriate if there is any delay in sample delivery (which should, in any case, not exceed 48 hrs).
Record of site visit	Record site/sample information on field sheet*
Instrument calibration	Monthly for temperature probes; Day of use calibration for DO and pH probes
Measurement of visual clarity of water	Use of black disc technique. *In-situ* measurement during base flows; trough technique on stream bank for high flows

(*) Development of a 'national' field sheet is recommended to promote consistency across councils and as a QA measure (Refer Section 6.5 below).

3.2.1 Timing of SoE sampling visits

We recommend that river water quality sampling be carried out monthly at all SOE sites. Although an argument could be made for less (e.g., bi-monthly) or more (e.g., fortnightly) frequent sampling, monthly sampling is a well-proven compromise that would enable long-

term trends to be detected (over time periods of between 5-10 years) while not being overly influenced by autocorrelation[1] (Davies-Colley et al. 2011a). For this reason also, it would be counter-productive to sample water quality at SOE sites on a revolving basis. Collecting data from sites for a few years would provide a 'snapshot' only and the data would be of little or no use, for detection of long-term trends.

Water quality sampling should also occur at sites at approximately the same time on each visit. In the NRWQN, field staff aim to collect samples within ±1 hour from a given site on each monthly 'run' (Davies-Colley et al. 2011b) and we suggest that this protocol be adopted for council SOE water quality sampling. Consistency over time of time-of-day of sampling is very important if data is to be used for long-term trend analysis particularly for variables that fluctuate diurnally (e.g., temperature, DO). Inconsistent sampling times would add a level of noise to the dataset that could compromise long-term trend detection (e.g., water temperature change driven by climate change). In order to achieve time-of-day consistency, sites should be visited in the same order on each 'run' and SOE field staff should not be required to carry out miscellaneous tasks (e.g., hydrometric site maintenance) if these could significantly delay their progress. Many councils are already achieving this ±1 hour target. We recommend that sampling within 1 hour be adopted as the national protocol.

3.2.2 Site metadata

Site metadata is an important component of field data quality assurance as recognised by most councils. The same SOE sites are typically visited by the same field staff on each sampling occasion for whom sample site identification and access is not an issue. However, unforeseen staff absences or staff turnover, make it important that a detailed record of each sampling location is available. We are aware that some councils already have metadata files to hand, and recommend that all councils collate metadata containing the following information.

- site name and ID number
- site directions (including map(s) and GPS coordinates)
- specific access notes (e.g., whether there are locked gates or private land to cross)
- photographs of the site, preferably under different flow states, including a photo of the precise sampling location(s) (annotations on photographs are desirable)
- a sketch map or satellite view (or another photograph) showing suitable locations for visual clarity measurement and for related river biomonitoring (macro-invertebrates and periphyton visual assessment) (refer Chapter 4)
- notes on a nearby hydrometric site from which flow can be estimated or else information on how to otherwise estimate flow at the SoE site (e.g., nearby staff gauge)

[1] Autocorrelation is the tendency of a new data point to be influenced by the magnitude its predecessors, even after trend and seasonality have been accounted for (Davies-Colley et al., 2011a)

- any relevant health and safety issues about the site, including safe access points at high flow.

3.2.3 Flow ‘’stamping’ of water quality data

We note that currently few councils are able to estimate flow at all of their river water quality monitoring sites. Measuring flow in an important aspect of long-term water quality monitoring, being needed to interpret water quality (notably for flow adjustment in trend detection) and for calculation of material fluxes such as annual loads of pollutants. Most important is flow at time of SoE sampling (‘flow stamping’), but a continuous record of flow is important for interpreting water quality (and biomonitoring data) as well as for load estimation.

Continuous flow recording (at nearby hydrometric sites) is ideal for full hydrological underpinning of water quality data – providing both flow stamping of particular visits and continuous data for antecedent flow interpretations and calculation of flow indices. However it is probably not necessary or possible to underpin all SOE monitoring sites with continuous flow records. Some councils are measuring flow on each SoE sampling occasion, but this is onerous and becomes increasingly redundant (given that regular sampling mostly intercepts baseflows) with time. Furthermore, the time taken for gaugings (probably exceeding 30 minutes on-site even for a simple wading gauging) may compromise the ± 1 hour target timing of SoE site visits (Davies-Colley et al. 2012).

An alternative is the installation of a staff gauge at a nearby site where there is reasonable hydrological control to allow a water level ‘stamp’ to be put on all SoE visits. Establishment of stage-discharge ratings (e.g., from gaugings at times of some water quality sampling, augmented with some special high flow gaugings) should eventually permit water levels to be converted to approximate flows for ‘flow stamping’ of SoE visits to these sites. Furthermore flow correlation with more distant hydrometric sites, usually on the same main-stem river, can then be undertaken with a view to applying a continuous record as an indicator of hydrological ‘context’ (e.g., flow distribution) and for load estimation. Finally, it may eventually become worthwhile, given a valuable long-term water quality record, to upgrade the SoE site with installation of a nearby continuous level sensor. Experienced hydrologists should be consulted regarding hydrometric under-pinning of SoE sites.

3.2.4 Stormflow sampling

SOE water quality monitoring is regular in time (‘pseudo-random’; Davies-Colley et al. 2011a), therefore the majority of samples are taken during base flow conditions (which occur the majority of the time). Because large quantities of contaminants (notably sediment, phosphorus and microbial pollution) can be transported during flow events it is also important that information on higher flows be obtained when possible. Deliberately not sampling the high flow events that occur on scheduled sampling days will introduce bias into datasets (Richards 1998) – because they are then no longer ‘pseudo-random’. We recognise that health and safety or physical access issues may occasionally preclude sampling during (very) high flow events, however, where possible, the scheduled sampling should be carried out irrespective of flow state. Where access to the regular sampling site is hazardous during (very) high flows it is recommended that nearby bridges (where available) be used for sampling (by bucket). Obviously substituting a traffic safety hazard for a hydraulic hazard is not good practice! So care should be taken when water sampling from bridges that do not

cater specifically for pedestrians. The use of a back-up (flood flow) access point, such as a bridge should, of course, be noted on the field form.

3.2.5 Water sampling

Most of the 'core' variables listed in Table 1-1, that need to be measured in the laboratory (including nutrient analyses; also the 'supporting' analyses of turbidity and conductivity) can all be done on the one (usually 500 ml) bottle. However, *E coli* are typically measured on a separate sterile vial (100 ml) filled at the same time. Measurement of TSS with reasonable precision on rivers at baseflow requires a much larger volume to provide sufficient residue on glass fibre filters (5 litres or more from clear rivers). No acid or other preservatives (as used, for example, for trace metal sampling) should be present in bottles for SoE sampling.

Many councils collect their water quality samples in bottles (usually 500 ml is sufficient) that have been provided by their laboratory service provider. Others use a variety of bottle cleaning techniques. Some recent analysis by NIWA-Hamilton's Chemistry Laboratory has indicated that acid washing new laboratory-rated sampling bottles (used for nutrient sampling) is unnecessary (pers comm. Mike Crump, NIWA-Hamilton), and experience of Hill Laboratories is similar (Peter Robinson, pers. comm.). However, in situations where sample bottles are reused it is desirable that bottles are scrupulously clean to avoid sample cross-contamination (UNEP/WHO 1996). Sample bottles from a general pool should probably *not* be used, because a particular bottle might have previously contained a wastewater or contaminated stormwater, and there is a risk that, even with thorough washing and rinsing, some contamination, particularly of nutrients, may carry over. SoE sampling bottles, if reused, should be kept separate. UNEP/WHO (1996) recommended that if polyethlene bottles are to be reused for nutrients then they should be cleaned with chromic acid before rinsing multiple times with distilled water. This is resource and labour-intensive, and an alternative, that seems increasingly common, is to use new laboratory-rated sample bottles on each sampling occasion.

Fourteen councils stated in their questionnaire responses that they currently used a sample bottle filling technique similar to that described in the NRWQN field protocols: '*Prior to immersion in the water, remove the sample bottle cap. Facing upstream, fill the bulk sample bottle after rinsing twice with stream water. Submerge bottle to mid-depth with the open orifice pointing down. At mid-depth, gently tip the bottle up into the flow, allowing water to enter the bottle.'.* While there are a number of variations of this technique, the method described above is similar to those recommended by UNEP/WHO (1996) and USGS (2009). For the NRWQN, the sampling sites (that are well-characterised in site metadata) are sometimes accessed by wading to the middle of the river. However, if flow conditions make this inadvisable a reliable sample can usually be obtained using a telescopic sampling pole from the river bank or else from a nearby bridge. Samples should be collected as close to the centre of the river channel as possible.

We recommend that water quality samples are obtained from either riffles or runs. Sampling from pools should be avoided if possible because of the potential occurrence of 'dead zones' of poorly mixed water in which water quality characteristics differ from the river main-stream. We recognise that there is a degree of subjectivity in identifying riffles, runs and pools in some rivers, and the following simple identification guideline (as proposed by Jowett 1993) may be helpful.

- Riffle = swiftly flowing with a high proportion of its water surface broken.
- Pool = slow flowing with a smooth water surface.
- Run = intermediate between pool and riffle with a wavy water surface. Runs often occur immediately downstream of a riffle.

3.2.6 Sample handling

Currently all councils store samples in a chilled storage container after sampling. Samples should be placed in the storage/shipping container promptly and not left on the river bank (exposed to sunlight and subject to warming and possible photochemical degradation), while other tasks are completed on-site. There is some diversity between regions in the method of chilling water samples. USGS (2009) recommend that samples should be chilled to below 4℃. Rapid chilling seems unlikely to be achieved b y using reusable freezer pads or frozen bottles of water. Therefore we recommend that all samples should be placed in an ice slush inside the storage bins so that samples are evenly and quickly chilled. To avoid icing inside bacterial vials (which might compromise *E. coli* analysis) these may be wrapped in bubble wrap (Peter Robinson, Hill Laboratories, pers. comm.). Care must be taken to ensure that labels on bottles are water-proof and not washed off by meltwater contact. (If paper labels are used enclosing bottles in plastic bags may help avoid label wash-off.) Chilling of samples is obviously much more important in warm than in cold seasons, and special care to achieve cooling is recommended in summer months.

The USGS (2009) states that water quality *'samples should be packaged and shipped to the laboratory for analysis as soon as possible. Generally, the shorter the time between sample collection/processing and sample analysis, the more reliable the analytical results will be'.* To ensure consistency between regional council datasets we give the following guidelines for delivery of samples to laboratories. NIWA (for the NRWQN – see Davies-Colley et al. 2011b), USEPA (1993) and Environment Canada (EC 1981) all require that samples are delivered to their laboratories within 24 hours of collection. Similarly, we recommend that councils aim to deliver samples to their laboratory service providers within 24 hours of collection. If that is not possible interim refrigeration (to < 4^{o}C, but *not* frozen) is recommended, but samples should, in any case be delivered within 48 hours. Note that 48 hrs is outside usual recommendations for bacterial analysis, but tests by Crump (2011) have shown that *E. coli* assays are comparatively stable over at least this time. Sampling should not be scheduled on Fridays or directly prior to public holidays, to ensure laboratories can receive and process samples promptly.

3.2.7 Field forms

The recording of site/sample information on field sheets for water quality monitoring site visits is a standard field procedure and commendably all councils currently complete them. We have sighted many of these field sheets (kindly forwarded by council science staff) and they seem broadly suitable in that they prompt for important on-site measurements and attendant observations. The river monitoring field sheets supplied by many of the councils indicate that there is considerable commonality in the information collected. The current council field sheets seem adequate to record the necessary field measurements and prompt for desirable ancillary observations such as weather conditions. However, it is important that staff be

trained on how to correctly fill out the field sheets. In particular it is crucial that, when sampling occurs in a non-standard way (e.g., water samples taken from bridge during a high flow event instead of at the regular sampling location), those details are recorded. As discussed below (Chapter 6) we think it highly desirable that a 'national' field form be developed (1) to promote consistency across councils and (2) to facilitate QA.

Some water quality variables (e.g., visibility and temperature) can only be measured at the time of a site visit. While pH and conductivity (and perhaps DO – by Winkler titration, APHA 2005) can be measured in the laboratory, many councils measure these variables in the field, either instead of or as well as laboratory measurements. Accordingly, councils have a suite of field instruments that they use to measure conductivity, temperature, DO and pH. To ensure accurate measurements, these instruments should be calibrated frequently, ideally with checks on each day of use. We recommend that DO and pH probes are calibrated on each day of use, ideally at *both* the start and end of the daily 'run' so as to bracket the field measurements. Temperature and conductivity probes are generally relatively stable, but calibrations should still be checked on each day of use.

3.2.8 Visual clarity

Most councils already measure water clarity using the black disc visibility method described in MfE (1994) and Davies-Colley (1988). Common perception is that the method is subjective, but optical theory and measurements paired with optical instruments (Zanevald and Pegau 2003) show that visibility is both insensitive to subjectivity and fairly precise (about 5% CoV of independently replicated measurements). (These attributes of visual clarity by the black disc method also apply to the Secchi depth observation recommended for lakes – Chapter 5).

Because no standard can (easily) be applied to routine visibility observations in the field, care needs to be taken that the visual range is not biased – most commonly from shadows across the path of sight in the water or contamination by plumes of disturbed fine sediment.

- **The path of sight must be uniformly lit.** Under clear sun conditions, the path of sight must be uniformly sunlit or uniformly in shadow. No partial shadowing, such as by riparian trees, should be affecting the path of sight.
- **Plumes of fine sediment must be avoided.** Obviously the water mass in which visibility is measured must be representative of the river. Any plume of fine sediment caused by wading in the channel must be fully flushed from the water volume encompassing the path of sight before visibility is recorded. (Such plumes will cause a low bias in visibility. This kind of bias might account for the outlier point in Figure 2-1.)

Simple, but effective methods have been developed for in-channel deployment of black disc targets by only one operator, such as fixing disc and tape measure to a tripod or a waratah stake driven into the stream bed. A particularly promising approach, developed by the NIWA-Hamilton field team (Margaret Bellingham, pers. comm.), has the three sizes of black disc (20mm, 60 mm and 200 mm diameter) mounted on a telescoping rod with a small black float attachment. The (single) operator simply pushes the floating rod away until the disc

disappears, and the extinction point is then refined by noting disappearance/reappearance distances and averaging in the usual way.

Anecdotally we are aware that some councils use only one size of black disc target irrespective of visibility – apparently because they have been misled by the MfE (1994) guideline that the visibility *of a 200 mm disc* (our emphasis) should exceed 1.6 m for contact recreation. However, the apparent size of the disc should be kept approximately constant (at 2-10^{o} of arc), which implies that the physical size of the disc should increase with visual range. This need to change disc size, depending on visibility, is not emphasised in MfE (1994), but Davies-Colley et al. (2003: see Figure 3.13) suggest routine use of three sizes: a 20 mm disc from (about) 0.05 up to 0.5 m visibility, a 60 mm disc between 0.5 and 1.5 m and a 200 mm disc above 1.5 m up to (about) 15 m visibility – which is the protocol used in the NRWQN. The smallest of these discs is usually permanently mounted on a graduated rod.

The black disc method usually requires operators to enter the water to take measurements. But wading in the channel during high flows may be hazardous. A survey of the site at low flow can reveal where access to the channel is safe for black disc observations at relatively high flow, such as on a point bar. However, wading is usually not advisable at very high flow. At such times the visibility is likely to be unusually low because of the typical inverse relationship of clarity and flow (Smith et al. 1997) and the preferred method for water clarity measurement is to take observations on a sample (obtained by bucket) contained within a trough (Davies-Colley and Smith 1992). A further benefit of the trough measurement method is that, if visibility is very low (say < 100 mm) *irrespective of flow conditions*, then the sample can be volumetrically diluted in the trough so as to eliminate interference (i.e., bias) from shadowing by the viewing periscope at very close ranges. (A very low visibility at baseflow would probably indicate some gross disturbance upstream such as a stock crossing or earthmoving machinery in the channel).

We recommend visibility be measured in a trough (Davies-Colley and Smith 1992) when entry to the channel for *in situ* visibility observations is hazardous or (with appropriate dilution of the water sample) when the visibility is < 100 mm. Use of a trough for visibility observations, rather than *in situ*, should be noted on the field form.

4 River Bio-monitoring Protocols

Appendix C gives the questionnaires used to elicit metadata on current regional biomonitoring of rivers in New Zealand.

4.1 Benthic macro-invertebrates

4.1.1 Current practice

Among the 15 (all but one) councils that do routine biomonitoring, there was reasonably high consistency with respect to macro-invertebrate protocols (Table 4-1). All fifteen councils monitor benthic macro-invertebrates as the core of their biomonitoring programmes. Thirteen councils conduct their macro-invertebrate monitoring once per year during the spring or summer period (Table 4-2), and all except one use the standard mesh size of 0.5 mm. In hard-bottomed streams, 13 councils follow sampling protocol C1 more or less as described in Stark et al. (2001). All except one focus on riffle (shallow, fast-flowing with broken water surface) habitats where available, or runs (moving water, deeper than riffles with typically smooth water surface) where riffles are not present. The remaining council conducts “all-habitat sampling”, i.e., sampling all habitats (riffles, pools, runs, stream margins) that are present in a reach. The data from all-habitat sampling is probably not comparable with that from riffle-only sampling.

There is considerable variability among councils in dealing with soft-bottomed streams. Only six of the councils use a standard soft-bottom protocol for soft-bottomed streams, though in some regions where no separate protocol is used it was not clear whether soft-bottomed streams actually occur in the monitoring network. All 15 councils identify macro-invertebrates to MCI-level (genus for most insects and molluscs, various levels for crustaceans, annelids and “minor phyla”), allowing all the recommended invertebrate indices to be calculated. However, the number of councils using coded abundance (semi-quantitative categorical) methods is almost equal to the number using fixed counts or fully quantitative methods. Regrettably, some indices recommended by the Variables Expert Panel, e.g., Quantitative MCI and %EPT abundance, cannot be calculated from coded abundance data. All 15 councils report their macro-invertebrate results using more than one index, but four limit their reporting to biotic indices (various versions of the Macro-invertebrate Community Index).

Anecdotally, it appears that almost all councils restrict their SoE biomonitoring to *wadeable* streams, therefore consistency of protocols for non-wadeable streams may not be an issue at present.

4.1.2 Recommendations

In general we recommend that the majority practice be adopted (Table 4 3), but for field protocols we recommend a change. In hard-bottomed streams, substituting Surber sampling for kick sampling requires minimal extra effort and provides a number of benefits. First, strongly attached and heavy taxa (e.g., limpets, snails and cased caddisflies) are more likely to be collected in Surber sampling than kick sampling because rocks are scrubbed directly into the net. This results in more accurate assessments of community composition. Second, by providing greater consistency in terms of sample area, Surber sampling promotes greater precision in richness measures and MCI. Third, Surber sampling provides areal density data that can be used to understand the amount of invertebrate food supply for fish and

invertebrate grazing control on periphyton, answering important management and scientific questions. Thus for hard-bottomed streams we recommend collecting Surber samples from five positions in the sample reach according to Protocol C3 of Stark et al. (2001), and compositing the five samples into one. In the laboratory, a 200-count with scan for rare taxa (Protocol P2) would be followed, as described below. For soft-bottomed streams, Protocol C2 remains the most practical option.

We recognise that councils may be reluctant to change from their current practice of kick sampling (Protocol C1). However, in our experience, Surber sampling requires only 5-10 minutes additional time in the field and no additional time in laboratory processing. Councils that choose to continue using Protocol C1 can still calculate all of the macro-invertebrate indices recommended in this report. If councils choose to continue with C1, we recommend they aim to sample as precise a total area of stream bed as possible within the range 0.6-0.8 m^2 – a narrower range than described in Stark et al. (2001). This is to maximise precision in macro-invertebrate indices, which are affected by sample area though they are not reported in "per unit area" terms. Whether the total area is made up of five or seven replicates is less important.

In our previous report (Davies-Colley et al. 2012) we recommended annual sampling and that previous recommendation is endorsed here. Most councils currently sample macro-invertebrates in summer, but two sample only in spring, and one only in spring and autumn (Table 4-2). Consistency across councils would argue for summer sampling (at which time stable baseflows are also more likely to be encountered). However, as noted by the Variables Expert Panel, in southern regions, macro-invertebrates may be larger and more easily identified in spring, so potentially yielding more accurate results. However, such seasonality in life history implies that invertebrate abundances (and relative abundances among taxa) also may differ between spring and summer, thus results from spring and summer sampling may not be comparable. In (more northerly and more maritime) Taranaki streams, Stark and Phillips (2009) concluded that seasonality of life histories does not affect non-quantitative biotic indices, but that quantitative indices are significantly different in summer due to more stable flow conditions. Therefore, timing of macro-invertebrate sampling is likely to have a significant effect on results in most regions. We suggest that the effects of season on invertebrate size and abundance be investigated nationally. ***Provisionally*** **we recommend that all councils consistently sample macro-invertebrates during late summer,** with January to March being the 'target' months.

Rotating sites allows a monitoring network to gain greater spatial coverage, but reduces the ability to detect trends in ecosystem health at individual sites. The Network Design panel recommended that councils monitor fixed sites rather than rotating among sites, and we have adopted this recommendation.

Currently 14 councils collect macro-invertebrate samples from riffle habitats where possible, while one conducts "all-habitat" sampling. All-habitat sampling provides a more complete description than riffle-only sampling of the invertebrate community at a stream site. However, it can be more difficult to compare ecosystem health among sites using all-habitat data, as variation arises from the range of habitats present. Thus the choice of method depends on the primary objective of the monitoring. Although the Variables Expert Panel recommended further discussion on this point, we note that adoption of all-habitat sampling would require a

change in method for all except one council. Since data from riffle-only sampling are not comparable to those from all-habitat sampling, such a change would create a discontinuity in the majority of biomonitoring datasets. For this reason **we recommend sampling from riffles only** (or runs if no riffles are available).

At present it is difficult to recommend a specific stand-down (no sampling) period after floods due to the paucity of relevant data for New Zealand streams on which to base a recommendation. Previous studies from New Zealand and other countries indicate that invertebrate densities may take from 30 to 140 days (Scrimgeour et al. 1988, Matthaei et al. 1996) to recover after a flood, depending on the size of the flood, though taxon richness typically recovers more rapidly (e.g., 3-6 days in Matthaei et al. 1996). Clearly it is not practical for councils to wait 30-140 days after a flood, so the stand-down period will remain a compromise between scientific robustness and practicality. Furthermore, characteristics of each stream, such as flood frequency, food base, and substrate size and type, may influence the recovery rate of invertebrates (Death 2008), making it difficult to recommend a simple guideline that is appropriate to all stream types. Currently almost all councils specify at least two weeks' stand-down after a flood (defined as an event >3x median flow). For now, **we recommend that all councils adopt two weeks as a minimum stand-down, but aim for three weeks if possible.** We note again that obtaining a 2 or 3 week 'window' after a flood is generally easier in late summer than in other seasons. Councils should record the time since the last >3x median flow event in order to interpret effects of recent high flows. We also recommend further research into the recovery trajectories of macro-invertebrate richness and density following floods of different magnitudes in different stream and river types. In this way, even if stand-down periods cannot be increased for practical reasons, the potential effects of floods on invertebrate data sets may be considered.

For sample processing **we recommend that all councils adopt the quantitative P2 protocol** (200-count with scan for rare taxa; Stark et al. 2001), noting in particular the importance of tracking the proportion of the sample that has been processed. Tracking can be done using a grid of squares, as described by Stark et al. (2001) or by using a subsampler. Protocol P2 does not require much more effort than the semi-quantitative (coded abundance) P1 protocol, while the quantitative data can be used for calculating quantitative metrics (e.g., QMCI) and for a wider range of analyses than coded abundance data. A 200-count is recommended over the slightly more rapid 100-count because of the information gains described in Stark et al. (2001). While standardising the use of protocol P2 requires a change in practice for seven councils that currently follow P1, this change does not create problems in data comparability, as quantitative data can be converted back into semi-quantitative categories for time-trend analysis.

Currently all councils identify invertebrates to MCI-level or better. Councils may choose to identify invertebrates to a finer level (e.g., species), but should use MCI-level when forwarding results for national reporting, so that indices such as taxon richness and MCI are nationally consistent. We recommend a workshop to address some problems with the current guidelines for MCI-level identification, as outlined in Davies-Colley et al. (2011a).

The indices recommended for reporting by the Variables Expert Panel included MCI, QMCI, EPT* richness, total taxa richness and %EPT* abundance (*indicates that members of the caddisfly family Hydroptilidae are excluded). These five indices are all relatively

straightforward to calculate from the data provided by the P2 processing protocol, and are recommended.

Table 4-1: Macro-invertebrate biomonitoring protocols: current practice among 15 councils.

Line		Question	Response 1		Response 2		Response 3		Response 4		Total
1		Frequency	annual	13	Twice/year	2					15
2		Timing	See table below								
3		Sites fixed* or rotating?	fixed	13	rotating	0	combination	2			15
4		Net mesh size	0.5mm	14	1 mm	1					15
5	Hard-bottomed streams	Sampling protocol	C1*	11	modified C1*	2	C3*	1	combination C1 and C3	1	15
6	Hard-bottomed streams	Sample area per replicate	0.1-0.2m^2*	12	0.3m^2	1	timed	1	fixed sample volume	1	15
7	Hard-bottomed streams	No. of Replicates per site	ten	1	five-seven*	11	three	1	one	2	15
8	Hard-bottomed streams	Total sample area per site	0.6-1.0m^2*	13	not defined	2					15
9	Hard-bottomed streams	Replicates composited into 1 sample?	yes	14	no	0	some yes some no	1			15
10	Hard-bottomed streams	Mesohabitat type(s)	riffles (if none then runs)	14	runs (if none then riffles)	0	all habitat	1			15
11	Hard-bottomed streams	Follows protocol C1*?	yes	11	modified	2	no	2			15
12	Soft-bottomed streams	Sampling protocol	C2*	5	modified C4*	1	HB protocol**	8	slight modification to HB protocol	1	15
13	Soft-bottomed streams	Sample area per replicate	0.3m2	5	timed	2	N/A	8			15
14	Soft-bottomed streams	No. of replicates per site	ten	3	five-ten	2	ND	1	N/A	9	15
15	Soft-bottomed streams	Replicates composited into 1 sample?	yes	7	no		N/A	8			15
16	Soft-bottomed streams	Mesohabitat types	wood/bank/macrophytes	7	other	0	N/A	8			15
17		Definition of HB vs. SB* streams	>50% of stream bed naturally composed of silt, sand or pumice	8	other	2	ND or N/A	5			15
18		stand down period after floods*	>3 weeks after flood	1	3 weeks after flood	4	2 weeks after flood	7	<2 weeks after flood	2	14
19		Sample processing protocol	P1*	7	P2**	6	P3*	2			15
20		Taxonomic level of ID	MCI level	14	species	1					15
21		Indices used for reporting	MCI*	13	SQMCI*	10	QMCI	5	total taxon richness	3	31
22		Number of indices used for reporting	MCI only	0	MCI and SQMCI	4	MCI and SQMCI/QMCI and EPT richness	3	MCI and SQMCI/QMCI and EPT richness and %EPT	5	12

Line	Question	Response 1		Response 2		Response 3	Response 4	Total
23	Field QA/QC	10% of field assessments repeated	1					1
24	Lab QA/QC	Stark et al. (2001) QC procedures (or better)*	14	none	1			15

Notes

ND=not determined; N/A=not applicable.
Line 3: *fixed=same set of sites visited every sampling occasion.
Line 5: * from Stark et al. (2001).
Line 6: *of these, 4 councils specify exact area; 8 councils give range.
Line 7: *of these, 8 councils specify an exact number; 3 councils give range 5-7.
Line 8: *of these, 4 councils specify exact number, 9 councils give range.
Line 11: *C1=5-7 kick samples of 0.1-0.2 m^2 each, giving total of 0.6-1.0 m^2 .
Line 12: * from Stark et al. (2001). **of these, 4 councils specified no soft-bottomed streams occur.
Line 17: HB=hard-bottomed, SB=soft-bottomed.
Line 18: 1 council said "variable". *Various definitions of flood: 8 councils used "3x median flow", 1 council used "3x previous baseflow" 2 councils used "1 year return period", 1 council used "bed-moving flow", 2 councils did not define.
Line 19: * from Stark et al. (2001). **P2: 1 used 100-count, 4 used 200-count, 1 used 300-count.
Line 21: In addition, EPT richness (9), %EPT abundance (6), other (5). Abbreviations: MCI=macroinvertebrate community index, QMCI=quantitative MCI, SQMCI=semi-quantitative MCI, EPT=Ephemeroptera, Plecoptera and Trichoptera. *May include soft-bottomed equivalent.
Line 22: in addition, 2 councils did MCI/SQMCI/QMCI and %EPT, 1 did SQMCI and total richness and EPT richness and total abundance.
Line 23: other procedures include staff training, same staff members used for all sites.
Line 24: *1 council does this every second year; 1 council alternates internal and external QC annually.

Table 4-2: Timing of macro-invertebrate sampling. Dark grey shows planned sampling period. Light grey shows actual, if delays occur.

Council	June	July	August	September	October	November	December	January	February	March	April	May
1												
2												
3												
4												
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												

Table 4-3: Summary of recommended protocols for macro-invertebrate monitoring.

Line	Attribute		Recommendation	Current majority practice?	Rationale/notes
1	Frequency		Annual	y	Reasonable trade-off between cost and data richness.
2	Timing		Jan-Mar	y	Status quo for most councils. Relatively stable flow conditions in summer likely to result in higher invertebrate densities and less postponing of sampling due to floods. But review needed as individual invertebrates may be larger in spring.
3	Sites fixed or rotating?		Fixed	y	Recommended by Network Design panel.
4	Net mesh size		0.5 mm	y	Recommended by Stark et al. (2001).
5	Hard-bottomed streams	Sampling protocol	C3	y	Although Stark et al. (2001) recommend C1 as suitable for SoE, there issignificant gain for little extra effort by using 5 pooled Surber samplers instead of 5-7 pooled kick nets.
6		Sample area per replicate	$0.1m^2$	y	
7		No. of Replicates per site	5	y	
8		Total sample area per site	$0.5m^2$	n	
9		Replicates composited into 1 sample?	Yes	y	Cost of separate replicates too high.
10		Mesohabitat type(s)	Riffles	y	Status quo for most councils, keeping continuity with previous data records. Riffles-only gives greater comparability among sites. But expert panel recommended further discussion.
11	Soft-bottomed streams	Sampling protocol	C2	y	Separate SB protocol recommended by Variables panel. C2 is the soft-bottomed equivalent of C1.
12		Sample area per replicate	$0.3m^2$	y	As specified in C2.
13		No. of replicates per site	10	y	As specified in C2.
14		Replicates composited into 1 sample?	Yes	y	As specified in C2.
15		Mesohabitat types	wood/bank/macrophytes in proportion to occurrence (proportions recorded)	y	As specified in C2.
16	Definition of soft bottomed vs. hard bottomed streams		>50% of stream bed naturally composed of silt, sand or pumice	y	

Line	Parameter	Recommendation	Current majority practice?	Rationale/notes
17	Stand-down period after floods	Interim: 2 weeks minimum after flow >3x median flow AND record time since last 3x median flow event		A trade-off between robustness and practicality will be required, but requires data on which to base the decision, and to understand impacts of floods on invertebrate data. Thus, further research needed.
18	Sample processing protocol	P2 (200-count with scan for rare taxa), tracking the proportion of sample processed	y	Fixed count data can be used for QMCI and other metrics. Fixed counts give much less variability than coded abundance (Duggan et al. 2003). 200-counts give significant information gains over 100-counts for modest increase in effort (Stark et al. 2001 and references therein). Recommended by Stark et al. (2001).
19	Taxonomic level of ID	MCI-level		Reasonable trade-off between detail and cost.
20	Indices used for reporting	MCI (sb-MCI), QMCI (sb-QMCI), %EPT abundance, taxon richness, EPT richness		List recommended by Variables expert panel. QMCI as well as MCI, because QMCI more sensitive to small or medium changes in invertebrate community. MCI (and other indices?) ideally would be indexed against reference condition, but may be hard to obtain reference values.
21	Field QA/QC	?		Not yet discussed.
22	Lab QA/QC	QC2	y	Every second year should be sufficient, provided sample processor remains the same and passes QC.

4.2 Periphyton

4.2.1 Current practice

Responses to the periphyton part of the biomonitoring questionnaire are collected in Table 4-4. Ten of the 15 councils returning the survey recognise excessive periphyton as an issue of concern in their region. Twelve councils monitor all forms of periphyton routinely at all SoE sites, and one monitors at some sites. Of these 13 councils, all except three monitor more than once per year, but only three monitor all sites monthly. Differences in the frequency of monitoring affect the relative dependability of different datasets, but do not preclude combining data from different regions. Nine of the councils monitoring periphyton use standard visual assessment protocols based on the Stream Periphyton Monitoring Manual (Biggs and Kilroy 2000), while three use quantitative methods only, and one uses "casual observations" only. These differences *do* preclude combining data from different regions. Among the nine councils using visual % cover estimates, all except one focus on "run" habitats, which means high consistency among councils, but that periphyton data cannot be used with great confidence to interpret invertebrate data, as invertebrates are collected from "riffle" habitat. A variety of indices are used to report visual estimate results, but this should not cause problems provided that the raw % cover results are available.

Quantitative (biomass) measures of periphyton (as chlorophyll-*a* or AFDM – ash-free dry mass) are more time-consuming and expensive than visual % cover estimates. Seven councils measure biomass routinely, all using the standard protocols described in the Stream Periphyton Monitoring Manual (Biggs and Kilroy 2000). Differences with respect to number of replicates and whole rock vs. defined area sampling should not reduce comparability among councils. However, differences in the extraction method for periphyton chlorophyll *a* (Chl *a*) may reduce comparability, as ethanol may be up to twice as efficient as acetone in extracting Chl *a* (Biggs and Kilroy 2000). All six councils report on Chl *a* and all except one also report on AFDM. A minority of councils also identify periphyton taxa.

4.2.2 Recommendations

Recommendations for periphyton monitoring are summarised in Table 4-5. The primary objective(s) for periphyton monitoring were not agreed by the Variables Expert Panel, making decisions on protocols somewhat difficult. Our first recommendation, therefore, is to agree on one to three primary objectives. The most likely choices are a) indicating nutrient enrichment, b) interpreting invertebrate data and c) assessing impacts on aesthetics and recreation.

Periphyton should be assessed monthly on visits to water quality sites during the warmest six months of the year (typically November to April), as and when conditions of water clarity, depth and current allow. This meets objective a) above. Monthly assessments provide sufficient data to account for temporal variability, so allowing comparisons among years and sites, and providing context for the annual assessments at macro-invertebrate sites. Focusing on the warmest six months of the year brackets the period when periphyton biomass is maximal. Periphyton should also be assessed annually at macro-invertebrate monitoring sites where these are separate from water quality sites, so meeting objective b) above.

We recommend visual assessments, following a modification of rapid assessment method 2 (RAM-2) in Biggs and Kilroy (2000). Instead of picking rocks from the stream bed, we recommend making observations of a 0.5 m diameter circle of stream bed, preferably using an underwater viewer or bathyscope, as this reduces bias when the bed consists of rocks too large to pick up (C. Kilroy, NIWA, pers. comm.). Twenty observations are made at evenly spaced intervals on transects across the stream, with four transects recommended for small streams (<5 m wide), while fewer transects, each with more observation, can be used for larger streams if required (e.g., three transects for medium streams (5-10 m wide), two transects for medium-large streams (10-20 m wide) and one transect for large rivers (20 m wide)). The rationale for this is mainly logistical – better representation of a stream reach is gained by sampling more transects, but it takes considerable time and effort to wade multiple transects across large rivers. Instead of the growth categories described in RAM-2, we recommend those described in Table 4-6 below (Kilroy 2011). These categories make no distinction between colours, which are often hard to determine, and combine "medium mats" with "thin films", as these also can be hard to distinguish. New categories are created for Didymo and cyanobacteria so data can be used to assess invasions and health risks of these organisms, respectively. Another new category, "sludge", describes a loose mucus-like growth. Kilroy (2011) provides a visual identification guide to these growth forms. Despite some new categories, it should be possible to transfer data recorded in the RAM-2 categories to the system in Table 4-6, allowing continuity between future and historic data.

Indices for summarising and reporting periphyton data are still being developed. Hopefully a strong correlation between visual assessments and quantitative Chlorophyll *a* measurements will emerge, so that a "Chlorophyll *a* equivalent" measure can be calculated from visual assessments[2]. However, more data are needed from a wider geographic range of sites to develop this correlation. The "periphyton enrichment index" described in RAM-2 and suggested for use by the expert panel has not been thoroughly tested for scientific robustness and is not recommended for SoE monitoring (C. Kilroy, NIWA, pers. comm.). In the meantime, data from individual categories can be reported, and data from "thick mat" and "filamentous" categories compared to the MfE guidelines for aesthetics and recreation (Biggs, 2000).

At present most councils assess periphyton in run habitats. This is appropriate for assessing periphyton as an indicator of nutrient enrichment, as periphyton growth tends to be more responsive to changes in nutrient concentrations in runs than riffles (C. Kilroy, NIWA, pers. comm.). Thus we recommend using run habitats at water quality sites. But at macro-invertebrate monitoring sites, where the aim is to interpret invertebrate data, periphyton must be assessed in the same habitat as invertebrates, i.e., riffles.

[2] We note that seven of fifteen councils already conduct quantitative periphyton assessments. Annual quantitative measurements at the same time and location as visual assessments would calibrate the visual cover assessments against an objective measure of periphyton biomass. If strong, the correlation between the two measures could be used to develop an "Chl *a* equivalent" index for reporting visual assessment data. Therefore, standardising the field and lab protocols among those councils that currently do quantitative assessments seems advisable. We recommend following protocol QM-1b in Biggs and Kilroy (2000). We suggest allowing individual discretion in choosing between whole rock and fixed area sampling methods, noting that if whole rocks are used then rock surface area should be multiplied by 0.65 to get exposed surface area and that exposed rock surface area is typically about 30% greater than the stream bed area. Chlorophyll *a* (extracted with hot ethanol) is the recommended measure, with ash-free dry mass (combusting at 400 ℃ for 4 hours) an optional extra. Biomass measurements can be done annually in association with visual assessments.

These recommendations for visual assessment are made in consultation with Cathy Kilroy (NIWA) and are consistent with new periphyton assessment guidelines being developed as an Envirolink Tools project (F. Matheson, NIWA), as suggested by the Variables Expert Panel. They are also consistent with new guidelines for the National River Water Quality Network (John Quinn, NIWA pers. comm.)

Table 4-4: Periphyton biomonitoring protocols: current practice among 15 councils.

Line	Question		Response 1		Response 2		Response 3		Response 4		Total
1	Is periphyton growth an issue of concern in your region?		yes	10	rarely	1	cyanobacteria only	1	no	3	15
2	Is periphyton monitored regularly?		yes	12	some sites only	1	cyanobacteria only	1	no	1	15
3	Main objective(s) for periphyton monitoring		link to nutrients/WQ*	9	interpret invertebrate data	4	aesthetics/recreation	7	human health	3	23
4	Frequency		annual	3	monthly	3	combination annual/monthly*	4	2-4x per year	4	14
5	Timed with		invertebrate biomonitoring	3	WQ	6	combination invert biomonitoring/WQ	4	other	1	14
6	Visual (% cover) assessment	field protocol	RAM-2*	7	MfE/MoH cyanobacteria*	1	Kilroy (2011)	1	none	4	13
7		mesohabitat type	runs (if none then riffles)	8	riffles (if none then runs)	0	random	1	N/A	5	14
8		sampled area per replicate	whole rocks	5	bathyscope views	3	other set area	1	N/A	6	15
9		no. of replicates	Twenty - twenty five	6	fifteen	1	ten	2	N/A	6	15
10		reporting indices	PEI*	3	%cover long filaments + thick mats	2	% coverage of mat, filamentous, film and clean	2	Not defined	2	9
11	Quantitative (biomass)	Are quantitative (biomass) assessments done?	yes	7	no	8					15
12		field protocol	QM-1b*	6	QM1-a*	1	N/A	8			15
13		sampled area per replicate	disk (5-6 cm diam)	6	whole rocks	1	N/A	8			15
14		no. of replicates	twenty	1	ten	2	five	2	five-ten pooled	2	7
15		reporting indices	Chl a*	1	AFDM*	0	Chl a and AFDM	4	ND	1	6
16		Extraction/analysis methods	Chl a: acetone	2	Chl a: ethanol	4	AFDM: 500 ℃ for 1 hour	2	AFDM: 400 ℃ for 4 hours	2	10
17	Identification of periphyton taxa		yes	5	no	10					15
18	Other analyses		cyanobacteria	6	Didymo presence	1	iron bacteria	1			8
19	QA/QC		procedures in Biggs and Kilroy (2000)	4	other	2	none/not specified	8	N/A	1	15

Notes

Line 3: other objectives include Didymo presence (1 council), link to/contribute to overall ecosystem health (3 councils), monitor effects of riparian planting (1 council).
Line 3: *WQ = water quality.
Line 4: *2 councils do quick visual assessment monthly and more comprehensive annually; 2 councils do selected sites monthly and all sites annually.

Line 6: In addition: 1 council uses RAM-1, 2 councils use coarse estimates. *RAM-1, RAM-2 are from Biggs and Kilroy (2000). MfE/MoH cyanobacteria is from Ministry for the Environment and Ministry of Health (2009).
Line 7: In addition: 1 council uses runs or riffles, depending on site.
Line 8: In addition: 1 council does whole reach estimates for rapid method.
Line 10: In addition: 1 council reports % total cover, 1 council reports % long green filaments. *PEI=periphyton enrichment index (Biggs and Kilroy 2000, p44).
Line 12: *Biggs and Kilroy (2000).
Line 15: *chlorophyll *a*, Ash-Free Dry Mass.

Table 4-5: Summary of recommended protocols for periphyton monitoring.

Line	Parameter		Recommendation	Rationale/notes
1	Main objective(s)		Indicator of nutrient enrichment, interpret invertebrate data, assess against aesthetic/recreation and human health guidelines	Needs further discussion to identify most important objective(s).
2	Visual (% cover) assessments	Frequency	monthly at WQ sites (during warmest 6 months of year when conditions allow) and annually at invertebrate sites	Monthly monitoring captures temporal variability, allowing more accurate time-trend analysis and putting annual data in context. Monitoring at WQ sites captures response to nutrients; monitoring at invertebrate sites enables interpretation of invertebrate data.
3		Timed with	WQ and invertebrate biomonitoring	
4		Field protocol	Modified RAM-2, using field of view instead of selected rocks and using Kilroy (2011) growth categories - see Table 3 below	Kirloy (2011) categories are mostly a simplification of those in RAM-2, so historic data using RAM-2 can be translated into new categories
5		Mesohabitat type	Runs for monthly sampling; riffles when linking with invertebrate sampling	For linking to nutrients, sampling runs is preferred, as periphyton growth is more responsive in runs. But for linking with invertebrates, better to sample in riffles where invertebrates are sampled.
6		Sampled area per observation	0.5m diam, preferably using underwater viewer	“views” recommended over picking rocks, as picking excludes very large and small rocks
7		No. of observations	20	Suggest small streams (<5m) do 5 observations x 4 transects; medium streams (5-10 m) do 7 obs x 3 transects; med-large streams (10-20m) do 10 obs x 2 transects; large rivers (>20 m) do 20 obs x 1 transect
8		Reporting indices	work in progress; % thick mats and filaments can be reported against MfE guidelines for aesthetic/recreation objectives (Biggs 2000)	A correlation between visual estimates and Chl a is being developed, which may allow reporting of visual estimates as "Chl a equivalent". The Periphyton Enrichment Index (Biggs and Kilroy (2000) recommended by expert panel) is not well enough validated to be used for national reporting.
9	QA/QC		QM-1b QA (Biggs and Kilroy 2000)	

Table 4-6: Proposed periphyton scoring sheet for a small stream.

Instruction – For each observation (column), record % for all appropriate variables within a 0.5 m diameter as a "bird's-eye" view, i.e., percentages should add up to 100 (if filamentous algae overlay mats then record only the cover you see from above). If bed is not observable, or is inaccessible, or the variable is not applicable, put a dash (-) into relevant box. See Kilroy (2011) for photographic guide to variables (periphyton growth categories).

*V=using underwater viewer; CE=close eyeball assessment (observer is in stream); DE=distant eyeball assessment (observer is on bank, bridge, etc.). Recommended protocol is using underwater viewer. DE is least preferable option. BG0 = best guess is that % cover = zero, applied in circumstances such as (i) the river is too deep/murky to make observations at all sites across the transect but observations at shallower depths and experience makes you pretty confident that cover would be zero, so record that rather than a dash for missing data, or (ii) when the river is at high flow and is unwadeable, but experience at the site would inform you that cover would be zero.

	Transect 1					Transect 2				
Observations	1	2	3	4	5	1	2	3	4	5
V, CE, DE or BG0*										
Filamentous										
Sludge										
Didymo										
Cyanobacteria mats										
Other Mats > 2mm thick (excluding Didymo & Cyano)										
Thin Films										
Bare Area										

Ditto for transects 3 and 4.

Predominant Periphyton colour (Filamentous): ..... (Mats): ....(Film):……

Observation site (circle one): Open / Shaded

4.3 Fish

4.3.1 Current practice

Eight councils currently monitor freshwater fish routinely in their SoE monitoring programmes, an increase of two since 2008 (Hudson et al. 2011). Current protocols are summarized in Table 4-7. Six of the eight councils monitor annually during summer months (December to March), the period recommended by David and Hamer (2010). However, only two councils appear to visit every site annually, while most others visit a core set of sites annually and sample the other sites on a 2-5 year rotation. All the councils monitoring fish routinely aim to compile a complete species list for each site, while four of the councils also aim to determine species' relative abundance, and two councils aim to collect size data. Most councils appear to be following the protocol described by David et al. (2010), based on single-pass electric fishing of a 150 m stream reach.

4.3.2 Recommendations

The report by Davies-Colley et al. (2011a) outlined the debate among fish biologists regarding the merits of obtaining a full species list vs. quantitative estimates of abundance, and recommended a workshop to settle this issue. Since the choice of protocol depends on the primary objective of the monitoring, no updates to recommendations in Davies-Colley et al. (2011) can be made here.

Nine councils identified obtaining a full fish species list as a primary aim, while none identified obtaining quantitative fish abundance data as an aim. Two councils record fish size. The protocol used by eight councils is single-pass electric fishing over a 150 m reach (following David et al. 2010), an appropriate protocol to approach the aim of obtaining a full species list and an abundance:area ratio. This protocol is also sufficient for calculating scores on the Fish IBI (Index of Biotic Integrity; Joy and Death 2004), the index recommended for national reporting by the expert panel. Davies-Colley et al. (2011a) related a concern by some councils that sampling 150 m may require too much time or effort, and that use of the electric fishing method may be restricted by the cost of equipment or lack of expertise. However the results of our survey show that councils doing fish monitoring have found a way to overcome these limitations. Councils appear to be overcoming time constraints by sampling sites on a 2-5 year rotation, although this is questionable from the perspective of time-trend analysis.

Although councils have clearly identified the types of data they are currently seeking, we still recommend a workshop discussion among fish biologists from councils, Department of Conservation and research institutions (NIWA and universities) to agree on (a) the most useful objectives for fish monitoring and (b) how to optimise sampling effort to best achieve these objectives[3]. There are direct trade-offs in sampling effort between reach length, the number of passes, recording extra data on fish size and the number of sites that can be sampled in a season. For example, a single pass over a 150 m reach may achieve a near-

[3] This workshop has now taken place. It was held on 13 August 2012, and involved fish biologists representing eight regional councils, three universities, one environmental consultancy, NIWA, Fish and Game, and Department of Conservation as well as MfE staff. The workshop addressed a wide range of issues regarding objectives, definitions, what should be measured, methodologies and reporting indices. The main current protocols (David et al. 2010) and indices (Fish IBI, Joy and Death 2004) were presented and discussed.

complete species list with semi-quantitative abundance data, whereas three passes over a 50 m reach may achieve fully quantitative population density estimates for up to 80% of the fish species (P. Franklin, NIWA, pers. comm.). If the overall objective is to monitor the response of the fish community to changes in the environment, it may be more appropriate to focus efforts on obtaining fully quantitative data than on obtaining complete species lists, because species presence/absence is likely to be relatively insensitive to changes in environmental conditions, and may show a greatly delayed response, compared to changes in population densities. Semi-quantitative data may provide a useful middle road, approaching the sensitivity of fully quantitative data without requiring the same sampling intensity. Further, fish size data may be very informative, indicating the types of pressures (e.g., migration barriers vs. over-fishing) experienced by fish populations. The benefits of recording fish size data may justify sacrificing precision in other parameters.

The trade-off between sampling frequency and site coverage also requires discussion. Because fish tend to respond more slowly to changing environmental conditions than macro-invertebrates, annual sampling may be more frequent than necessary to assess population trends. Sampling sites on a 2- or 3-year rotation allows the number of sites monitored to increase by two or three times, but longer rotations inevitably increase the time period required to detect temporal trends. Furthermore, recruitment can vary greatly from year to year, creating significant “noise” in data sets that can only be separated from longer-term trends by frequent sampling. One way to account (partially) for such temporal variability while maintaining wide site coverage is to sample a core set of reference sites every year while rotating among non-reference sites every two or three years.

Clear protocols on the most appropriate fishing method to use in streams with different characteristics need to be agreed. Electric fishing is the preferred method where stream characteristics allow, as it produces the least biased results. However, conditions of very low or high electrical conductivity, macrophyte growth, stream depth, etc., may inhibit use of the electric fishing method. Guidelines are required for consistently deciding when to use another method, and what method(s) to use. In particular, protocols for non-wadeable rivers are currently lacking and need to be developed.

Finally, agreement on data requirements, including both fish and habitat data, and production of ‘national’ field forms for use by all councils would help minimise the chance of missing or inadequate data.

The overall objectives of fish biomonitoring at both national and regional scale should be decided through robust discussion, and with a clear understanding of the conclusions that can be drawn from different types of data. The protocols that optimise sampling effort to best achieve the stated objectives can then be chosen. We recommend a workshop involving fish scientists from regional councils and research institutions to (1) determine objectives of on-going, routine fish monitoring and (2) recommend protocols consistent with those objectives.

Table 4-7: Fish biomonitoring protocols: current practice among 15 councils.

Line	Question	Response 1		Response 2		Response 3		Response 4		Total
1	Are fish monitored regularly?	yes	8	no	7					15
2	Frequency	Annual*	6	variable	1	not decided	1	N/A	7	15
3	Timing	Dec-Mar	6	Oct-Jun	1	all year	1	N/A	7	15
4	Sites fixed (same set of sites visited every sampling occasion) or rotating?	fixed	2	some fixed some rotating	3	rotating	2	some fixed, some one-off	1	8
5	Timed with other biomonitoring?	yes	1	no	7	N/A	7			15
6	sample method*	EFM*	10	spotlight	7	minnow traps	5	Fyke nets	4	26
7	reach length	150 m	8	100m	1	variable	1	N/A	5	15
8	no. of passes	one	9	other	0	variable	1	N/A	5	15
9	objective of data collection*	taxa list	9	relative abundance	4	fish size	2	N/A	5	20
10	data forms	NZFFD* forms	6	WRC* form	2	other (not NZFFD or WRC)	1	N/A	6	15
11	QA/QC	Responses include training and refresher courses in EFM and fish identification, 2 qualified staff members on site, experienced staff members on site								

Notes

Line 2: *not necessarily same sites each year.

Line 6: in addition, one council uses Seine nets. *most councils use more than one method, hence total is greater than 15. EFM = electric fishing method.

Line 9: *most councils have more than one objective, and some councils that sample fish non-routinely also answered the question, hence total is greater than 15.

Line 10: NZFFD = NZ Freshwater Fish Database; WRC = Waikato Regional Council.

4.4 Macrophytes

4.4.1 Current practice

Eight of the fifteen councils returning the survey recognise macrophytes as an issue of concern in at least some sites in their region. Ten councils monitor macrophytes routinely (Table 4-8), either on their own or as part of a physical habitat assessment. Most councils monitor macrophytes annually at the same time as other biomonitoring or habitat assessments. The objectives for monitoring vary, but the most common objectives are to detect negative effects on flow or dissolved oxygen, and as a general measure of stream health, while only two councils explicitly aim to detect invasive species. In line with these objectives, almost all councils restrict their measurements to broad categories, i.e., 2-3 different growth forms, and only one council identifies species.

4.4.2 Recommendations

The expert panel did not recommend incorporating macrophytes in national environmental monitoring for two reasons. First, they were unsure whether macrophyte growth is of widespread concern across New Zealand. However, in our survey, eight of fifteen councils indicated that macrophyte growth was of concern in at least some parts of their region, thus it appears that the concern is wider than the panel believed. Second, the panel felt that protocols for assessing macrophyte growth were lacking. An Envirolink tools project (Matheson et al. in prep.) is currently developing guidelines that define nuisance macrophyte growths and recommend standard protocols for quantifying macrophytes. In addition, our survey revealed that all but four councils currently monitor macrophytes in some way in at least some sites. Therefore standardising protocols for assessing macrophytes would not seem to add significant extra effort to monitoring programmes. Further, the national database on macrophytes that would result from nationwide monitoring would provide very valuable information for understanding the primary factors stimulating macrophyte growth and quantifying relationships with dissolved nutrients (F. Matheson, NIWA, pers. comm.).

Despite these valuable research applications of macrophyte data, and the availability of protocols, we do not recommend all councils assess macrophytes routinely and indefinitely. Some councils may however, choose to do so if macrophytes are a particular concern in their region.

The main concerns identified by councils regarding macrophytes were effects on flow, dissolved oxygen, general stream habitat quality and aesthetics. The key parameter underlying all these concerns is the percentage of stream channel volume occupied by macrophytes. Therefore we recommend, following Matheson et al. (in prep.), assessing macrophyte growth according to a slightly modified version of the protocols in Collier et al. (2007) (Table 4-9). Four evenly-spaced transects are located along the sampling reach (50-100 m long). To save time, the same transects can be used for macrophytes as for periphyton assessments. On each transect, the total macrophyte % cover is estimated in plan view (i.e., from directly above) occupying a 1 m wide belt upstream of the transect across the entire wetted width of the stream (Table 4-10). The % cover of emergent, surface-reaching and below-surface macrophytes is then assessed (these should add to the total % cover). Finally, for the submerged macrophytes, their height as an average proportion of the depth of the stream is estimated (this last step is additional to Collier et al. 2007).

Table 4-8: Macrophyte biomonitoring protocols: current practice among 15 councils.

Line	Question	Response 1		Response 2		Response 3		Response 4		Total
1	Is macrophyte growth an issue of concern in your region?	yes	6	invasive species only	1	a few sites only	1	no	7	15
2	are macrophytes monitored regularly?	yes	5	for habitat only	5	some sites only	1	no	4	15
3	Main objective(s) for macrophyte monitoring*	effects on flow	4	effects on dissolved oxygen	3	interpret invertebrate data	2	invasive species	3	20
4	Frequency	annual	5	monthly*	2	N/A	9			16
5	Timing	summer	6	N/A or ND	9					15
6	Timed with other biomonitoring?	yes	4	no	2	N/A or ND	9			15
7	field protocol	SEV*	1	SHAP*	5?	WRC* protocol	2	other	4	7
8	categories recorded	emergent/submerged/surface reaching	5	native/introduced	1	species ID	1	presence/absence	1	8
9	QA/QC	Responses included training in species identification								

Notes

Line 3: in addition: general habitat quality (4 councils), particle retention (2 councils), aesthetics (1 council), link with nutrients (1 council). *most councils have more than one objective, so total is greater than 15.

Line 4: 1 council does both monthly and annual.

Line 7: SEV = Stream Ecological Valuation (Storey et al. 2011), SHAP = Stream habitat Assessment Protocols (Harding et al. 2009), WRC = Waikato Regional Council (Collier et al. 2007).

Table 4-9: Summary of recommended protocols for macrophyte monitoring.

Line	Parameter	Recommendation	Rationale/notes
1	Frequency	annual	
2	Timing	With invertebrate/periphyton biomonitoring or physical habitat assessments	
3	Field protocol	Estimate % cover of submerged, surface reaching and emergent macrophytes within a 1 m-wide band across 4 evenly-spaced transects (based on Collier et al. 2007)	As recommended in the new NZ instream plant and nutrient guidelines (Matheson et al. in prep.). Could use same transects as periphyton monitoring, requiring little extra effort.
4	Indices for reporting	Macrophyte channel clogginess (based on Collier et al. 2007)	As recommended by Matheson et al. (in prep.)

Table 4-10: Field data sheet for assessing % cover of aquatic macrophytes (based on Collier et al. 2007).

Transect	Wetted width (m)	Channel width (m)	Vegetation cover (as % of 1 m band across wetted width)				
			Total plants	Emergent plants	Surface-reaching plants	Below-surface plants	Average height of below-surface plants as proportion of channel depth
1							
2							
3							
4							

With these data, the “Macrophyte Channel Clogginess” (MCC) is calculated:

$$= \frac{\sum\{(\%emergent + \%surface\ reaching) + (\%below\ surface\ \times proportion\ of\ channel\ depth\ occupied)\}}{number\ of\ transects}$$

4.5 Physical habitat assessment

4.5.1 Current practice

Thirteen of the 15 councils assess stream physical habitat, and 11 of these monitor at least once per year (Table 4-11). All except one council assesses physical habitat at the same time as biomonitoring. A wide variety of variables are measured, though some, such as % shade and % cover of different substrate types, are almost universal. However, some of the common variables may be measured in different ways, because only five councils base their methods on the Stream Habitat Assessment Protocols (Harding et al. 2009). Physical habitat data is probably used mainly to interpret patterns in biomonitoring data. Thus, national assessments of physical habitat condition are not possible at present, and achieving these would require significant changes to current protocols, particularly the development of a scoring system. However, for a few variables, such as % shade, comparable data from different regions may be available.

4.5.2 Recommendations

The Variables Expert Panel workshops (October and November 2011) noted the importance of assessing stream physical habitat. However, little progress was made towards recommending protocols. Thus our main recommendation is to convene a physical habitat working group focused on developing suitable protocols.

Two systems in current use by councils provide potential frameworks for a ‘national’ monitoring protocol.

Protocol P2 of the Stream Habitat Assessment Protocols (SHAP; Harding et al. 2009) is currently used by five councils for SoE monitoring. SHAP P2 characterises a stream reach by measuring a range of hydrological, morphological, instream habitat and riparian characteristics. The protocol involves measuring these variables quantitatively, which has the advantage that the values of individual attributes can be extracted for a wide range of purposes, including creating stream/river typologies, modelling, and interpreting biological data. The disadvantage of this approach is that measurements are made without reference to natural or pristine condition, therefore for many attributes for which the pristine condition is not obvious, stream health cannot be easily calculated. In order to develop a scoring system, values representing the pristine condition must be defined for each stream type, and this may require considerable work. Also, SHAP P2 includes a large number of attributes, which would be onerous to monitor frequently and indefinitely, so a subset of key variables would have to be selected for the purpose of routine, on-going SoE monitoring.

The Stream Ecological Valuation (SEV; Storey et al. 2011) is currently used by Auckland Council for SoE monitoring. SEV quantifies stream health by assessing the ability of a stream reach to perform a number of ecological functions. To calculate function scores, SEV combines different attributes, each of which indicates a different type of human alteration to

the natural condition. The advantage of this system is that it is designed to measure deviations from the natural condition, i.e., it specifically measures degree of degradation using a built-in scoring system. The overall score can be broken down into individual function scores and then further into "variable" (attribute) scores to examine the underlying data. However, the "variable" scores themselves are often composites of several environmental characteristics, and in these cases quantitative measures of the individual attributes cannot be extracted. For example, % cover of large woody debris contributes to the bully spawning and channel modification "variables" within the fish spawning and natural flow regime functions, respectively. While scores of those functions and variables are available, the % cover of woody debris is not. Therefore, while SEV is useful for the purpose of reporting on the ecological health of streams, its constituent data may not be suitable for the other purposes described above for SHAP P2.

Other habitat assessment systems available, or in current use by councils, have characteristics in between SHAP and SEV. For example, Waikato Regional Council uses a system based on the USEPA Rapid Bioassessment protocols (Barbour et al. 1999). This system measures "state" rather than "function" variables, but as with SEV, the variables measure deviations from natural condition. Thus they are conducive to scoring but some may not be as versatile for uses other than reporting on ecological health.

We recommend a working group to select or tailor a suitable habitat assessment system from among the available protocols. The working group might focus on the relative merits of a protocol involving simple quantitative measurements of individual habitat variables *versus* one that specifically measures human alterations to the natural condition, or a combination of these.

Table 4-11: Physical Habitat monitoring protocols: current practice among 15 councils.

Line	Question	Response 1		Response 2		Response 3		Response 4		Total
1	frequency	annual	10	less than annual	2	quarterly	1	none	2	15
2	timing	summer	12	quarterly	1	N/A	2			15
3	Timed with	biomonitoring	12	WQ	0	neither	1	N/A	2	15
4	hydrology/morphology parameters measured	wetted width	7	% riffle/run/pool	7	discharge	5	depth profile	4	23
5	instream parameters measured	% substrate types	11	% organic matter	9	% fish cover	7	substrate compaction	3	30
6	riparian parameters measured	% shade	11	dominant veg type	9	bank stability	9	stock access	8	37
7	protocol based on	SHAP*	5	USEPA RBP*	2	internal	4	SEV*	1	12

Notes

Line 4: In addition: mean velocity (4 councils), channel XS shape (4 councils), undercuts (3 councils), active/non-veg channel width (2 councils), channel alteration (2 councils), floodplain XS shape (2 councils), pool max depth, crest depth, sediment depth (2 councils), mean depth (1 council), frequency of riffles (1 council), channel sinuosity (1 council), pool variability (1 council).

Line 5: In addition: substrate embeddedness (3 councils), sediment deposition (3 councils), obstructions to flow (3 councils), habitat diversity (2 councils), scouring/depositional zone area (2 councils), armouring (1 council), resuspendible sediment (1 council).

Line 6: In addition: buffer width (8 councils), longitudinal intactness (7 councils), soil permeability (5 councils), groundcover (4 councils), rills/channels (3 councils), soil denitrification potential (3 councils), land slope (3 councils), fencing (1 council), adjacent land use (1 council).

Line 7: *SHAP = Stream habitat Assessment Protocols (Harding et al. 2009), USEPA RBP = United States Environmental Protection Agency Rapid Bioassessment Protocols (Barbour et al. 1999), SEV = Stream Ecological Valuation (Storey et al. 2011).

4.6 Hydrology

4.6.1 Current practice

It is unclear what flow indices councils calculate currently, but indices of human alterations to flow are probably not included.

4.6.2 Recommendations

The hydrological regime of a stream is a vital attribute affecting its ecological health, and is incorporated in a number of stream health indices (e.g., the Victoria Index of Stream Condition; State of Victoria DSE 2005). Although the Variables Expert Panel recognised the importance of hydrology, it deferred making recommendations on protocols to future discussions. Therefore, as for habitat assessment and freshwater fish, **we recommend a working group to research and agree on suitable protocols for assessing the condition of stream hydrology**.

Here we note two different approaches that may inform and guide the working group in their choice of protocol. The Stream Ecological Valuation (Storey et al. 2011) scores four "hydraulic functions" of streams indirectly, by assessing alterations to physical parameters of a stream reach that will result from altered hydrology. Attributes include the presence of natural and artificial obstacles to flow such as large wood, boulders, macrophytes or culverts, straightening or lining of the channel, and outlets of stormwater pipes. The Victoria Index of Stream Condition (State of Victoria DSE, 2005) measures alterations to natural hydrology directly, by extracting a number of statistics from flow records. These statistics include a flow variability index, a high flow, low flow and zero flow index and a seasonality index. This approach requires both current actual hydrological records and estimates of a natural or unimpacted record for the same site. The latter can be difficult to obtain, but records from nearby undeveloped catchments and/or streamflow models might be used to derive the required flow distribution.

4.7 Ecosystem processes

4.7.1 Current practice

Ten of the fifteen councils expressed a definite interest in monitoring ecosystem processes such as Gross Primary Productivity (GPP), Ecosystem Respiration (ER) and nutrient processing (Table 4-12). However, only two councils believe they currently have the capacity to do such monitoring. Councils perceive three key barriers to adding ecosystem processes to their SoE monitoring: the cost of equipment, staff time involved in deploying and retrieving equipment, and training in use of the methods. The only council that currently measures ecosystem processes identified equipment problems as limiting.

4.7.2 Recommendations

The Variables Expert Panel recommended that ecosystem processes not be included in national environmental monitoring. However some of the current barriers to ecosystem process monitoring may be overcome in the future if the cost of equipment goes down, its performance in the field improves, and once methods have been more thoroughly validated in New Zealand.

Table 4-12: Monitoring of ecosystem processes (includes gross primary production, ecosystem respiration and nutrient processing).

Line	Question	Response 1		Response 2		Response 3		Response 4		Total
1	Is there interest in your council to monitor ecosystem processes?	yes	10	maybe	2	no	3			15
2	Does your council have the capacity to measure ecosystem processes?	yes	2	partly	4	no	8	don't know	1	15
3	What are the main limitations to measuring ecosystem processes?*	cost of equipment	10	staff time	9	training	7	equipment performance/security	2	28

Notes

Line 3: *most councils listed more than one limitation, hence total is greater than 15.

5 Lake Monitoring Protocols

In order to determine long-term changes in lakes driven by changing land use or by changing climate, long term records with stable protocols for the sampling of water quality are required. For instance, New Zealand's air temperatures have warmed by about 1℃ since 1900 (Mullan 2008), but expected consequent warming (and strengthening stratification), and its effects on the lake ecosystem, have yet to be demonstrated in any New Zealand lake. Detection of changes in water quality resulting from changes in land use is usually easier because these changes tend to be more rapid, but still demands rigorously maintained, stable sampling protocols. In order to examine whether consistency of sampling of water quality in lakes may be improved, both at the national level and for individual lakes, we compare the procedures used by each of the (eleven) regional councils that currently monitor water quality in lakes. The following summarizes responses of the regional councils to a questionaire (Appendix D) about protocols for lake sampling.

LakeSPI (Lake Submerged Plant Index, based on SCUBA survey of submerged aquatic plants in lakes; de Winton et al. 2012) is considered complementary of water quality monitoring in lakes (Davies-Colley et al. 2012; Table 1-2). LakeSPI requires special expertise applied at comparatively low frequency of 1-3 years in lakes subject to change or invasion, and about every 10 years in lakes considered stable (Davies-Colley et al. 2012). So most regions are unlikely to do their own LakeSPI monitoring (although that might be feasible in future – Mary de Winton pers. comm.) not least to main quality assurance (QA). Therefore, protocols for LakeSPI monitoring are not discussed here.

5.1 Comparison of lake monitoring protocols between regions

The questionnaire responses clearly indicate that regional lake sampling protocols are currently appreciably diverse in New Zealand. This inevitably compromises our ability to report nationally (e.g., Verburg et al. 2010).

Sampling frequency is either monthly (5 regions), bimonthly (3 regions), quarterly (2 regions) or only in the summer months, December to April (1 region). One of the regions that samples lakes monthly operates 'rolling' sites, by monitoring one of four groups of lakes for 2 years in a row, after which another of the four groups is monitored for the next two years, and so on – such that each group is re-visited for 2 years after a gap of 6 years. Lake sampling is usually carried out from early morning to mid-afternoon. In all but two regions lakes are sampled in the same order on each sampling day so that lakes are generally sampled at more or less the same time on each sampling occasion.

Sampling is done from boats in all but one region. Lakes are accessed by helicopter in two regions, with the sampling done from the helicopter in one while an inflatable boat, carried with the helicopter and inflated at each lake with compressed air from a dive tank, is used by the other. When bad weather occurs, sampling is usually delayed until suitable conditions return. In three regions sampling is attempted anyway unless bad weather is deemed to make lake access unsafe.

Sampling sites are usually near the deepest point (or include a deep site if multiple sites are sampled in one lake) except in one region where shore-based sampling is conducted for

ease of access. In another region, in one particular lake, sampling is done at the lake edge. In six regions at least some of the lakes are sampled at more than 1 site.

The water level is usually recorded during lake sampling except in two regions. In regions where water levels are recorded as part of the sampling this was done in all monitored lakes (5 regions) or in a selection of lakes where staff gauges where present (4 regions).

All but two of the regions also collect water samples in some of the inflowing rivers, usually as part of another programme, but not at the same time as the lake sampling and not for all monitored lakes. Sampling frequency of the inflows tends to be lower than for the lake sampling. Up to 11 inflows are sampled at up to 20 sites.

The sampling depth within the epilimnion is quite variable between regions and not necessarily because of the differences in epilimnion depths between lakes in different regions. In two regions the epilimnion is sampled at the 'surface' (not further specified), in one region at 20 cm depth, in another region at 1 m depth and in three regions depth-integrated samples are collected between 0-10 m depth (1 region), between 0-25 m depth (1 region), and across a depth range equal to the average epilimnion depth (1 region). Four regions cite the Burns et al. (2000) protocol for epilimnion sampling in stratified lakes. One of these regions mixes two water samples collected at ¼ and ¾ of the epilimnion depth, the two other regions collect a sample at the middle of the epilimnion. In the fourth region the epilimnion sample is a mixture of samples from 3 or 4 different depths (the protocol description and responses to the questionnaire were contradictory regarding the number of depths but it includes at least water collected at ¼, ½ and ¾ of the epilimnion depth) depending on the depth of the thermocline. In these four regions the epilimnion sampling depth therefore may vary from month to month. It appears that none of the four regions which cite Burns et al. (2000) for the selection of the epilimnion sampling depths actually follows the Burns recommendations exactly.

In the region where the epilimnion sample is a mixture of water from ¼, ½ and ¾ of the epilimnion depth, an additional chlorophyll *a* sample is collected as a 10 m depth-integrated sample. In 7 regions samples are taken at least at 1 additional depth below the epilimnion (usually 1 or 2 depths additional to the surface sample), depending on the lake, its depth and whether it is stratified.

In two regions acid-washed bottles are used. In all regions samples are placed immediately in a chilly bin with ice. Nine regions pre-rinsed bottles with lake water (usually twice) except when sampling by helicopter. Water samples are usually delivered to the laboratory within 24 hours. Only 2 regions indicated that sometimes the samples are late and are sometimes refrigerated if they cannot be delivered to a laboratory within 24 hours. A field sampling sheet is always completed for each site visited, usually a hardcopy but sometimes electronically.

In eight regions temperature and dissolved oxygen (DO) depth profiles are recorded. Only 1 regional council uses a continuous profiler (a Seabird, 0.25 second sampling frequency) to collect depth profiles of temperature and dissolved oxygen. Another council has used a Eureka Manta D-opto multi-probe (3 second sampling frequency) in the past and is currently looking for a replacement. Two more regional councils use continuous profilers (Hydrolab and YSI Castaway) but only in surface water, therefore no full depth profiles are taken. In the other seven regions temperature and dissolved oxygen are measured at discrete depths in

recovered water samples, usually with YSI probes, and in one region with a Hach DO meter, while another uses a WTW Oxi 197 instrument. Depending on the type of instrument, temperature, DO and conductivity probes, if used, are calibrated at least once each year.

Secchi depth is recorded in 10 regions. In one of these 10 regions the use of a helicopter for the lake sampling makes recording Secchi depth difficult (sampling is done from the helicopter) and Secchi depth is only recorded in those lakes where the sampling is not done by helicopter. Viewers are used in four regions when recording Secchi depth and four councils specified that Secchi depth is recorded on the sunny side of the boat.

5.1.1 Sampling depths as recommended by Burns et al. (2000)

In four regions sampling depths are decided more or less, but not quite, as recommended by Burns et al. (2000). For stratified lakes, Burns et al. (2000) recommended mixing equal volumes from depths of 20 cm, ¼, ½ and ¾ of the epilimnion depth. These four sampling depths in the epilimnion must be newly determined for each sampling event, depending on the epilimnion depth. The epilimnion depth is determined by eye from a plot of temperature versus depth, and can therefore be somewhat arbitrary, especially when there is no obvious abrupt change with depth, from a stable temperature in a fully mixed and therefore isothermal surface layer to a rapidly decreasing temperature in the thermocline, as is often the case. Furthermore, diurnal thermoclines may be mistaken for seasonal thermoclines in lakes, especially when sampled in the middle of the day or in the afternoon. Whether a lake is deemed stratified is determined by the temperature profile as well, with a larger than 3 ℃ difference between top and bottom taken, operationally, to indicate a stratified water column. When the top to bottom temperature difference is less than 3℃ the lake is considered isothermal and Burns et al. (2000) recommends collecting separate water samples at ¼ and ¾ of the full lake depth. One quarter of the full lake depth may be well below the usual mean sampling depth of the epilimnion. However, when the top to bottom temperature difference is less than 3℃ there may still be substantial differences in nutrient concentrations between top and bottom water layers. Nearly isothermal lake waters are not necessarily fully mixed. In fact, the water column in lakes is rarely fully mixed except in shallow lakes. As a result, nutrient and chlorophyll *a* concentrations are incomparable between months, if large differences in the sampled depth occur because in some months the lake is considered isothermal and stratified in the other months. Therefore we recommend a change from the Burns (2000) protocol, with sampling to be carried out at specific depths which are lake-specific, but do not vary with time in any one lake.

5.1.2 Reporting of the lake water quality results

The actual reporting of the water quality results and the calculation of the compound TLI (Trophic Level Index, an index to compare water quality over time and between lakes) was not part of the protocol questionnaire and was not discussed in the materials provided by the councils. Burns et al. (1999; 2000) described the methods of the calculation of the TLI for lake water quality.

The TLI method logarithmically transforms total nitrogen (TN), total phosphorus (TP), Secchi disk depth (ZSD) and chlorophyll *a* (Chla) data to give a trophic level score for each parameter on the same scale. The formulae for these four component indices (abbreviated as TLn, TLp, TLs and TLc respectively) are:

- TLn = -3.61 + 3.01log10 (TN).
- TLp = 0.218 + 2.92log10 (TP).
- TLs = 5.10 + 2.27log10 (1/ZSD – 1/40).
- TLc = 2.22 + 2.54log10 (Chla).

The method then averages the four component scores into an index of trophic condition, the TLI. Generally the TLI is calculated in two ways, as TLI4 (using TN, TP, Chl*a* and Secchi data) and as TLI3 (without using Secchi disc data). TLI3 is used to allow comparison between a larger number of lakes, because there are a number of monitored lakes for which there are no Secchi depth data, and in other lakes Secchi depth is controlled mainly by sediments or coloured organic matter rather than phytoplankton. Trophic state classes are defined based on the TLI range (Verburg et al. 2010)

A logarithmic transformation is part of the TLI calculation (Burns et al. 1999; 2000; 2005). Burns et al. (1999; 2005) or any of the other publications describing the TLI method, did not mention that individual data of TP, TN, Secchi depth and chlorophyll *a* should be used for the calculation of the TLI. On the other hand, they also did not suggest that some sort of pre-processing of the data such as time averaging should be used prior to applying the equations above to calculate TLI. Burns (1999; 2005), however, did describe how they arrived at their equations which are used to calculate the TLI. The TLI formulations, in particular the linear functions of the log transformed variables, were arrived at by fitting to annual mean input data for TN, TP, chlorophyll *a* and Secchi depth, i.e., data averaged *before* log-transformation. It appears that this may have been taken by some users to mean that in the general application of the TLI calculation to their data also annual mean input data should be used. In other words, from TLI results reported by various councils it seems that TN, TP, chlorophyll *a* concentrations and Secchi depth are averaged over a year before calculating the annual value of the TLI.

However, TN, TP, chlorophyll *a* and Secchi depth data are typically highly positive-skewed (Sorrell 2006) (i.e., their distributions have an asymmetric tail extending toward higher values, with skewness being often highest in chlorophyll *a* data). Therefore a logarithmic transformation is desirable to achieve data symmetry *before* calculating averages used to compare between lakes (Sorrell 2006). The TLI should be calculated for every sampling event before calculating the mean annual TLI for a lake and the annual mean TLI should not be calculated using mean untransformed values of TN, TP and chlorophyll *a* concentrations and Secchi depth. Calculating annual mean values for TN, TP, chlorophyll *a* concentrations and Secchi depth before log-transformation and before calculating the TLI defeats the purpose of the log-transformation and produces values of TLI that are different from TLI values calculated from data that are individually log-transformed. The differences in the results between the two methods are sufficiently large to require a recommendation aligning TLI calculations between regions. Annual means overestimate TLI on average by about 7% (but this may vary widely between lakes and years; Hudson et al. 2011).

In the example in Table 5-1, TLI for 2007-2008 is overestimated by 10% when annual mean TP, TN and chlorophyll *a* concentrations are used.

It must also be noted that TLn is negative when TN < 15.82 mg m^{-3}, TLp is negative when TP < 0.84 mg m^{-3}, and TLc is negative when chlorophyll *a* < 0.13 mg m^{-3}. Such low concentrations, resulting in negative TLI components, are occasionally recorded (e.g., negative values occur twice for TLp in in Table 5-1) and should not be discarded. The present detection limits at the NIWA water quality laboratory are 10 mg m^{-3} for TN, 1 mg m^{-3} for TP and 0.1 mg m^{-3} for chlorophyll *a*. The concentration of TP in Table 5-1 was therefore probably recorded as half the detection limit. The detection limits for TN and chlorophyll *a* are lower than the values that result in negative TLI components.

Table 5-1: An example illustrating the calculation of TLI3, using data from an existing lake.

Date	TN	TP	CHLA	logTN	Log TP	Log Chla	TLn	TLp	TLc	TLI
21/08/2007	189.5	11	2.7	2.28	1.04	0.43	3.25	3.26	3.32	3.27
18/09/2007	126.5	2	0.8	2.10	0.30	-0.10	2.72	1.10	1.97	1.93
16/10/2007	162.5	0.5	1.2	2.21	-0.30	0.08	3.04	-0.66	2.42	1.60
13/11/2007	139	13	0.7	2.14	1.11	-0.15	2.84	3.47	1.83	2.71
11/12/2007	339.5	11	0.9	2.53	1.04	-0.05	4.01	3.26	2.10	3.12
7/01/2008	233.5	4	0.6	2.37	0.60	-0.22	3.52	1.98	1.66	2.38
19/02/2008	150.5	4	1	2.18	0.60	0.00	2.94	1.98	2.22	2.38
18/03/2008	145	0.5	0.6	2.16	-0.30	-0.22	2.90	-0.66	1.66	1.30
21/04/2008	145	5	1.2	2.16	0.70	0.08	2.90	2.26	2.42	2.53
20/05/2008	307.5	10	1.4	2.49	1.00	0.15	3.88	3.14	2.59	3.20
20/06/2008	171	4	1.6	2.23	0.60	0.20	3.11	1.98	2.74	2.61
									Mean TLI:	**2.46**
Incorrect results are achieved by using annual means:										
2007/2008	191.8	5.9	1.2	2.28	0.77	0.06	3.26	2.47	2.38	**2.70**

Also the final TLI can be negative. For example, where TP, TN and chlorophyll a are exactly at the detection limits the TLI is -0.2. At half of detection limits TLI is -1.1. Lakes with persistent negative TLI exist in New Zealand, but we don't monitor any of those. In theory, TLI can range from infinitely negative to a maximum that is only limited by the amount of P, N and chlorophyll a that could theoretically fit in one unit volume of water. However, in reality, monitored lakes in New Zealand are usually found in 6 bands of trophic state ranging from microtrophic (TLI 1-2) to hypertrophic (TLI >6, but lakes seldom exceed TLI = 7).

5.2 Lake monitoring recommendations

In general good field sampling practise (rinsing of bottles with lake water, delivery to laboratory within 24 hours, chilled and dark storage, use of a viewer for Secchi depth with observations on the sunny side of the boat) is followed in each region. Some protocol aspects were probably considered so obvious that they were not always specified in the questionnaire responses, nor in the copious protocol materials provided.) However, there were major differences between regions regarding where (within the lake) and when and which lake samples are collected. Here we recommend national methods for various aspects of the lake sampling protocol.

- Monthly sampling of lake water quality is recommended. Monthly is preferable for time-trend analysis, so is strongly recommended for 'sentinel' lakes for climate change effects, and lakes threatened by catchment land use changes. Bimonthly or other less frequent sampling is not ideal for tracking environmental changes in lakes, and should only be contemplated where a lake would otherwise not be sampled at all.
- Lake sampling should be done from a boat. Helicopters may be used to access the lake, but not as a sampling platform. An inflatable boat carried by the helicopter should be used to carry out the lake sampling. This permits Secchi depth observations and full depth profiles of temperature and dissolved oxygen to be obtained, as well as being preferable for sampling (better depth control and permitting multiple rinsing of sample containers before collection). Where lakes are too large for an inflatable dinghy to be convenient or safe, access by vehicle with a larger vessel is a preferred, although more expensive, option. Float planes (as used in some other countries) are an alternative to helicopters that allow sampling in the same fashion as from a vessel.
- Sampling should be done remote from the shore, preferably where the lake is deepest. Additional nearshore sampling sites can be useful, for instance to assess the effect of local runoff from cities into large lakes or to gauge littoral water quality as experienced by recreational lake users, but there should be at least one deep 'reference' site.
- Constant sampling depths are preferable, irrespective of thermal conditions. The depth of the sample collected nearest the surface of the lake, whether stratified or unstratified, may vary by lake, depending on lake depth and average epilimnion depth, but should not vary with time. The sampling should be at least sufficiently below the surface, to avoid any material floating on the surface. The average depth of the epilimnion would be a good choice to sample the surface layer.
- Where depth-integrated samples across fixed depth ranges or samples at discrete fixed depths in the epilimnion are collected, sampling should be continued as in the past, in order to preserve consistency of sampling for individual lakes.
- A full depth profile of temperature and dissolved oxygen should always be recorded.

- Secchi depth should always be recorded – whether or not Secchi data are used in TLI calculation. Observations should be made using an underwater viewer (from the sunny side of the boat to avoid the boat shadow under clear sun), following the protocols of Smith (2001). Lighting conditions during Secchi observations should be noted on the field form. Laboratory turbidity is a valuable supporting variable to back-up Secchi observations.
- Water samples can be refrigerated (at $4^{o}C$) if there is any delay in sample delivery (say up to 48 hrs), but should never be frozen.
- Lake levels should be recorded in all monitored lakes. (A level gauge may need to be installed in some monitored lakes).
- For reporting, calculate the TLI index for each sampling occasion separately (from log-transformed TP, TN, Chla and Secchi data), *before* averaging the TLI values into an annual mean TLI.
- Metadata files should be collated on all monitored lakes and lake sites, including pragmatic aspects such as lake access, boat launching ramps and wind wave exposure. Metadata should extend to -
  - The volume of each lake for nutrient budgetting (for many monitored lakes this information is not available).
  - Lake outflow rates (in $m^3 s^{-1}$) are also needed for nutrient budget analysis.

Some of the above-mentioned protocols will require a change by some councils, at least in matters of detail, in order to achieve national consistency. However, a change in protocols always comes with risks of a step-change in long-term monitoring records which can confound trend detection, so great care will need to be exercised with moving towards national protocols. We recommend that a **specialist working group** of limnologists be convened, to consider the above detailed recommendations in turn and consider whether data continuity (and thus the ability to detect trends) is likely to be compromised in particular regions and for particular lakes. Special funding may be needed for a transitional 'overlap' period wherein both 'old' and 'new' protocols are followed in particular lakes of national importance.

The opportunity should also be taken with a limnological workshop to consider quality assurance (QA) issues in lake monitoring, and recommendations given below (Chapter 6) for a 'universal' field form and a field auditing programme.

6 Quality Assurance

6.1 Principles of Quality Assurance

Quality assurance (QA) and quality control (QC) refer to a range of principles and practices which, if followed strictly, will yield 'dependable' monitoring data (APHA 2005; Ward et al. 1990; Weiner 2008). For current purposes we interpret 'dependable' as meaning *accurate* – that is free of bias and of sufficiently high precision for the purpose of state-of-environment reporting. Time-trend analysis is particularly 'demanding' as regards freedom from bias and consistency over time. A well-designed quality assurance programme eliminates bias so far as possible.

Quality assurance has been referred to as "monitoring of the monitoring programme" – to ensure it performs as designed (Ward et al. 1990: p 125). There are many components to a QA programme, ranging from training and certification of monitoring staff competency through to adherence to international quality standards (Ward et al. 1990; Weiner 2008). However, probably the most important aspect of QA is the deliberate *duplication or replication* of measurements. This has two major applications:

- Duplicated measurements that agree confer confidence in the data in the *whole* programme including measurements that are *not* duplicated. Conversely, if duplicate measurements do not agree, the monitoring supervisor is alerted to a problem and steps can be taken to identify the source of the bias.
- The statistics of deviations provide a measure of precision, usually expressed as coefficient of variation of replicates (%). If, furthermore, the duplication is done *independently*, the statistics of deviation give an indication of (joint) accuracy (in the absence of an absolute point of comparison).

Duplication of effort is usually regarded as a bad thing. But in SoE monitoring (a limited amount of) duplication is an essential step in quality-assuring monitoring data. And the duplication should be done not only by the same agency (providing a measure of precision), but by a *different agency working independently* so as to give an indication of accuracy, including possible systematic error (otherwise known as bias).

The single most important principle of QA in monitoring can be stated as follows: **a subset of monitoring measurements should be *independently* duplicated so as to confer confidence in all data.** That means that a subset of water monitoring measurements in each region should be duplicated by an *independent* agency. As a guide, following US EPA (1998), we suggest that perhaps 5% of SoE measurements should be independently duplicated so as to confer confidence in all data[4].

It is convenient to distinguish field and laboratory components of QA (e.g., Weiner 2008). Field QA is intended to (1) obtain accurate field measurements (e.g., of DO and visual clarity) and (2) deliver to the laboratory 'dependable' samples that are representative of the water body being monitored and field conditions (i.e., the sample properties are representative of

[4] For example, established practice is that 10% of macro-invertebrate samples should be independently analysed as a QA measure, and 10% of analyses are typically re-run by reputable water quality testing laboratories.

the water mass as a whole and remain unchanged during transit to the laboratory). Laboratory QA is designed to produce accurate measurements of those sample properties. Here we emphasise field QA aspects, because most reputable laboratories have internal QA/QC procedures, such as duplicating a subset of all analysed samples, maintaining QC curves, and engaging in inter-laboratory comparison rounds (as indeed they may be required to so as to maintain IANZ laboratory registration). Overall, it seems reasonable to us that field sampling is a greater QA concern than laboratory analysis – as has been asserted by Robinson (2010). However, note that some attention was given to laboratory issues in Chapter 2, particularly as regards nutrient partitioning and data 'censoring'.

6.2 Summary of current regional QA

It is apparent from the responses to the QA questions (Appendix B-D) that most regional councils are aware of the need for QA, but only in certain aspects of regional monitoring is formal QA built in. In the river and lake water quality sub-domains, no councils appear to be routinely sending duplicate samples (unknowns) in each sample batch to laboratories – as is recommended in Section 6.4. Unknown duplicates ('blind' samples) would provide for a check on stated laboratory precision and is to be recommended, say at least one duplicate per sample batch. Returns on the QA questions (Appendices) sometimes refer to training of field workers and international standards (ISO) principles. These activities are laudable for generally encouraging a 'culture' of quality. However, they do not of themselves *assure* data accuracy.

In the river biomonitoring sub-domain, QA protocols for laboratory procedures (sorting and identification for macro-invertebrates, chlorophyll-*a* and ash-free dry mass analyses for periphyton) are well described in sampling protocol handbooks (Stark et al. 2001 and Biggs and Kilroy 2000, respectively). With few exceptions, councils are following the QA procedures appropriate to the analyses they are performing. However, QA protocols for field sampling and on-site visual assessments are not well-defined in the published handbooks. Only one council indicated that it follows a specified QA procedure for macro-invertebrate sampling, and none indicated a specific QA procedure for periphyton, fish, macrophyte or physical habitat field assessments.

There is a clear need for major upgrading of QA within regional monitoring, particularly as regards fieldwork. This goes for all three sub-domains of river water quality, river bio-monitoring and lake monitoring. It is not sufficient to rely on laboratory QA because no amount of laboratory finesse can correct for biased or unrepresentative sampling. Nor can laboratory measurements back up some field instrumental measurements on variables that cannot be preserved, such as temperature and pH. However, laboratory measurements *can* 'backup' certain field measurements including visual clarity (Section 6.4).

More generally, based on the questionnaire survey, we think there is a need to comprehensively upgrade regional QA in SoE monitoring of waters. An 'overview' of regional monitoring by some centralised agency, say, either SWIM or MfE or both, is needed, probably centred on 'audits' of water monitoring. Short of formal auditing (considered further below), it is highly desirable that regional council staff engage frequently in joint field exercises with staff of neighbouring councils and with NIWA field staff (as is also considered below).

6.3 Quality issues in SoE monitoring

6.3.1 River and lake water quality

Errors in field work for river and lake water quality monitoring may arise both with field instruments and with water sampling. Laboratory errors in water quality work are generally more easily detected and traced (Robinson 2010), and can sometimes be corrected by a repeat analysis on preserved samples, which is why we emphasise fieldwork here. Both field instrumental measurements and water sample collection are potential sources of error.

Great care must be taken to ensure that all important field measurements and supporting observations are taken. **A field form** (or its electronic equivalent) that prompts for all important data is needed. Davies-Colley et al. (2011a) recommend the development of 'national' field forms that would be suitable for use by all councils, and we reiterate that recommendation here.

Field instrumental measurements must be taken with particular care to ensure accuracy. It is all too easy to simply dip a probe into water and note down the numbers without any checks that those results are meaningful. The (base or laboratory) calibration of field instruments should be checked in the field with suitable standards - such as a standard buffer for pH. Visual clarity observation is an 'absolute' measurement, so no standard is applicable and even more care is needed here than with calibration of instruments[5].

Water samples must be representative of the water mass being sampled. In rivers that means flowing water in riffles or runs must be accessed, rather than pools which may contain 'dead zones'. In lakes the epilimnion must be reliably sampled. Integrity of water samples (Weiner (2008: p353), involves the following considerations.

- **Sample contamination must be avoided**. Sample containers and collection vessels (e.g., buckets, van Dorn samplers) must be scrupulously clean, but in any case, these vessels should be washed twice in ambient water before sample collection. Care should be taken that sample bottles are stoppered correctly and are water-tight.
- **Samples must be unambiguously identified**, with (at least) water body name, site code, date and time (with time qualification, i.e., NZST or NZDT) of sampling. An alternative approach, with merit in terms of field simplicity (Peter Robinson, Hill Laboratories, pers. comm.), is to use a unique ID number on each bottle and record this ID number along with the other needed information (site, date, time) on the field form. 'Chain-of-custody' (CoC) protocols for compliance monitoring are probably overly 'bureaucratic' for SoE samples, but some general principles of CoC are worth considering.

[5] In principle visibility could be substituted by measurements with a beam transmissometer, but these are expensive instruments more suited to research or special investigations than routine SoE monitoring (Davies-Colley and Smith 2001). Visibility measurements are sometimes perceived as 'rough' and appreciably subjective, but both optical theory and replicate measurements suggest good repeatability and minimal subjectivity. For example, Smith (2001) reported data showing excellent precision in Secchi depth observations taken with good protocols. The black disc method has been well proven *versus* beam transmissometers – Zanevald & Pegau 2003; http://www.opticsexpress.org/abstract.cfm?URI=OPEX-11-23-2997).

- **Samples must be properly preserved.** Samples for the core variable set recommended by Davies-Colley et al. (2012) (Table 1) should be chilled to 4^{o}C or less (but not frozen) to slow biochemical processes that could otherwise result in changed sample properties. Samples should be stored in the dark to minimise photochemical degradation. In practice, water samples should promptly (within a few minutes) be placed in ice chests with slush ice. Water samples should not be left exposed to the sun on the river bank or boat deck while other field tasks are completed.
- **Sample transit time to the laboratory should be minimized**, and, in general overnight couriers should be used so that samples are, ideally, only 24 hrs and at most 48 hours old at time of laboratory processing.

6.3.2 River biomonitoring

Opportunities for error are somewhat different in river biomonitoring than in water quality monitoring.

- There is high natural spatial variability in macro-invertebrate communities, periphyton and macrophytes. Therefore, repeated sampling of the same stream reach will produce slightly different results, even if the field staff are using good technique and randomly locating their replicates. The level of precision that results from such natural variability is rarely quantified in SoE monitoring, and is not traceable when replicate samples from a site are pooled into one composite sample (as is usual practice).
- Fish sampling has a particular set of issues. Sampling technique and level of expertise/experience can make an appreciable difference to the number and types of fish caught. It can also make a difference in correct identification of fish, especially juveniles. Unlike macro-invertebrates which are collected and can be sent to another biologist for checking, fish are usually identified in the field, thus identifications are often made quickly, sometimes under non-ideal conditions, and without opportunity for later checking by experts.
- Visual assessments for periphyton, macrophytes and physical habitat involve a degree of subjectivity that introduces both systematic and random differences (bias and imprecision, respectively) between observers. Error can be introduced when transects, periphyton scrapes, etc., are located non-randomly, or when an observer is inaccurate in estimating attributes such as % cover. Potentially much larger errors can occur if the definition of a particular attribute is misunderstood. The amount of inter-observer variability can be reduced with experience.

6.4 General QA recommendations

6.4.1 Laboratory backup of field water quality measurements

Laboratory measurements on water samples can (sometimes) be repeated if there are grounds for thinking bias has arisen. For example, nutrient sub-samples are routinely frozen, so providing a backup should repeat analysis be indicated. In contrast to laboratory

measurements, there is (usually) no opportunity for repeating field measurements such as temperature. However, certain field measurements can, and should be, backed up, where possible.

An important example is as regards nephelometric turbidity backup in the laboratory of field measurement of visual clarity by the black disc method or, traditionally in lakes, by the Secchi depth method. For a particular water body, a plot of turbidity versus visual clarity will usually be well fitted in log-log space by a power-law relationship (e.g., Figure 2-1). (Note, however, that the regression line may vary appreciably between different waters, and somewhat between different makes and models of turbidimeter, such that there is no 'universal' relationship.) Outlier points on this type of plot may indicate bias or transcription errors. Of course, although nephelometric turbidity is a useful backup to visual clarity measurement, it cannot and should not *replace* visibility measurement which is (1) more precise, (2) better defined in optical terms and (3) more environmentally relevant (refer the NEMaR "Variables" report by Davies-Colley et al. 2012).

6.4.2 Duplication of samples

Sending duplicate (or even multiple replicate) samples to laboratories provides for a very useful check, over time, on laboratory precision. The statistics of deviation of the duplicates can be compared to stated laboratory precision for a particular test. In the long run the statistics of deviations of paired analyses should converge on the stated laboratory precision. Duplicate 'blind' samples should be obtained from the same (large, well-shaken) collection bottle as the 'official' sample – so as to avoid discrepancies in analytical results arising as an artefact of sampling variability.

Somewhat more sophisticated laboratory checks, which are fairly standard QA practice for trace constituent testing, include 'spiked' water samples (to which a known amount of analyte is added) to check recovery efficiency, and 'blank samples' in which distilled water is added to sample bottles to check for any background container or handling effects (e.g., Weiner 2008). We do not see the need for spikes and field blanks in routine SoE monitoring, but such special 'samples' might be appropriate as part of investigations when a particular chronic bias has been detected in regional datasets, say by comparison with field audit sampling.

We suggest that regional councils routinely include one duplicated water sample in each batch of water quality samples (or on each SoE sampling 'run'). The duplicated sample should not be identified to the laboratory staff. The choice of sample to be duplicated could be chosen at random or else a sample from a 'reference' site with rather low levels of most analytes split routinely.

A similar philosophy applies to duplication of macro-invertebrate biomonitoring samples. Such duplication would capture both natural spatial variability and the variability in laboratory processing. Although it would be difficult to separate these two components, the resulting data would give an indication of how accurately the true relative abundance of invertebrate taxa is being estimated using current field and lab protocols.

Almost all councils currently follow the standard protocol (Stark et al. 2001) of sending 10% of macro-invertebrate samples to a second laboratory to confirm the accuracy of processing and identification. We recommend continuing this practice, but note that if the same staff or

consultants consistently pass QA, then relaxing to every *second* year seems justifiable instead of every year as most councils currently do. This is consistent with an overall 5% measurement duplication rate for QA.

6.4.3 Visual assessments in biomonitoring

Visual assessments are made of periphyton, macrophytes and some physical habitat variables, which can be appreciably subjective. The following steps are recommended to minimise such subjectivity.

- **Observer training**: Because of the inherent subjectivity of visual assessment methods, data quality depends strongly on high-quality and consistent training of observers (Roper and Scarnecchia 1995). This also applies to assessments where observers describe features by selecting from qualitative categories (Hannaford et al. 1997). Training the observers in the specific protocols to be used often has a greater effect on data quality than the observers' general experience in ecological assessments (Wang et al. 1996; Herlihy et al. 2009). Therefore the first requirement for quality assurance is staff field training in the specific protocols, preferably by a single trainer (or 'auditor' – see Section 6.5 below).

- **Discussion of differences on site**. In field audits, differences between the auditor's and council staff's results may be due to differences in interpreting a particular attribute or to individual biases between observers (consistent over- or under-estimates). To correct misinterpretations and reduce biases, marked differences should be discussed between the auditor and field staff.

- **Photographs are very valuable**. To enable broad confirmation of data and correction of gross (e.g., typographic) errors, photographs should be taken showing the assessed features. Ideally photographs should show the entire site from different angles, as well as close-up views of measured attributes. In particular, photos should be used to confirm extreme or atypical values of any attributes that may later be regarded as dubious. Photographs may be seen as the equivalent of laboratory backups for visual assessments.

- **Number of categories affects resolution,** and therefore quality, of results. When designing scoring systems, thought should be given to the width of scoring categories for semi-quantitative and descriptive variables. Narrower categories may result in greater inter-observer variability than wider ones (Roper and Scarnecchia 1995).

6.4.4 Joint field exercises

We think there is considerable potential to 're-invigorate' water monitoring activity by promoting (say, through SWIM) joint field exercises between field staff of adjacent councils and local NIWA field stations. (We are aware that some regional staff and NIWA field staff already engage in local joint field exercises.) These joint field exercises could range from simply taking interested staff from neighbouring agencies along (to help where they can) on SoE sampling runs, through to duplicated field measurements and sampling by different field teams for later comparison of data.

Such joint field exercises are time-consuming (and so cost money), and it may be difficult, at first, to justify such exercises given pressure on monitoring budgets generally. But the cost may be set against the following multiple potential benefits.

- Increased collegiality across staff of different agencies, so promoting future collaboration.
- Potential for in-built training of new staff (and refreshment of existing staff).
- A platform for formal training programmes.
- Improved 'engagement' of field staff, who will feel valued for being recognised as crucial to an important enterprise.

As a guide we suggest that annual joint field exercises might be appropriate – probably in late summer when days are relatively long and air and water temperatures more conducive to prolonged outdoor activities. Joint field exercises might usefully be combined with annual sampling of some important components, notably macro-invertebrates. For example, joint fieldwork by two staff from two different agencies could cut through the heavy fieldwork burden of invertebrate sampling piggybacked on water quality sampling at 'integrated' sites.

The principle of collegiality developed in joint field exercises applies to field staff activities beyond SoE monitoring. Regional field staff who are temporarily over-burdened with say a spill event or an urgent compliance monitoring issue, would feel more willing and 'trusting' to call for help from staff of neighbouring councils or other agencies with whom they had worked in joint field exercises in SoE monitoring. That is, there are clear benefits to joint field exercises between different agencies in terms of wider collaboration and development of a climate of mutual respect.

Joint fieldwork during flood conditions may be particularly valuable. Field staff are often heavily burdened at such times and may have to make invidious prioritisations regarding high flow gaugings, sediment sampling or servicing of event-triggered autosamplers. Clearly storm flow sampling is a situation where overburdened regional staff, who have an existing relationship with staff from other agencies, might feel more able to call on those agencies for help during such events.

6.5 Field audit

Ensuring dependability of freshwater SoE monitoring seems to demand some sort of on-going national field audit system. Although the term 'audit' may at first sound somewhat daunting, we think that council field staff would have nothing to fear if this auditing process was initiated and 'owned' by regional councils via the SWIM group, and if it was introduced sensitively and implemented in a non-judgemental way. On the contrary, auditing of water (and aquatic biota) SoE monitoring emphasises the great importance attached to this activity and the integrity of the data being collected.

6.5.1 NRWQN field auditing system

A field auditing system has been developed for the NRWQN that is used on biannual visits to NIWA field stations by NRWQN supervising staff (Davies-Colley et al. 2011b). The field auditing system involves a site visit, once every two years, during which the auditor

accompanies NIWA field station staff on a NRWQN sampling 'run'. The auditor, usually one of the NRWQN supervisors based at the NIWA-Hamilton laboratory, observes field practice with reference to a checklist. Field staff are given a copy, a few days later, of the audit report. This endorses good practice and points out areas for improvement in a non-judgemental way. The NRWQN auditing system is explicitly designed to improve practice, and not be threatening to field staff.

6.5.2 A pan-regional field audit system

A broadly similar auditing system could be instituted for regional SoE monitoring. We envisage the audit team (1 or 2 people) visiting each region about once every two years and accompanying field staff to a subset of their river and lake SoE sites, probably a one-day field 'run' on each of river water quality (including periphyton assessment) and lakes. The auditor(s) would observe regional staff during their field measurements and sampling, noting field procedures, and would later send a brief report to the regional staff. Such reports might best be confidential to the field staff themselves.

Such a system could be extremely useful, if not actually 'essential' to ensure regional SoE monitoring data dependability. For biomonitoring, this system may be the best practical means to achieve independent quality assurance (see next section for further discussion). The auditing might best be implemented by, or at least commissioned by, SWIM, although other agencies (MfE, NIWA, university scientists) could assist with design.

The field audit system needs to be centred on a common ('national') **field form** for the sampling of each sub-domain: river water quality, river bio-monitoring and lakes. Therefore a set of national field forms should be developed consistent with the protocols outlined earlier in this report. We have sighted several examples of current field forms used by regional council staff and they are all broadly suitable with many important commonalities, but differ in detail. Rather than trying to 'impose' a particular design of field form (and its electronic equivalent for use with PDAs or field computers), we think that a workshop approach is desirable to develop suitable 'universal' field forms acceptable for use by all councils. Because these field forms would be central to the auditing system the workshop or workshops developing a national auditing system would also design national field forms.

6.5.3 Duplicate sampling as part of field audit

We have stated above (Section 6.1) that the best overall QA measure is duplicate sampling by an independent agency. Auditing visits provide a built-in opportunity to independently sample water bodies and duplicate the field measurements by regional staff. We envisage this duplication of measurements and sampling eventually becoming a routine part of the auditing activity. The auditor's water and biological samples would be processed *separately* from the regional SoE samples (ideally by a different laboratory) and a comparison of measurements (ideally an X-Y plot for each 'national' variable such as that in Figure 6-1) would be provided as part of the auditor's report. (We chose TN for illustration purposes, being a variable that is recommended in our previous report as a 'core variable in both lakes and rivers – Davies-Colley et al. 2012).

Far from being threatening, we think a field audit system as described here should be positively welcomed by most councils and their field staff. This includes duplication of measurements, because dedicated field staff will be interested to know how 'repeatable' their

measurements are. If their data agree closely with the auditor's, then that confers confidence in *all* their SoE data. If some data plots are noisy or biased, then measures can be instituted to track down why.

For now we present Figure 6-1 merely to illustrate the concept of QA of regional monitoring data by comparison with independent measurements. However, the 'degree of agreement' (precision and bias) is amenable to statistical analysis – to address the question: "is the regional method equivalent to the auditor's (or the neighbouring region's) method?" This could be included in a recommended analysis protocol that would need to be developed for comparing duplicated measurements, perhaps using Lin's Concordance Correlation procedure (see http://www.niwa.co.nz/online-services/statistical-calculators/concordance).

Water quality – rivers and lakes

For river water quality, the auditors would carry out their duplicate measurements and sampling immediately *after* the regional staff had finished, and immediately *upstream* so as to avoid any plumes caused by wading. Care would need to be taken not to unduly delay the regional field staff on site, so that the target time at subsequent sites on the 'run' was not compromised. When laboratory and field data had been collated and internally checked, an exchange of data would be made between the auditor and regional field staff.

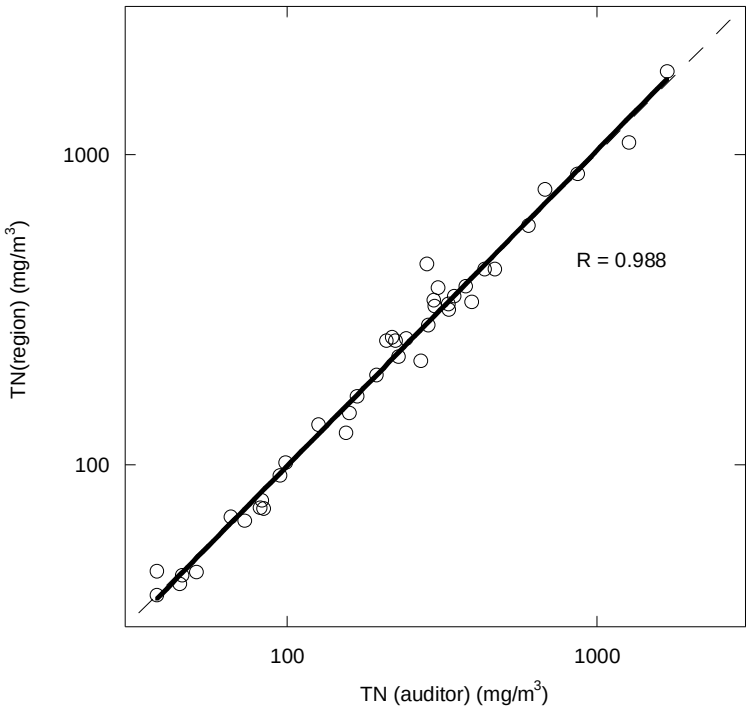


Figure 6-1: Indicative data for total nitrogen in several regions compared with that obtained by auditors on the same monthly 'runs'. There is good overall agreement (with about 10% RMS deviation) and the regression line is nearly co-incident with the 1:1 line of perfect agreement, indicating minimal bias. One point appears to be an outlier.

Field auditing, including duplication of measurements, is particularly important for lakes for which there is no national programme of monitoring by one agency, like the NRWQN, that might otherwise provide for duplication of monitoring effort as a QA measure (Section 6.6

below). Duplicate sampling of lakes will usually require boat sampling of the same water mass. The auditor(s) would travel in the *same* boat as the regional field staff so as to closely note on-site field procedures. The auditor(s) would conduct duplicate sampling *after* the regional staff, again taking care not to delay field work by unduly increasing on-site time.

Biomonitoring

As outlined in Section 6.3.2, river biota and habitat are subject to much greater spatial variability than river water. This natural variability is not easily separated from that due to differences in sampling technique when two sets of independently-collected samples or assessments are compared.

For visual assessments of periphyton, macrophytes and physical habitat, natural spatial variability may be reduced if the auditor uses the same reach and the same transect locations as the council (provided the council has not disturbed the flora or habitat during their assessments). For macro-invertebrates, and even more so for fish, separating spatial variability from inter-observer variability would require large amounts of data, including duplicates by both council staff and the independent auditor.

For invertebrates, the duplicates could be taken immediately upstream to avoid areas of stream bed disturbed by a council's kick sampling, provided that the upstream area has comparable habitat flow, depth, substrate and riparian conditions to the area sampled by the council. Alternatively, the auditor might sample the same reach as the council after waiting about 6 days for the macro-invertebrate community to recover from the sampling disturbance. The latter approach would be more accurate provided that no high flow events occur between the two sampling dates, but more costly because a separate field trip is implied.

For fish, duplicate assessments are unlikely to produce meaningful data because of the need for recovery of the fish community after electrofishing (which is not well known: M Bonnett, NIWA, pers. comm.), and several factors that may affect the fish present in a reach at a later date. Therefore, for fish (and possibly invertebrates if the same transects cannot be re-used), the best practical QA may be simply for an auditor to evaluate and provide feedback on regional field staff practice as described above, without any attempt to duplicate sampling. It can reasonably be assumed that if sample placement and technique are similar and consistent across the country (verification of which would ideally involve a single 'national' auditing team covering all regions), then fish data should be comparable.

6.5.4 Summary of field auditing

To summarise, we envisage field audit centred on a universal ('national') field form with two main components, both of which would be reported on in the field audit report (provided to the regional field staff) to foster good practice, as follows:

(1) **Commentary on regional practice** for field measurements and supporting observations, and water and biological sampling and sample handling.

(2) **Duplicate field measurements, sampling and analysis** to provide comparative data for identifying bias and quantifying accuracy. (We recognise that duplicate sampling may not be possible for fish and might sometimes be difficult for macro-invertebrates).

Such a field audit system would provide a powerful QA measure covering all aspects of field, and laboratory sources of bias and random error for water quality, although with important qualifications as regards biomonitoring. A regional council could cite its engagement in such auditing (e.g., in SoE reporting) as demonstrating the 'dependability' of its SoE data and subsequent accuracy-sensitive data reduction, notably time-trend analysis.

6.6 Duplication of NRWQN sampling

In the previous section we recommend that duplicate sampling be combined with audit of field procedures. For lakes, for which no national programme of monitoring exists, that may be the only opportunity for duplication. However, for rivers the National Rivers Water Quality Network (NRWQN; 77 sites that are visited monthly, Davies-Colley et al. 2011b) provides the opportunity for duplicate sampling as a quality assurance measure for relatively low extra cost separate from (occasional – e.g., biannual) auditing visits. There is at least one NRWQN site in 15 out of the 16 regions and TLA areas (Nelson City is the exception). The NRWQN includes all of the 'core' variables identified for river water quality plus biomonitoring of periphyton (visual assessment) and annual macro-invertebrate monitoring.

Some councils already duplicate sampling at NRWQN sites. That is, some SoE sites are co-incident (or very nearly so) with NRWQN sites. Generally, however, the NRWQN sites are not sampled at the same time (i.e., the same water mass within the river and the same communities of biota). This means that the resulting datasets are not 'paired' and only the overall distributions and statistics (e.g., medians) of the data can be compared (Figure 6-2). This is still a very valuable duplication of effort, but to be fully effective in a QA sense the *same water mass* needs to be sampled independently. Similarly for periphyton, the visual assessments would ideally be paired. For macro-invertebrates, the pairing in time is not quite so compellingly important.

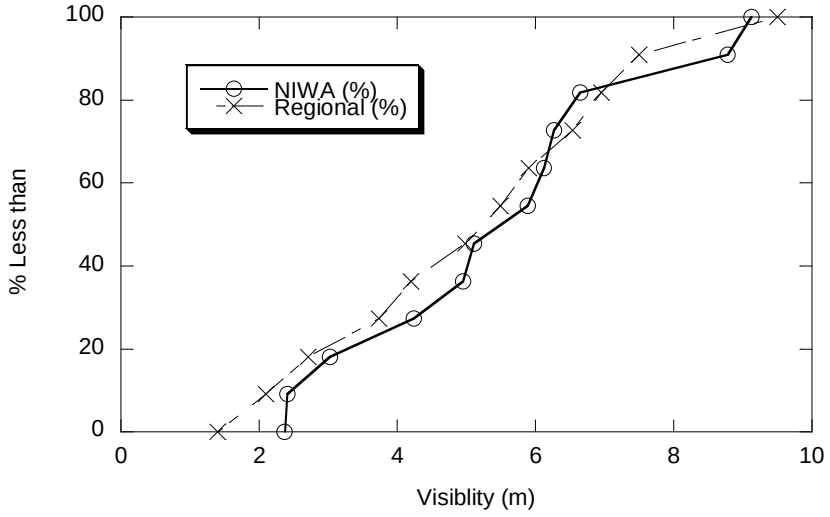


Figure 6-2: Indicative distributions of 12 monthly visibility observations at a NRWQN site. The distribution of data obtained by Regional field staff compares fairly closely with that obtained by NIWA field staff. (Note that, because the monthly data were obtained on *different* occasions they are not paired and cannot be compared point-by-point).

We are aware of only one council that currently duplicates NRWQN sampling at the same time – by having field teams rendezvous with NIWA staff on site such that both field and laboratory data are paired. This duplication has been maintained since the early 90's and has proven a very valuable QA measure.

We recommend that **regional field staff independently duplicate NIWA sampling at NRWQN sites** (at least one in their region) on the same day and at (nearly) the same time, say within 1 hour. Such duplicate sampling is of essentially the same water mass (and periphyton assemblage) so that each regional measurement can be paired with a NIWA measurement for all the common variables. Figure 6-3, illustrates the kind of comparison that this pairing makes possible. Regional data for visibility is plotted versus NIWA data for this variable at a NRWQN site for 12 monthly visits. In this case there is one outlier point (the same datapoint appearing as an outlier in Figure 2-1). Inter-agency data plots like this can show at a glance with reference to the 1:1 line of perfect agreement whether datapoints are biased and the overall accuracy of the measurement can be analysed.

Similarly Figure 6-4 shows a comparison of the 'core' water quality variables measured at a NRWQN site on the one date-time by both regional and NIWA field staff. Agreement is fairly good for all variables plotted apart from for *E. coli* which is intrinsically rather imprecise. A plot like this, showing good overall agreement for most core variables, confers confidence in *both* datasets.

We recognise that there would be the possibility of collusion of NIWA and regional staff regarding field measurements such as DO and visual clarity if the two teams actually met on site. We might ask both NIWA and regional field staff not to divulge field measurements, but of course, we have no way of ensuring that they adhere to this injunction! Collusion would, of course, compromise the independence of duplicated field measurements. However the water sample testing would still be independent. Collegiality of regional field staff and NIWA staff is to be encouraged in every other regard in order to promote interagency co-operation *including* on data quality. Therefore the need for strict independence for assessment of accuracy can probably be relaxed, in this case, in the interests of long-term collaboration.

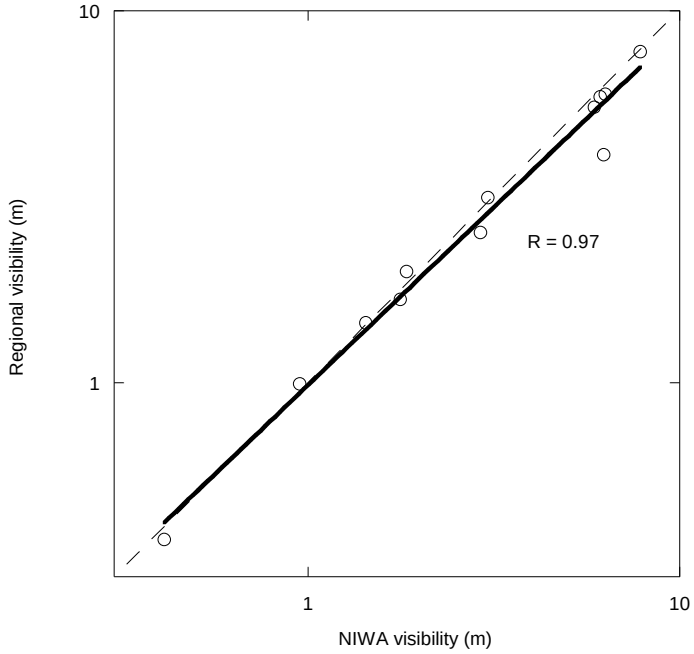


Figure 6-3: Paired observations of visibility by regional field staff compared to that by NIWA field staff at a NRWQN site. 12 months of monthly duplications is displayed. The measurements agree well overall, but regional data are slightly lower on average and there is one outlier. This is the same visibility observation showing as an outlier in Figure 2.1 – that was judged to be biased low based on the turbidity-visibility relationship and plots versus flow.

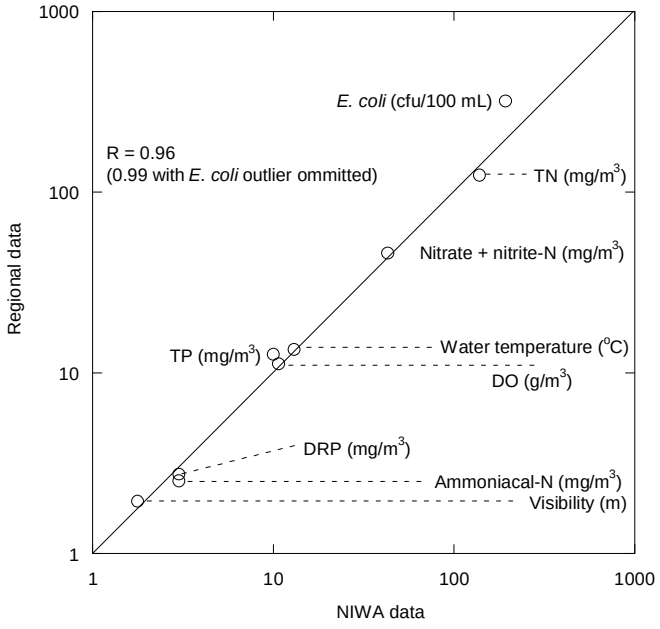


Figure 6-4: Comparison of paired Regional and NIWA data for different 'core' variables measured on one visit to a NRWQN site. Agreement of the variables is good overall, but less close for *E. coli* which is intrinsically more variable.

6.7 Data Batch Checking

6.7.1 Water quality – rivers and lakes

A very useful final QA step is data batch checking by regional SoE officers prior to uploading to databases. Data should be broadly consistent both across sites (with similar flow and weather history) for a particular sampling 'run', and within sites (between different variables). A range of simple, but powerful, tests on each data batch can be run to check internal consistency (e.g., Davies-Colley et al. 2011b).

For example, on a particular sample, total phosphorus (TP) should always be greater than DRP, and total nitrogen should always be greater than inorganic nitrogen (NOx + ammoniacal-N). Visual clarity should be consistent with turbidity (Figure 2-1). TP (also *E. coli*) correlates broadly with turbidity, visibility (and with flow) in rivers, so unusually high TP or *E. coli* datapoints may be plausible if visibility was low and flow was high at the time of sampling. Both TP and *E. coli* could be regionally high during storm flows. (This highlights another valuable application of 'flow stamping' of all river samplings.)

This kind of data batch checking for internal consistency is probably best done graphically, both by plotting data for related variables in X-Y graphs for all sites in a region, and by overlaying new data on established plots for historical data site-by-site. A data exploration package such as DataDesk ® (Velleman 1996) is probably preferable to spreadsheet-based statistical packages like EXCEL because the latter are usually designed for formal data presentation. Such internal data checks take time and expertise (which takes time to develop), further highlighting the need for QA in water monitoring to be properly resourced.

Ideally such data batch checking would be conducted promptly so that analysis repeats may be done in the laboratory if any (laboratory) data are suspect. Checking data only every 12 months for annual reporting is not sufficiently 'responsive'. Some councils already have a semi-automated interface with their laboratory service provider that screens data by comparison with historical records for individual SoE sites.

6.7.2 Biomonitoring

Before being entered into databases, fish data should be checked by an expert fish biologist who is familiar with the fish fauna of the region. Data checking can detect some errors in identification (e.g., migratory species being recorded upstream of known barriers, or species recorded outside of their known geographic range) and in data quality (e.g., level of identification too coarse; "eel" instead of "shortfin eel", for example). For records entered into the NZ Freshwater Fish Database, such data checking is done routinely and is regarded as an essential QA step.

Macro-invertebrate data should be less prone to misidentifications than fish data given that about 10% of sample analyses are duplicated for laboratory QA. Rare species, however, may be missed by this method. As with fish, misidentifications may be detected if species are recorded outside of their known range or habitat preferences. Another taxonomic "error" that may be detected by data batch checking is recording a taxon at two taxonomic levels (e.g., *Paralimnophila* and Hexatomini), resulting in skewed values for indices such as richness or MCI. Errors in abundance will probably be detected only if they are very large.

6.8 Consideration of QA in workshops

In order to rapidly advance regional practices on QA for water monitoring we recommend a workshop series be commissioned by SWIM and charged with the following tasks. Because there are important differences between the sub-domains of river water quality, river biomonitoring and lake monitoring, we think the workshops should concentrate on bringing together, respectively, specialists in those sub-domains. Each QA workshop series would:

- Introduce and discuss the general QA recommendations given here, including joint field exercises, duplicate sample analysis, and data batch checking.
- Develop a 'national' field form for each of river water quality, river bio-monitoring (different components) and lake monitoring.
- Establish a field auditing system for regional monitoring. The auditing system would centre on a checklist harmonised with the 'national' field form.
- Encourage councils to duplicate sampling of at least one NRWQN river site within their regions and discuss the practicalities of that.

7 Summary and Recommendations

7.1 General principles

- Freshwater monitoring must be stable over time as regards variables, sites and protocols in order that data is suitable for tracking trend in freshwater conditions over time.
- However, the need for change in methods and protocols can arise due to scientific advances and desirability of consistent protocols (which is important for comparisons between regions and for national summaries).
- Changing to consistent ('national') or improved protocols and methods needs to be balanced against the risks of causing discontinuities in historical long-term data records.
- Redesign of monitoring programmes, however innovative, must be implemented with great care to avoid data discontinuities. Usually an overlap period is required (say of 12 months) to accommodate a changed protocol – which itself is a disincentive to change. (An improvement in precision (without bias), in contrast, has no spurious trend ramifications.) Such redesign can be supported by stable, national 'benchmark' sites which provide a platform for method inter-comparison and calibration.
- Precision of most water quality variables should generally be better than 10% over most of the data range of interest. Lesser precision is acceptable for highly time-variable attributes such as *E. coli* concentration and periphyton biomass. Low relative precision may need to be accepted when analytes are very low, such as for DRP when it is being driven very low by algal bloom uptake in a P-limited lake.
- Laboratory protocols are usually a smaller source of error in water monitoring than field procedures. However, some issues with laboratory practice in water monitoring need to be addressed, as follows:
    - SoE data should not be censored. That is, no "<detection limits" should be reported by laboratories. Laboratories should be specifically contracted to deliver their best estimate of concentration, even if very low, along with an indication of precision.
    - Nutrient analyses should be done directly according to the recommended variables in an earlier NEMaR report. In particular, TN is best analysed directly rather than calculated as the sum of oxidised-N and TKN – which results in 'noisy' data because of the compounding of errors.

7.2 River WQ recommendations

- River monitoring should be monthly irrespective of flow conditions, and at the same time of day with a standard deviation < 1 hour. All river SoE site visits should be 'flow stamped', that is, flow should be estimated on each sampling occasion.
- Detailed metadata should be compiled on all SoE river sites indicating: location, flow estimation, and safe access to the channel over a range of flow conditions. At very

high flow when direct access to the channel may not be advisable, sampling should be carried out (by bucket) from a nearby bridge or from the bank using a telescoping pole sampler.

- Visual clarity should always be measured. If channel access is inadvisable at very high flow, visibility can be measured on a sample contained in a reflective trough. If visibility is < 100 mm, observations should, in any case, be made in the trough on a volumetrically diluted sample.
- Water sampling and sample handling should follow nationally agreed protocols (e.g., the NRWQN protocols). A field form with prompts for all important on-site measurements and attendant observations, should be completed for each river SoE visit.
- A **specialist workshop** of river water quality experts is recommended to confirm national protocols, and consider how to cope with the change in protocols that may be required at some river sites where long-term records already exist.
- The specialist workshop would also consider QA issues in river water quality, including design of a 'national' river water quality monitoring field form suitable for use by all councils, duplication of NRWQN sampling at one or more sites per region, and implementation of a field audit system designed to ensure consistency between councils and over time.

7.3 River bio-monitoring recommendations

7.3.1 Benthic macroinvertebrates

- Sampling should be annually, in late summer (January to March).
- For hard-bottomed streams, the semi-quantitative protocol C1 (Stark et al. 2001) should be followed, in riffles, with a total sampled area of 0.6-0.8 m^2.
- For soft-bottomed streams (>50% of the stream bed naturally composed of silt, sand or pumice), follow semi-quantitative protocol C2, in wood, stream bank and macrophytes, with a total sampled area of 3 m^2.
- Sample processing should follow quantitative protocol P2 (200-count with scanning for rare taxa).
- Some important details need to be considered by a **specialist benthic ecology working group**, including stand-down period after floods and QA measures in macro-invertebrate monitoring, including field audit.

7.3.2 Periphyton

- The main objectives of periphyton monitoring need to be decided by a **specialist benthic ecology working group**, and therefore the field protocols. QA measures in periphyton monitoring would also be considered by the working group, including field audit and a 'national' field form.
-

- *In the interim:*
    - Conduct visual assessments[6] monthly for the six warmest months of the year at water quality sites, and also during annual macro-invertebrate sampling.
    - Make a total of 20 x 0.5 m-diameter observations over four transects (fewer transects in large rivers). Record data using the categorisation of Kilroy (2011).

7.3.3 Fish

- A **specialist fish working group** is recommended to decide on:
    - The primary objectives of fish monitoring.
    - The optimal trade-off between completeness of taxa list, quantitative data, inclusion of size data, site coverage and sampling frequency.
    - A decision support system for deciding on the most appropriate fishing method(s) for different stream conditions.
    - QA measures in fish monitoring, including field audit and a 'national' field form.

7.3.4 Macrophytes

Assessment of macrophytes is not needed for national reporting, although macrophytes, if present, would normally be assessed as part of habitat survey. Protocols will soon be available to guide regional field staff.

7.3.5 Physical habitat

- A specialist **physical habitat working group** is recommended to decide on:
    - A short list of 'key' attributes from the many that could potentially be assessed.
    - Whether to take simple quantitative measurements or score alterations from natural condition, or use a combination of these.
    - QA measures in physical habitat assessment, including field audit and a 'national' field form.

7.3.6 Hydrology

A specialist **eco-hydrology working group** is recommended to research appropriate indices and decide on direct *versus* indirect measures of altered hydrology.

[6] Councils that can conduct quantitative (biomass) assessments are encouraged to do so in association with visual assessments. Measure Chlorophyll *a* following protocol QM-1b (Biggs and Kilroy 2000) and using hot ethanol extraction.

7.4 Lake monitoring recommendations

Although most regional lake monitoring appears to be competently done, there is considerable diversity of protocols in detail. The Burns (2000) protocols are only broadly followed in four (our of eleven) regions. Lake monitoring protocols need to be appreciably tightened to achieve national consistency. We realise however that change in protocols for lake monitoring may compromise the continuity of valuable long-term records in certain NZ lakes, so great care will have to be taken in recommending and implementing national protocols.

We recommend as follows.

- Sampling of lakes should be monthly, by boat, from the same depths in the epilimnion irrespective of thermal conditions and mixing depth.
- DO and temperature profiles should always be measured.
- Secchi depth should be measured irrespective of whether Secchi data are to be used in TLI calculation. The protocols of Smith (2001) should be followed, with an underwater viewer deployed on the sunny side of the boat.
- Sample handling should follow NRWQN protocols (unfiltered water samples chilled rapidly and delivered for laboratory processing within 24 hours).
- A deep lake site should be the 'reference' site, even where other sites or near-shore sites are also occupied in the one lake.
- A **specialist working group** of limnologists is recommended to confirm detailed protocols, and decide how to cope with the *change* in protocols that may be required in some lakes and regions for which valuable long-term records exist.
- Such a specialist working group would also consider QA issues, with design of a 'national' lake monitoring field form suitable for use by all councils, and implementation of a field audit system designed to ensure consistency between councils and over time.

7.5 QA recommendations

The most important underlying principle of QA in environmental monitoring is that a limited subset of measurements should be independently duplicated. Therefore, some fraction (we suggest of order 5%) of each sub-domain or component of regional water monitoring should be duplicated by an independent agency.

With this QA principle in mind we recommend as follows:

- **The stated precision of contract laboratories should be checked** by arranging for the analysis of one duplicate sample (unknown to the laboratory, i.e., a 'blind' sample) with each sample batch. Over time (say 12 months of monthly sampling) the statistics of replicate deviations should converge on the precision claimed by the laboratory.

- **Joint field exercises are to be encouraged** between different agencies engaged in SoE monitoring (including field staff from adjacent councils and nearby NIWA field stations). Such joint field exercises will promote collegiality around water monitoring leading to improved practice. Joint field exercises should also have a wide range of other benefits leading to improved interagency co-operation notably during flood events.

- **A programme of field audit is recommended,** probably organised and commissioned by SWIM on behalf of all the regional councils. An auditor or auditing team would visit each regional council regularly (we recommend once every 2 years) and accompany regional field staff on one day of a routine monitoring 'run' for both rivers and lakes. An auditor's report (confidential to field staff?) would (1) comment on field procedures, commending good practice and documenting areas for attention, and (2) *duplicate* regional field and laboratory measurements so as to provide an assessment of overall accuracy. '**National' field forms** need to be developed to promote consistency and facilitate auditing.

- **NRWQN sampling should be duplicated** routinely by regional field staff, at one NRWQN site per region, as a built-in QA measure for river water quality (and associated periphyton) monitoring. As with field auditor duplication, this duplication of monitoring effort will identify bias (if any), will assess accuracy, and, over time, will confer confidence in both the NRWQN and regional SoE datasets.

- **Each batch of data should be checked for internal consistency** before uploading to the data archive. The checks include both at-a-site (e.g., TP > DRP) and across-site checks on a sampling 'run' (e.g., TP and *E. coli* might plausibly be relatively high at most river sites over a region during storm flow conditions).

- **Workshops need to be convened to introduce and establish QA measures in regional water monitoring.** The working groups need to be convened along sub-domain lines, because there are appreciable differences between river and lake water quality, and between different biotic components. The workshops would initiate a field audit system centred on use of national field forms designed to be acceptable to all the councils.

7.6 Summary table

Finally, we provide a summary table (Table 7-1) that collects the main recommendations of this report as regards protocols and QA. This includes implementing *change* in protocols when valuable historical data is needed to track water condition over time. Much of the future action and implementation of protocols and QA towards a NSWMP (National Surface Water Monitoring Programme; Davies-Colley et al. 2011a) would be done most usefully in *workshops* organised along sub-domain lines with regional council scientists working with CRI and university science advisors.

Table 7-1: Summary of recommendations on surface freshwater monitoring protocols and quality assurance (QA).

Monitoring activity	National protocol of guideline existing?	Required work
River water quality	No (but NRWQN provides a model?)	Protocol details suggested herein; **Workshop on river water quality** needed, particularly on QA (joint exercises, field audit, NRWQN duplication)
Bio-monitoring *Macro-invertebrates*	Yes (C1, C3 in Stark et al. 2001)	Some protocol details need to be developed (incl. non-wadeable streams), QA **River benthic ecology workshop?**
Bio-monitoring *Periphyton*	Yes (RAM-2 in Biggs & Kilroy 2000; Kilroy 2011)	Some protocol details need to be developed (incl. non-wadeable streams), QA **River benthic ecology workshop?**
Bio-monitoring *Fish*	Yes (David et al. 2010), but still requires agreement among all experts.	Broad agreement to use David et al. 2010 protocol as basis. This protocol likely to require further modifications. QA may be difficult? **A further river fish workshop to resolve outstanding issues?**
Bio-monitoring *Physical habitat (*)*	No, although SHAP (Harding et al. 2009) & SEV (Storey et al. 2011)	Protocols need to be developed (with scoring), QA **River habitat workshop?**
Lake water quality	Yes in part (Burns et al. 2000)	Protocol details recommended herein; **Workshop of specialist limnologists**, particularly on QA (joint exercises, field audit)
QA protocols, resourcing & measurement duplication	No, but guidelines suggested herein (10% of budget; 5% of measurements)	**Workshops** to establish QA principles and discuss implementation (within different sub-domains)
Changing protocols	No, but guidelines suggested herein (12 months overlap)	**Workshop** to establish principles and discuss ramifications of protocol change for time trend analysis (within different sub-domains)

(*) Macrophyte monitoring is not needed for national reporting, but would normally be done as part of habitat survey. Protocols will be available very soon (Matheson et al. 2012 in prep.)

8 Acknowledgements

We thank Lucy Baker and Lian Potter of MfE for on-going support of the NEMaR project and review comments. Graham Sevicke-Jones (GWRC; chair of the SWIM group) facilitated feedback from regional scientists and SoE officers on monitoring protocols. Special thanks to Robert Ward (Emeritus Professor, Colorado State University) and USA colleagues for commentary on laboratory 'censoring' of water quality data. Peter Robinson (Hill Laboratories) is thanked for a particularly comprehensive and incisive review.

Jochen Schmidt (NIWA) is thanked for project oversight including report review. Deborah Ballantine (chair of the Indicators Workstream) and Graham McBride, both of NIWA-Hamilton, reviewed an advanced draft of the report (and Graham acted as a sounding board on statistical matters, notably laboratory data censoring). Doug Booker and hydrological colleagues at NIWA-Christchurch provided valuable commentary on hydrological underpinning of SoE sites. Mike Crump (NIWA-Hamilton) commented on laboratory protocols, particularly for TN; and John Quinn (NIWA-Hamilton) on macro-invertebrate sampling protocols. Mary de Winton and John Clayton (both NIWA-Hamilton) updated the authors on LakeSPI monitoring. Alison Bartley (NIWA-Hamilton) assisted with report compilation and formatting.

The project team thanks the NEMaR Variables expert panel (listed below) for their valued contributions in workshops and subsequent discussion. Regional council water resources scientists and monitoring staff from all the councils (listed below) are thanked for forwarding regional water monitoring protocol documents and for completing questionnaires on protocols and QA. Special thanks to Chris Fowles for commentary on historical comparisons with the NRWQN, and to Juliet Milne for helpful review comments.

Variables Expert Panel.

Adrian Meredith	Environment Canterbury
Alex Connolly	Taranaki Regional Council
Bill Vant	Waikato Regional Council
Brenda Baillie	Scion
Fleur Tiernan	Marlborough District Council
Jon Harding	University of Canterbury
Juliet Milne	Greater Wellington Regional Council
Jon Roygard	Horizons Regional Council
Marc Schallenberg	University of Otago
Maree Clark	Horizons Regional Council
Mike Eades	Marlborough District Council
Rachel Ozanne	Otago Regional Council
Trevor James	Tasman District Council

Regional Authority Contacts on Water Monitoring Protocols and QA.

Emma Simpson	Northland Regional Council
Martin Neale	Auckland Council
Kevin Collier, Bill Vant (also Ian Buchanan and Mark Hamer)	Waikato Regional Council
Paul Scholes, Alistair Suren	Bay of Plenty Regional Council
Dennis Crone	Gisborne District Council
Maree Clark	Horizons Regional Council
Adam Uytendaal	Hawkes Bay Regional Council
Alex Connolly, Chris Fowles	Taranaki Regional Council
Alton Perrie	Greater Wellington Regional Council
Trevor James	Tasman District Council
Paul Fisher	Nelson City Council
Fleur Tiernan	Marlborough District Council
Jonny Horrox	West Coast Regional Council
Adrian Meredith	Environment Canterbury
Rachel Ozanne	Otago Regional Council
Kirsten Meijer	Environment Southland

9 References

APHA (2005). Standard methods for the examination of water and wastewater. American Public Health Association.

Ballantine, et al. (2010). 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.

Barbour, M.T.; Gerritsen, J.; Snyder, B.; Stribling, J. (1999). Rapid bioassessment protocols for use in streams and wadeable rivers: Periphyton, Benthic Macroinvertebrates and Fish. *EPA 841-B-99-002*.

Biggs, B. (2000). New Zealand Periphyton Guidelines: Detecting Monitoring and Managing Enrichment of Streams. Prepared for Ministry for the Environment. 122 p.

Biggs, B.J.F.; Kilroy, C. (2000). Stream Periphyton Monitoring Manual. Prepared for Ministry for the Environment. 228 p.

Burns, N.; Bryers, G.; Bowman, E. (2000). Protocol for monitoring trophic levels of New Zealand lakes and reservoirs. Ministry for the Environment.

Burns, N.; McIntosh, J.; Scholes, P. (2005). Strategies for Managing the Lakes of the Rotorua District, New Zealand. *Lake and Reservoir Management 21*: 61–72.

Burns, N.M.; Rutherford, J.C.; Clayton, J.S. (1999). A monitoring and classification system for New Zealand lakes and reservoirs. *Journal of Lakes Research & Management 15*: 225–271.

Collier, K.J.; Kelly, J.; Champion, P. (2007). Regional Guidelines for Ecological Assessments of Freshwater Environments: aquatic plant cover in wadeable streams. *Environment Waikato Technical Report 2006/47*. 25 p.

Crump, M. (2011). The Effects of Transport and Storage on *Escherichia coli* Concentrations in Water Samples. *Water(NZ) (September)*: 49–53.

David, B.; Hamer, M. (2010). Regional Guidelines for Ecological Assessments of Freshwater Environments: Standardised Fish Monitoring for Wadeable Streams. *Regional Guidelines for Ecological Assessments of Freshwater Environments. Environment Waikato Technical Report 2010/09*. 21 p.

David, B.O.; Hamer, M.P.; Collier, K.J.; Lake, M.D.; Surrey, G.M.; McArthur, K.; Nicholson, C.; Perrie, A.; Dale, M. (2010). A standardised sampling protocol for robust assessment of reach-scale fish community diversity in wadeable New Zealand streams. *New Zealand Journal of Marine and Freshwater Research 44(3)*: 177–187.

Davies-Colley, R.; Larned, S.; Unwin, M.; Verburg, P.; Hughes, A.; Storey, R.; McBride, G.; Ballantine, D.; Hudson, N. (2011a). Dependable monitoring of

freshwaters for national scale environmental reporting. *NIWA Client Report HAM 2011-055.* Prepared for the Ministry for the Environment.

Davies-Colley, R.; Verburg, V.; Hughes, A.; Storey, R. (2012). Variables for regional monitoring underpinning national reporting. Summary of recommendations. *NIWA Client Report* For the Ministry for the Environment. February 2012 (Interim report). 30 p.

Davies-Colley, R.J.; Smith, D.G. (1992). Offsite measurement of the visual clarity of waters. *Water Resources Bulletin 28*: 1–7.

Davies-Colley, R.J.; Smith, D.G. (2001). Turbidity, suspended sediment, and water clarity: A review. *Journal of the American Water Resources Association 37*: 1085–1101.

Davies-Colley, R.J.; Smith, D.G.; Ward, R.C.; Bryers, G.G.; McBride, G.B.; Quinn, J.M.; Scarsbrook, M.R. (2011b). Twenty years of New Zealand's river water quality monitoring network: benefits of careful design and consistent operation. *Journal of the American Water Resources Association 47*: 750–771.

Davies-Colley, R.J. (1988). Measuring water clarity with black disk. *Limnology and Oceanography 33*: 616–623.

de Winton, M.D.; Clayton, J.S.; Edwards, T. (2012). Incorporating invasive weeds into a plant indicator method (LakeSPI) to assess lake ecological condition. *Hydrobiologia 691*: 47–58. http://dx.doi.org/10.1007/s10750-012-1009-0

Death, R. (2008). Effect of floods on aquatic macroinvertebrate communities. *In*: Lancaster, J.; Briers, R. (eds). Aquatic Insects: challenge to populations. Proceedings of the Royal Entomological Society's 24th Symposium, pp. 103–121. CAB International, Wallingford, UK.

Duggan, I.C.; Scarsbrook, M.R.; Quinn, J.M. (2003). Comparison of coded abundance and fixed count rapid assessment techniques for biomonitoring in New Zealand streams. *New Zealand Journal of Marine and Freshwater Research 37(1)*: 23–29.

EC (1981). Analytical Methods Manual. Water Quality Branch, Inland Waters Directorate, Environment Canada, Ottawa, Canada.

Hannaford, M.J.; Barbour, M.T.; Resh, V.H. (1997). Training reduces observer variability in visual-based assessments of stream habitat. *Journal of the North American Benthological Society 16(4)*: 853–860.

Harding, J.; Clapcott, J.; Quinn, J.; Hayes, J.; Joy, M.; Storey, R.; Grieg, H.; Hay, J.; James, T.; Beech, M.; Ozane, R.; Meredith, A.; Boothroyd, I. (2009). Stream Habitat Assessment Protocols for wadeable rivers and streams in New Zealand. University of Canterbury, Christchurch. 133 p.

Herlihy, A.; Sifneos, J.; Bason, C.; Jacobs, A.; Kentula, M.; Fennessy, M. (2009). An Approach for Evaluating the Repeatability of Rapid Wetland Assessment Methods: The Effects of Training and Experience*. Environmental Management 44(2)*: 369–377.

Hudson, N.; Ballantine, D.; Gibbs, M.; de Winton, M.; Storey, R.; Verburg, P.; Hamill, K. (2011). Investigation of single indicators for water quality assessment and reporting. *NIWA Client Report HAM 2011-066* prepared for Ministry for the Environment.

Jowett, I.G. (1993). A method for objectively identifying pool, run, and riffle habitats from physical measurements. *New Zealand Journal of Marine and Freshwater Research 27*: 241–248.

Joy, M.K.; Death, R.G. (2004). Application of the Index of Biotic Integrity Methodology to New Zealand Freshwater Fish Communities. *Environmental Management 34(3)*: 415–428.

Kilroy, C. (2011). Categories of periphyton for visual assessment. Prepared for Environment Canterbury periphyton monitoring programme.

Matheson, F.; Quinn, J.; Hickey, C.W. (in prep.). Review of the New Zealand instream plant and nutrient guidelines and development of an extended decision making framework: Phase 1 and 2 final report. Prepared for the Ministry of Science & Innovation Envirolink Fund.

Matthaei, C.; Uehlinger, U.R.S.; Meyer, E.; Frutiger, A. (1996). Recolonization by benthic invertebrates after experimental disturbance in a Swiss prealpine river. *Freshwater Biology 35(2)*: 233–248.

McBride, G.B. (2005) *Using Statistical Methods for Water Quality Management: Issues, Problems and Solutions*. John Wiley & Sons, New York.

McKergow, L.M.; Davies-Colley, R.J. (2010). Storm dynamics and loads of faecal pollution in a large mixed land use catchment. *Hydrological processes 24*: 276–289.

MfE (1994). Water quality guidelines No. 2: Guidelines for the management of water colour and clarity. Ministry for the Environment, Wellington, 77 p.

Ministry for the Environment and Ministry of Health (2009). New Zealand Guidelines for Cyanobacteria in Recreational Fresh Waters – Interim Guidelines. Prepared for the Ministry for the Environment and the Ministry of Health by SA Wood, DP Hamilton, WJ Paul, KA Safi and WM Williamson. 82 p.

Mullan, B.; Wratt, D.; Dean, S.; Hollis, M.; Allan, S.; Williams, T.; Kenny, G. (2008). Climate Change Effects and Impacts Assessment. A guidance manual for local government in New Zealand. 2nd Edition. Ministry for the Environment.

Porter, P.S.; Ward, R.C.; Bell, H.E. (1988). The detection limit. *Environmental Science & Technology 22*: 856–861.

Richards, R.P. (1998). Estimation of pollutant loads in rivers and streams: a guidance document for NPS programs. US Environmental Protection Agency, Denver.

Robinson, P.G. (2010). Review of regional council and NIWA analytical methods for surface waters. *Hill Laboratories Report to the Ministry for the Environment (Contract 13263),* 30 p.

Roper, B.B.; Scarnecchia, D.L. (1995). Observer Variability in Classifying Habitat Types in Stream Surveys*. North American Journal of Fisheries Management 15(1)*: 49–53.

Scrimgeour, G.J.; Davidson, R.J.; Davidson, J.M. (1988). Recovery of benthic macroinvertebrate and epilithic communities following a large flood, in an unstable, braided, New Zealand river. *New Zealand Journal of Marine and Freshwater Research 22(3)*: 337–344.

Smith, D.G. (2001). A Protocol for Standardizing Secchi Disk Measurements, Including Use of a Viewer Box. *Journal of Lake and Reservoir Management 17*: 90–96.

Smith, D.G.; Davies-Colley, R.J.; Knoef, J.; Slot, G.W.J. (1997). Optical characteristics of New Zealand rivers in relation to flow. *Journal of the American Water Resource Association 33*: 301–312.

Smith, D.G.; McBride, G.B.; Bryers, G.G.; Wisse, J.; Mink, D.F.J. (1996). Trends in New Zealand's river water quality network. *New Zealand Journal of Marine and Freshwater Research 30:* 485–500.

Sorrell, B. (2006). Lake water quality in New Zealand. Status in 2006 and recent trends 1990–2006. 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 No. 1 No.* 57 p.

Stark, J.D.; Phillips, N. (2009). Seasonal variability in the Macroinvertebrate Community Index: Are seasonal correction factors required? *New Zealand Journal of Marine and Freshwater Research 43(4)*: 867–882.

State of Victoria Department of Sustainability and Environment (2005). Index of Stream Condition: The Second Benchmark of Victorian River Condition.

Storey, R.G.; Neale, M.W.; Rowe, D.K.; Collier, K.J.; Hatton, C.; Joy, M.K.; Maxted, J.R.; Moore, S.; Parkyn, S.M.; Phillips, N.; Quinn, J.M. (2011). Stream Ecological Valuation (SEV): a method for assessing the ecological functions of Auckland streams. *Technical Report 2011/009*. 66 p.

UNEP/WHO (1996). Water Quality Monitoring - A Practical Guide to the Design and Implementation of Freshwater Quality Studies and Monitoring Programmes. Published on behalf of United Nations Environment Programme and the World Health Organization.

USEPA (1993). Method 180.1: Determination of turbidity by nephelometry. Environmental Monitoring Systems Laboratory Office of Research and development U.S. Environmental Protection Agency, Cinccinnati, Ohio.

USEPA (1998). Guidance for data quality assessment. Practical methods for data analysis. EPA/600/R-96/084. U.S. Environmental Protection Agency, Cinccinnati, Ohio.

USGS (2009). U.S. Geological Survey, variously dated, National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations. Book 9, chaps. A1-A9 Available at http://pubs.water.usgs.gov/twri9A.

Verburg, P.; Hamill, K.; Unwin, M.; Abell, J. (2010). Lake water quality in New Zealand 2010: status and trends. *NIWA Client Report HAM2010-107*. For the Ministry for the Environment, 48 p.

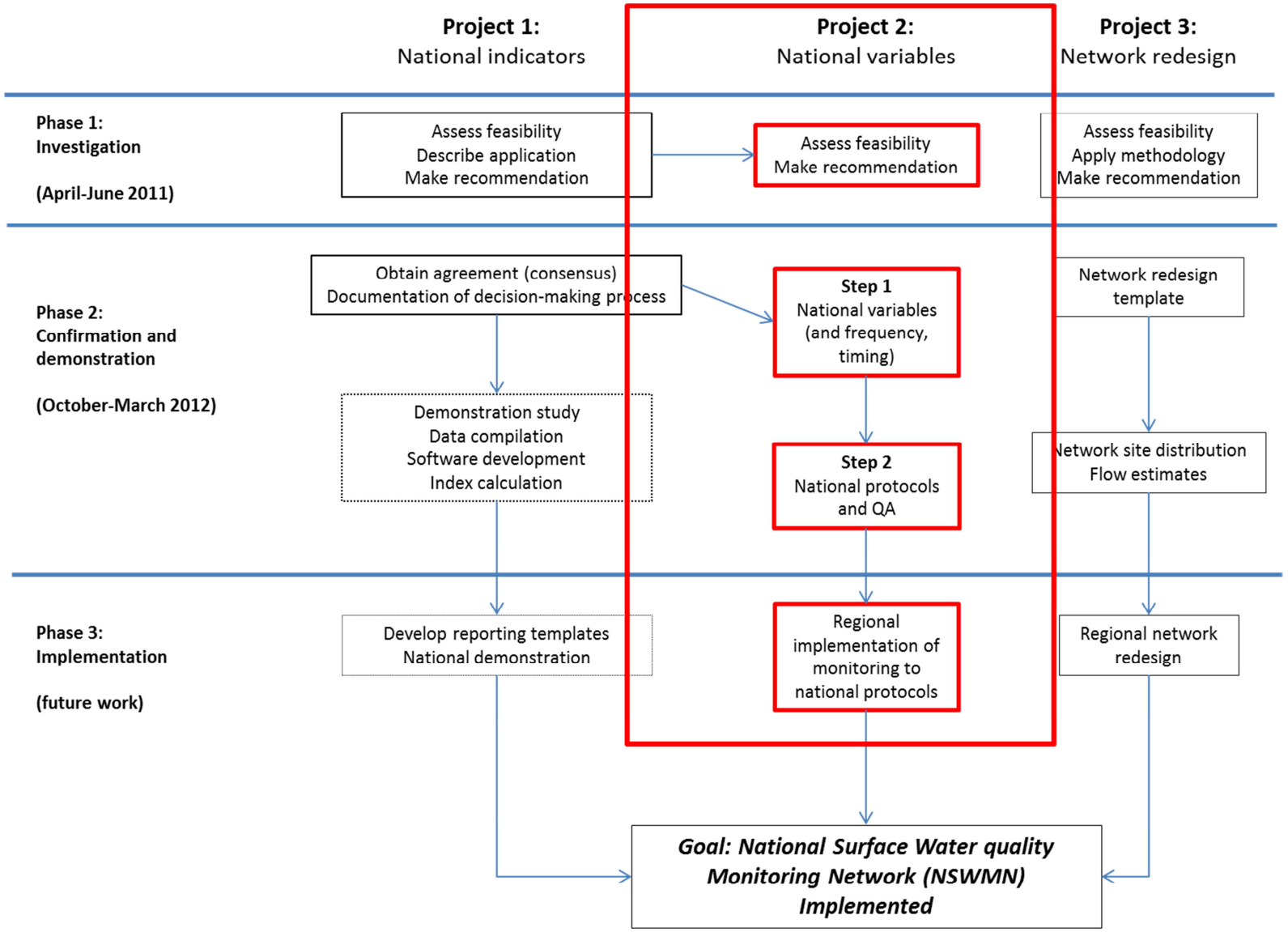
Wang, L.; Simonson, T.D.; Lyons, J. (1996). Accuracy and Precision of Selected Stream Habitat Estimates. *North American Journal of Fisheries Management 16(2)*: 340–347.

Ward, R.C.; Loftis, J.C.; McBride, G.B. 1990. *Design of Water Quality Monitoring Systems*. Van Nostrand Reinhold, New York. 231 p.

Weiner, E.R. (2008). *Applications of Environmental Aquatic Chemistry.* 2nd Edition. CRC Press, Boca Baton. 436 p.

Zanevald, J.R.V.; Pegau, W.S. (2003). Robust underwater visibility parameter. *Optics express 11*: 2997–3009.

Appendix A – Flowchart for the NEMaR Variables Workstream


Project 1:
National indicators
Project 2:
National variables
Project 3:
Network redesign
Phase 1:
Investigation
(April-June 2011)
Assess feasibility
Describe application
Make recommendation
Assess feasibility
Make recommendation
Assess feasibility
Apply methodology
Make recommendation
Phase 2:
Confirmation and
demonstration
(October-March 2012)
Obtain agreement (consensus)
Documentation of decision-making process
Step 1
National variables
(and frequency,
timing)
Network redesign
template
Demonstration study
Data compilation
Software development
Index calculation
Step 2
National protocols
and QA
Network site distribution
Flow estimates
Phase 3:
Implementation
(future work)
Develop reporting templates
National demonstration
Regional
implementation of
monitoring to
national protocols
Regional network
redesign
Goal: National Surface Water quality
Monitoring Network (NSWMN)
Implemented

Appendix B – River Water Quality Questionnaire

River Water Quality Questionnaire	**Example:**
Council name (Name and position of staff member responding).	**Somewhere Regional Council** (Joe Bloggs, Water Resources Scientist)
What is the sampling frequency.	Monthly, sampling occurs on the third Monday-Wednesday of every month
Are sites visited at approximately (± 1hr) the same time on every visit?	Yes, the same route is taken during each sampling run so that each site is visited at approximately the same time
Are the same sites visited on every cycle?	Yes. We no longer have "rolling" sites.
Do you have metadata on each site available to guide field staff?	Yes. Metadata includes photos, annotated maps, and notes on access, staff gauges, etc.
Do you estimate streamflow on each sampling visit? (How?)	Yes, if the site is close to a hydrometric station. Otherwise no.
Does sampling occur during high flows/floods	No, for H&S reasons
How do you cope with bad weather on the target day/time?	Sampling is deferred to first safe day after target
Are samples collected in bottles that have been acid-washed?	No, bottles have been rinsed but not acid-washed
Where in streams are samples collected?	Point bars usually used for safe access. In small streams the centre is accessed by wading and sample bottles are dipped into the stream by hand. On large streams or rivers a telescopic bottle holder (Mighty Gripper) is used to reach as far as possible out to the free-flowing water.
Are bridges or other artificial structures used for access?	No, because water access is needed for visibility
Describe water sampling procedure.	Facing upstream, the bottle is submerged to mid-depth with the open orifice pointing down. At mid-depth, the sample bottle is gently tipped up, allowing water to enter the bottle. Bottle is rinsed twice with stream water before collection
Are samples collected from pools or riffles?	Samples are collected from or just downstream of riffles where present
Are samples placed into a chilled/light sealed storage bin?	Yes, samples are placed immediately in a chilly bin with ice
Are samples delivered to lab within 24 hours of collection?	Yes, except when samples are collected late on Friday and are then stored in freezer over weekend
Is a field sampling sheet completed for each site visited/sample taken?	Yes. An example of a completed field sheet is attached.
If in-field measurements are taken (e.g., DO, temperature, pH, conductivity), how often are instruments calibrated?	All field instruments are calibrated every 12 months
Is visual clarity (e.g., black disc) measured?	Black disc visibility is measured whenever access to the channel (usually at a point bar within 200 m of the designated sampling point) is judged to be safe

***River Water Quality Questionnaire* (ctd)**	**Example:**
If yes, describe visual clarity protocol.	Black disc protocol follows MfE (1994) "Guidelines for colour and clarity of waters."
Do you QA your river water quality field procedures?	No, we are not sure how this can be done practically
If yes, describe QA approach.	No QA as such, but staff have attended training courses

Appendix C – River Bio-monitoring Questionnaire

River Biomonitoring Questionnaire ***Macro-invertebrates***	**Example:**
Council name (Name and position of staff member responding)	**Somewhere Regional Council** (Joe Bloggs Water resources scientist)
Frequency	Annual
Timing	Jan-March
Sites fixed (same set of sites visited every sampling occasion) or rotating?	Fixed
Net mesh size	0.5mm
Sampling protocol (Hard Bottomed streams)	C1 (Stark et al. 2001)
Sample area per replicate (Hard bottomed streams)	0.1-0.2m2
No. of Replicates per site (hard bottomed streams)	5 per site
Replicates composited into 1 sample?	Yes
Mesohabitat type(s) (e.g., riffles/runs/all habitat) hard bottomed streams	Riffles
Sampling protocol (Soft bottomed streams)	C2 (Stark et al. 2001)
Sample area (Soft bottomed streams)	0.3 m2
No. of replicates per site (soft-bottomed streams)	Ten
Replicates composited into 1 sample?	Yes
Mesohabitat types (e.g., stable substrates, wood, macrophytes) soft bottomed streams	C2 (Stark et al. 2001)
Definition of soft bottomed vs. hard bottomed streams	Soft bottomed is: >50% of stream bed naturally composed of silt, sand or pumice
Stand down period after floods	3 weeks after flood >3x median flow
Sample processing protocol	P2, i.e., 200-count with scan for rare taxa (Stark et al. 2001)
Processing&ID done by	External consultant
Taxonomic level of ID	MCI-level (genus or higher)

***River Biomonitoring Questionnaire* (ctd)** ***Macro-invertebrates* (ctd)**	**Example:**
Indices used for reporting	MCI, SQMCI
Field QA/QC	10% of samples repeated by a second staff member
Lab QA/QC	QC2 (Stark et al. 2001): 10% of sorted samples including residue re-examined by external expert. Passes QC if <=10% of taxa identified differently.
Lab QA done by	NIWA
Other relevant notes	

River Biomonitoring Questionnaire* (ctd)** ***Periphyton	**Example:**
Council name (staff member responding)	**Somewhere Regional Council** (Joe Bloggs Water resources scientist)
Is periphyton growth an issue of concern in your region?	Yes
Main objective(s) for periphyton monitoring (e.g., nuisance growths, nutrient enrichment, effects of regulated/reduced flows, interpret invertebrate data, other)	Nuisance growths for angling and aesthetics, indicator of nutrient enrichment
Frequency	Monthly
Timing	First week of the month
Timed with water quality/biomonitoring/both/other	water quality
Visual (%cover) assessment: field protocol	RAM-2 (Biggs and Kilroy, 2000)
Visual (%cover) assessment: no. of categories	RAM-2 (Biggs and Kilroy, 2000)
Visual (%cover) assessment: mesohabitat type	Runs (depth 0.2-0.6 m, velocity 0.3-0.8 m/s)
Visual (%cover) assessment: sampled area per replicate	5 rocks per transect
Visual (%cover) assessment: no. of replicates	4 transects
Visual (%cover) assessment: reporting indices	Periphyton enrichment index
Quantitative (biomass): field protocol	QM-1b (Biggs and Kilroy, 2000)
Quantitative (biomass): sampled area per replicate	5 cm diameter scrape per rock
Quantitative (biomass): no. of replicates	10 rocks
Quantitative (biomass): sample storage	Frozen
Quantitative (biomass): reporting indices and units	Chlorophyll a (mg/m2), ash-free dry mass (mg/m2)
Quantitative (biomass): extraction and analysis methods	Chl a: acetone solvent, spectrophotometer analysis; AFDM: ignition at 400C for 4 hours
Identification of periphyton taxa	Yes
Other analyses	Nutrient diffusing trays, cyanobacteria
QA/QC	Procedures in Biggs and Kilroy (2000)
Other relevant notes	

River Bio-monitoring Questionnaire* (ctd)** ***Fish	**Example:**
Council name (staff member responding)	**Somewhere Regional Council** (Joe Bloggs Water resources scientist)
Frequency	Annual
Timing	Jan-March
Sites fixed (same set of sites visited every sampling occasion) or rotating?	Fixed
Timed with other bio-monitoring?	No
Sample method (e.g., Electric Fishing Method, spotlight, minnow traps)	EFM where conditions allow
Reach length	150 m
No. of passes	One
Focus of data collection: complete species list/absolute abundance/relative abundance	Complete species list
Data forms	NZ Freshwater Fish database form filled in
QA/QC	Training course in fish identification required for all field staff. Two qualified staff members present for all sampling.
Other relevant notes	

River Biomonitoring Questionnaire* (ctd)** ***Macrophytes	**Example:**
Council name (staff member responding)	**Somewhere Regional Council** (Joe Bloggs Water resources scientist)
Is macrophyte growth an issue of concern in your region?	Yes
Main objective(s) for macrophyte monitoring (e.g., clogging of streams, invasive species, indicator of altered light/nutrients/flows)	Invasive species, clogging of streams
Frequency	Annual
Timing	Jan-March
Timed with other biomonitoring?	Macroinvertebrates
Field protocol	% cover in a 1 m wide transect across stream
Growth form categories used or species identified	Submerged/emergent/floating; presence/absence of invasive species
QA/QC	Training in identification of invasive plants required for all field staff
Other relevant notes	

River Bio-monitoring Questionnaire* (ctd)** ***Physical Habitat	**Example:**
Council name (staff member responding)	**Somewhere Regional Council** (Joe Bloggs Water resources scientist)
Frequency	Annual
Timing	Jan-March
Timed with WQ/biomonitoring/other?	Biomonitoring
Hydrology/morphology parameters measured	Discharge (ADCP gauging), channel shape (V/U/square/wide), wetted width, non-vegetated width, undercuts, % riffle/run/pool
Instream parameters measured	%substrate types, %organic matter cover, % cover for fish, bank stability
Riparian parameters measured	Width, longitudinal intactness, dominant vegetation type, % shade, livestock access, soil permeability
Protocol based on (reference to source protocols)	USEPA Rapid Bioassessment Protocol
Other relevant notes	

River Biomonitoring Questionnaire* (ctd)** ***Ecosystem processes	**Example:**
Council name (staff member responding)	**Somewhere Regional Council** (Joe Bloggs Water resources scientist)
Is there interest in your council to monitor ecosystem processes (e.g., gross primary productivity, ecosystem respiration)?	Yes
Does your council have the capacity to measure ecosystem processes?	No
What are the main limitations to measuring ecosystem processes?	Cost of equipment, time required to deploy equipment, lack of training in methods

Appendix D – Lake Water Quality Questionnaire

Lake Water Quality Questionnaire	**Example:**
Council name (Name and position of staff member responding)	**Somewhere Regional Council** (Joe Bloggs, Water Resources Scientist)
What is the sampling frequency?	Monthly, sampling occurs on the third Monday of every month
What is the time of day of sampling?	Around midday
How do you cope with bad weather on the target day/time?	Sampling is deferred to first safe day after target
Are sites visited at approximately (± 1hr) the same time on every visit?	Yes, the same route is taken during each visit and each site is visited at approximately the same time
Where in the lake do you sample?	Center (or near shore, or at the outlet?)
How was the sampling site chosen?	Easy access, we can't use a boat etc.
Do you sample more than 1 site in the same lake (how many?)	Mostly just 1 central site. 3 sites in Lake Wherever, one central and one in each of two main side arms
Is lake water level recorded?	Yes, staff gauges are installed on all monitored lakes, usually near the outlet river.
Do you sample also rivers flowing into lakes for water quality?	Yes - in Lake Wherever. Otherwise no.
How many inflowing rivers do you sample (do not count outflow)	3 rivers flowing into Lake Wherever
At what depth are samples taken for water quality?	Surface or mid-epilimnion depth (discrete - using van Dorn sampler), or 0-10 m depth (integrated)
How is this depth decided?	Middle of epilimnion is Burns protocol
Does the sampling depth vary between sampling occasions?	Yes - depending on whether stratified and epilimnion depth
Is the sampling at a discrete depth or integrated tube sampling?	Surface or mid-epilimnion depth; Integrated tube 0-10 m depth in Lake Wherever
How much volume do you collect and mix before subsampling?	Up to 20 L in large bucket
Describe water sampling procedure	Bottles (2 X 1 L) rinsed twice with lake water.

***Lake Water Quality Questionnaire* (ctd)**	**Example:**
Are samples collected in bottles that have been acid-washed?	No, bottles have been rinsed but not acid-washed
Are samples placed into a chilled/light sealed storage bin?	Yes. Samples are placed immediately in a chilly bin with ice
Are samples delivered to lab within 24 hours of collection?	Yes, except when samples are collected late on Friday and are then stored in freezer over weekend
Is a field sampling sheet completed for each site visited/sample taken?	Yes (Standard lake field sheet is attached)
Do you take profiles for oxygen (DO) and temperature (T)?	Yes
Are DO and T samples taken at discrete depths or continuously using a profiler?	Profiler. Discrete depths were measured (usually at 1 m intervals) until 2006.
Which brand and type of profiler or other instrument do you use?	Seabird SBE 19plus V2 (Version 2) SEACAT Profiler with SBE 43 DO sensor
What profiler sampling frequency do you use?	1 sec. Profiler is lowered at less than 10 cm per second so that data is acquired every 10 cm.
If in-field measurements are taken (e.g., DO, temperature, pH, conductivity), how often are instruments calibrated?	All field instruments calibrated every 12 months
Is visual clarity (e.g., Secchi depth) measured?	Secchi is routinely measured - except in some remote lakes where helicopter sampling is used
If yes, describe protocol	Secchi extinction depth is observed with a bathyscope or other viewer on sunny side of boat
Do you QA your lake field procedures?	No, we are not sure how this can be done practically
If yes, describe QA approach	No QA as such, but staff have attended training courses