Freshwater Science and Technical Advisory Group

Supplementary report to the Minister for the Environment

April 2020

Contents

Executive summary and recommendations

Role and process

The Freshwater Science and Technical Advisory Group (STAG) was established to support officials by providing science and technical advice on the work programme of the Water Taskforce.

- In June 2019 the STAG submitted a report to the Minister for the Environment containing 15 recommendations to clarify national direction and improve freshwater management in Aotearoa/New Zealand.
- In September 2019 the Government released a public discussion document titled 'Action for healthy waterways: A discussion document on national direction for our essential freshwater', accompanied by a series of draft national regulations, standards and policies.
- From November 2019 to January 2020 the STAG met as a group, or in some cases as sub-groups, to consider questions from officials on the scientific rationale for thresholds and national bottom lines, perspectives on the technical feasibility of different policy options, and comments on points raised in submissions to the Government's draft policy and regulatory proposals.

This report summarises these discussions and notes several amendments to the recommendations in the June 2019 report that STAG members consider necessary following further analysis and consideration.

This report should be treated as supplementary to the primary report provided to the Minister for the Environment in June 2019 and read in conjunction with it.

No changes to overarching recommendations from primary report

Biophysical Ecosystem Health Framework

Freshwater systems are complex – there isn't always a clear linear or mechanistic relationship between different variables in an ecosystem, and freshwater indicators of habitat or aquatic life have complex relationships with multiple stressors.

The integrated approach of Te Mana o te Wai underscores the importance of taking a much more holistic view of the things we need to measure and manage. This is necessary if we are to effectively protect and enhance our shared values for water – avoiding the perverse outcome of managing to 'targets' for individual attributes and achieve what appears to be positive results in silos but leave the overall health of the ecosystem no better off.

Members consider that to understand and manage the health of freshwater ecosystems effectively we need to know more about aquatic life, physical habitat, water quality, water quantity, ecological processes, longitudinal river connectivity and connectivity between rivers, lakes, wetlands and groundwater.

Members agree it is technically feasible to generate a comprehensive picture of ecosystem health without having to measure and monitor the state of all metrics of ecosystem health in all monitoring locations within a given management area. Guidance from central government, supported by worked examples, is required to clarify what level of monitoring will be necessary to inform local management and implementation decisions.

Mātauranga Māori

It is essential that more work is done to introduce and integrate mātauranga Māori into the national framework of freshwater policy and regulation, and to enhance scientific assessment, decision-making and policy implementation with mātauranga-based frameworks and monitoring tools.

Requirement to 'maintain or improve'

Matters of data-availability, practicality and economics will influence exactly what data are considered sufficient to determine the current state of a waterbody under the framework of the National Policy Statement for Freshwater Management (NPS-FM). The historical data record available to environmental managers in Aotearoa/New Zealand was not generated with the objective of establishing management limits to preserve the health and functioning of freshwater ecosystems – especially prior to 2011 and the introduction of the NPS-FM.

If the decision is taken to estimate the health of a waterbody at a point in the past and use that to set the state at which it needs to be maintained or improved at, members consider that the earliest there may be sufficient data to attempt to estimate the prior state of a waterbody would be around the turn of the century – some five to seven years after regional councils started monitoring and after the RMA had forced changes to the conditions of resource consents managing 'point source' discharges (noting that point source discharges remain an issue in some locations).

Accounting for environmental variability

For some attributes included in the Government's proposed changes to the national objectives framework (NOF) there are likely to be instances where natural variation across ecosystems means the state of 'reference sites' may exceed national bottom lines. The STAG considers that these instances will be relatively rare and straightforward to identify and, in most instances, can be dealt with reasonably by exception(e.g. fish IBI above natural barriers to migration). In a small number of instances separate river classes may be required, i.e. for suspended sediment in glacial watersheds and dissolved reactive phosphorus (DRP) in catchments with naturally acidic volcanic geology.

In the STAG's view, where apparently-healthy rivers exist with high levels of contaminants (i.e. high MCI scores at elevated nutrient concentrations for instance) other indicators which may not have been measured will likely reveal stress. In addition, and using the same example, even if elevated nutrients are not having an obvious impact on MCI in a specific stretch of river, environmental managers still need to have good information to establish the effects of those nutrients on downstream receiving environments such as lakes and estuaries.

No changes to the following recommendations from primary report relating to specific management categories

DO – rivers (recommendation 5 from primary report)

DO – lakes (recommendation 6 from the primary report)

Ecosystem metabolism (recommendation 7 of the primary report)

Changes to the following recommendations from primary report relating to specific management categories

Periphyton – rivers

Amend recommendation 8 to reword the typo error in footnote 8 to read "Based on a monthly monitoring regime. The minimum record length for grading a site based on periphyton (chl-a) is 3 years." As per the relevant footnote in the NPS-FM (2017).

Fish biotic integrity

Amend recommendation 9 to:

- Provide for two tables specifying numeric biophysical values for fish biotic integrity using the Fish Index of Biotic Integrity (Fish IBI) and specifying different thresholds for bands and national bottom lines – one including salmonids and one excluding them.
- Clarify that decisions on which table to apply should be made locally and that national direction is required to clarify the circumstances in which the tables are applicable.

Macroinvertebrates

Amend recommendation 10 from the primary report to clarify that:

- MCI and QMCI should be assessed together, and the lower of the two results should apply.
- ASPM is a separate metric and should be assessed separately.
- The national management framework should rely on MCI scores derived using the most recent update of the MCI methodology.

Macrophytes – lakes

Amend recommendation 11 from the primary report to:

- Modify proposed tables for lake submerged plant index to reflect that it is considered appropriate to conduct an assessment every three years for at-risk lakes, and every five to ten years for lower risk lakes.
- Clarify that eradicating exotic macrophytes entirely is unnecessary to reach the bottom lines proposed and active management of exotic species would be suitable in many instances.
- Provide for an exception that allows environmental managers to leave exotic species in place in circumstances where this is necessary to maintain ecosystem health, i.e. where exotic species make up the entire (or close to the entire) macrophyte community or increase resilience to the effects of storms by stabilising the bed of shallow lakes.

Sediment

Amend recommendation 12 of the primary report to clarify that:

For suspended sediment

- The attribute tables should be described in terms of visual clarity and should allow for the site-specific inter-conversion of suspended sediment, turbidity and visual-clarity measures.
- A lower level of classification (e.g. Level 2) should be used for developing attribute tables, with thresholds derived using the community deviation method.

For deposited sediment

- A lower level of classification (e.g. Level 2) should be used for developing attribute tables, with thresholds derived using the community deviation method.

Nutrients – rivers

A majority of members consider:

- Recommendation 13 should be retained without amendment – the methodologies and data sets used to derive the proposed criteria, bottom lines and thresholds for DIN and DRP for rivers are scientifically rigorous, well explained and well justified, have been discussed at length by the STAG and peer reviewed independently by Professor David Hamilton who generally supported the approach adopted.
- A note should be added to Recommendation 13 acknowledging that, while some rivers in acid-volcanic geological terranes may have naturally high levels of DRP, these rivers are readily identifiable, equate to 17% of national stream length (70,899 km) and, where identified, can reasonably be dealt with by exception – although it would also be technically feasible for the management framework to treat these rivers as a separate class.

A minority of members consider:

- the evidence provided to establish nationally applicable bands and bottom lines is insufficient to provide confidence that a given DIN or DRP concentration will achieve the desired improvement in ecosystem health or ensure that the target of a specific ecosystem health metric will be met.

- There are concerns about the reliability and effectiveness of nationally-applied nutrient criteria in managing for ecosystem health, given they have been derived from weak relationships that vary spatially. This could have the effect of not triggering a management response in rivers where this is necessary to protect ecosystem health and vice versa.

The minority sub-group recommends that recommendation 13 should be deleted and that controls in the current NPS-FM on the effects of nutrients in rivers should be strengthened by:

- Giving effect to recommendation 8 of the primary report and strengthening the periphyton attribute in the current NPS-FM by providing a default nutrient table with spatially variable bottom lines and band thresholds, via guidance, as recommended in the primary report of the STAG.
- Increasing the level of protection from toxicity by making the current bottom of the ‘B band’ the national bottom line for ammonia and nitrate The current national bottom line provides for 80% species protection from chronic toxicity and the sub-group’s recommendation is to raise this to 95% species protection from chronic toxicity which is more consistent with other ecosystem health protection measures recommended by the STAG.
- Introducing national monitoring requirements for DIN and DRP in rivers that, where increasing trends are detected in a freshwater management unit, would trigger the requirement to develop a management plan for reducing nutrient concentrations.

Wetland extent and condition

Amend recommendation 14 of the primary report to specify that the most recent version of the wetland condition index should be used when evaluating the condition of wetlands.

Recommendations for further work

The primary report of the STAG identified a series of topics that urgently require additional work. In addition to underscoring the importance of filling these gaps in the current national policy and regulatory framework for freshwater management, members would like to highlight the need to:

- Complete work currently underway to refine and clarify the WCI (Wetland Condition Index) methodology.
- Provide guidance on how to deal with waterways that qualify as an 'exception' to the general requirements specified under the NPS-FM and NOF.

Part 1: Introduction

1.1 Role of the STAG

The Terms of Reference (see **Appendix 1)** for the Science and Technical Advisory Group (STAG) establish that the STAG is to support officials by providing science and technical advice on the work programme of the Water Taskforce, as requested by government officials.

The core roles of the STAG are to: "… advise on scientific evidence for freshwater policy development by:

- reviewing science that underpins the National Policy Statement for Freshwater Management's (NPS-FM) National Objectives Framework attributes and other freshwater policy options
- identifying any gaps in the science
- improving the NOF attribute development process
- improving protocols to better manage incorporating science into the policy process
- providing overarching scientific advice and guidance as it relates to freshwater policy development
- contributing to science and technical related guidance for councils to implement the NPS-FM
- providing science advice on issues raised in public submissions on proposed Appendix 2 attributes and wider freshwater policy."

The STAG was not given the specific task of developing additional attributes *per se* for the NPS-FM. Its role was to determine a range of ecologically-meaningful thresholds and break points (including bottom lines) for a range of variables, or management categories, relevant to aquatic ecosystem health. When drafting the primary report the STAG chose, however, to follow the same format and use the same terminology as the NPS-FM. This was to signal that the management categories, measures and thresholds recommended by the STAG should be treated as having equal importance to the existing categories, measures and thresholds in the NOF for managing freshwater.

How the STAG's recommendations are translated into national policy and regulatory tools is outside the scope of the STAG's mandate. The question of whether, or indeed if, these recommendations should be translated into national policy and regulatory tools as 'attributes', 'monitoring guidance', 'research priorities' or in some other guise sits with government informed by the advice of officials.

Similarly, when developing its recommendations, the STAG was asked not to consider the economic implications of potential management categories, measures and thresholds – these implications are to be considered by others in the Essential Freshwater programme, including government officials and the Freshwater Leaders Group. STAG members were, however, invited to provide a perspective on the scientific and technical feasibility of proposed management categories, measures and thresholds, including matters such as the availability (but not cost) of technology, sensitivity of measurement methods relative to the delineation of break points, access to critical expertise and, potentially, other matters related to capability and capacity – provided they weren't directly related to cost.

In this regard, members note that kaitiaki, regional councils, sector groups and environmental NGOs hold substantial expertise in the practical application of freshwater science and regulation. These parties will have much to contribute to the objective of maintaining a strong connection between national policy and regulatory settings and their effective implementation.

1.2 Scope and focus of the STAG's work programme

In June 2019 the STAG submitted a report to the Minister for the Environment containing 15 recommendations to clarify and improve the national framework for freshwater management, and identify and address knowledge gaps which currently constrain our ability to manage freshwater and the health of freshwater ecosystems in Aotearoa/New Zealand.[1]

In September 2019 the Government released a public discussion document titled 'Action for healthy waterways: A discussion document on national direction for our essential freshwater', accompanied by a series of draft national regulations, standards and policies.

The Government received approximately 17,500 submissions in response to its proposals for stopping further degradation of freshwater resources and beginning to reverse past damage. Officials requested that the STAG reconvene to respond to specific questions in relation to identified topic areas.

The questions from officials – largely seeking further elaboration on the scientific rationale for thresholds and national bottom lines, and perspectives on the technical feasibility of different policy options – provide the mandate for STAG's continuing involvement in the policy development process.

In this regard, questions from officials largely determine the scope and focus of the STAG's considerations in this cycle of its work. In some cases when answering officials' questions the STAG's discussions referred to or traversed matters that were raised in its primary report but not explicitly queried by officials. This supplementary report provides an overview of the STAG's general discussions to aid understanding and highlights points of agreement and disagreement.

1.3 Process followed

The STAG met as a full group on Wednesday 27 November 2019 to discuss a series of topics and questions posed by officials, reflecting matters that have arisen during the process of public consultation on the Government's proposals.

Following this discussion, officials facilitated the development of advice in areas of interest to them by: forming the STAG into subgroups and charging those groups with answering a series of questions in targeted topic areas; commissioning independent peer-reviews of key pieces of research

[1] https://www.mfe.govt.nz/publications/fresh-water/freshwater-science-and-technical-advisory-group-report-minister-environment

underpinning the primary recommendations of the STAG; providing the STAG with a summary of key submissions; and inviting the STAG to make further comments.

Several STAG members also met with the Independent Advisory Panel appointed to provide a perspective on public submissions and participated in discussions with the Freshwater Leaders Group relating to technical issues arising from the submissions.

The STAG met again for a second time on 22/23 January 2020 to work through a final set of questions developed by government officials and circulated to members with supporting material prior to the meeting. In each topic area members were asked to reach a position on whether anything raised by government officials in response to public submissions on policy proposals justified any changes to recommendations contained in the STAG's primary report.

1.4 Structure of supplementary report and relationship to primary report

This report should be treated as supplementary to the primary report provided to the Minister for the Environment in June 2019 and read in conjunction with it.

The STAG 'supplementary report' follows the same basic structure as the primary report – addressing general overarching recommendations and specific proposed management categories, measures and thresholds.

Each section of this supplementary report will:

- identify what, if any, changes are required to recommendations provided in the primary report
- highlight key points raised by members during discussion
- identify any caveats that need to sit alongside the STAG's discussions and recommendations
- express any perspectives that individuals or sub-groups of the STAG would like to record.

Part 2: Comments on overarching recommendations from primary report

2.1 Biophysical ecosystem health framework

Supplementary comments on relevant recommendation from primary report

No change to recommendation 1 from the primary report.

STAG discussion

In our primary report we made the key recommendation that the national policy and regulatory framework for freshwater management in Aotearoa/New Zealand needs to prioritise ecosystem health as the primary outcome. This reflects the STAG's understanding that Te Mana o te Wai requires upholding Te Hauora o te Taiao, Te Hauora o te Tangata and Te Hauora o te Wai when managing freshwater systems.

In response to submissions, officials asked STAG to consider whether some of the attributes included in the Government's reform proposals were redundant.

Freshwater systems are complex – there isn't always a clear linear or mechanistic relationship between different variables in an ecosystem, and freshwater indicators of habitat or aquatic life have complex relationships with multiple stressors.

The integrated approach of Te Mana o te Wai underscores the importance of taking a much more holistic view of the things we need to measure and manage. This is necessary if we are to effectively protect and enhance our shared values for water – avoiding the perverse outcome of managing to 'targets' for individual attributes and achieving what appears to be positive results in silos, but leaving the overall health of the ecosystem no better off.

Members consider that in order to understand and manage the health of freshwater ecosystems effectively we need to know more about aquatic life, physical habitat, water quality, water quantity, ecological processes, longitudinal river connectivity and connectivity between rivers, lakes, wetlands and groundwater. The STAG's primary report recommends the addition of further management categories, measures and thresholds to ensure environmental managers have the information they need to make robust decisions.

Caveats

The STAG acknowledges these requirements are likely to have significant cost implications. It is not the STAG's role to make judgement calls balancing the value of information against the cost of generating it (and any associated opportunity costs). Indeed, STAG anticipates that environmental managers will make judgement calls about what information is required to be generated where, to promote the best outcome for the health of freshwater ecosystems.

In the primary report, members indicated that guidance from central government would be required to determine what level of monitoring would be necessary to inform local management and implementation decisions, and that this guidance would need to be supported by worked examples of how this should be done.

Without seeking to pre-empt this guidance, members consider that it is technically feasible to gain a comprehensive picture of ecosystem health within a management area by generating information on aquatic life, physical habitat, water quality, water quantity, ecological processes and longitudinal river connectivity without having to measure and monitor the state of all management categories in all monitoring locations within that management area. In other words, members expect that the focus of measurement and monitoring will vary across a management area depending on the nature of the waterbodies, the values associated with them, their current state and the pressures they are experiencing.[2]

STAG member perspectives

There is broad consensus amongst members.

2.2 Mātauranga Māori

Supplementary comments on relevant recommendation from primary report

No change to recommendation 2 from the primary report.

STAG discussion

Officials did not raise any questions for STAG to consider. STAG members would like to reiterate their view that the integrated approach of Te Mana o te Wai requires a better understanding of the relationship between Māori attributes of freshwater health and the numeric biophysical attribute states and regulatory measures dealt with in the primary report.

Caveats

Our work has been informed by kaupapa Māori approaches. During our discussions we have benefitted from the support and insights of the group's Māori members, particularly Ra Smith who did much to deepen our understanding. Members do acknowledge, however, that not all the Māori

---

[2] This is consistent with the original conception of the National Objectives Framework, which anticipated focussing on managing – and monitoring – a subset of attributes based on the values present in a particular location, as explained in the discussion document: Freshwater Reform, 2013 and beyond pp 29-31 https://www.mfe.govt.nz/sites/default/files/freshwater-reform-2013.pdf

members of the STAG have been able to attend all meetings. The weight of the STAG's membership as well as the general biophysical/biochemical nature of questions asked by officials has meant the supplementary work has focussed primarily on questions relating to western biophysical science.

Members are reassured that other groups in the Essential Freshwater programme are working to ensure that Te Ao Māori and mātauranga Māori are integrated appropriately into the design of policy and regulatory proposals.

Members would like to reinforce the point made in the primary report: it is essential that more work is done to introduce and integrate mātauranga Māori into the national framework of freshwater policy and regulation, and to enhance scientific assessment, decision-making and policy implementation with mātauranga-based monitoring.

STAG member perspectives

There is broad consensus amongst members.

2.3 The requirement to 'Maintain or Improve'

Supplementary comments on relevant recommendation from primary report

No change to recommendation 3 from the primary report.

STAG discussion

In response to points raised by submitters, officials asked the STAG for views on data requirements for determining the current state of a waterbody, and whether there is enough reliable data available to allow for the calculation of the state of waterbodies in the past.

In a general scientific sense, the data required to establish the current state of a waterbody can be derived by combining the requirements of the NPS-FM with the additional requirements associated with the management categories, measures and thresholds recommended by STAG in its primary report.

In its primary report STAG recommended (Recommendation 3c) the development of 'guidance on how to determine what level of monitoring is enough to inform analysis and reporting, supported by worked examples of how this should be done'. The STAG would like to reinforce the importance of this recommendation and notes that guidance would need to recognise that the level of complexity across the range of current and proposed management categories, measures and thresholds, which makes a 'one size-fits all' approach inappropriate.

A sub-group of the STAG met to discuss this topic on 21 January 2020. A memo outlining the nature of these discussions has been produced by officials (see **Appendix 2**).

Caveats

STAG understands that matters of data-availability, practicality and economics will influence what data are considered sufficient to determine the current state of a waterbody under the NPS-FM.

STAG notes that the historical data record available to environmental managers in Aotearoa/New Zealand was not necessarily generated with the objective of establishing management limits to preserve the health and functioning of freshwater ecosystems – especially prior to 2011 and the introduction of the NPS-FM.

Even in instances where scientific information is robust and accessible, it is unlikely to cover all the components of freshwater ecosystems necessary to gain a reliable understanding of their current health and functioning, let alone at some point in the past.

Members consider that the earliest there may be sufficient data to attempt to estimate the state of a waterbody would be around the turn of the century, some five to seven years after regional councils started monitoring and after the RMA had forced changes to point source discharge consents (prior to the introduction of the RMA, several of Aotearoa/New Zealand's largest rivers and many of its small ones were in a far worse state than at present because of the prevalence of unmanaged point source discharges – while that situation has improved under the RMA, point source discharges are still an issue in some locations).

If the decision is taken to estimate the health of a waterbody at a point in the past and use that to set the state at which it needs to be maintained or improved, the STAG has some reservations about the use of models to establish this 'previous state'. Modelled estimates of the state of a waterbody will contribute constructively to policy decisions on where to set limits, so long as they are appropriately robust and validated and if bounds of uncertainty are clear and transparent. But not all attributes have tight relationships with land use, so trying to model a "state" in the past for each attribute based on the estimated impacts of assumed land use at the time will be highly problematic.

STAG member perspectives

One member considers it would be appropriate to clarify that the 'maintain or improve' requirement applies from the date of gazetting the policies that introduce, or have introduced, attributes into the NPS-FM.

Some members query the value or necessity of seeking to set the state at which a waterbody needs to be maintained or improved at some point in the past. The framework provided by the NPS-FM allows communities to set management objectives at the A or B band – probably equating to the state of water quality at some point in the past. Under this framework the technical information is available to inform local decision-making to determine the environmental state communities aspire to, what is required to achieve this state, and the length of time allowable for managers and resource users to achieve that state.

2.4 Accounting for environmental variability

Supplementary comments on relevant recommendation from primary report

No change to recommendation 4 from the primary report.

STAG discussion

In response to submissions, officials asked STAG to consider whether it is appropriate that in some instances the state of 'reference sites' will be worse than national bottom lines for some attributes included in the Government's proposed reforms.

The STAG recognises natural variation across ecosystems may mean that the state of 'reference sites' exceeds the national bottom lines for some attributes included in the Government's proposed changes to the national objectives framework. The STAG considers that these instances will be relatively rare and straightforward to identify and, in most instances, can be dealt with reasonably by exception. An example would be the effects of natural barriers to migration on Fish IBI. In a small number of instances, separate river classes may be required, i.e. for suspended sediment in catchments with glacial influence and DRP in streams draining catchments with naturally-acidic volcanic geology.

It is important to note that, in the STAG's view, it is likely that apparently-healthy rivers with naturally-high levels of contaminants – streams with high MCI but elevated nutrient concentrations for instance – will have other indicators that reveal stress. In addition, and using the same example, even if elevated nutrients are not having an obvious impact on MCI in a specific stretch of river, environmental managers still need to have good information to establish the effects of those nutrients on downstream receiving environments such as lakes and estuaries.

The question posed by officials prompted discussion regarding the appropriate definition of a reference condition or state:

- Reference conditions or states that seek to estimate and reflect the natural state prior to human impact may not always be appropriate for the contemporary management context, as ecosystems may have adapted to changed conditions and developed resilience around a stable and healthy equilibrium that meets community aspirations.
- Reference conditions or states based on an estimate of pre-human conditions could set up unrealistic expectations about the potential future state of a waterbody – even with active management it may not be possible to remove sediment from some streams, for instance.
- Models used to derive some reference conditions or states may be influenced by the extent to which systems used in modelling have adapted to more contemporary context. The extremely dominant and longstanding influence of human activities on the health of freshwater environments and the highly-modified state of many of Aotearoa/New Zealand's freshwater ecosystems mean that estimates used to derive reference states might not provide a true reflection of an ecosystem in a 'natural state'.

- Modelling a reference state and using it to help derive a bottom line (sometimes from modelled stressor data) adds uncertainty to the establishment of bottom lines. If bottom lines can be derived from ecological principles this avoids the uncertainty created by having to model reference conditions – but this may not always be possible.

Ultimately the STAG notes that the question of where to set ‘reference state’ is contestable scientifically – individuals’ perspectives will be influenced by their philosophical standpoint, by the policy framework in place, and the objectives of the affected communities.

Caveats

The STAG’s discussion on this topic raises a general question: at what point does it become unreasonable to account for the extremes of natural variation by providing for exceptions to a framework that is appropriate for most instances? Ultimately the decision on what proportion of waterbodies can be managed by exception is a policy decision.

The STAG is of the view that more consideration needs to be given to where and how precaution is built into all attributes introduced into the NPS-FM to account for environmental variability. Ideally this would provide a consistent framework for application across attributes, and be accompanied by guidance detailing how and when to apply exceptions and what degree of precaution is appropriate in given circumstances.

STAG member perspectives

Some members query whether some of the apparent discomfort of submitters with some of the proposed bottom lines – especially those relating to DIN and DRP – could be due to technical issues arising from over-reliance on the relationship between these metrics and one (proxy) measure of ecosystem health (i.e. MCI) and/or derive from research that relies on a very small sample size and infrequent sampling.

Part 3: Comments on recommended additional ecosystem metrics

3.1 Dissolved oxygen – rivers

Supplementary comments on relevant recommendation from primary report

No changes to recommendation 5 from primary report.

STAG discussion

In response to submissions, and in relation to dissolved oxygen (DO) in rivers, officials asked the STAG whether:

- there is enough natural variation and a suitably robust evidence base to warrant the creation of different attribute states for different river types, and
- percent saturation would be a more suitable measure of dissolved oxygen than concentration.

There is natural variation, but it is complex and there are many natural processes that involve oxygen production or consumption. Levels of DO will depend on geography, as well as on flow, periphyton and macrophytes, fine sediment deposition, productivity, decomposition and reaeration. There is no redundancy among these attributes; instead they provide complementary information for assessing the state and potential drivers of ecosystem health (for more STAG discussion on this topic see **Appendix 3**).

STAG members disagreed with the point made by some submitters that saturation is more relevant, and noted that this issue had been addressed when the DO attribute was introduced. Temperature changes saturation, therefore, concentration would be more relevant than saturation at higher temperatures (for more discussion on this topic see **Appendix 4**).

Caveats

None recorded.

STAG member perspectives

Some members consider that the minimum requirement of one, seven-day period per year for assessing this dynamic process is insufficient to provide a robust picture of oxygen demand in many rivers. Variability in monthly measures, potentially over shorter timespans, should be evaluated prior to providing definitive guidance on monitoring needs.

3.2 Dissolved oxygen – lakes

Supplementary comments on relevant recommendation from primary report

No changes to recommendation 6 from the primary report.

STAG discussion

The STAG considers that we need to measure and manage DO in lakes and, while the current state of information is lacking, members agree there is value in national policy and regulation that encourages monitoring and the generation of information on lake DO and consequent improvement where poor states are observed.

National guidance should provide direction on how to use temperature profiles to work out hypolimnetic boundaries and midpoints. This guidance should specify the use of agreed (and published) methods and clarify whether a volumetric or depth-related midpoint should be used.

Caveats

The STAG noted several caveats relating to the measurement of DO in lakes in its primary report. These caveats have not been reflected in the proposed NPS-FM released for public consultation. Members would prefer these caveats to be associated with any changes to the NPS-FM in this area – at minimum these caveats should be reflected in national guidance on implementation.

STAG member perspectives

There is broad consensus amongst members.

3.3 Ecosystem metabolism

Supplementary comments on relevant recommendation from primary report

No change to recommendation 7 of the primary report.

STAG discussion

Although officials did not raise any questions for STAG consideration, members queried the rationale for proposing to include an attribute in the draft NPS-FM with no attribute bands.

In its primary report the STAG recommended introducing a management category, measure and thresholds for ecosystem metabolism but suggested deferring the matter of where to set a bottom line until more data were available.

Caveats

No new caveats recorded – noting that several caveats were attached to this management category in the primary report.

STAG member perspectives

One member considers that, while this is a promising area, it is too soon for the measures and thresholds associated with this proposed management category to be translated into attributes in the National Objectives Framework – further scientific analysis and a more comprehensive national-scale database is required before this step can be taken with confidence.

Although there is some evidence to suggest that a single week of monitoring data taken during the 'most stressed' time (summer/autumn) can be indicative of metabolic state,[3] some members consider that the minimum requirement of one, seven-day period per year for assessing this dynamic process is insufficient to provide a robust picture of oxygen demand in many rivers. One member considers that variability in monthly measures, potentially over shorter timespans, should be evaluated prior to providing definitive guidance on monitoring needs.

Some members note that the proposed 'D band' represented a substantially-degraded state and, if a more precautionary approach were to be applied, there would be technical justification for defining the national bottom line at the point of the proposed 'C band'. These members note it is highly unlikely that further research would find evidence to support setting a bottom line at a level less stringent than the bottom of the proposed 'C band'.

3.4 Periphyton – rivers

Supplementary comments on relevant recommendation from primary report

Amend recommendation 8 to reword the typo error in footnote 8 to read "Based on a monthly monitoring regime. The minimum record length for grading a site based

on periphyton (chl-a) is 3 years." As per the NPS 2017 footnote.

[3] Clapcott JE, Young RG, Neale MW, Doehring KAM, Barmuta LA 2016. Land use affects temporal variation in stream metabolism. Freshwater Science 35(4): 1164-1175.

STAG discussion

Officials did not ask any specific questions, but did note that some regional councils suggested they should be able to manage periphyton in some locations using stream shading and that the level of stream shading could be a useful proxy indicator for periphyton risk.

Members do not support the suggestion that the nature and extent of stream shading can be relied upon as an indicator of the risk of periphyton growth and the key management response. Stream shading is a valid control measure in some circumstances, but it is not universally-applicable (i.e. it is not possible to shade large rivers) and nor does the nature/extent of stream shading provide insight into the nutrient imbalances in freshwater systems that may affect other aspects of ecosystem health.

Members do support the use of established generalised models for estimating nutrient concentrations in the management of periphyton, in the absence of more specific regional models.

Caveats

Although the note to the relevant attribute table in the NPS-FM signals that three years of monthly monitoring is the minimum record length for grading a site based on periphyton, five years would provide a much more accurate estimate of the 92nd percentile.

STAG member perspectives

The major studies feeding into the existing periphyton attribute in the NPS-FM were conducted 15 to 20 years ago. A significant amount of periphyton data has been collected since then. Reviews of the periphyton attribute in the NPS-FM were undertaken in 2016 by Fleur Matheson[4,5] and in 2019 by Ton Snelder, Cathy Kilroy and others[6] focussing particularly on the levels for chlorophyll-a in Appendix 2 of the NPS-FM. Recommendation 8 from the primary report was informed by this more recent analysis, which formed the evidence base and rationale for the STAG's recommended changes to the existing periphyton attribute in the NPS-FM – including the introduction of a default nutrient table.

[4] Matheson, T., Quinn, J., Unwin, M. 2016. Instream plant and nutrient guides. NIWA report HAM2015-064 to BIE. 117 pp

[5] This review found convergence between MCI critical values and the 200 mg/m2 chla periphyton bottom line.

[6] Snelder, T.H., Moore, C., Kilroy, C. 2019. Nutrient concentration targets to achieve periphyton biomass objectives incorporating uncertainty. Journal of the American Water Resources Association: 1-21. Doi/10.1111/1752-1688.12794

3.5 Fish biotic integrity

Supplementary comments on relevant recommendation from primary report

Amend recommendation 9 to:

- Provide for two tables specifying numeric biophysical values for fish biotic integrity, using the Fish Index of Biotic Integrity (Fish IBI) and specifying different thresholds for bands and national bottom lines – one including salmonids and one excluding them.
- Clarify that decisions on which table to apply should be made locally, and that national direction is required to clarify the circumstances in which the tables are applicable.

STAG discussion

Officials did not ask any specific questions, but did note that many submitters made points in relation to whether introduced salmonids should be included in the measures of Fish IBI.

Salmonids fill ecological niches that require a high level of ecosystem health. Since Fish IBI is a measurement of ecosystem health – not nativeness – salmonids could be included in the attribute.

Despite filling ecological niches, salmonids may still compete with and/or predate on native species. There is potential for conflict between providing habitat for salmonids and for indigenous species.

If salmonids were part of the metric, this could complicate management at those sites where both salmonids and threatened native fish were present.

In the primary report the STAG's approach to the Aquatic Life component of the Biophysical Ecosystem Health Framework was to focus on indigeneity. That was reflected in the original recommendation to exclude salmonids from the IBI. Members acknowledged that the decision to focus on indigeneity could be interpreted as straying into the territory of policy.

This issue could be addressed by providing for two columns in tables specifying bands and bottom lines for the Fish IBI metric. This would enable flexibility for managers to respond to contextual factors – in some ecosystems the presence of salmonids will be considered positive because it indicates healthy habitats, while in others it will be considered negative because of the impacts of predation on indigenous aquatic species.

Caveats

Members are unsure of:

- what requirements must be met before sport fishery management plans are signed off,
- whether the framework requires the 'weighing-up' of potentially competing values of sports fisheries and indigenous ecology, and

- the implications of new indigenous fish conservation regulations and their relationship to decisions on when and where to apply different columns ('salmonids included' versus 'native species only') when specifying bands and bottom lines using the Fish IBI metric.

National direction will be required to specify the circumstances when each column is applicable and to clarify the different responsibilities of regional councils, Department of Conservation, and Fish & Game when making these decisions. These organisations could combine efforts/plans and work together constructively to help identify which aquatic ecosystems are managed for sport fishing and which are managed for fish biodiversity.

STAG member perspectives

One member notes that amending the management measures and thresholds for Fish IBI to provide for salmonids potentially exacerbates the risk of conflict between those seeking to protect indigenous fish and those seeking to enhance the exotic sports fishery.

Some members register concerns regarding the proposal to introduce Fish IBI into the NOF as an attribute, owing to:

- naturally-low site-specific fish diversity and the prevalence of migratory fish species in rivers in Aotearoa/New Zealand
- the inherent subjectivity of the methodology proposed, and
- the need for more detailed and independent evaluation of the methodology and rationale used to derive the proposed numeric attribute states for the fish IBI.

3.6 Macroinvertebrates

Supplementary comments on relevant recommendation from primary report

Amend recommendation 10 from the primary report to clarify that:

- MCI and QMCI should be assessed together, and the lower of the two results should apply.
- ASPM is a separate metric and should be assessed separately.
- The national management framework should rely on the updated MCI scores.

STAG discussion

In response to submissions, officials asked STAG to consider whether the proposed bottom line of 90 for MCI is achievable in urban streams and how much more rehabilitation would be required to get 90 as opposed to 80.

STAG is not able to answer questions on the achievability and level of rehabilitation required to attain an MCI score of 90 in urban streams, as these will both vary and depend on site-specific factors. Habitat rehabilitation and contaminant management needs will vary between sites. Some streams would require habitat restoration, some would not. Some would require significant hydrological reengineering, some would not.

As expressed in the primary report, the key arguments for STAG's recommendation of establishing a national bottom line for MCI of 90 were:

- By definition, an MCI score below 90 indicates the waterbody is approaching a 'severely degraded' state. Members considered that a narrative description of 'severely degraded' was not appropriate for the threshold between 'C' and 'D' bands. In other words, members do not support establishing a management framework that allows communities to maintain a waterbody in a state approaching 'severely degraded'.
- The discriminatory power of the MCI deteriorates as the value drops below 90, reducing the technical effectiveness of the metric.
- Wherever possible, management aspirations for urban and rural environments should be consistent.

Caveats

Invertebrate scores for soft-bottomed streams should only be used in streams with naturally soft sediments.

Some regional councils have their own versions of the MCI, which is not helpful for consistent national evaluation, data-aggregation and reporting. On the other hand, regional councils are tasked with evaluating the effectiveness of their policies, plans and rules – which potentially makes it appropriate for regional councils to develop local variants of some tools to match their context.

This issue needs to be resolved. The STAG considers that that national environmental monitoring standards for macroinvertebrates are required urgently, while acknowledging that this may cause some issues with regards to data continuity in some regions.

These standards should clarify that:

- MCI and QMCI should be assessed together, and the lower of the two results should apply.
- ASPM is a separate metric and should be assessed separately.
- The national management framework should rely on the updated MCI scores.

STAG member perspectives

There is broad consensus amongst members.

3.7 Macrophytes – lakes

Supplementary comments on relevant recommendation from primary report

Amend recommendation 11 from the primary report to:

- Modify proposed tables for lake submerged plant index to reflect that it is considered appropriate to conduct an assessment every three years for at-risk lakes, and every five to ten years for lower-risk lakes.
- Clarify that eradicating exotic macrophytes entirely is unnecessary to reach the bottom lines proposed, and active management of exotic species would be suitable in many instances
- Provide for an exception that allows environmental managers to leave exotic species in place in circumstances where this is necessary to maintain ecosystem health, i.e. where exotic species make up the entire (or close to the entire) macrophyte community or increase resilience to the effects of storms by stabilising the bed of shallow lakes.

STAG discussion

Officials did not ask any specific questions. They did note that some councils may want to leave exotic macrophytes in place as they may have ecological benefits – and that some submitters raised concerns regarding the practicality of measuring vegetation coverage on lake beds.

Members note that raising the Lake Submerged Plant Index (LakeSPI) score usually involves increasing the percentage of native plants at a site. That said, eradicating exotic macrophytes entirely is unnecessary to reach the bottom lines proposed, and active management of exotic species would be suitable in many instances.

Members consider it appropriate to conduct an assessment every three years for at-risk lakes, and every five to ten years for lower risk lakes. Members are comfortable with proposals to include an additional measure of vegetation composition.

In response to comments on the practicality of measuring vegetation coverage on lake beds, members noted that LakeSPI includes a measure of percentage cover based on transects and doesn't require a vegetation cover map for the entire lake. Members also note they are aware of an Envirolink proposal to examine the potential for remote sensing (ie, underwater camera systems) to aid the assessment of lakebed vegetation cover.

Caveats

Members agree that caution is required, as the proposed attributes may have unintended consequences for some lakes where it may be preferable to leave exotics in place for ecosystem health. This is only likely to occur in a few instances, i.e. where exotics make up the entire (or close to the entire) macrophyte community or increase resilience to the effects of storms by stabilising the bed of shallow lakes. In such cases, an improvement would require de-vegetation of the lake, which may not be desirable. This could be dealt with by providing for exceptions in extreme circumstances.

Members consider that more research into restoring native plants in lakes is necessary, and that national guidance will need to be revisited when further information is available.

STAG member perspectives

There is broad consensus amongst members.

3.8 Sediment

Supplementary comments on relevant recommendation from primary report

For suspended sediment

- the attribute tables should be described in terms of visual clarity and should allow for the site-specific inter-conversion of suspended sediment, turbidity and visual-clarity measures.
- a lower level of classification (e.g. Level 2) should be used for developing attribute tables, with thresholds derived using the community deviation method.

For deposited sediment

- a lower level of classification (e.g. Level 2) should be used for developing attribute tables, with thresholds derived using the community deviation method.

STAG discussion

In response to submissions, officials asked STAG to consider the following questions:

a. Do members consider:
    - the proposed band thresholds to be scientifically robust?

- that the extirpation method (Franklin *et al* 2019 Appendix H) is an appropriate alternative to the community deviation method?
- that the GLM method using sediment MCI (Franklin *et al* 2019 Appendix I) is an appropriate alternative to the community deviation method?

b. Is there a clear alternative indicator and monitoring method to percentage areal fine coverage, as determined using the SAM2 in-stream visual assessment method, that would be appropriate in soft-bottomed streams?

c. What might be the negative consequences of allowing for inter-conversion of turbidity and visual-clarity measures?

To assist STAG to address these questions, officials arranged for two independent peer-reviews of the research commissioned by the Ministry for the Environment and submitted to the STAG in early 2019, which underpinned the recommendations on sediment management in the STAG's primary report (see **Appendix 5**). Members were supplied with the results of these peer-reviews and participated in a round-table discussion with the original authors of the research and the peer-reviewers at the STAG meeting in January. The key points raised in this discussion are presented below in separate sections relating to suspended fine sediment and deposited fine sediment.

Suspended sediment

The visual effects of suspended fine sediment on aquatic species occur at low levels and are due to a reduction in visual clarity, which makes it more difficult for visual feeders to capture food. This means that median visual clarity is a relevant indicator of these effects as it is a time-averaged measure of what the visual feeders are 'seeing'.

Suspended sediment also has direct impacts on biota (clogging gills and causing abrasion) when levels are high, which typically occurs during high flow events. There are three ways to measure the amount of suspended material in a water column: turbidity; water clarity; and the concentration of suspended sediment. Conversion between the three measures can be made, as long as their relationship to each other has been established for the freshwater body being monitored. Both turbidity and visual clarity can be used to establish levels of suspended sediment. Visual clarity is a direct measure of the effect on aquatic species (sighting range of visual feeders). As noted, site-specific validation of indicators is essential because visual clarity depends on both suspended sediment characteristics and water colour (absorbance), which varies from site to site.

Members have a strong preference for continuous monitoring. Continuous monitoring is extremely important for establishing a reliable understanding of the nature and causes of elevated levels of suspended fine sediment.

In most instances there is a tight relationship between turbidity and visual clarity, and between turbidity and suspended sediment concentration – as turbidity goes up, suspended sediment concentration goes up and clarity goes down. Continuous sensors for turbidity are available, but can deliver variable results – it has been noted that two sensors of the same make and model side-by-side in a river can return different measurements. These variations in measurement do, however, appear to be consistent, which means this issue can be addressed by calibrating each sensor at a site with measurements of visual clarity and suspended sediment taken at the same site.

Allowing for site-specific inter-conversion of suspended sediment, turbidity and visual-clarity measures would be advantageous. It would allow for the complementary use of monitoring techniques to deliver a continuous annual record in this management category.

For the suspended sediment attribute, two methods for deriving sediment attribute tables were considered: community deviation (used to derive current thresholds); and an extirpation method.

The differences between the two modelling approaches are summarised in Table 1 below.

Table 1: Summary of different modelling approaches for deriving attribute tables for suspended fine sediment.

Model	Community deviation	Extirpation
Training data	Fish community	Macroinvertebrate community
Thresholds derived by	Deviation from reference predictions	Absolute change in community
Ecological meaningfulness	Probabilities of relative change less clear to determine ecological endpoint	Ecological endpoint clear
Spatial classification	Applicable at all levels of aggregation	Limited to Level 2 aggregation
Outcomes	More protective bottom lines	More permissive bottom lines

Uncertainty in the community deviation method occurs at multiple steps: error in sediment data collection as recorded in the New Zealand Freshwater Fish Database (NZFFD) used in spatial classification; error in the River Environment Classification (REC) network as used in the spatial classification; error in probability of occurrence of fish taxa; error in data from turbidity meters; and error in relationships between fish probability and sediment measures within each of 12 classes.

Uncertainty in the extirpation method is limited to use of the spatial classification and turbidity meter variance.

STAG is supportive of the use of the community deviation method in principle but concerned with the potential for propagated error at Level 4 class aggregation. The error may be higher than the band differentiation. There was enough data at the Level 2 class aggregation to be confident that ecological relationships were meaningful. However, lowering the level of aggregation will reduce accuracy and lead to a greater number of modelled under and overestimates when these thresholds are applied in the field.

The proposed NOF attribute table results in very fine-grained differences between bands and STAG is concerned it may not be technically feasible to distinguish between them. The monitoring methods/tools that would be necessary to make this approach work in practice are not available. STAG recommends use of a lower level of classification (e.g. Level 2) for attribute tables for suspended fine sediment, with thresholds derived using the community deviation method.

Deposited sediment

Deposited fine sediment has a clear impact on benthic communities in hard bottomed streams but we don't currently have a good understanding of the effect of excess sediment deposition on naturally soft-bottomed streams.

Methods for measuring deposited sediment (e.g. SAM2) were developed for hard-bottomed streams and are not meaningful when applied to soft-bottomed streams.

It is not practical to measure deposited sediment in these cases It may be better to treat soft-bottomed streams (7% of the digital river network) as if they are a 'receiving environment', similar to a lake or estuary, in which case sediment delivery and deposition rates may be more meaningful.

However, while there are methods for measuring rates of sediment deposition in streams, these methods have not been standardised or explored in relation to ecological measures to date. STAG suggests exclusions may be acceptable in naturally soft-bottom streams.

Further, in many instances where sediment is being introduced to soft-bottomed streams it is the impact on downstream receiving environments that we will be most concerned about, in which case measures of suspended sediment are likely to be more relevant/appropriate.

Members generally agreed with peer reviewer concerns about the implications of introducing so many river classes into the NOF for the management of deposited fine sediment.

The peer review process suggested an alternative approach of adopting a bottom-line threshold for deposited sediment of 20% fine sediment cover and dispensing with management bands. Key rationale for this approach were:

- It would provide a simple threshold beyond which impacts would be significant for approximately 93% of rivers in Aotearoa/New Zealand and would make it very clear when and where management intervention will be required.
- A single bottom line would be easy to interpret and apply and would avoid the risk of creating a false expectation that it will be possible to progress through management bands – in many instances it will extremely difficult to remove deposited sediment from rivers.

An obvious downside of this approach is that it would dramatically reduce management resolution. STAG has confidence in the proposed bottom lines and A bands and considers it is useful to have aspirational A/B and B/C bands, even if there is less statistical confidence in the numbers relied upon to determine thresholds – as more data are collected confidence will increase.

For the deposited sediment attribute, two methods for deriving sediment attribute tables were considered: community deviation (used to derive current thresholds) and a general linear model.

The reviewers considered the evidence for bottom lines based on the community deviation method to be adequate, clear, and reflective of current knowledge (notwithstanding the concern raised about the classification system). However, in relation to the community deviation method at least, both reviewers and numerous submitters stated that the evidence for differential ecological effects between bands was weak.

The differences between the two modelling approaches are summarised in Table 2 below.

Table 2: Summary of different modelling approaches for deriving attribute tables for deposited fine sediment.

Model	Community deviation	GLM
Training data	Macroinvertebrate community	Macroinvertebrate metrics
Thresholds derived by	Deviation from reference predictions	Deviation from reference predictions
Ecological meaningfulness	Probabilities of relative change less clear to determine ecological endpoint	Probabilities of relative change less clear to determine ecological endpoint
Spatial classification	Applicable to all levels of aggregation	Applicable to all levels of aggregation
Outcomes	Slightly more permissive bottom lines	Slightly more protective bottom lines

Uncertainty in the community deviation method occurs at multiple steps: error in sediment data collection as recorded in the NZFFD used in spatial classification, error in the REC network as used in the spatial classification, and error in relationships between macroinvertebrate and sediment measures within each of 12 classes.

Uncertainty in the GLM method is limited to use of the spatial classification.

STAG was supportive of the use of the community deviation method in principle but was concerned with the potential for propagated error at Level 4 class aggregation. There was enough data at the Level 2 class aggregation to be confident that ecological relationships were meaningful, although this will lead to more model under and overestimates when applied in the field. STAG recommend the use of a lower level of classification (e.g. Level 2) with thresholds derived using the community deviation method.

Caveats

In relation to suspended sediment some members note that:

- Visual clarity relies on manual sampling methods which can be influenced by individual factors (i.e. operator experience) and ambient light conditions and can suffer from practical constraints (i.e. in deeper and faster-flowing streams or during high flows).
- Relying on monthly median measurements of visual clarity taken over a two-year period might fail to capture the events managers are seeking to observe.

In relation to deposited sediment, no caveats were recorded.

STAG member perspectives

In relation to suspended sediment, some members note that:

- The community change metric is based on a regression incorporating predicted reference state, which is taken from the median model output indicator value. Some members consider this is not precautionary enough as it would mean half of the measurements would be worse than the median, and consider that the 20% community deviation figure should be reviewed.
- Continuous measurement of turbidity is the ideal method and metric for the monitoring and management of suspended fine sediment. Turbidity can be converted directly to suspended sediment and to water clarity for specific water bodies, once a calibration has been established. This will be more reliable than converting water clarity to turbidity, and then turbidity to suspended sediment. Water clarity and suspended sediment often do not show a tight relationship because of the uncertainty in water clarity measurements, or non-sedimentary material affecting water clarity (e.g., algae).
- There may be a need for exclusions in some circumstances to account for the following factors:
    - The relationship between turbidity and visual clarity is probably truer within rivers than among rivers because dissolved humic matter can strongly affect visual clarity but doesn't affect turbidity.
    - Glacial flour can strongly affect turbidity but is natural.
    - Rivers downstream from large lakes are likely to have a naturally very low turbidity.

In relation to deposited sediment, there is generally broad consensus amongst members although one member considered it would be preferable to adopt a bottom-line threshold for deposited sediment of 20% fine sediment cover and dispense with management bands.

3.9 Nutrients – rivers

Supplementary comments on relevant recommendation from primary report

A majority of members consider:

- Recommendation 13 should be retained without amendment – the methodologies and data sets used to derive the proposed criteria, bottom lines and thresholds for DIN and DRP for rivers are scientifically rigorous, well explained and well justified, have been discussed at length by the STAG and peer reviewed independently by Professor David Hamilton who generally supported the approach adopted.
- A note should be added to Recommendation 13 acknowledging that, while some rivers in acid-volcanic geological terranes may have naturally high levels of DRP, these rivers are

readily identifiable, equate to 17% of national stream length (70,899 km) and, where identified, can reasonably be dealt with by exception – although it would also be technically feasible for the management framework to treat these rivers as a separate class.

A minority of members consider:

- the evidence provided to establish nationally applicable bands and bottom lines is insufficient to provide confidence that a given DIN or DRP concentration will achieve desired improvement in ecosystem health or ensure that the target of a specific ecosystem health metric will be met.

- There are concerns about the reliability and effectiveness of nationally-applied nutrient criteria in managing for ecosystem health, given they have been derived from weak relationships that vary spatially. This could have the effect of not triggering a management response in rivers where this is necessary to protect ecosystem health and vice versa.

The minority sub-group recommends that recommendation 13 should be deleted, and that controls in the current NPS-FM on the effects of nutrients in rivers should be strengthened by:

- Giving effect to recommendation 8 of the primary report and strengthening the periphyton attribute in the current NPS-FM by providing a default nutrient table with spatially variable bottom lines and band thresholds, via guidance, as recommended in the primary report of the STAG, noting that these default bottom lines are nearly all more stringent than those proposed in recommendation 13 and would be applicable to at least 72% of national river length.
- Increasing the level of protection from toxicity by making the current bottom of the 'B band' the national bottom line for ammonia and nitrate. The current national bottom line provides for 80% species protection from chronic toxicity and the minority group's recommendation is to raise this to 95% species protection from chronic toxicity which is more consistent with other ecosystem health protection measures recommended by the STAG.
- Introducing national monitoring requirements for DIN and DRP in rivers that, where increasing trends are detected in a freshwater management unit, would trigger the requirement to develop a management plan for reducing nutrient concentrations.

STAG discussion

In response to submissions, officials asked STAG to consider the following questions:

- Is sufficient information and justification provided in the supplementary technical report on the development of DIN and DRP attributes to resolve questions and issues raised by STAG members?

- Is any further peer review needed, and if so, what should the focus be?

The STAG's primary report noted the agreement amongst members that elevated nutrient concentrations are widespread and can alter ecological communities and processes through multiple complex pathways. The primary report also noted that almost all members:

- agreed the current provisions for managing nutrients in rivers in the NPS-FM were insufficient for maintaining or improving ecosystem health in rivers in which there is no conspicuous periphyton,
- agreed the bottom lines currently set by the NPS-FM for ammonia and nitrate toxicity are not sufficient for protecting ecosystem health, and
- supported the introduction of management categories, measures and thresholds for nitrogen and phosphorus for ecosystem health protection because both impact the structure and functioning of healthy ecosystems, and, while there may not always be a direct link and/or well understood mechanisms by which nutrients interact with components of a healthy ecosystem, ecosystems are dominated by indirect relationships and the framework for managing the health of freshwater must account for this.

For this reason, the primary report included the recommendation[7] to:

"Amend the national framework for freshwater management to introduce numeric biophysical tables for dissolved inorganic nitrogen (DIN) and dissolved reactive phosphorus (DRP) and specifying national bottom lines of 1 mg/L DIN as an annual median (and 2.05 mg/L as a 95th percentile) and 0.18 mg/L DRP as an annual median (and 0.054 mg/L as a 95th percentile)."

The primary report recorded the view held by a minority of members that the controls within the NPS-FM should:

- focus on managing the direct impacts of contaminants on ecosystem health,
- set management thresholds based on the percentages of species protection that is provided for at differing concentrations of target nutrients, and
- be amended to clarify the process for setting nutrient limits for ecosystem health.

The introductory section to the STAG's primary report expressed members' understanding that:

- the Ministry for the Environment intended to consider the group's recommendations prior to providing advice to the Minister for the Environment on possible changes to the NPS-FM and broader framework for freshwater management, and

[7] Recommendation 13, page 39, with proposed 'attribute tables' on pages 39 and 40.

- any proposed changes would be subject to a public submission process.

When developing the recommendations included in the primary report, members where reassured that the process of public submissions would bring public and practitioner experience to bear, as well as enable the contribution of scientists employed in the various sectors of the economy impacted by the STAG's recommendations. Members were aware that many of the group's recommendations were based on scientific judgements and, in the primary report, noted that the analysis and recommendations should be subject to peer review prior to being incorporated into national policy or guidance.

Since completing the primary report, members have had an opportunity to consider the perspectives expressed by submitters in response to the Government's proposed programme of regulatory change.

In addition, one member was tasked with compiling and expanding on the analysis that underpinned the STAG's recommendations in relation to the introduction of numeric biophysical tables DIN and DRP to the NOF.

The key findings of this process of interrogation were presented to the entire STAG at its meeting on 22/23 January 2020. A draft of the supplementary technical report on the development of DIN and DRP attributes was circulated to members on 3 February for review. A final version of that report can be found in **Appendix 6** to this supplementary report.

After having reviewed the report circulated on 3 February:

- all members agreed they have been provided with a suitably clear and detailed explanation of the approach taken to derive the proposed bottom lines and thresholds and have had appropriate opportunities to discuss this topic in-depth. Members agree that no further peer-reviews are required for members to be able to form their view on the derivation and the robustness of the DIN and DRP tables included in the government's proposed changes to the NOF.
- a majority of members reconfirmed their support for recommendation 13 from the primary report.
- A minority of members concluded that:
    a. they do not currently consider the justification of DIN and DRP thresholds and bottom lines sufficient – the thresholds and bottom lines set out in the primary report are dependent on weak (in some case are non-existent) relationships which are highly spatially variable, and
    b. the methodology used to derive the nutrient criteria set out in recommendation 13 of the primary report is not sufficiently robust to support the inclusion of a management category, thresholds and bottom lines for DIN and DRP for rivers in national regulatory tools. Further comment on this is provided in **Appendix 7**. There are questions about what, if any, gains in ecological health will be achieved if investment is made in catchments to achieve these targets.

A majority perspective on the commentary of the minority can be found at the end of appendix 7.

3.10 Wetland extent and condition

Supplementary comments on relevant recommendation from primary report

Amend recommendation 14 of the primary report to specify the most recent version of the wetland condition index should be used when evaluating the condition of wetlands.

STAG discussion

In response to submissions officials asked the STAG to consider the following questions:

- The Wetland Condition Index (WCI) methodology was published in 2004, but some councils use more recent iterations of it adapted especially for their regions, which version should prevail at a national scale?
- What are the differences between the regional versions, and if these were used in preference to the original 2004 version of the WCI methodology, how would this affect the consistency of indicators and scores when comparing versions to establish a national picture? There is a council project underway revising and clarifying the indicators, score descriptions and record sheets. When this is finalised it could contribute to version-control issues.
- Some councils have found variation in the way different experts have been scoring the WCI indicators and are uncertain the method works well with large wetlands with different ecological areas, which limits ability to detect real change. Similarly, some submitters queried whether there were instances where a relatively degraded wetland could achieve a high score using the WCI methodology. How subjective are the scores and how far does this influence the results?
- Given the WCI requires fieldwork and that access to wetlands on private land may not always be possible, are there other methodological options that use remote assessment to assess wetland condition?
- Some submitters noted that the extent of remaining wetlands varies greatly from region to region. Is using a threshold of wetland remaining within regions viable and, if so, what thresholds would be appropriate, and should there be different thresholds for different wetland types? Is an analysis based on political regions appropriate, or would something like biogeographical units be preferable?

The STAG noted that less than 10 percent of Aotearoa/New Zealand's original inland wetlands remain – in 2008 there were approximately 250,000 ha remaining of an estimated 2.4 million ha pre-human settlement.[8] This historical destruction affected all inland wetland types, although the impact

[8] FENZ database

on swamps was greatest, mainly because they were most prevalent prior to settlement and typically occurred on fertile lowlands.

Recent studies show that the extent of Aotearoa/New Zealand's remaining inland wetlands is continuing to decline despite national direction encouraging their protection. For example, a national study using 2001 – 2016 data shows a total of 214 wetlands (nearly 1,250 ha) were lost, with a further 746 wetlands declining in size.[9]

National direction requiring the generation of effective information on the location, extent and condition of wetlands will help environmental managers identify areas that require active protection and areas that would benefit from enhancement and contribute to objectives including water quality improvement, ecosystem functioning and flood attenuation.

There are currently two versions of the WCI. The 2004 methodology was updated to address new council reporting requirements and is being used by some councils. In practice, the two versions should be reasonably consistent as the same five indicators of ecological condition are used (changes in hydrology, physico-chemistry, area, flora, and fauna, compared to reference state). The council project currently underway on refining the indicators would clarify the best approach. In the meantime, the most recent version of the WCI method should be used.[10]

The WCI is semi-quantitative. If a user has undergone adequate training, then the measure is sufficiently consistent and able to track changes in condition reliably – similar to MCI and LakeSPI. In addition, consistency improves with experience as wetland field ecologists become more familiar with the system. The method is not the cause of subjectivity – capability is. The robustness of using similar rapid survey techniques in assessing wetland condition has been peer-reviewed internationally and has been found to be appropriate for environmental management.[11]

There were no examples provided of relatively-degraded wetlands achieving a high WCI score. In practice, any degradation from reference condition would be reflected in the individual total for the relevant indicator.

The WCI method accounts for large-scale, complex wetlands by separating out different wetland types (e.g., bog, swamp) within the larger area of the wetland and scoring these separately. This follows a well-established approach from the USA.[12] If a single score is required for the wetland, the individual scores can be averaged according to area of each constituent type across the entire wetland.

[9] Belliss, S., Shepherd, J., Newsome, P., Dymond, J. (2017). An analysis of wetland loss between 2001/02 and 2015/16. Landcare Research Contract Report LC2798 prepared for the Ministry for the Environment. Landcare https://www.mfe.govt.nz/sites/default/files/media/Fresh%20water/analysis-of-wetland-loss.pdf

[10] Clarkson B, Bartlam S 2017. State of the Environment monitoring of Hawke's Bay wetlands: Tukituki Catchment. HBRC Report No. RM 17-06. HBRC Publication No. 4228. 41 p. https://www.landcareresearch.co.nz/_data/assets/pdf_file/0008/181349/Clarkson_Bartlam_2017-LC2713_HBRC-wetland-monitoringHBRCcover.pdf

[11] Dorney J., Savage R., Tiner R., Adamus P. eds (2018). Wetland and stream rapid assessment: development, validation, and application. Elsevier, USA. ISBN: 978-0-12-805091-0.

[12] CWMW (2013) California Rapid Assessment Method (CRAM) for Wetlands, Version 6.1. California Wetlands Monitoring Workgroup. http://www.cramwetlands.org/sites/default/files/2013-04-22_CRAM_manual_6.1%20all.pdf

It is possible to use aerial photos to conduct an initial evaluation if access to private land is an issue. Condition can be inferred based on vegetation coverage and species makeup and the nature of drainage ditches.

It may also be possible to use models (e.g. LUCI) to estimate what ecosystem services and biodiversity outcomes wetlands could support if they are restored and remediated. This could feed into objective and limit-setting.

To manage New Zealand's wetlands effectively we need far better information on the extent and condition of wetlands. This will allow environmental managers to establish a good picture of the spatial distribution of different wetland types within a management area. This will then allow environmental managers to develop a workable framework for identifying management objectives and priorities. Priorities will also be guided by assessing the representativeness of current wetland extent, distribution and type compared with former wetlands.

Members do not consider that the relative extent of wetlands and wetland types between political regions is biologically relevant given the magnitude of loss at a national scale. If the policy relating to monitoring and assessing the extent and condition of wetlands applies nationally then it doesn't matter whether the management is based on biogeographical or political areas.

Caveats

None recorded.

STAG member perspectives

There is broad consensus among members.

Part 4: Recommendations for further work

The primary report of the STAG identified a series of topics that urgently require additional work.[13] While noting that all these areas of work were important, members were particularly concerned that the current framework for freshwater management has important gaps relating to:

- Ecological flows (variability and minimum flows) for rivers and levels for lakes, wetlands and groundwater,
- Microbiological guidelines for the management of recreational waters,
- Toxic cyanobacteria in rivers, monitoring methods, tools for and evaluating risks, and thresholds for management action,
- Understanding and protecting groundwater quality, which is a need that goes well beyond simply preventing nitrate-nitrogen elevation in spring-fed streams and rivers,
- Nationally consistent methods for monitoring compulsory values, guidance on the design of systems for data generation and analysis, and
- Applied science to describe what is required to lift ecosystem health to meet community objectives and support adaptive management.

Members would like to underscore the importance of undertaking this work and would like to highlight the importance of completing work currently underway to refine and clarify the WCI (Wetland Condition Index) methodology.

Members agree that national guidance and worked examples are urgently required to aid the consistent interpretation and application of the NPS-FM and NOF. In particular, members consider that guidance should be provided to clarify:

- Exactly what constitutes the ‘current’ state of freshwater and the extent of variability that will be acceptable when defining this state,
- Which statistical tests should be applied to determine whether measures of ecosystem health and water quality have been maintained, improved, or have declined – along with protocols on these statistical tests and how to interpret them to ensure consistent use and reporting,
- How to test for nutrients when they fluctuate due to biological uptake,
- What monitoring and statistical methods and reporting tools should be used in what circumstances, to ensure consistency,
- The value of collecting other information that will allow interpretation of any ecosystem health and water quality changes observed,
- How to determine what level of monitoring is enough to inform analysis,
- How to apply mātauranga Māori and Māori indicators of ecosystem health,

[13] See Part 4 and Recommendation 15, pp 47-49

- How to account for unavoidable or predicted declines due to past management activities (i.e., lag effects) – this issue exists with the current 'band test' but would become more acute with the more stringent test being proposed here,
- How to account for the influence of natural barriers (i.e. waterfalls) and non-natural barriers (i.e. dams) to fish passage, where exclusions to bottom lines may be appropriate, and how to balance the benefit of removing barriers to indigenous fish passage with the possible drawback of creating pathways for invasive exotic species, and
- How to deal with waterways that qualify as an 'exception' to the general requirements specified under the NPS-FM and NOF.

Appendices

1. STAG Terms of reference.
2. Ministry for the Environment memo: STAG sub-group meeting on 'maintain or improve', 21 January 2020.
3. STAG technical paper: 'Are ecosystem health attributes redundant?' circulated on 22 January 2020.
4. STAG technical paper: Subgroup paper on DO (Lakes) percentage saturation vs concentration.
5. Independent peer reviews of the NIWA work underpinning the proposed sediment attribute.
6. STAG technical paper: Nutrients in New Zealand rivers and streams: An exploration and derivation of national nutrient criteria.
7. Addendum to STAG technical paper: Summary of perspectives of STAG members who have suggested deleting Recommendation 13 of the primary report and a commentary in response.

Ministry for the
Environment
Manatū Mō Te Taiao

TERMS OF REFERENCE

FOR THE FRESHWATER SCIENCE AND TECHNICAL ADVISORY GROUP

November 2018

Version 1
Status: Adopted by Group
Document: TP 10133284

1. Purpose of the Document

This document defines the Terms of Reference (TOR) for the Freshwater Science and Technical Advisory Group (the Group). The document provides:

- contact details of key Water Taskforce staff for freshwater policy development
- information on the role of the Group and standards of conduct (Appendix 1)
- Conflict of Interest declaration (Appendix 2).

2. Contacts for Freshwater Science and Technical Advisory Group members

Director

- Martin Workman – Director – Water. Email: Martin.Workman@mfe.govt.nz

Managers

- Lucy Bolton – Manager Freshwater Policy – Responsible manager for the Group. Email: Lucy.Bolton@mfe.govt.nz
- Jo Burton – Manager Freshwater Policy. Responsible manager for the Group. Email: Jo.Burton@mfe.govt.nz

Freshwater Science and Technical Advisory Group Secretariat

- Jennifer Price – Senior Analyst. Email: Jennifer.Price@mfe.govt.nz

3. Background

The Freshwater Science and Technical Advisory Group (STAG) has been established to support the work of the Water Taskforce at the Ministry for the Environment (MfE).

The Water Taskforce is comprised of officials from MfE, the Ministry for Primary Industries (MPI) and other central and regional government organisations.

Since early 2018, the Water Taskforce has been examining options for broad and narrow reform of the National Policy Statement for Freshwater Management (Freshwater NPS) and on furthering the guidance available on the Freshwater NPS. To inform advice to Ministers on these matters, the Taskforce is building a scientific evidence base for freshwater policy options.

The Group has been established to support the Water Taskforce for the next two years.

4. Purpose and functions of the Freshwater Science and Technical Advisory Group

The purpose of the Group is to support officials on the Water Taskforce with science and technical advice on the Water Taskforce work programme, as requested by Water Taskforce officials throughout 2018 - 2020.

This supporting role is critical to ensuring the interpretation of science for policy development is accurate and to help improve Taskforce protocols to better manage the incorporation of science into the policy process.

The Group will:

- have a solid understanding of the fundamental purpose of the Freshwater NPS and the guiding principles of policy development
- advise on scientific evidence for freshwater policy development by:
  - reviewing science that underpins Freshwater NPS National Objectives Framework (NOF) attributes and other freshwater policy options
  - identifying any gaps in the science

  - improving the NOF attribute development process
  - improving protocols to better manage incorporating science into the policy process
  - providing overarching scientific advice and guidance as it relates to freshwater policy development.
- contribute to science and technical related guidance for councils to implement the Freshwater NPS
- provide science advice on issues raised in public submissions on proposed Freshwater NPS Appendix 2 attributes and wider freshwater policy.

Final decisions on policy advice, working with Ministers, management and provision of funding, budgets and financial aspects of the programme and the management of procurement processes remain the sole responsibility of the Water Taskforce and not the Group.

The Group will be supported by a secretariat from the Water Taskforce, who will:

- lead all administrative actions associated with the smooth operation of the Group including:
  - leading communication with the Group
  - organising meetings, including catering and arranging reimbursement of travel expenses for non-government members
  - distributing papers to members prior to meetings and minutes after the meeting.

Officials from the Water Taskforce will provide feedback to the Group on how science advice is integrated into policy.

5. Freshwater Science and Technical Advisory Group Membership

The Group includes members with varied expertise across a range of fields including data, science and technical matters related to freshwater and estuarine water quality, ecosystem health and processes.

Members and the Chair will be appointed by the Water Taskforce managers responsible for the Group. Members will be appointed until 2020, and may be reappointed for a subsequent term/s.

If the Chair is absent from a meeting, the Chair may designate an Acting Chair for that meeting. If the Chair does not designate an Acting Chair, then the Acting Chair shall be elected by simple majority of those members present at the next meeting when an election is required.

The Chair or another member may resign from the Group by notifying the Water Taskforce managers responsible for the Group in writing.

A member will lose their position if they miss two consecutive meetings without prior approval of the Chair.

Because members and observers are appointed in their personal capacity rather than as representatives of organisations, no proxies will be permitted to attend in place of members, except at the discretion of the Chair.

Water Taskforce officials may co-opt expertise for particular meetings at their sole discretion.

The Group members for 2018 - 2020 (the Members) are:

- Ken Taylor (Agresearch) (Chair)
- Dr Adam Canning (Fish & Game NZ)
- Dr Bev Clarkson (Landcare Research)
- Dr Bryce Cooper (NIWA)
- Dr Clive Howard-Williams (NIWA)
- Dr Chris Daughney (GNS)
- Dr Dan Hikuroa (University of Auckland)
- Graham Sevicke-Jones (Environment Southland)
- Prof. Ian Hawes (University of Waikato)
- Prof. Jenny Webster-Brown (University of Canterbury, Lincoln University)
- Dr Joanne Clapcott (Cawthron Institute)
- Dr Jon Roygard (Horizons Regional Council)
- Dr Marc Schallenberg (University of Otago)
- Dr Mike Joy (Victoria University of Wellington)
- Rawiri Smith (Kahungunu ki Wairarapa)
- Prof. Russell Death (Massey University)

6. Meetings

Face to face Group meetings will be one or two days long, held in Wellington, at least six times per calendar year, with catering provided. Water Taskforce officials will attend all meetings. If required additional meetings will be held on an ad hoc basis; these meetings may be face to face or held via telephone/video conference.

The secretariat will endeavour to organize meetings on a day that suits most Members. Members are asked to keep the secretariat informed if they are unavailable for particular dates. If a Member is unavailable for a meeting they may provide advice on a topic via email – preferably prior to the meeting.

The deliberations of the group will be recorded as meeting minutes and with the agreement of the Chair made available publically on the Ministry for the Environment website, to increase transparency.

7. Roles and responsibilities

Chair

The Chair has the following roles and responsibilities:

a) set meeting agendas, with the assistance of the secretariat, and approve meeting minutes
b) chair meetings, encouraging and modelling open communication where all members contribute effectively
c) determine, with assistance from the Water Taskforce managers, what action is appropriate if a member has a potential conflict of interest
d) seek written approval from the Water Taskforce before incurring any expenditure or financial commitment on behalf of the Group.

Members

All Group members have the following roles and responsibilities:

a) make every effort to attend each meeting and report anticipated absences to the Secretariat

b) prepare adequately prior to each meeting, review any papers provided prior to meetings and participate actively in meetings, contributing to actions when agreed
c) bring matters of significance to the attention of the Group and use professional perspectives to undertake analysis or prepare advice as required
d) contribute to email discussion amongst the Group about relevant technical issues
e) approve minutes of meetings
f) comply with the Standards of Conduct in Appendix 1
g) complete the conflict of interest declaration form in Appendix 2 and return it to the secretariat.

8. Interaction with other advisory groups

Cabinet recently agreed to establish a Freshwater Leaders Group. This group will be appointed by Ministers to test freshwater policy as it is developed. Cabinet has also agreed to establish Kahui Wai Māori as a key forum for engagement with Māori on freshwater issues.

Conversations and engagement with the Freshwater Leaders Group and Kahui Wai Māori are encouraged. However, Freshwater Leaders Group and Kahui Wai Māori will not direct or commission work from the Group. Freshwater Leaders Group and Kahui Wai Māori may pose questions to the Group in relation to their consideration of freshwater policy.

Communication between Science and Technical Advisory Group, Freshwater Leaders Group and Kahui Wai Māori will be facilitated by:

- Water Taskforce officials will provide an online portal allowing information sharing between the groups
- minutes from all the groups will be circulated to members of all groups
- there will be a regular newsletter update sent to the three groups
- meetings will include a standing agenda item allowing for updates from the other groups.

9. Remuneration and reimbursement of expenses

No remuneration is payable to members, and where members are employees of central government their employer is responsible for meeting all cost associated with their membership on the Group.

For university and non-central government employees or members not in paid employment, all reasonable travel costs will be paid for by the Ministry for the Environment. The Ministry for the Environment will, as a general rule, book all accommodation and travel for members. Where members book their own accommodation or travel, that person is entitled to have to have the actual and reasonable costs of Expenses for travel and accommodation (Expenses) reimbursed by MfE, if:

- MfE has given prior written consent to the Supplier incurring the Expense
- the Expense is charged at actual and reasonable cost
- the claim for Expenses is supported by GST receipts.

10. Confidentiality

Members are expected to maintain confidentiality of matters discussed at meetings, where specified by Water Taskforce officials. After Ministers have made and announced decisions on issues considered by the Group, then members may comment as they see fit.

11. Conflict of Interest

The Freshwater Science and Technical Advisory Group members will be asked to formally declare real or possible conflicts of interest with the development of freshwater policy (see Appendix 2). These will be noted in the members' records and will be reviewed and accepted by the Water Taskforce Manager responsible for the Group.

Disclosure of interest can be:

- self-initiated
- raised by the Water Taskforce
- raised by other members.

Members should operate on the understanding that "if in doubt, disclose the interest". The appearance and perception of a conflict is just as important to manage as an actual conflict. The Water Taskforce Manager responsible for the Group will decide if there is a relevant interest and determine appropriate action.

Appendix 1 to Freshwater Science and Technical Advisory Group Terms of Reference:

The Freshwater Science and Technical Advisory Group Standards of Conduct

All members are expected to adhere to the following principles:

Conflicts of interest

A conflict of interest will occur when a member's private interest interferes, or appears to interfere, with an issue that faces the Group. A conflict of interest will also occur when there is a possibility that a benefit may apply to a sector, industry or organisation that they represent. A conflict of interest may be real or perceived.

Any situation that involves or may be expected to involve any real or potential conflict of interest must be declared immediately to the Water Taskforce Manager responsible for the Group, as soon as the conflict arises, using the form in Appendix 2.

At the discretion of the Water Taskforce, members may participate in discussions about issues in which they have declared a conflict of interest.

Guidelines for completing the Conflict of Interest Declaration Form:

Members of the Group may have direct or indirect dealing with organisations or persons, both commercial and other, which could lead to a perceived or actual conflict of interest. By disclosing interests, members ensure that they are accountable and that the integrity and public confidence in the Group is maintained.

Members should be pragmatic about disclosing interests and are not required to include an interest that is remote or insignificant so that it cannot reasonably be regarded as likely to influence the member from carrying out his or her responsibilities. In deciding whether a member is interested he or she should consider whether it would be reasonable to see the interest as likely to influence decision-making.

As a guide, an interest may be financial, professional, personal, direct or indirect and may include:

- you or your spouse, de facto partner, child, or parent may derive a financial benefit from the matter
- you may have a financial interest in a person to whom the matter relates
- you are a partner, director, officer, council member, or trustee of a person who may have a financial interest in a person to whom the matter relates
- *you are otherwise directly or indirectly interested in the matter*.

For example, the following types of interest might be relevant:

- employment/directorship within an institution applying to MfE or the Water Taskforce for funding
- interests in business enterprises or professional practices
- sharing ownership/beneficial interests in a trust
- existing professional or personal associations with MfE or the Water Taskforce
- professional and personal associations with organizations in the environmental sector
- a family relationship (including member with shares/ benefits in trusts etc).

Members may be concerned about the privacy of such information. Information held by the Water Taskforce is subject to the Official Information Act. Officials from the Water Taskforce the Water Taskforce will consult with the person who provided the information before making a final decision on release. If that person cannot be located, the Water Taskforce will consult with the Chair on behalf of that person. Other members of the Group will be aware of disclosed interests and have a duty to notify the Water Taskforce of any failure of any member to comply with obligations to disclose interests.

Confidentiality and media

In order for the Group to operate effectively, members must maintain the confidence of the group, including maintaining confidentiality of matters discussed at meetings, and any information or documents provided to the group. Water Taskforce staff will identify whether Information provided to the Group is confidential. With the agreement of the Chair, members and observers may share information about the business of the Group.

Where information is already in the public domain (through no fault of a member or observer), the confidentiality requirements do not apply to that information.

Members and observers must refrain from representing the Group, or commenting on the business of the Group, to the media.

Where information is not already public;

1. The Chair may seek agreement from the Water Taskforce for the Group to release a media statement.
2. A Member may only participate in a media interview or public statement about the business of the Group if they have obtained the prior written approval of the Water Taskforce.

Privacy Act 1993

Members must at all times comply with the requirements of the Privacy Act 1993 and keep information about identifiable individuals confidential.

Official Information Act 1982

All information provided to the Group or by the Group to the Secretariat will be treated as official information under the Official Information Act 1982 and, subject to the requirements of that Act, may be released to the public if there are no grounds for withholding it.

If the Water Taskforce is considering releasing information about Group meetings or Group-authored documents under the Official Information Act 1982, the Water Taskforce will consult with the person who provided the information before making a final decision on release. If that person cannot be located, the Water Taskforce will consult with the Chair on behalf of that person.

Corporate opportunities

Members must not exploit any opportunity that is discovered through access to information within the Group for their own personal gain or that of any industry, sector or organisation that they represent.

Respect for others

Members and observers will treat each other and the opinions of others with respect at all times. Members will not take unfair advantage of anyone through manipulation, concealment, abuse of privileged information, misrepresentation of material facts or any other unfair dealing practices.

State Services Standards of Integrity and Conduct

State servants have statutory demands under the State Services Standards of Integrity and Conduct. In the case of any conflict between the obligations outlined there and the ones in this document, those of the Standards and Integrity of Conduct shall preside.

Appendix 2 to Freshwater Science and Technical Advisory Group Terms of Reference:

Conflict of Interest Declaration Form

An actual conflict of interest arises in a situation where a candidate's private interest interferes or appears to interfere with an issue that faces the Freshwater Science and Technical Advisory Group (the Group). Perceived or potential conflicts of interest exist in situations where a candidate of the Group, a family member or a close personal relation has private interests that interfere or appear to interfere with an issue that faces the Group (see Appendix 1 for further information).

A conflict of interest arises in a situation where there is a possibility that a benefit may apply to a sector, industry or organisation that a candidate may represent.

Name:	
⌋ I declare that there are no conflicts of interest could compromise my objectivity, judgement, integrity or ability to perform the responsibilities of the Group.	
⌋ I declare the following situation(s) that would cause a conflict of interest to exist	
Please describe how this conflict of interest will be managed:	
⌋ I declare the following situation(s) that may be perceived as a conflict of interest	

Please describe how this conflict of interest will be managed:

Date:
Signed:

Memo: STAG sub-group meeting on 'maintain or improve', 21 January 2020

This memo records outcomes of the STAG sub-group on 'maintain or improve', held on 21 January 2020.

Attendees

Nik Andic (MFE); Clive Howard-Williams; Adam Canning; Mike Joy.

Context

The purpose of this meeting was to have a frank discussion about issues raised in submissions, focussing on those that would benefit from technical input. This is intended to inform officials' advice to Ministers towards the end of February 2020.

Sub-group members were provided a draft summary of submissions relating to 'maintain or improve' proposals; and a table listing issues to focus on, initial options, pros and cons, and early thoughts/comments.

Officials did not have a settled position on any of the issues raised. Rather, the meeting was intended to be an open and frank discussion, and record comments that could assist officials to reach a position.

Crucially, the discussions were held before the Independent Advisory Panel's report and recommendations, or commentary from other advisory groups and departments were received. As such, the views expressed in this document are likely to change.

The remainder of this document is structured according to specific issues raised in submissions, and minutes of discussion about those issues. Detailed descriptions of the issues raised are not repeated here, and readers should refer to the summary of submissions directly.

Comments on specific issues raised in submissions

Current state should be defined relative to a past date

- Submitters suggested a range of dates as the appropriate time to 'maintain' from, although it was accepted this is a decision for Ministers.
- The core issues are likely to be: giving Ministers an accurate sense of what past water quality was (i.e. how much of an improvement would changes require); and the feasibility of determining the state of water in the past, particularly where data is not available.
- Long term trends are a sound way to give Ministers some indication of past state, although historical data is limited.
- Pegging the definition of a current state to a past date would require improvements (assuming past state is better than present state).
- It is possible (if not likely) that quality in larger rivers was worse pre-RMA due to point-source discharges. Any policy would have to account for the risk that past state was worse than present.
- Basing recommendations on the availability of data was suggested as a way forward, with the earliest realistic date being 2000 (note that Clive subsequently suggested 2011). For

example, ensuring the date is set at a point in time where there is sufficient data to provide a national/regional-scale assessment. Note that earlier dates are not impossible, but will have greater error and require assumptions and modelling.

Relationship between values, attributes, and maintenance

- Concerns raised by submitters do not fundamentally change our advice – the proposed set of attributes represent what the STAG considers to be necessary measures of ecosystem health, and there is no sub-set of these that is sufficient to assess the health of the ecosystem.
- The underlying rationale does not appear to have been communicated as clearly as intended. It will be helpful if the STAG report can include a clear and unambiguous statement about how and why the proposed set of attributes was selected.

Maintenance, improvement and degradation should be more clearly defined

- Discussions focussed on how the NPSFM can provide more direction and certainty on when a trend is sufficient to constitute improvement or degradation.
- There are two Envirolink projects already funded and underway that should be considered as part of any advice:
    - Trend assessment guidelines (Status: funding confirmed, project due to commence in early 2020); and
    - Guidance on Assessment of Maintain or Improve (Status: under negotiation with Envirolink, funding not yet approved.
- The scope of any definition is limited by our ability to describe maintenance, improvement and degradation at the national level (e.g. a quantitative test that would apply in all cases). However, it is still desirable to provide as much direction (and certainty) as is possible.
- Where we can specify how individual attributes should be measured (e.g. appropriate time period, sampling frequency, and distribution of sampling dates), then this should be specified within the attribute itself – it is desirable to provide this kind of direction, one of the reasons attributes were included is to provide increased certainty and avoid debate where possible.
- More generally, it may be desirable to specify a backstop (e.g. specify a 10 year time period for trend assessment unless otherwise specified in attributes).
- Scott Larned (NIWA) also provided a useful flow chart illustrating the process and decision points underlying any trend analysis (attached as Appendix 1 to this document). This could form the basis of more specific process requirements for assessing trends in the NPSFM. Council decisions as part of this process (i.e. defining time periods, minimum sampling frequency, distribution of sample rates, and desired confidence intervals) would be informed by the above Envirolink reports (subject to them being reviewed and fit for purpose), which could be published as guidance as soon as they are available.
- Responses to a trend (e.g. a decline) should be proportionate to the magnitude of the trend, and confidence interval applied (i.e. a matrix approach). A policy direction to increase certainty would be desirable to avoid perverse incentives to reduce certainty.
- It may be desirable to specify councils must use a range of confidence intervals (per IPCC guidance on consistent treatment of uncertainty), acknowledging management decisions should not be delayed because of uncertainty.
- A crucial part of this process not covered by the flow chart, is what the response should be if an environmentally significant trend cannot be identified. If it is a result of inadequate monitoring, then monitoring should be increased. If there is adequate monitoring, then a

deeper investigation is needed (i.e. the appropriate response is not straightforward and will vary according to circumstance).

Robustness of information to be used in determining current state

- This issue was seen as largely the same as the above, that maintenance, improvement and degradation should be more clearly defined.
- Again, the scope of any information requirements is limited by our ability to describe them at the national level (e.g. that would apply in all cases). Refer to bullet 4 in section above.
- It would be desirable to include policy direction that monitoring should seek to reduce uncertainty and avoid bias. These are fundamental criteria for any monitoring system that could be specific, even if councils are not given closer direction on how to design a monitoring system.
- Sub-group members provided useful examples of good monitoring network design, which could be referenced in information boxes within the NPSFM (i.e. Unwin et al 2014, and Larned and Unwin 2012). This was discussed as a possible approach to including greater detail in the NPSFM – attendees provided an exemplar monitoring network, a published paper and noted an R package for implementation is already used by two councils.
- While it is not feasible to address broader issues with NZ's monitoring system through the NPSFM now, there is a role for the broader STAG to comment on the importance/urgency of doing so through future work – including in response to the PCEs recent report calling for a more representative national water quality monitoring network.

Representative monitoring and spatial scale

- Underlying concerns behind submissions was the potential for localised declines to be hidden (i.e. if management is undertaken at too coarse a scale). Note some submitters were also concerned about the perverse consequences of too fine a management scale (e.g. inability to move resource use to better soils, etc).
- Reasoning for not regulating the scale of management was tested. In general, there was an acceptance that the appropriate scale of management couldn't be narrowly defined. There are situations where it's desirable to group similar catchments where the objectives and limits are likely to be the same. Similarly some catchments may be too large a scale.
- Policy direction to avoid perverse outcomes is desirable (e.g. ensuring that monitoring and management is at a scale that does not disguise localised degradation). Similarly, principles expressed in guidance on setting FMUs could be incorporated into the NPSFM as criteria.
- It is important to note that, because requirements to set target attribute states are now so specific and site-scale, the scale of the FMU is largely irrelevant – it has importance for the planning process but doesn't impact on what needs to be maintained. While regional councils and community can still determine the number and location of sites, water quality has to be maintained at the site (i.e. there is no scope for aggregation of sites or attributes). Councils still have the opportunity to comment on what changes across sites and attributes mean overall when reporting on results (see proposed reporting requirements in NPSFM), but it will not change whether they have met requirements to set target attribute states.

Target attribute states (to maintain, or otherwise) should not have to be set at physical monitoring sites

AND

The role of modelling

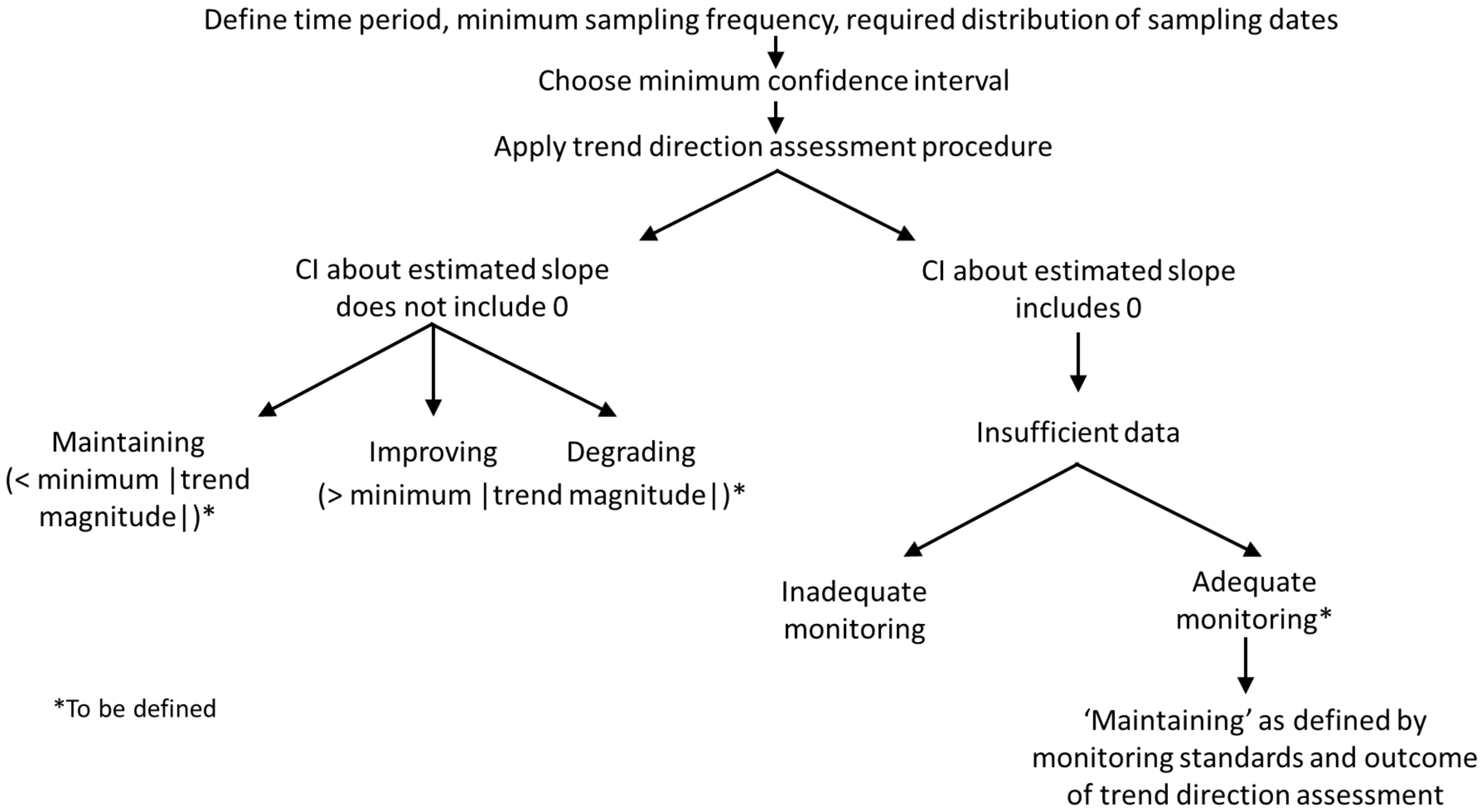
- Requiring target attribute states to be set for all physical monitoring sites was a drafting error. The policy intent was for target attributes to be specific about where they apply (i.e. where success will be measured, alongside how, when, etc), and that modelling state at a site is sufficient.
- Modelling is generally less certain than measurement, however discussions indicate it is still appropriate and desirable overall.
- Policy direction to improve certainty and validate models over time could mitigate risks associated with modelling.

Lag times and the 'load to come'

- This issue is about making sure we have assessed the impacts of proposals properly. Submissions were generally not requesting permission to degrade, rather, they considered the RIS did not adequately describe the impact of lag times which is more than just an opportunity cost.
- Officials will seek additional information from STAG members and regional councils to better assess where lags are expected, and the scale of impact (i.e. where we expect improvement will be required to maintain present state, and how long that lag might be).
- Discussions questioned whether short term lags (i.e. less than one regional planning cycle or less than the time it takes to establish a trend) pose a significant problem.
- Advice should convey the expected impact of climate change, which is analogous to a 'load to come' over a long timeframe.

Appendix 1

Decision flowchart for determining whether an attribute has improved, degraded or been maintained



*To be defined

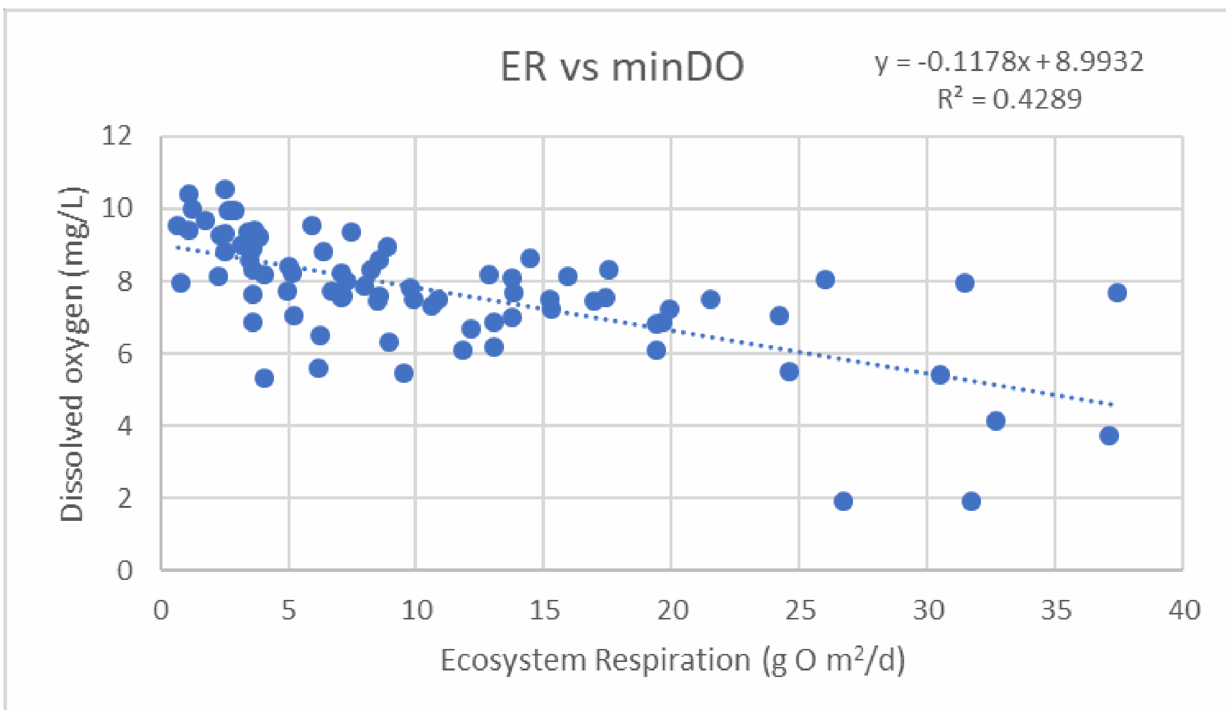
Are Ecosystem Health attributes redundant?

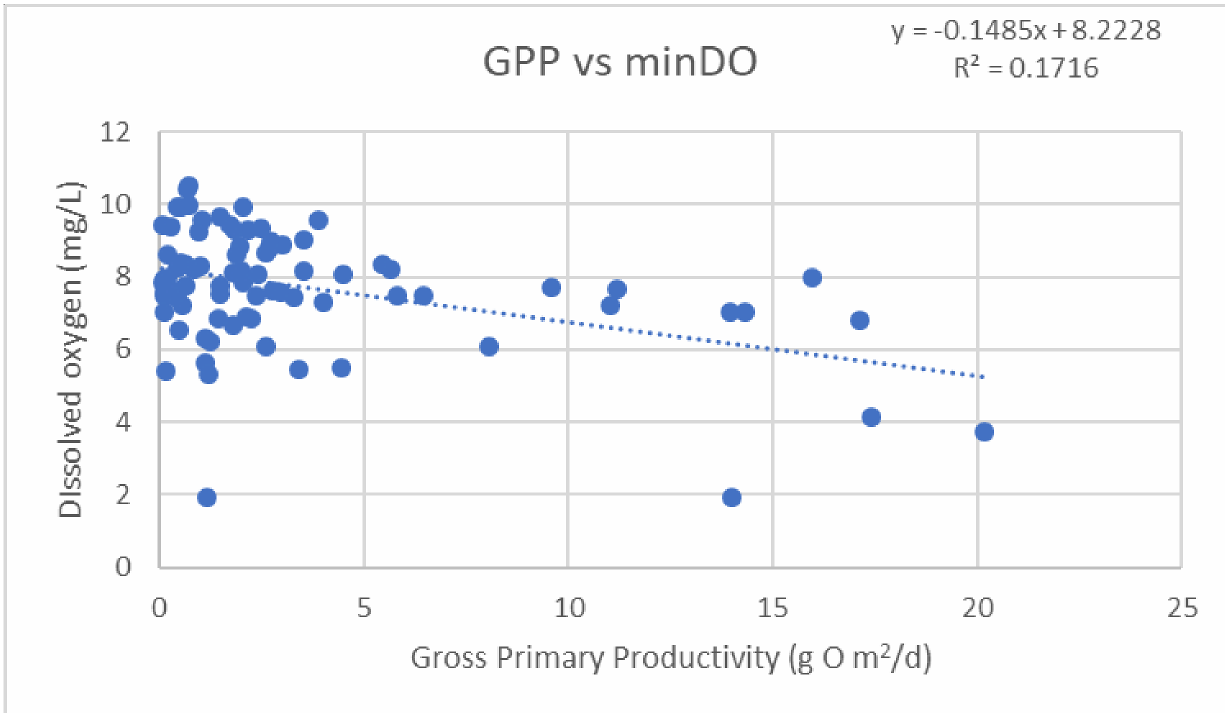
From: Joanne Clapcott – 22 January 2020

Background: The STAG meeting on November 27th reviewed the technical content of submissions relating to the attributes in the proposed revision of the National Objectives Framework (NOF). A question was posed to STAG in response to submissions: Are any of the ecosystem health measures redundant? How could this be assessed; do we currently have the data available to do this?

In partial response to this question (putting aside STAG support for the EH framework which outlines the need for five core components!), I explored the relationship between minimum DO (a measure of Water Quality) and the ecosystem metabolism measurements of ecosystem respiration (ER) and gross primary productivity (GPP) (measures of Ecological Processes). I used data from a single study of 82 stream sites from across New Zealand where dissolved oxygen was recorded for 48 hours in summer/autumn (Clapcott et al 2010). These attributes are likely to be the most closely related because they are calculated from the same data – continuous dissolved oxygen. The below graphs show the relationship between 1-day minimum DO and ER and GPP.

The comparison suggests a predictable relationship between 1-day DO minimum and ecosystem respiration (R^2 = 0.43), if data are collected on the same day. It is worth noting that DO below the national bottom line (< 4 mg/L) occur at very high levels of ER (poor ER ≥ 9.5 g O m^2/d). There is no consistent relationship between 1-day minimum DO and gross primary productivity. Low DO (< 4 mg/L) occurs at excellent (< 3.5 g O m2/d) and poor (> 7 g O m^2/d) levels of GPP. Results suggest there is no redundancy among these attributes and instead they provide complementary information for assessing the state and potential drivers of ecosystem health.





Reference: Clapcott JE, Young RG, Goodwin EO, Leathwick JR. 2010. Exploring the response of functional indicators of stream health to land-use gradients. Freshwater Biology, 55: 2181-2199

Essential Freshwater – Dissolved Oxygen (DO) attribute(s)

Notes to STAG regarding the units for dissolved oxygen in the National Objectives Framework

From: Marc Schallenberg, Clive Howard-Williams, and Jon Roygard – November 30th 2019

Background:

The STAG meeting on November26th reviewed the technical content of submissions relating to the attributes in the proposed revision of the National Objectives Framework (NOF). The meeting noted the substantive submission from the Hawkes Bay Regional Council regarding dissolved oxygen. This note addresses the suggestion from the council to: "Base the NPSFM on oxygen saturation, rather than concentration".

The submission argues that 'Using concentration to measure oxygen only serves to confuse and complicate resource management'.

Information:

While we agree that temperature and salinity (and altitude) play a minor role in oxygen availability to organisms, we note that measures of dissolved oxygen concentration in the field are usually carried out in concert with measurements of temperature (and sometimes in concert with measurements of conductivity) and that it is very easy to convert dissolved oxygen concentration to % dissolved oxygen saturation if and when needed with these additional parameters.

The submission of the Hawkes Bay Regional Council cited a paper by Verberk et al. (2011) as providing a rationale for monitoring % oxygen saturation instead of dissolved oxygen concentration in the NOF. Verberk et al. (2011) used a parameter termed the Oxygen Supply Index (OSI) (mol $m^{-1} s^{-1}$) that relates oxygen availability to both the partial pressure and solubility of oxygen in water. In the paper, the OSI was more strongly correlated to species richness in Ecuadorean streams than either oxygen concentration or percent saturation was. While we acknowledge the study of Verberk et al. (2011), we note that the OSI has received little (if any) further attention in the scientific literature.

The issue of reporting DO concentration (mg/L) in the NOF attribute tables was given careful consideration by Davies-Colley et al. (2013) in the report to MfE on the potential inclusion of temperature, dissolved oxygen and pH in the NOF(https://www.mfe.govt.nz/publications/fresh-water/national-objectives-framework-temperature-dissolved-oxygen-ph). That report provided considerable background detail on the DO requirements of New Zealand freshwater organisms.

The report states: "*Furthermore, specifying limits in the form of dissolved oxygen concentration (mg L^{-1}) was considered more appropriate than using saturation (as is currently used in the RMA). There is a generally good fit between dissolved oxygen concentration and ecological response thresholds in the existing literature. By defining a standard as a percentage of maximum saturation, the threshold dissolved oxygen concentration decreases as water temperature increases (i.e., 80% saturation at 10°C is 9.0 mg L-1 and at 25°C is 6.6 mg L-1). This seems counter-intuitive for ecosystem protection purposes given that the oxygen demand of aquatic fauna generally increases with increasing temperature*".

A recent report to Horizons Regional Council (Graham and Franklin 2017) argued: "*Continuous DO data recorded over the past six years from six HRC monitoring sites were analysed to determine whether current One Plan minimum DO targets are being exceeded, and if so, how often and for how long. The DO data were also compared to the dissolved oxygen limits recommended under the National Objectives Framework (NOF) for the protection of ecosystem health to determine which set* *of limits was more conservative. This comparison was of interest because the NOF guidelines are specified in terms of dissolved oxygen concentration (mg L^{-1}), which decreases with increasing water temperature, while the One Plan target limits are defined in terms of dissolved oxygen percent saturation. Less oxygen dissolves in water at higher temperatures, and therefore percent saturation can remain high or increase in warmer water, even as the concentration of oxygen declines. Therefore, standards expressed in terms of percent saturation have a higher risk of being under-protective for aquatic organisms. Additionally, the majority of studies reported in the literature on dissolved oxygen thresholds for New Zealand species investigated changes in oxygen concentration, rather than percent saturation*."

We discussed this issue with Paul Franklin (one of the authors of Graham & Franklin 2017) and he stated that the use of DO concentration is easier to understand than OSI, has a wealth of relationships with measures of organismal and ecosystem health backing it up, and that the effect of including partial pressure in the OSI was small compared to the changes in oxygen demand by organisms with increasing temperature. So his suggestion was to continue to use DO concentration.

Recommendation:

We recommend that the units of Dissolved Oxygen on the NOF remain as concentration units for simplicity in regulation with the proviso that temperature (and conductivity if this is at high levels) are also simultaneously measured so that DO concentration can be converted to DO % saturation is managers consider this desirable. This advice should be in the Guidance document for the NOF together with the equations allowing for the conversion of concentration to percent saturation if needed.

Peer review of community deviation method shown in Appendix J of Franklin et al. (2019)

Context

Franklin et al. (2019) provide a research framework that:

1.) characterises the relationship between fine sediment indicators and indicators of ecosystem health through a range of analyses, and

2.) uses a formal weight of evidence approach to combine multiple lines of evidence and define regulatory thresholds for National Policy Statement – Freshwater Management (NPS-FM) attribute states.

Through the weight of evidence process, the researchers concluded that the community deviation method (hereafter, the method) produced the weightiest evidence, and, by extension, its results would define the preferred regulatory thresholds.

Submissions on the proposed NPS-FM raised concerns that there was insufficient peer-review, testing, and validation of the method. In response to those concerns, the Science and Technical Advisory Group (STAG) requested a peer-review of the method. This review will support STAG's assessment of the robustness of the method for use in development of regulations.

Documents requiring review

The primary focus of the review is the method described in full in Appendix J of Franklin et al. (2019). The review will require general familiarity with the environmental classification system results shown in Appendix D of the same paper. The reviewer should not require more information on the general context and application of the research than that provided in the executive summary of Franklin et al. (2019).

The review only pertains to the method as described in Appendix J of Franklin et al. (2019). However, in order to understand the development and use of the method in prior research, the reviewer may wish to read Appendices DD and EE from Depree et al. (2019). That research was the forerunner and starting point for Franklin et al. (2019) and is the first publication to describe the method's development.

Purpose of commissioned review:

Review the method robustness, comparable to review for a journal submission.

In addition to any comments arising from the review, provide authors with suggestions on how they could more clearly articulate the following:

1. What ecological outcomes the proposed bottom lines protect/provide.
2. Descriptions of that level of protection in comparison to existing threshold values (e.g. Australia and NZ Guidelines for Freshwater and Marine Water Quality Default Guideline Values).

3. Appropriate statistical tests for, and descriptions of, sensitivity and uncertainty analyses that would improve the transparency of the method and results as well as ease understanding of the method and results in the research community.

The review will be provided to the authors for response. Ultimately, the review is intended to support the STAG when they deliberate on the robustness of the method and its application for the purposes of providing water quality regulatory thresholds.

Outputs and timeframes

1. The review will consist of a stand-alone memorandum of less than 3,000 words as well as comments and, if appropriate, track changes on the text of Appendix J. The review will be provided to the Ministry, STAG, and the authors by close of business 10 January 2020. The review memorandum will be made public on the Ministry's website at a time of the Ministry's choice. Payment for the review will be fixed.

2. Subsequent to the delivery of the review, the reviewer will be available for up to 4 hours to discuss and clarify comments with the authors and the Ministry. The reviewer will also be available to meet with STAG at a meeting date of STAG's choice in January or early February 2020. The reviewer's attendance may be in-person in Wellington (with travel expenses paid for by the Ministry) or via skype. Payment for reviewer attendance at meetings after delivery for the review will be invoiced according to billable hours.

School of Biological Sciences

College of Science
Tel: +64 3 364 2061, Fax: + 64 364 2590
Email: angus.mcintosh@canterbury.ac.nz

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Referee Report

"Appendix J, Deriving potential fine sediment attribute thresholds for the National Objectives Framework, Franklin et al 2019"

NIWA report for Ministry for the Environment

Overview
My brief was to assess the method described in full in Appendix J of Franklin et al. (2019). It was also suggested that undertaking that job would require general familiarity with the environmental classification system results shown in Appendix D and that I should not require more information on the general context and application of the research than that provided in the executive summary of Franklin et al. (2019). I found the Executive Summary difficult to follow and that it lacked justifications for various aspects of the approach I wished to understand, so I ended up reading much more of the full report.

I was also reluctant to take on this review job because when I first read the Executive Summary I thought that the whole approach taken was problematic. Essentially, I didn't want to review the details of a methodology that I thought was based on a flawed approach in the first place. However, I was encouraged to undertake the review in the spirit of providing useful feedback to the Science and Technical Advisory Group (STAG) and the authors. I have structured my report to document my fundamental problem with the approach followed by a critique of the methodology in Appendix J and a number of other more specific points about parts of the Franklin et al (2019) report. I couldn't adequately review Appendix J without dealing with the other issues mentioned.

Adjusting limits based on spatial variation in a sediment classification
Appendix J (and D) is based on the assumption that it is necessary to adjust fine sediment limits nationally according to landscape variations in fine sediment which result in natural variations in ecological communities. The rationale given in the report is:
"*Because the tolerance of different species to fine sediments varies, it can be expected that biological communities will vary in space concurrently with natural variations in sediment state, if the magnitude of natural variation in sediment exceeds the tolerance range of different species. Consideration must, therefore, be given to accounting for natural spatial variations in ecosystem structure and function associated with natural variations in sediment state when defining sediment attributes.*"

I agree those ecological communities are a good indicator of the impacts fine sediment, that fine sediment varies across the landscape, that macroinvertebrate and fish communities vary in response, but I don't agree that means limits should be adjusted to reflect those variations in natural communities in the way suggested. It is clear that more than high rates of deposited fine sediment cover degrade macroinvertebrate communities (appendix A) and that measures like %EPT are a good indicator of that degradation. In fact, a main conclusion of Appendix A is that, "… EPT metrics were a good indicator of deposited fine sediment effects" (p118). Similarly there is good evideince for effects on fish (Appendix C). However, just because a flat region of the country tends to accumulate more fine sediment, the impact on those ecological values (and life supporting capacity) isn't less! It means a flat area will be *more* prone to degradation, and it certainly doesn't mean the limits set there should be discounted.

Of all the stressors in the NOF, my experience suggests the ecological effects of fine sediment are the most clear and obvious, and the shape of the stressor-response relationship is the best defined. Various authors have called it a 'master stressor'. I think the approach taken is making what should be a comparatively simple process overly complicated. By applying a spatially dependent approach the methodology is also out of line with most other limits set for other NOF attributes. My recent experience at the coal face of stream restoration with managers and farmers also suggests twelve spatially varying sets of limits will be very difficult to implement or for the public to understand. Finally, taking this approach forces a reliance on a modelling approach for justifying limits which is very difficult to grasp and justify (see below). I found Table 1.1 incomprehensible after just reading with Executive Summary, for example.

Methodology in Appendices D & J

Determining a spatially-varying reference state for communities and thereby calculating deviation from that reference state is problematic for a number of reasons:

1. If I'm interpreting the classification and clustering of landscape-dependent fine sediment classes correctly (Appendix D), it is based on data which includes those from already impacted locations. If the classification is to accurately define classes, then it would need to be based on data from non-degraded sites. Thus, I worry that the classification is not a fair representation of the fine sediment state associated with locations because I can't see an attempt to deal with the already degraded state of many of the location used.

2. The absence of actual reference sites for both fish and macroinvertebrate data for a number of spatial classes considerably reduces the rigour because reference state has had to be 'estimated'. I also found it hard to distinguish the rationale for what was regarded as a reference site or to determine what the effect of 'reference site estimation' had on the relationships. It would be useful to see some comparison of sites that were estimated versus those not, and a more comprehensive justification of what constitutes a reference site.

3. There is an assumption (appendix D) that the REC framework represents the underlying processes that govern fine sediment distribution and this has not been adequately justified. I was expecting an analysis of actual fine sediment measures – did I miss it? For example, springs are an example of a habitat which are highly sensitive to fine sediment, but the REC is notoriously poor at identifying them in the landscape.

4. The Community Deviation (Appendix J) method had to 'fill in' missing data in some (many?) cases using adjacent sites. In my experience fine sediment is highly spatially variable within streams and most spatial variability is accounted for by within reach alterations. For example, unpublished analysis of multiple years of fine sediment data from 6 CAREX (www.carex.org.nz) sites in lowland Canterbury indicates that 62% of variation in sediment cover occurs at reach scales and between stream variation accounts for only 28%. This gets worse if local 'hot spots' of sediment are included. Thus, I'm sceptical that extrapolating from adjacent reaches (and certainly not different streams) would provide reliable measures of sediment cover.

5. One community assembly process was considered (competition/predation with trout) when creating models of community structure, but there are many other community assembly processes which are not included in the processes of determining which species might occur at a site. What about flood disturbance which is a main driver of community structure in both fish and invertebrates and we know it is not adequately represented by landscape variables?

6. The fish species modelled to produce the fish limits do not include a non-migratory galaxiid. This is an important gap given that they incorporate the most threated group of freshwater fish in New Zealand. They are also likely to be particularly vulnerable to fine sediment accumulations because they use interstitial spaces as refuges, especially during low flow (see comments on Appendix C below).

Overall, I remain unconvinced about the rigour of the Community Deviation approach. The methodology is complex and difficult to evaluate. My biggest concern is that it appears to rely on layer upon layer of prediction. Each stage seems to involve large amounts of extrapolation or estimation. In that case, one would want to see some verification using data which were actually independently measured rather than training data sets. Statistical measures of uncertainty would also be useful. There are many levels of prediction, but it is hard to evaluate what level of confidence we could have in them.

Comments on the Executive Summary (ES) of Franklin et al 2019

I found the terminology used confusing in places. 'Disturbance' is commonly used to refer to natural alterations to freshwater ecosystem structure (e.g., from floods), so to interpret the NPS-FM values as referring to "structure and function… expected under minimally disturbed conditions" is ambiguous. The term disturbance is being confused with 'altered'.

The 2nd sentence of paragraph 3 does not make grammatical sense, so the statement about the mechanisms leading to accumulation of fine sediment is not clear.

Para 4 – species and life stages of what?

The derivation of the sediment state classification is not explained.

The rationale for adjusting limits based on spatial variation in a sediment classification is not explained. This is an important omission because anybody wanting to quickly understand the derivation of the limits cannot do so based on the ES.

Comments on Appendicies A-C

Appendix A is difficult to evaluate scientifically because not enough results information is provided to allow an evaluation of rigour or outcomes. An important limitation is that it doesn't provide a clear mechanistic basis for proceeding with limit-setting because the mechanistic hypotheses shown in Figure 1-A are not explicitly tested, at least as shown by Table A-2. It also doesn't appear to have used NZ work which has examined the mechanistic drivers of fine sediment effects (e.g., Burdon 2013 PhD thesis, another chapter recently published as Burdon et al 2019).

The lack of helpful mechanistic analysis in Appendix A also makes Appendix B less useful because cause and effect relationships between suspended sediment and macroinvertebrates have not been rigorously determined. This work needed some insightful analysis utilising a smaller number of high quality datasets to determine the *shape* relationships between the stressors and ecological responses.

Appendix C presents a much more useful analysis because it links specific cause and effect mechanisms with both deposited and suspended fine sediment. However, one of the potentially most important relationships, the effects of fine sediment of access to low-flow refuges used by non-migratory galaxiids looks to have been missed. Work by Nicholas Dunn (2003) for alpine and Canterbury galaxias shows that fine sediment accumulation restricts the ability of alpine and Canterbury galaxias to burrow into the substratum intersticies, potentially affecting drought survival. Nicholas may has possibly carried out similar experiments on other threatened galaxiids (e.g., lowland longjaw galaxias). Glenjarman (2017) has also been overlooked.

Refrences

Burdon, F. J., McIntosh, A. R., Harding, Jon S. (2019) Mechanisms of trophic niche compression: evidence from landscape disturbance. Journal of Animal Ecology, doi: 10.1111/1365-2656.13142.

Burdon F.J. (2013). Impacts of sedimentation on the structure and functioning of agricultural stream communities. *Unpublished PhD thesis, University of Canterbury*.

Glenjarman, N. (2017). The impact of suspended and deposited fine inorganic sediment on New Zealand freshwater fishes. *Unpublished MSc thesis, University of Canterbury*.

Dunn, N. (2003). The effects of low flow on Alpine (*Galaxias paucispondylus*) and Canterbury (*G. vulgaris*) galaxiids. *Unpublished MSc (hons) thesis, University of Canterbury*

Angus McIntosh, 10 January 2020

Review of Appendix J, Franklin et al. 2019 – Gerry Closs

General Comments

My review of this document is based on reading selected parts of Depree et al. 2019, particularly Appendices DD and EE, and selected parts of Franklin et al. 2019, particularly the Appendix J, Executive Summary and relevant sections of most chapters.

My expertise relates to knowledge of the data sets on which the analyses are based, and a general understanding of the methods used to conduct biomonitoring. However, I am not a statistician - whilst I have a general understanding of the various statistical and modelling approaches used to derive the fine sediment attribute thresholds, I do not have the expertise to fully assess the appropriateness of the specific statistical/modelling approaches methods used.

Overall, as a theoretical exercise in determining relationships between landscape, sediment and community, the analysis clearly represents a methodologically cutting-edge approach to defining landscape-specific thresholds for detecting community-level responses to fine sediment inputs in streams.

However, the strongest community-level responses to sediment inputs mostly occur at relatively low levels of sediment input – presumably due to the loss of highly sensitive EPT taxa. In practice, this means that for many landscape classes, the range of sediment values from reference to the C/D threshold is relatively narrow. It is difficult to see how each A/B, B/C and C/D threshold could be accurately assessed across such a narrow range of values.

To sum up (and to play Devil's Advocate to some degree), the complex and rigorous analysis more or less supports what we already know – that highly sensitive EPT taxa are lost at relatively low inputs of sediment, with other elements of the community being less responsive to further inputs. I can see how the methodology could be used to set a National Bottom Line threshold – for most landscape classes, the difference between the reference and C/D threshold is sufficiently wide to be measurable using current sampling protocols. However, setting and assessing intermediate thresholds would be challenging in many areas.

Further, the lack of information on uncertainty in the models also means the uncertainty around the proposed thresholds is poorly understood. Whilst that may not be a problem for comparing reference with C/D thresholds, this again presents a problem in setting finer scale intermediate thresholds, particularly where the difference between the reference and C/D threshold is relatively narrow.

It may be just simpler to accept the wider message of the modelling – that at ~20% deposited fine sediment coverage (or equivalent turbidity or visual clarity value), sensitive taxa are lost from stony streams. In theory, the detail can be modelled, but in the messy real world of field sampling, assessing community responses to small changes in sediment input will be far more challenging.

Specific Comments

The first part of AppendixJ assesses the availability and quality of suitable datasets. As far as I am aware, the authors have accessed and made appropriate use of available relevant datasets. Overall, clear understanding of the data sets is demonstrated, and the strengths and limitations of each data set are clearly explained. Clearly problems exist with respect to the complementarity of the various data sets, particularly NZFFD data in relation to various measurements of sediment. The visual clarity and turbidity data have been 'in-filled' using modelled data – it is hard to assess the overall accuracy of this approach, although I accept it represents the best available option at present. The use of NZFFD sediment data would also seem to be the best available approach, although I would query the accuracy/consistency of the deposited sediment data in the NZFFD – at 'coarse scales', I would be confident in the data, but would query its consistency at finer scales given the wide range of expertise of the people recording data.

Selection of taxa contributing to the Community Deviation Method seems logical and appropriate. Understanding of the cofactors contributing to distribution also seems logical.

Whilst I do not have the expertise to comment on the specific details of the methods used to assess community change resulting from changes in the ESV state, the approach seems logical as far as I can follow it. I do not have sufficient expertise to suggest alternate approaches. The results that flow from this analysis intuitively correspond with the taxa-specific responses that I would expect to see.

The selection of a 20% deviation from reference community integrity values is appropriate, representing what would seem to be a fairly significant deviation from the reference condition. That said, I suspect that the 20% deviation in most communities is driven by the loss of EPT taxa, which would likely limit the responsiveness of communities to further change (see subsequent comments).

I note in Chapter 4, statements such as:

p. 56: However, it is noted that approximately 70% of the data used to build the model occur in the range of 0% to 25% deposited sediment cover.

p. 57: Data are again unevenly distributed across the deposited sediment gradient (as indicated by percentile rug plots on the x-axis) with approximately 70% of data below 30% cover in class L2.2

Does this suggest that rapid and substantial community responses to increased sediment occur at relatively low levels of sediment input, with relatively limited changes as further inputs occur?

This also suggests that if changes in community are occurring across a relatively narrow band of sediment input, then deriving A/B, B/C, C/D thresholds within that narrow band will be difficult.

Tables J2-J4, which present potential band thresholds for the SSC classes would appear to support the previous statement – the spread of values from the reference to C/D threshold can, in some cases, be quite narrow. It is hard to see how intermediate thresholds within the narrow range of many SSC classes could be reliably be assessed.

Nutrients in NZ Rivers and Streams

AN EXPLORATION AND DERIVATION OF NATIONAL NUTRIENT CRITERIA

A, D. Canning | 04-2020

This report should be cited as:

Canning, A. D. (2020) *Nutrients in New Zealand Rivers and Streams: An exploration and derivation of national nutrient criteria.* Report to the Minister for the Environment. Essential Freshwater Science and Technical Advisory Group, Wellington, New Zealand (131 pp.). DOI: 10.6084/m9.figshare.12116460

The Science and Technical Advisory Group (STAG) oversees the scientific evidence for freshwater policy development. The establishment of this group draws on useful discussions between freshwater scientists about policy development and the science behind the National Policy Statement for Freshwater Management (Freshwater NPS).

This group provides scientific and technical advice on the Essential Freshwater work programme and other Ministry for the Environment work. It has a role in ensuring the interpretation of the science for policy development is accurate. It also helps improve protocols to better manage the incorporation of science into the policy process.

The views and opinions expressed in this publication are largely (but not all) those of the author and do not necessarily reflect those of the entire STAG or the New Zealand Government. Furthermore, many components of the derivation of the principal nutrient criteria and the supporting analysis have arisen as an amalgamation of views and opinions from STAG and are not necessarily those of the author.

Contents

Executive Summary

Many of New Zealand's rivers are enriched with nutrients, primarily nitrogen and phosphorus, which can reduce ecosystem health via eutrophication. According to *Environment Aotearoa* (Ministry for the Environment and Statistics New Zealand, 2019), approximately 82% of rivers, by length, in pastoral land do not meet the ANZG (2018) default guideline values for nitrate-nitrogen and 77% for dissolved reactive phosphorus (DRP). Eutrophication is the increased production of organic matter. The most visible form of eutrophication is excessive algal growth, i.e., where rivers become covered in green slime. However, eutrophication does not always involve excessive algal growth, nor is algal growth a requirement. Nutrient enrichment can also increase and alter microbial activity, which can: alter invertebrate and fish growth (via changes in nutrient or macromolecule availability); increase hypoxia (which can suffocate invertebrates and fish); disrupt food web cycling and stability; reduce greenhouse gas storage and abatement; alter disease transmission; and alter the availability of nutrients and energy for downstream systems. Nitrogen and phosphorus are critical components of ecosystems and the species that inhabit them. Anthropogenic nutrient enrichment should be minimised to maximise an aquatic ecosystem's health.

New Zealand's National Policy Statement for Freshwater Management (NPS-FM) contains grades (or bands) for various measures (termed attributes) of aquatic ecosystem health. When a measure receives a 'D' grade, or a grade lower than desired by the community (via a regional planning process), then the environment must be managed in a way that, over time (decided by the community), improves that measure. This report outlines the approach adopted by the Science and Technical Advisory Group (STAG) to recommend dissolved inorganic nitrogen (DIN) and DRP criteria for consideration during the Government's *Essential Freshwater* reform. Using the same format as the NPS-FM, DIN and DRP criteria are presented as attribute tables with grades ranging from A-D band. In addition to the nutrient criteria derivation, the report includes additional analyses that sense-check the criteria.

Nutrient criteria:

- Nutrient criteria were developed by averaging multiple lines of evidence. Each line of evidence represented a relationship between DIN and DRP concentrations and another metric of ecosystem health. These metrics included periphyton, macroinvertebrates, fish and ecosystem functioning. Nutrient criteria were not designed to guarantee a specific ecological state; it is neither possible to summarise the health of an ecosystem into one, all-encapsulating metric, nor to determine all ecosystem outcomes. Rather the nutrient criteria are intended to be benchmarked to be similar stringency to other attributes. Improvements in nutrient concentrations will increase the propensity or likelihood of a healthy ecosystem.

- **DIN derived nutrient criteria grades are: A ≤ 0.24 mg/L, B ≤ 0.50 mg/L, C ≤ 1.0 mg/L and D > 1.0 mg/L.**
- **DRP derived nutrient criteria grades are: A ≤ 0.006 mg/L, B ≤ 0.01 mg/L, C ≤ 0.018 mg/L and D > 0.018 mg/L.**
- If there is strong evidence that a river reach would naturally exceed the D-band, then it should be exempt from assessment against these criteria. Options for handling the volcanic acidic geology class are included.

Additional supporting exploration:

- Predicted river-specific change-points between nutrients and the macroinvertebrate community index (MCI), suggest the large majority of DIN change-points occur between 0.8-1 mg/L. Whereas most DRP change-points occur between 0.017-0.023 mg/L. Rivers with volcanic acidic (VA) geology have a tendency for change-points to occur at higher concentrations (median = 0.021 mg/L) than other geologies (typically 0.017 mg/L), though there is considerable spread within the VA geology class.
- Boosted regression tree explorations consistently suggest that nitrogen is a highly influential predictor of the macroinvertebrate metrics examined when a range of temperature, vegetation, landscape, meso-habitat and hydrological factors are accounted for.
- Quantile regression was used to examine the nutrient concentrations that would, in the worst case scenario, likely limit the proposed MCI, QMCI (quantitative MCI) and ASPM (average score per metric) bottom-lines from being achieved, assuming no other factors are interacting with nutrient limits. The proposed C/D thresholds for both DIN and DRP were more protective than these levels.
- Gradient forest modelling was used to examine the turnover of macroinvertebrate assemblages. Out of the 30 environmental variables included DIN and DRP were the top two most influential factors. Turnover of macroinvertebrate assemblages was largely predicted to occur between approximately 0-1 mg/L of DIN and 0-0.05 mg/L of DRP.
- An exploratory model of observed/expected (O/E) fish species presence suggests DIN and DRP are strong predictors of O/E scores, and that scores decline until a plateau between 1-1.3 mg/L before continuing to decline. For DRP a sharp decline occurs

between 0.015-0.02 mg/L, followed by a shallow decline between 0.02-0.025 mg/L and then a plateau.

- Binomial regressions were used to examine the proportion of the 60 most common macroinvertebrate taxa that experience more than minor changes in their predicted probability of occurrence (i.e., >20% change from no enrichment levels) with increasing DIN and DRP concentrations. The analysis suggests that for DIN, 5% of taxa experience substantial impacts on their probability of occurrence at 0.20 mg/, 20% of taxa at 0.44 mg/L, 40% at 0.67 mg/L and 60% at 1.14 mg/L. Whereas for DRP, 5% are affected at 0.008 mg/L, 20% at 0.011 mg/L, 40% at 0.018 mg/L and 60% at 0.029 mg/L.
- Negative correlations were observed in regions and river classes across the country between nutrient concentrations and the macroinvertebrate metrics examined. No positive correlations were observed. Where there were no statistically significant correlations, these relationships usually failed to meet the basic assumptions of the linear regression. Often they had too few sites, were highly non-random, did not cover a substantial range of nutrient concentrations, or had influential outliers. Whilst simple correlations may be noisy or fail to yield significance, this does not necessarily mean a relationship does not exist (as lack of correlation does not mean lack of causation); they are still useful for direction of change and can usually be improved by adding additional and interacting factors.
- Nutrient criteria were also compared against concentrations that are (1) predicted to occur in reference condition and (2) measured in pristine habitat for different river classes. Across both, the measured and estimated DIN reference conditions almost always fell within the proposed A-band and never exceeded the proposed bottom-line. Whereas with DRP, reference concentrations largely fell in the A-band, sometimes B band, and some occasions in the D band. River classes with VA geology appear to have considerably higher concentrations of DRP in reference state, with some sites or estimates near or exceeding the proposed DRP bottom-line.

Background

Eutrophication (the increased production of organic materials) is a major environmental issue globally. It is primarily due to high levels of nitrogen and phosphorus, which are driven largely by agricultural intensification and urban wastewater discharges (Smith, Tilman and Nekola, 1999; Dodds, 2007). New Zealand is no different. Across New Zealand's state of environment monitoring network, nitrate-nitrogen and dissolved reactive phosphorus (DRP) loads are predicted to be 159% and 18% above natural levels respectively (Snelder, Larned and McDowell, 2018). According to Environment Aotearoa (Ministry for the Environment and Statistics New Zealand, 2015), approximately 82% of rivers, by length, in pastoral land do not meet the ANZG (2018) default guideline values for nitrate-nitrogen and 77% for DRP.

The excessive growth of algae is probably the most visible effect of nutrient enrichment and occurs when nutrients are the limiting growth factor. In addition to being aesthetically displeasing, excessive algal growth can (1) skew food webs towards autochthonous sources, promote the growth of algae consumers and alter the community composition, and (2) drive large diurnal fluctuations in dissolved oxygen, particularly when dead algae decompose. Hypoxic stress and alterations in energy pathways can alter fish and macroinvertebrate condition and assemblages. Enrichment can drive a river from being dominated with mayflies, stoneflies and caddisflies to one dominated by worms, snails and midges. The latter tend to be less nutritious for fish relative to grazing effort.

In similar vein to algae, heterotrophic bacteria are often nutrient-limited and nutrient enrichment can promote excessive growth. This, in turn, results in excessive decomposition of detritus, such as leaf-litter, which can also alter food web structure, cause hypoxia and increase acidity. A large meta-analysis by Ferreira *et al.* (2015), found that across 840 case studies, the average litter decomposition rate increased by 50%, with rates being highest when large amounts of nutrients were added to naturally oligotrophic streams. Nutrient-stimulated decomposition was also greater for plant matter with low nutrient content and high lignin, and in colder regions (Ferreira *et al.*, 2015; Manning *et al.*, 2015; Rosemond *et al.*, 2015; Jabiol *et al.*, 2019). Furthermore, litter decomposition is often faster when both microbes and invertebrates are present compared with microbes alone (Ferreira *et al.*, 2015; Manning *et al.*, 2015; Tant *et al.*, 2015). Microbes can mine their environment for nutrients and pre-condition detritus, making it more palatable and nutritious for invertebrate detritovores, and potentially more selectively grazed, thereby exacerbating invertebrate decomposition, and altering activity and abundance

(Bärlocher and Kendrick, 1975; Anderson and Sedell, 1979; Arsuffi and Suberkropp, 1989; Graça, 2001; Pasco*al et al*, 2003; Gulis, Ferreira and Graca, 2006; Woodward *et al*, 2012).

The growth of many invertebrates is also often nutrient-limited (Elser *et al.*, 1996, 2000; Benke and Huryn, 2010; Hessen *et al.*, 2013), with typically little flexibility (i.e., strong homeostasis) to adjust body nutrient stoichiometry to accommodate environmental limitations (Persson *et al.*, 2010). Nitrogen limitation can arise from the need to replace nitrogen-rich chitin from moulting exoskeletons and produce protein and nucleic acids (Elser *et al.*, 1996; Frainer *et al.*, 2016). Whilst phosphorus limitation can arise where high rates of protein synthesis (requires P-rich ribosomal RNA) is required. Under favourable conditions, the intergenic spacer (IGS) regions of rDNA tandem repeat units increase, as does the number of repeat units and rRNA transcription rates – all of which are associated with increased growth and production (Elser *et al*>, 1996, 2000; Hessen, Ventura and Elser, 2008; Hessen and Persson, 2009; Hessen>*et al*>, 2010). As a result, increases in either nitrogen or phosphorus can alter ecological communities by permitting the growth of nutrient-limited invertebrates (Cross *et al.*, 2003, 2006; Cross, Wallace and Rosemond, 2007; Danger *et al.*, 2013; González, Romero and Srivastava, 2014; Demi *et al.*, 2018). Elser *et al.*, (1996) provides an excellent and more complete explanation of mechanisms for nutrient limitations. Whilst some invertebrates have relatively plastic nutrient stoichiometry (e.g., driven by alterations in digestive enzymes relative to dietary nutrient restrictions (McCarthy, Rafferty and Frost, 2010; Wojewodzic *et al.*, 2011)), most have homeostatic nutrient stoichiometry (Persson *et al.*, 2010; Feijoó *et al.*, 2014), requiring a relatively fixed ratio of N:P. Furthermore, often invertebrates do not increase consumption to compensate for nutrient limitations (Stelzer and Lamberti, 2002; Fink and Von Elert, 2006), rather the growth of some invertebrates responds to the nutrient stoichiometry of food (Singer and Battin, 2007; Evans-White *et al.*, 2009; Guo, Kainz, Valdez, *et al.*, 2016b; Demi *et al.*, 2019). The phosphorus content, specific growth rate, and RNA content all negatively correlate with body size (Sutcliffe Jr., 1970; Bdmstedt and Skjoldal, 1980; Gillooly *et al.*, 2005; Hessen *et al.*, 2013). An experimental example, by Cross *et al.*, (2005), examined the influence of long-term nutrient enrichment on the growth of stoneflies and chironomids in a detritus (not algal) based stream. Nutrients were increased in a pristine stream over a 2 year period, resulting in DIN increasing from approximately 0.03 mg/L to 0.3 mg/L and DRP from 0.004 mg/L to 0.051 mg/L; whilst there was no growth effect on the stoneflies, the growth rate of chironomids increased by ~50% and production (by area) increased 183%, changing whole ecosystem nutrient stoichiometry (Cross *et al.*, 2003) Nutrient enrichment can, therefore, result in a dominance of small-bodied, fast growing invertebrates, such as chironomids and snails (Elser *et al.*, 1996; Frost *et al.*, 2006; Back and King, 2013), which are less energetically rewarding for fish and may alter fish communities (Liao *et al.*, 1995; Schindler and Eby, 1997; Zimmerman and Vondracek, 2006; Vinson and Baker, 2008; Weber, Bouwes and Jordan, 2014; Shearer and Hayes, 2019).

In addition to direct nutrient limitation, invertebrate growth can also be limited by a lack of macromolecules, such as sterols, essential amino acids and polyunsaturated fatty acids (Mueller-Navarra, 1995; Goedkoop, Demandt and Ahlgren, 2007; Wacker and Martin-Creuzburg, 2012). Nutrient enrichment can alter the microbial community composition (Tuomi *et al.*, 1995; Leflaive *et al.*, 2008) that provide invertebrates with macromolecules (Martin-Creuzburg and Elert, 2009; Martin-Creuzburg, Beck and Freese, 2011; Guo, Kainz, Sheldon, *et al.*, 2016; Guo, Kainz, Valdez, *et al.*, 2016a; Sanpera-Calbet *et al.*, 2017). For example, Guo, Kainz, Valdez, *et al.*, (2016a) enriched nutrients in laboratory stream experiments and found that the microbial community composition attached to leaf litter increased the availability of polyunsaturated fatty acids. This, in turn, enabled greater somatic growth of shredder invertebrates. Within 17 days of N enrichment from background levels to 1 mg/L, the measured shredder dry biomass increased by approximately 50% relative to the controls without enrichment.

Alterations in nutrient stoichiometry not only affect invertebrates directly, but also indirectly through changing symbiotic relationships. Nutrient enrichment can alter the intensity and incidence of pathogenic infections (Frost, Ebert and Smith, 2008; Civitello *et al.*, 2018), often by exacerbating infections of generalist parasites with direct or simple lifecycles (Johnson *et al.*, 2010). Nutrient enrichment could also alter invertebrate gut microflora and consequently affect biochemical processes such as nitrogen fixation and enzyme activity (Harris, 1993; Camargo and Alonso, 2006), and alter fish gut microflora, affecting fish growth and health, and consequently their consumption of invertebrates (Gómez and Balcázar, 2008; Nayak, 2010; Sullam *et al*>, 2012; Romero, Ringø and Merrifield, 2014).

Nutrient enrichment can also be directly toxic to invertebrates and fish (Camargo, Alonso and Salamanca, 2005; Camargo and Alonso, 2006; Hickey and Martin, 2009). Numerous direct impacts can arise and include (but are not limited to): damage to gills causing asphyxiation; suppressing the Krebs's cycle and increasing glycolysis, reducing oxygen carrying capacity and causing acidosis; inhibiting ATP production; osmoregulatory upset; immune system suppression (Camargo and Alonso, 2006); increased metabolic cost from the excretion of excess nutrients (Hessen *et al.*, 2013); and chemosensory impairment, potentially affecting behavioural responses to predators (Turner and Chislock, 2010). In some cases, a stoichiometric knife edge can arise whereby nutrients initially increase the growth of invertebrates by alleviating nutrient limitations, but then reduce growth via toxicity effects (Hessen *et al.*, 2013). It should be noted that most, if not all, studies that measure toxic effects do so in a clean environment that are absent of detritus, algae and periphyton (Camargo, Alonso and Salamanca, 2005; Camargo and Alonso, 2006; Hickey and Martin, 2009). Therefore, the impacts of the nutrient enrichment from trophic or ecosystem mechanisms, along with environmental conditions and stress, are undetected – rather the toxicity tests are only illicit a response when nutrients disrupt biochemical processes under prescribed conditions. Toxicity derived nutrient criteria are highly unlikely to bear relevance in real ecosystems.

At the community scale, nutrient enrichment driven changes in the overall stoichiometry of the ecological community can disrupt food web functioning by altering nutrient cycling, food chain length, structural asymmetry and homogeneity of energy flows, connectance, transfer efficiency and the strength of trophic cascades (Neutel, Heesterbeek and de Ruiter, 2002; Post, 2002; Patrício and Marques, 2006; Hall, 2008; Ulanowicz *et al.*, 2009; Davis *et al.*, 2010; O'Gorman, Fitch and Crowe, 2012; Rooney and McCann, 2012; Kovalenko, 2019) – all of which can de-stabilise food webs or parts of them (Saint-Béat *et al.*, 2015; Mougi and Kondoh, 2016; Zhao *et al.*, 2016; Canning and Death, 2017, 2018). In the same long-term enrichment experiment described earlier, the magnitude of N, C and P flows experienced, on average, a 97% increase in flow (Cross, Wallace and Rosemond, 2007). Furthermore, Demi *et al* (2020) recently demonstrated that the 2-year experimental nutrient enrichment of five detritus-based streams resulted in a 150% increase in basal flows to primary consumers. If food chains present as linear chains and/or the transfer efficiencies increase, the disruptive oscillations can travel and disturb further, potentially leading to more volatile and less stable system dynamics; if the food chains present more omnivorous, then there is the potential for dampening of oscillations (Persson *et al.*, 2001; Steiner *et al.*, 2005; Attayde and Ripa, 2008; Canning and Death, 2017). Models of tri-trophic chains, with Holling type 2 functional responses, suggest that nutrient enrichment increases lower trophic levels and can lead to the reduction, even extinction, of top predators through alterations of cycling and increased chaos, and permit greater success of invasion (Abrams and Roth, 1994; Belgrano *et al.*, 2004). Very small inputs can be beneficial by providing a mild subsidy and providing an alternative nutrient supply when others are perturbed, but can quickly become detrimental when the food web is skewed towards the input (Huxel, McCann and Polis, 2002).

At the ecosystem scale, nutrient enrichment affects the storage and flux of carbon and nutrients, both locally and spatially (Benstead *et al.*, 2009). Driven primarily by an increase in microbial activity in heterotrophic streams and increased primary productivty in autotrophic streams (Dodds, 2006). More organic matter is processed, transformed and exported from systems with elevated nutrients. This has consequences for annual patterns in localised ecosytem metabolism, resulting in more variable and less resilient systems (Clapcott *et al.*, 2016), and consequences for downstream environments with for example, the ultimate cause of estuarine eutrophication being an increase in organic matter loading (Pinckney *et al.*, 2001). Furthermore, it can mean that nutrient enrichment can reduce the amount of greenhouse gas (or 'blue carbon') abated and stored by aquatic environments (Macreadie *et al.*, 2017).

Nutrients are clearly a key component of ecosystems and nutrient enrichment can result in numerous consequences, with New Zealand rivers and streams likely to be highly susceptible to the effects of nutrient enrichment. New Zealand has temperate and cold climates, with vegetation typically with low N:C (Wardle, Bonner and Barker, 2002; Bellingham *et al.*, 2013) and naturally oligotrophic streams (McDowell *et al.*, 2013); therefore, when enriched with nutrients the decomposition of organic matter is likely to be high, as microbes harness the added nutrients to compensate for the low nutrient content in detritus. Further, macroinvertebrate communities have a high reliance on energy from fine particulate organic matter (FPOM; Winterbourn, Rounick and Cowie, 1981; Winterbourn, 2000),which exacerbates the impacts of nutrient enrichment as microbes condition the FPOM and improve the nutrient availability to invertebrates – permitting growth that is otherwise nutrient limited. Thus, the management of nutrients to low levels needs to be a key component of any freshwater management plan that seeks to safeguard ecosystem health.

The New Zealand National Policy Statement for Freshwater Management contains a range of attributes that use ecological metrics to grade the aspects of ecosystem health and trigger management responses. This report documents the approach used by STAG (Science and Technical Advisory Group)[1] to develop national nutrient criteria (using nitrogen and phosphorus) for rivers that could be used in freshwater management.

[1] https://www.mfe.govt.nz/fresh-water/science-and-technical-advisory-group

Derivation of criteria

Principles adopted

In deriving the nutrient criteria, STAG identified and adopted a series of principles and methodologies by which appropriate criteria should be developed:

a) Multiple lines of evidence are used. Often in ecology, ecosystems and the ways we measure them are highly variable with multiple causality. As a result, most bivariate relationships between ecosystem health metrics and their stressors are weak. Criteria derived from one relationship alone may, therefore, be considerably uncertain. However, if multiple relationships are used then our confidence in the final criteria can be bolstered, particularly if they convey a consistent message.

b) Nationally derived datasets are used. Regionally focussed datasets may not adequately cover the range of environments found across the country. Nationally derived datasets tend to be larger and cover a greater range of environments. A drawback of nationally derived datasets is that they are often collected by different agencies using different techniques and disparate skills, which can cause considerable variation. Furthermore, nationally compiled datasets are usually collected for a range of different reasons and are not necessarily intended to be representative of all environment types. However, STAG considers these risks are outweighed by that of some environments being inadequately covered, and thus deemed that large, national datasets should be used where possible.

c) Nutrient criteria are to be derived by correlating nutrients with metrics of aquatic life and ecosystem metabolism. Clapcott *et al* (2018) identified five components of ecosystem health (water quantity, water quality, aquatic life, habitat, and ecological processes). The approach is to use correlation to ensure parity of stringency with other metrics of aquatic life and ecological processes. Given that ecosystems are inherently indeterminate (Ulanowicz, 1997, 2019; Fath, Patten and Choi, 2001), building a deterministic model that reliably assesses how the structure and function of riverine ecosystems across New Zealand responds to nutrient enrichment would be impossible. As an alternative, nutrient criteria will be derived by correlating nutrients with ecosystem health metrics with the view that there is a plausible mechanistic link between nutrients and the metric assessed.

d) Recognise that nationally correlative relationships do not always translate to site-specific thresholds. All ecosystem health metrics are driven by multiple, and often interacting, factors. By way of example, Figure 1 shows a Bayesian Belief Network developed by Death *et al.*, (2015) to predict QMCI, whilst nutrients are a key predictor, other water chemistry, meso-habitat and riparian vegetation are also influential. Managing nutrients to a particular concentration may, therefore, not always yield the same score for an ecosystem health metric at all sites. Furthermore, when multiple lines of evidence are used simultaneously to derive the final criteria, averaging across multiple relationships with different metrics decouples the criteria from a single metric. Reducing nutrients to a given level will improve ecosystem health, but it does not guarantee that the target of a specific metric will be achieved; rather it increases the probability of meeting the target.

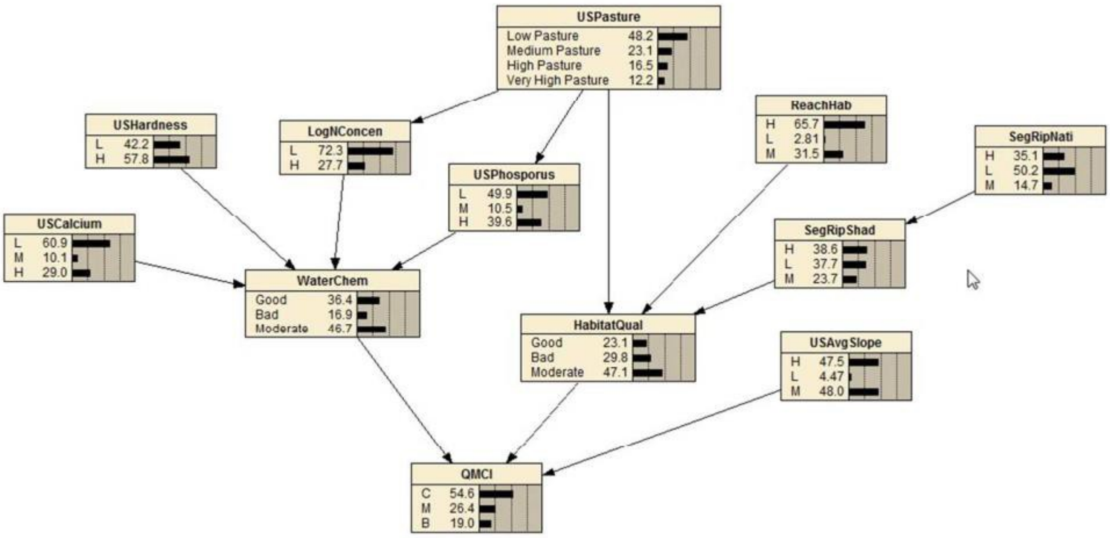


Fig. 1. A BBN predicting QMCI across the lower North Island. Reproduced from: Death, R. G., Death, F., Stubbington, R., Joy, M. K., & van den Belt, M. (2015). How good are Bayesian belief networks for environmental management? A test with data from an agricultural river catchment. Freshwater Biology, 60(11), 2297–2309. https://doi.org/doi:10.1111/fwb.12655

e) All trophic groups are weighted equally. It was not deemed appropriate to value one trophic group more than another, weighting each trophic group differently would impose a deemed value hierarchy that may not be realised by all. Ecosystems are composed of multiple trophic groups and all are necessary for ecosystem health; giving one group more attention (weighting) than another may mean the protection for another is overlooked. A downside of equal weighting is that trophic groups with little and/or poor-quality data inform the final criteria on equal par with groups with better quality data. However, using large, national datasets where possible, helped to reduce the influence of data disparity on the final criteria. Furthermore, a relationship with high scatter compared to one with low scatter does not necessarily mean that nutrients are of lesser importance in the one with high scatter; rather the one with high scatter is likely to have other interacting or mediating factors.

f) Nutrient bands are to be broadly equivalent to those of other attributes. There was a commonly held view that different attribute bands should be harmonised to represent a similar level of stringency/degradation as this allows communities and decision makers to compare like with like. For example, an A grade for one metric should approximately correlate to an A grade for another metric. Nutrient bands are not designed to achieve a particular target of a given metric; they are to be designed to be benchmarked against a range of bands from other attributes.

g) Site-specific criteria may need to be more stringent than the general national criteria. Local conditions may mean that the general nutrient criteria developed here may be insufficient to achieve the community's aspirations for another attribute and more stringent criteria may be required. For example, a community may aspire to have very low periphyton to enhance swimming and fishing values, which may require stricter nutrient criteria to achieve.

h) Nitrogen and phosphorus criteria are derived for dissolved inorganic nitrogen (DIN) and dissolved reactive phosphorus (DRP) respectively. Consideration was given to other measures of the nutrients, for example nitrate-nitrogen and total phosphorus; however, DIN and DRP are preferred as they both represent the biologically available forms. In the case of nitrogen, DIN was preferred over nitrate-nitrogen because it also includes nitrite and ammonia to capture the limited circumstances where these are high and also contribute to biological growth. Total forms, particularly TP, tend to be more variable as they capture molecules bound to sediment which may be suspended in the water column at the time of sampling. Whilst using TN and TP may be more convenient for network accounting (lake and estuarine criteria use TN and TP), the draft technical guidance for implementing the periphyton note exemplifies how to reconcile the differences (NIWA, 2018) .

i) Nutrient criteria are prescribed both as medians and 95th percentiles. Whilst traditionally nutrient management in NZ has focused on the median of monthly samples, much of the variation between samples is climate driven. Global climate change is set to increase variability and result in more extreme concentrations (Kunzewicz *et al*, 2008; Whitehead *et al*, 2009; Moss *et al*, 2011). Including the upper quantile into nutrient management means that changes in extreme concentrations are regulated even when the median remains unchanged (Scharf, Juanes and Sutherland, 1998; ANZG, 2018), and managed by adjusting the quantum and timing of discharges. Including quantiles as a complement to means or medians has also proven meaningful in managing other ecosystems elsewhere (e.g., Chamaillé-Jammes, Fritz and Murindagomo, 2007; Schmidt, Clements and Cade, 2012; Brennan, Cross and Creel, 2015; Rijal, 2018).

j) Where possible, use measured rather than modelled data. Where sites have measured and modelled nutrient data then the measured data is preferred (so long as the dataset is sufficiently large) to reduce the risk of an inaccurate prediction. However, members recognise that measured data is often also subject to error and bias at different scales and well-developed models have an advantage of smoothing messy data.

k) Nationally applicable criteria, rather than regional or river class, is preferred. Given the explorations in Appendices A-H, there was little convincing evidence that large differences occur throughout the country. Conceptually, given that the array of common species persists throughout the country, the extent to which 0.5 mg/L of DIN encourages growth in an otherwise nutrient limited chironomid species in the Manawatu would be unlikely to differ drastically from the same species experiencing the same DIN concentration in Southland. Their stoichiometric ratios are likely to be fairly homeostatic. Unlike periphyton biomass, the frequency of flushing will not change the biological demand for nutrients by an individual invertebrate, with detritus and bacteria unlikely to ever be completely flushed out of the system. Detritus is likely to be replenished as quickly as it is removed, and bacteria will substantially colonise detritus within hours (if they were not already present). Furthermore, members recognise that the policy still has ample flexibility to allow regional councils to set nutrient targets more stringent than the bottom line (or current where better than the bottom line) should their communities aspire for better ecosystem health or to achieve other objectives (e.g., low periphyton for swimming).

Like all other attributes, if a site can be shown to naturally exceed the bottom line, then it may be desirable to exempt it from the general criteria. For example, recognising that many river reaches within the Volcanic Acidic (VA) river class typically have naturally high DRP concentrations (Appendix K) and are consequently exempt, an option with VA river class specific nutrient criteria is provided. It is a policy call as to the extent to which the number of exemptions is acceptable.

Collation of nutrient-ecosystem health relationships

Having established the guiding principles given above, STAG collated data on relationships between nutrient concentrations and a range of ecosystem health metrics, covering periphyton, invertebrates, fish and ecosystem processes. All relationships were derived from national datasets and weighted equally to produce criteria for each trophic level & processes (Table 1). Nutrient criteria derived for periphyton, invertebrates, fish and ecosystem processes were then averaged without weightings to derive a single national nutrient criterion (Figure 3), providing multiple lines of evidence.

Table 1. The ecosystem health metrics using the in MLoE derivation of nutrient criteria, the number of sites and the relationship used.

	Ecosystem health metric	Number of sites	Relationship
Periphyton	Chlorophyll *a* (Matheson et al, 2016)	871-981	Quantile regression
	Chlorophyll *a* (Biggs, 2000)	30	Log regression
Macroinvertebrates	Macroinvertebrate community index (MCI)	388	Piecewise regression
	Quantitative macroinvertebrate community index (QMCI)	293	Piecewise regression
	Macroinvertebrate average score per metric (ASPM)	388	Piecewise regression
Fish	Fish index of biotic integrity (F-IBI)	2922	Quantile regression
Ecosystem processes	Ecosystem respiration	83	Log-log regression
	Gross primary production	83	Log-log regression
	Cotton decay	83	Log-log regression

Periphyton-nutrient relationships

Two periphyton-nutrient datasets were used. The relationships derived from both sources were used to provide nutrient criteria for bands that aligned with the periphyton attribute table in current NPS-FM that has A, B, C and D criteria of 50, 120 and 200 mg chlorophyll *a* m^{-2}.

The first relationship was sourced from Biggs (2000), who collected a variety of periphyton and nutrient measures from 30 rivers throughout New Zealand and derived regression equations for maximum chlorophyll a as predicted by nitrate-nitrogen (N) or dissolved reactive phosphorus (DRP).

The second set of relationships were sourced from Matheson et al (2016), whereby upper quantile regression was used to relate nutrient concentrations with periphyton biomass from 871 and 981 sites for N and P respectively. With large datasets, an advantage of quantile regression is that it can elucidate relationships between variables without needing to control for other limiting factors where data may not be available. They do not, however, account for interactions by other factors (Cade and Noon, 2003). This is better explained graphically (Fig. 2.), with a figure reproduced from Cade & Noon (2003).

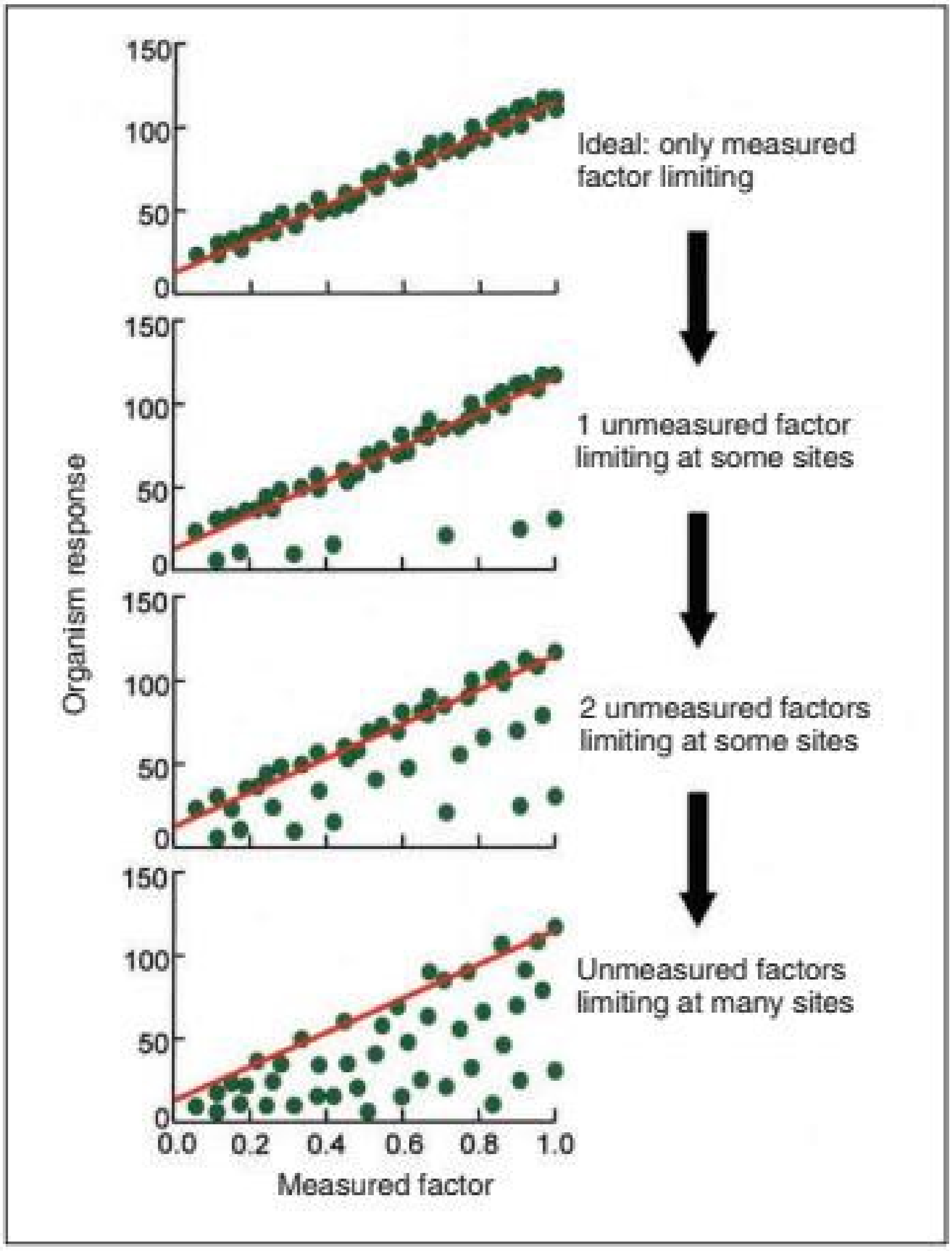


Figure 2. The top graph represents the ideal statistical situation where an organism response is driven primarily by the measured factor(s) included in the linear regression model; ie all other potential limiting factors are at permissive levels. As we proceed from top to bottom, an increasing number of factors that were not measured become limiting at some sample locations and times, increasing the heterogeneity of organism response with respect to the measured factor(s) included in the regression model.

Fig. 2. Hypothetical example of quantile regression detecting relationships in large datasets when there are other limiting variables. Copied from Cade & Noon (2003).

Invertebrate-nutrient relationships

STAG has recommended two invertebrate attribute tables are included in the proposed NPS. The first is based on QMCI & MCI (macroinvertebrate community index) (Stark and Maxted, 2007), with band thresholds of A/B = 6.5 & 130, B/C = 5.5 & 110, and C/D = 4.5 & 90 respectively. The second is based on Collier's (Collier, 2008) ASPM (average score per metric), with band thresholds of A/B=0.6, B/C=0.4 and C/D=0.3. In deriving invertebrate-nutrient relationships, STAG has used annual macroinvertebrate survey data collected for state-of-environment monitoring across the country. Scores used were calculated by Clapcott et al (2017) to ensure the consistent calculation of MCI, QMCI and ASPM across regions. The raw macroinvertebrate data was collected by different agencies, with variation in sampling technique and intensity, taxonomic and count resolution, and skill in identifying invertebrates. Considering these drawbacks and accepting the inevitable scatter that may arise as a result (perhaps explaining why regional correlations were stronger than river class specific ones in Appendix L), the dataset was still preferred as it is the most comprehensive and with the greatest spatial coverage.

At sites with paired monthly nutrient monitoring (388 sites), the average annual scores for each of the three metrics were correlated with the median DIN and DRP concentrations over the same five-year period (2012-2016). Given that it is inevitable for large datasets to have outliers with disproportionately high influence, piecewise regression using Crawley's (2013) iterative procedure was used to 'break' off the few extreme points until the lowest residual MSE was achieved. The points excluded were also identified as extreme outliers using the 'rule of thumb' that suggests an extreme outlier is Q3+3*IQR, where Q3 is the third quartile and IQR is the inter-quartile range. All relationships had significant correlations (Table 2, Figure 3) and were used to determine the nutrient concentrations that corresponded with the bands described above. Given that MCI and QMCI are very similar metrics and in the same attribute table, the concentrations derived for each band were averaged so that collectively they had the same weighting towards the invertebrate partition of overall criteria derivation as ASPM (Figure 4).

Table 2. Regression statistics between invertebrate metrics and nutrient concentrations.

Relationship	R^2	p-value
MCI vs N	0.13	<0.001
MCI vs P	0.10	<0.001
QMCI vs N	0.09	<0.001
QMCI vs P	0.06	<0.001
ASPM vs N	0.10	<0.001
ASPM vs P	0.10	<0.001

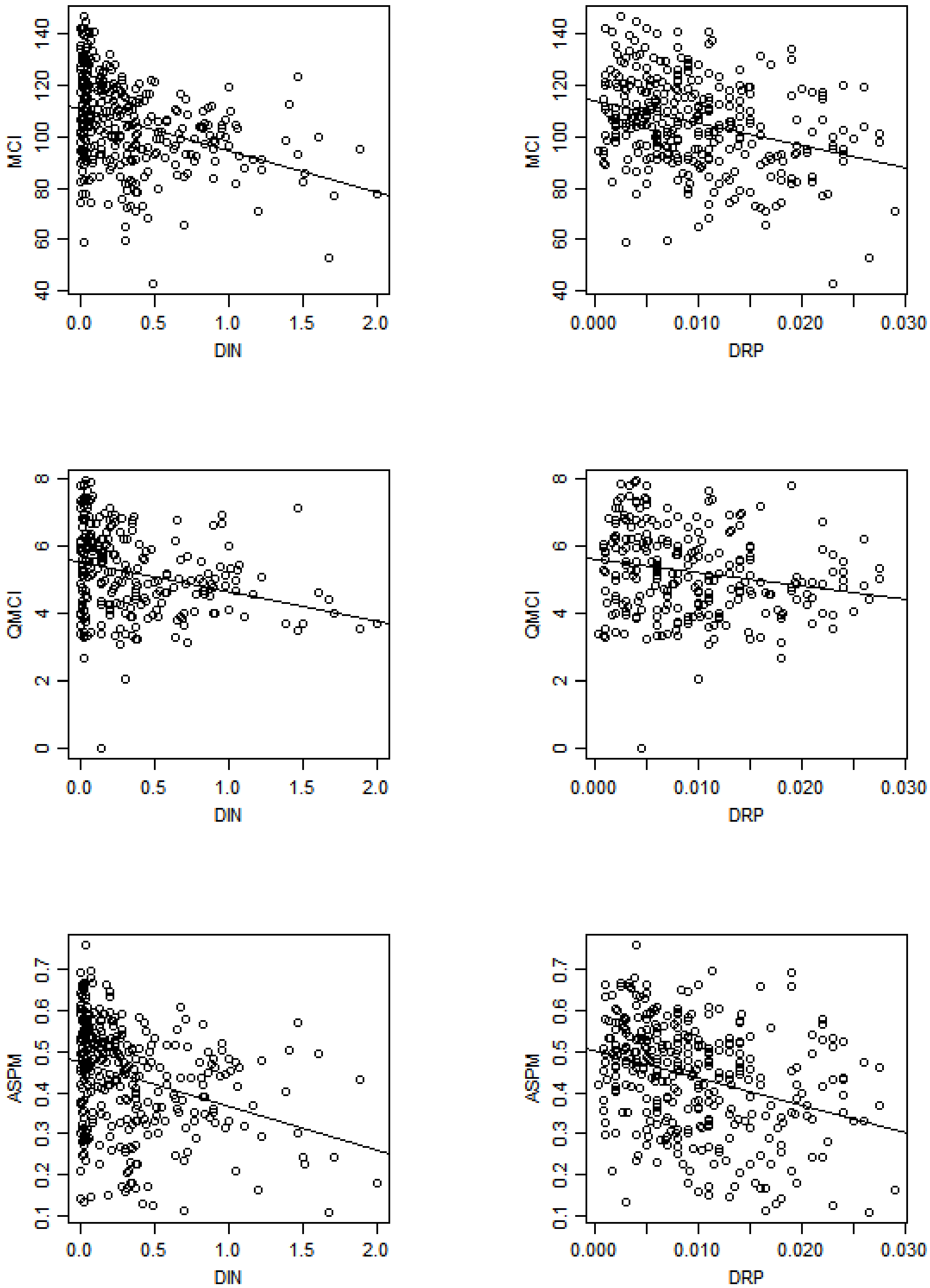


Fig. 3. The MCI, QMCI and ASPM versus DIN and DRP concentrations.

Fish-nutrient relationships

The Fish Index of Biotic Integrity (Joy & Death, 2004) is a popular, nationally applicable fish-based indicator of ecological health and has been recommended by STAG for inclusion in the proposed NPS. The fish IBI data was calculated for sites surveyed and collated in the NZ Freshwater Fish Database (NZFFD) between 2010 and 2017. Only survey records that covered at least 150m of river reach were used, as recommended in Joy, David, & Lake (2013). Where sites were surveyed on multiple occasions, a survey was selected at random, which amounted to 2923 sites. Regular fish monitoring is not common in regional council state of environment monitoring. As a result, most surveys in the NZFFD are one-off surveys and are unfortunately not paired with monthly nutrient monitoring. One-off surveys in NZFFD are likely to have considerable variability as survey intent, intensity and skill of operators differs, along with most surveys being one-off snapshots. Fish communities do, however, tend to be relatively more consistent over time than periphyton and invertebrate communities, and the IBI is based on species presence, rather than abundance, which adds a level of robustness against differences in sampling intensity. Given the lack of paired nutrient monitoring, modelled nutrient predictions for each site were used, as predicted by Larned, Snelder and Unwin (2017).

Given that the IBI is a holistic indicator that responds to a range of pressures, nutrients being one, and that surveys differ considerably in quality, quantile regression was used to relate fish IBI with nutrients to capture the relationship when nutrients are likely to be the limiting factor and not interacting with the many other factors. Consistent with the quantile used by Matheson et al (2016) for the periphyton-nutrient relationships, the 85^{th} percentile was chosen as this appeared a reasonable balance between capturing the upper quantile relationship yet not being driven by exceptionally high values. The quantile regression was used to determine the nutrient concentrations used for each band that corresponded with the 25^{th}, 50^{th} and 75^{th} percentiles of the fish IBI data.

Ecosystem process-nutrient relationships

Three metrics of ecosystem processing were used, being gross primary production (GPP), ecosystem respiration (ER) and cotton cellulose decomposition potential. The data used comprised 84 sites across three main bioregions of NZ, as described by Clapcott et al (2010). Bands for GPP and ER were derived from those proposed by Young et al (2008) and recommended by STAG in the ecosystem metabolism attribute. For cotton decomposition, there were no previously suggested bands, instead the 25th, 50th and 75th percentiles comprised the A, B and C bands respectively. Log-log transformations were applied to all metric and nutrient relationships, and all were statistically significant (Table 3).

Table 3. Regression statistics between ecosystem process metrics and nutrients.

Relationship	R^2	p-value
GPP vs N	0.15	0.0004
GPP vs P	0.06	0.02
ER vs N	0.13	0.001
ER vs P	0.13	0.0008
Cotton K dd vs N	0.16	0.0003
Cotton K dd vs P	0.10	0.004

Aggregating nutrient relationships

Where multiple nutrient-metric relationships were used to derive criteria for a single trophic level, these were averaged equally to produce nutrient band thresholds for each trophic level (summarized in Table 4). The nutrient criteria for each trophic group were then combined into a single criterion by averaging, so as to not value some biological groups more than others (Figure 4). Alternative aggregation methods examined included averaging with periphyton excluded (given that the NPS already requires nutrients to bet set where needed to limit excessive periphyton growth) and using the most stringent criterion across the four trophic groups for each band. Excluding the periphyton criteria from the averaging made very little difference to the final bands derived. The trophic group specific nutrient criteria and the averaged overall criteria are presented in Table 5. Suggested attribute tables and band threshold descriptions are provided in Tables G1 and G2 of Appendix G. A short comparison of this method with that of Death *et al.*,(no date) is provided in Appendix H.

Table 4. The bands used for each ecological metric used in nutrient band derivation.

Band	Chlorophyll a	MCI	QMCI	ASPM	IBI	GPP	ER	Cotton K dd
A	50	130	6.5	0.6	36	3.5	5.8	0.0009
B	120	110	5.5	0.4	28	5	7	0.0019
C	200	90	4.5	0.3	20	7	9.5	0.00395

Table 5. Nutrient criteria for each trophic group and the overall average (mg/L).

Nutrient	Band	Periphyton	Invertebrates	Fish	Ecosystem processes	**Average**	**Average (excluding periphyton)**	**Most stringent**
DIN	A	0.11	0.01	0.50	0.35	**0.24**	**0.29**	**0.01**
	B	0.53	0.33	0.63	0.50	**0.50**	**0.49**	**0.33**
	C	1.00	1.47	0.76	0.77	**1.00**	**1.00**	**0.76**
DRP	A	0.004	0.001	0.013	0.008	**0.006**	**0.007**	**0.001**
	B	0.009	0.009	0.016	0.009	**0.010**	**0.011**	**0.009**
	C	0.016	0.028	0.019	0.010	**0.018**	**0.019**	**0.010**

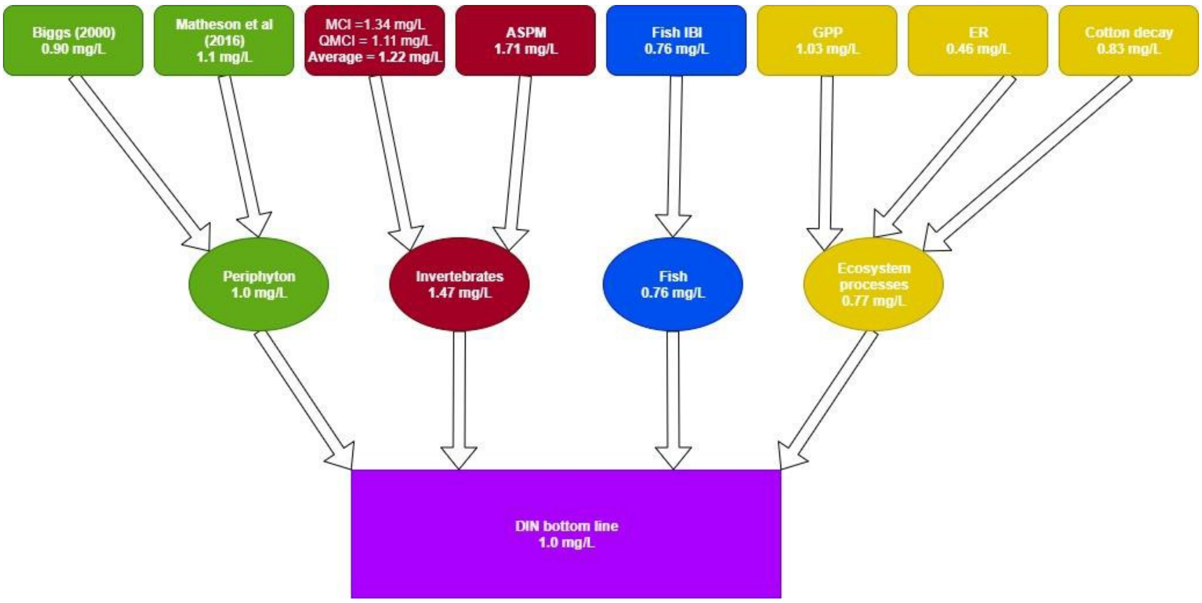


Fig. 4. Schematic showing the compilation of multiple lines of evidence to inform the proposed DIN national bottom line.

Alternative option for DRP

In Appendix K, the derived nutrient criteria were compared with reference condition by river class as predicted by McDowell *et al.* (2013) and as measured at sites considered to be in an undisturbed, or reference, condition. Estimated DIN at reference condition fell largely within the proposed A-band, as it should. DRP, however, was estimated to be more variable between and within river classes. The Volcanic Acidic (VA) river class was the most problematic as the estimated medians and spread for reference sites expanded across the C and D bands. Of the 79 measured DRP reference sites – eight (all within the VA class) – were in the D band. Approximately 15% of New Zealand's river network is in the VA geology class of the REC (Snelder, Biggs and Weatherhead, 2010). STAG suggest two options for consideration:

(1) Set a national DRP bottom line but allow Regional Councils to account for sites with naturally high DRP to breach the bottom line by use of the "exceptions clause".
(2) Include a separate DRP criterion for rivers in the VA geology class. There is insufficient data for this river class to recreate all the regressions used for this river class only. Of the 80 monitored nutrient reference sites, the 23 in VA geology have an average 5-year median DRP that is 0.007 mg/L greater than the average of the remaining 53 sites. When using predicted reference condition from McDowell *et al's* (2013) 3rd order classification, the average DRP concentration (weighted by river length) for the VA class is 0.004 mg/L higher than the average DRP concentration for all other river classes combined. A simple, albeit coarse, work-around is to increase each band criteria by 0.004-0.007 mg/L. If the bands are increased 0.007 mg/L then for the VA class the band thresholds become: A/B=0.013 mg/L, B/C=0.017 mg/L and C/D=0.025 mg/L. A possible alternative attribute table is Appendix G.

If option (1) is accepted then there may be somewhere between 5-15% of sites in New Zealand that requiring an exemption – particularly in parts of Auckland, Bay of Plenty, Taranaki, Manawatu and Banks Peninsula. Where those natural exceptions apply, there is a risk that councils may choose targets that are inconsistent with the stringency of the overall policy direction and could therefore fail to safeguard ecosystem health.

If option (2) is accepted, then the number of sites requiring exemption would be considerably reduced. However, the method used is coarse. If sufficient data ever becomes available in future, then VA criteria could be re-examined using the same methodology as the default criteria, though this may be unlikely as river classes tend to show bias towards different land uses that may not provide the complete gradient. There is also a policy question of whether natural baselines should be accounted for to allow the same 'headroom' for degradation, or whether an equivalent ecosystem health is required across the board and places with higher reference condition simply have less natural 'headroom' for development. An argument could also be made that aquatic life in VA geological classes could tolerate more liberal bottom-lines, as Duggan, Boothroyd and Speirs (2007) found even pristine sites lacked the sensitive macroinvertebrates observed elsewhere and were dominated by tolerant taxa. Furthermore, the VA geology class did have a higher median change-point between nutrients and MCI, though there was a large spread within the river class (Appendix K). However, both of those studies only focused on macroinvertebrates and did not assess periphyton, microbes, macrophytes, or fish. There is also considerable variability in reference DRP concentrations within the VA geology class, and the mechanisms predicting the release of DRP from sediment are not well understood, with links suggested with acidity, temperature, water hardness, drying and wetting, oxygenation, and even interactions with some aquatic life such as microbes, invertebrates and macrophytes.

Deriving 95th percentiles

Given that all relationships derived used average annual median concentrations, to determine the 95th percentile a typical standard deviation expected for each band was estimated and the 95th percentile was defined as two standard deviations from the median. For all SoE monitoring sites, the standard deviation of data collected between 2012 and 2016 was correlated with the average annual median (N: r^2=0.89, p<0.0001, Fig. 4.; P: r^2=0.95, p=<0.0001). Using these correlations, the 95th percentiles were derived for each band.

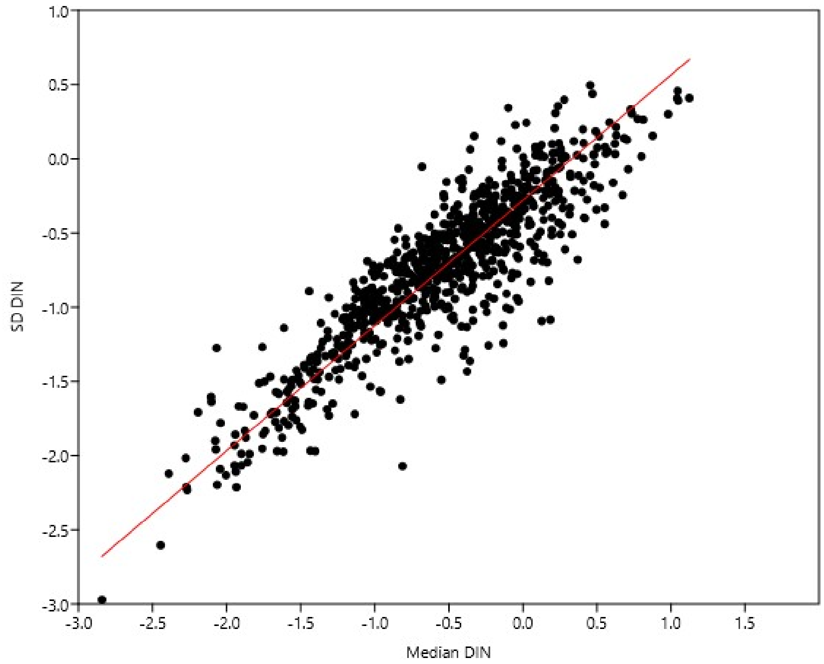


Fig. 5. Correlation between DIN median and standard deviation across SoE monitoring sites between 2012-2016.

Discussion

Using multiple lines of evidence, national-scale (not site specific) nutrient criteria have been proposed for the NPS-FM (Appendix G). Whilst many of the relationships had considerable prediction error, and thus some uncertainty is inevitable, using multiple lines of evidence provides strength – if one relationship is poor, then it is only a single line among numerous other lines. As emphasized earlier, the criteria are intended to be correlative and of similar stringency to other bands. They do not represent functional ecological thresholds or tipping points nor do they guarantee that a particular target of a given metric (i.e., an MCI of 90) is achieved – this a consequence of averaging multiple lines of evidence with different metrics together to produce criteria. Nutrients are a fundamental component of ecosystems and any enrichment increases the probability of a reduction in ecosystem health; likewise, reducing nutrients increases the probability of an improvement in ecosystem health. Councils may still need to reduce nutrients to more stringent levels than proposed here when communities aspire for a high level of ecosystem health or when needed to meet targets for other attributes (e.g., ecosystem metabolism or periphyton targets).

In deriving the criteria, linear regressions were used to model both the median of some relationships and a given quantile in others. An implication of this is that the relationships using an upper quantile are usually likely to arise in more liberal nutrient criteria than those derived from the median. However, quantile regressions were preferred for the relationships from Matheson *et al* (2016) and for the fish IBI because the metrics/datasets were highly influenced by other factors – using quantile regressions allowed the encapsulation of nutrients at limiting levels (assuming no other interactions). In contrast, the MCI was developed primarily to respond to organic enrichment – whilst other factors do influence the score, the relationships with nutrients could be reasonably encapsulated from regression of the mean. If all lines of evidence used regression of the median, then the final criteria would be more stringent than proposed here.

It is, however, encouraging that the nitrogen bottom-line is in line with Camargo & Alonso (2006), who conducted a global review of inorganic nitrogen pollution in rivers and suggested levels should be less than 0.5-1 mg/L to prevent eutrophication and protect against toxicity. Furthermore, the B-bands for both N and P are well aligned with the ANZECC (2000) trigger values for lowland rivers of 0.444 mg/L and 0.010 mg/L respectively. They are also sufficiently protective to ensure that there are no observable toxic effects on the few sensitive species tested (Hickey and Martin, 2009).

An exploration of spatially-explicit saturation change-points (Appendix A) demonstrated that across all river classes, DIN and MCI demonstrated a consistent change-point near ~1mg/L, above which the response of MCI to DIN enrichment was saturated. The change-points between DRP and MCI were, however, more varied, with the VA geology class having change-points at higher concentrations, perhaps reflecting more tolerant communities due to naturally high DRP concentrations – though there was still considerable variation of change-points within the VA class. As the proposed bottom-lines show strong alignment with the estimated change-points between nutrients and MCI for different river classes, improvements in nutrient state that are better than the bottom-line have greater propensity of also improving MCI scores. The criteria are also, and desirably so, more stringent than nutrient criteria derived from upper quantile regressions between macroinvertebrate metrics and nutrients in Appendix C. Nutrient bottom-lines that exceed the quantile regression criteria from Appendix C would highly likely be insufficiently protective to meet the proposed MCI, QMCI and ASPM bottom-lines, as was sought by some who consider nitrogen bottom-lines more appropriately set at the 95th, or even the 90th, percentiles for protection against nitrate-nitrogen toxicity.

The DIN bottom-line also aligns closely with the point where the cumulative change in macroinvertebrate community compositions exceeds 20% (of the change range assessed) with increasing DIN, as identified by the gradient forest analysis in Appendix D. Both DIN and DRP bottom-lines also align closely with the points where a fitted function of declining Fish Observer/Expected assemblage scores flatten out and reduce responding to increasing DIN and DRP (i.e., the responses are saturated), as shown in Appendix E. Based on the Macroinvertebrate Change Analysis in Appendix F, the proposed DIN bottom-line would protect approximately 40% of taxa, and DRP bottom-line approximately 60% of taxa, from experiencing more than minor change in probability of presence. To yield an 80% species protection, then DIN would be approximately 0.5 mg/L and DRP approximately 0.011 mg/L – aligning closely with the proposed B-bands. Regional explorations of macroinvertebrate-nutrient responses within the Manawatu-Whanganui region also found that the impact of nutrients on various ecosystem health metrics and assemblage turnover ceased at TN~0.5 mg/L – also aligning with the proposed B-band (Wagenhoff, Clapcott, *et al.*, 2017; Wagenhoff, Liess, *et al.*, 2017). The same studies also found nutrient impacts on macroinvertebrate metrics, assemblage turnover and cotton decay at TN~0.2 mg/L – similar to the proposed A-band; and bacteria assemblage turnover peaked at TN~0.9 mg/L – similar to the proposed bottom-line (Wagenhoff, Clapcott, *et al.*, 2017; Wagenhoff, Liess, *et al.*, 2017).

In the USA, the USEPA developed nutrient criteria for TN and TP for 13 different ecoregions (USEPA, 2019). For TN, other than one ecoregion that has a standard of 2.18 mg/L, the other twelve have standards ranging from 0.12-0.90 mg/L. The proposed DIN bottom-line of 1 mg/L slightly exceeds this range, but is not substantially higher than the upper range of standards in the USA. Particularly when considering that TN includes more forms of N than DIN, using data from LAWA, a DIN of 1 mg/L correlates to a TN concentration of approximately 1.3 mg/L, Also, worth noting that 1 mg/L is simply a bottom-line and councils should set more stringent standards if local conditions and goals necessitate. In terms of TP, the thirteen regions range from 0.01 mg/L to 0.076 mg/L. Also using data from LAWA, a DRP concentration of 0.018 mg/L correlates to an approximate TP concentration of 0.037 mg/L – which is mid-range of the USEPA standards. Whilst the USEPA use 13 ecoregions to prescribe criteria, a single nutrient criterion for NZ is still at a finer scale than the USA ecoregion approach - given that the USA is approximately 36 times the size of NZ, the average ecoregion is approximately three times larger than NZ.

China, since 2003, have used a 6-tier grading system to rate surface water quality. Similar to NZ, the national policy sets the minimum tier rivers must attain by a given date. Local management authorities must meet or do better than the prescribed grade. For TN, the grade thresholds are I=0.25 mg/L, II=0.5 mg/L, III=1.0 mg/L, IV=1.5mg/L, V=2.0 mg/L and inferior to V>2.0 mg/L. For TP, the grades are I=0.02 mg/L, II=0.1 mg/L, III=0.2 mg/L, IV=0.3 mg/L, V=0.4 mg/L and inferior to V>0.4 mg/L. By mid-century it is anticipated that all rivers should at least meet grade III, with some rivers having intermediate targets prescribed to them at either IV or V to ensure they are on a trajectory to improvement and reaching grade III – effectively a national bottom-line. For nitrogen, China's I-III grades align with the proposed A-C grades, with China's grades being slightly more conservative as they use TN not DIN. The TP standards used in China are, however, considerably more liberal than those proposed here for NZ, with proposed bottom-line here sitting broadly at the higher end of China's grade II. China has considerably different geology from NZ and has substantial phosphorus pollution from the wastewater treatment plants that treat China's 1.4 billion inhabitants. NZ rivers typically have phosphorus concentrations an order of magnitude less than China's.

In summary, DIN and DRP nutrient criteria are proposed, DIN A-D band thresholds are proposed at 0.24, 0.50 and 1.0 mg/l respectively, and DRP A-D band thresholds are proposed at 0.006, 0.01 and 0.018 mg/l respectively. The criteria were derived using multiple lines of evidence, using datasets that span the country, and the bands are benchmarked against the other proposed attributes. Nutrients are a fundamental component of any ecosystem and their enrichment and imbalance can cascade through ecosystems and substantially alter their structure and function. Reducing anthropogenic nutrient enrichment will highly likely result in improvements to the health of any river ecosystem – if a single metric does not detect improvement, then it is highly likely another metric will (if measured at all).

APPENDIX A: Exploration of spatially explicit nutrient criteria

In setting nutrient targets, it is desirable to ensure they are more stringent than any identified change-points in biological response. That way, tipping points are avoided, and nutrients levels are in the vicinity required for further improvements to have a greater probability of realizing measurable ecological benefit (i.e., more stringent than the level that saturates ecological response).

Here the aim is to identify river class specific change-points between nutrient concentrations and the macroinvertebrate community index (MCI).

Given that state of environment (SOE) monitoring data across New Zealand's river network is non-random, skewed and does not adequately span the entire nutrient concentration gradient for each river class, Boosted Regression Tree (BRT) models were used to predict the change in MCI across nutrient gradients for every river reach, accounting for a range of site-specific environmental variables that may influence the relationship. BRTs are capable of fitting complex interactions, non-linear predictors, and can handle non-normal error terms and missing values (Elith, Leathwick and Hastie, 2008).

Data used in the analysis included the median MCI, as calculated by Clapcott *et al* (2017), from annual surveys between 2012-2016, collected as part of the state of environment monitoring. DIN and DRP concentrations are medians at paired sites over the same five-year period. All environmental variables were sourced from the Freshwater Environments New Zealand (FENZ) geodatabase (Leathwick *et al.*, 2010).

Two BRT models were developed, both predicted MCI using the factors in Table A1, though the first used DIN only as the nutrient species, whereas the second used DRP only. Two separate models were preferred, given potential collinearity. BRTs were ran with a tree complexity of 10, learning rate of 0.001, and cross-validated using a bag fraction of 0.2, and carried out using the Dismo package (Hijmans *et al.*, 2012) in R (R Core Team, 2016). Both models performed excellently, with the DIN-based model having a cross-validated correlation coefficient of 0.77 (SE=0.018) and the DRP-based model having a cross-validated correlation coefficient of 0.76 (SE=0.015). The six most influential predictors (of decreasing influence) of the DIN-based model were: SegFlowVariability, DIN, SegJanAirT, SegMinTNorm. USHardness and ReachSed. The six most influential predictors (of decreasing influence) of the DRP-based model were: SegFlowVariability, USCalcium, SegMinTNorm, DRP, SegFlow and ReachSed. In both models there was a large, and almost identical, interaction between SegFlowVariability and SegJanAirT, that associates hydrologically stable rivers and warm Summer air temperatures with relatively lower MCI scores (Figure A1).

Table A1. Environmental predictors from FENZ (Leathwick *et al.*, 2010) used in BRT explorations.

SegJanAirT	Average January Air Temperature (°C)
SegMinTNorm	Average minimum daily air temperature (°C) normalised with respect to SegJanAirT
SegFlow	Mean annual flow (m^3/sec)
SegLowFlow	Mean annual 7-day low flow (m^3/sec)
SegFlowVariability	Ratio of annual low flow/annual mean flow – indicates long-term stability of flow through the year
SegRipNative	Proportion of native riparian vegetation within a 100 m buffer of the river
USCalcium	Average calcium concentration of rocks in the catchment, 1 = very low to 4 = very high
USHardness	Average hardness of rocks in the catchment, 1 = very low to 5 = very high
USPhosporus	Average phosphorus concentration of rocks in the catchment, 1 = very low to 5 = very high
USPeat	Proportion of upstream catchment covered by peat
USWetland	Proportion of upstream catchment covered by wetland
USLake	Proportion of upstream catchment covered by lake
USGlacier	Proportion of upstream catchment covered by glacier
ReachSed	Weighted average of proportional cover of bed sediment using categories of: 1 = mud; 2 = sand; 3 = fine gravel; 4 = coarse gravel; 5 = cobble; 6 = boulder; 7 = bedrock
ReachHab	Weighted average of proportional cover of local habitat using categories of: 1 = still; 2 = backwater; 3 = pool; 4 = run; 5 = riffle; 6 = rapid; 7 = cascade

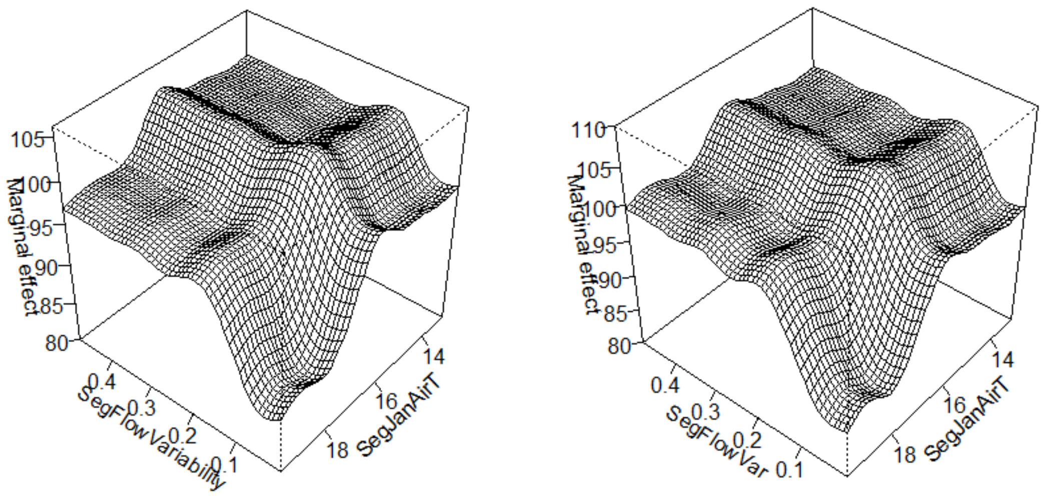


Fig. A1. Interaction plots between SegFlowVariability and SegJanAirT in predicting mean MCI, for the DIN-based model (left) and the DRP-based model (right).

For all river reaches across the entire network, as mapped in FENZ, the DIN-based model was used to predict the MCI across a DIN gradient (intervals=0.05 mg/L) and the DRP-based model was used to predict the MCI across a DRP gradient (intervals=0.002 mg/L). The dominant change-point was then detected for both nutrient species at every river reach. Change-points were identified using King and Richardson's (2003) non-parametric change-point analysis (nCPA) procedure.

The medians, along with the 20th and 80th percentiles, of change-points for all non-conservation dominated river reaches within the climate and geology classes, based on the River Environment Classification (REC) system (Snelder, Biggs and Weatherhead, 2010), are presented in Tables A2-5, and maps of the geology class median concentrations in Figures A2-3. The extremely wet climate classes were merged with the wet climate classes, given that they have few reaches.

As with all analyses using these datasets, potential sources of uncertainty include differences between councils in in their spatial coverage; representativeness; collection technique; and invertebrate identification intensity, taxonomic resolution and quality. To improve the consistency of scoring, MCI values derived by Clapcott et al (2017) were used instead of council derived scores. Nutrient grab sampling also presents considerable uncertainty as nutrients can fluctuate diurnally, seasonally, with rainfall patterns, and labs differ in their consistency between labs in determining concentrations. Many of the FENZ geodatabase attributes are also modelled or derived from coarse base layers, inherently bringing in error. The boosted regression tree modelling also presents its own error (indicated by the CV-correlations and standard errors), and the nCPA method also has uncertainty in identifying the change-point. However, despite these sources of uncertainty, FENZ is the best aggregation of reach-specific environmental variables for all river reaches available and the BRT models showed excellent performance in predicting the MCI. Furthermore, by using medians and percentiles to represent river-class concentrations, this overcomes the presentation of false-precision at the river-class scale.

The analysis suggests that DIN change points, across most river reaches and river classes, occur at around 0.98 mg/L, with a small portion having change-points at lower concentrations, suggesting those few reaches may require more stringent nutrient concentrations than the rest before improvements in MCI are likely to be realized if nutrients are the only driver to be improved.

The DRP change points, across most river reaches and river classes, occur at around 0.017 mg/L, with almost all river classes having reaches that had change-points at 0.023 mg/L. The volcanic acidic (VA) river class had a higher median threshold (0.021 mg/L) than all other geology classes (~0.017 mg/L); however, the range between the 20th and 80th percentiles were also larger, suggesting the higher change-point is not universal across the river class.

Table A2. The median, 20th and 80th percentile, change-points between DIN and MCI for the REC climate classes.

Climate	Median	20th percentile	80th percentile
CD	0.975	0.875	1.025
CW	0.925	0.375	0.975
WD	0.975	0.975	1.025
WW	0.975	0.725	1.025

Table A3. The median, 20th and 80th percentile, change-points between DIN and MCI for the REC geology classes.

Geology	Median	20th percentile	80th percentile
Al	0.975	0.975	1.025
HS	0.975	0.425	0.975
M	0.975	0.725	0.975
Pl	0.875	0.325	0.975
SS	0.975	0.725	1.025
VA	0.975	0.375	1.025
VB	0.625	0.325	0.975

Table A4. The median, 20th and 80th percentile, change-points between DRP and MCI for the REC geology classes.

Climate	Median	20th percentile	80th percentile
CD	0.017	0.017	0.023
CW	0.017	0.017	0.023
WD	0.017	0.017	0.023
WW	0.017	0.017	0.023

Table A5. The median, 20th and 80th percentile, change-points between DRP and MCI for the REC geology classes.

Geology	Median	20th percentile	80th percentile
Al	0.017	0.017	0.023
HS	0.017	0.017	0.023
M	0.017	0.017	0.021
Pl	0.019	0.017	0.023
SS	0.017	0.017	0.023
VA	0.021	0.011	0.023
VB	0.017	0.011	0.023

Fig. A2. Median DIN change-point based on the REC geology classes. Green=0.975 mg/L, blue=0.875 mg/L and purple=0.625 mg/L.

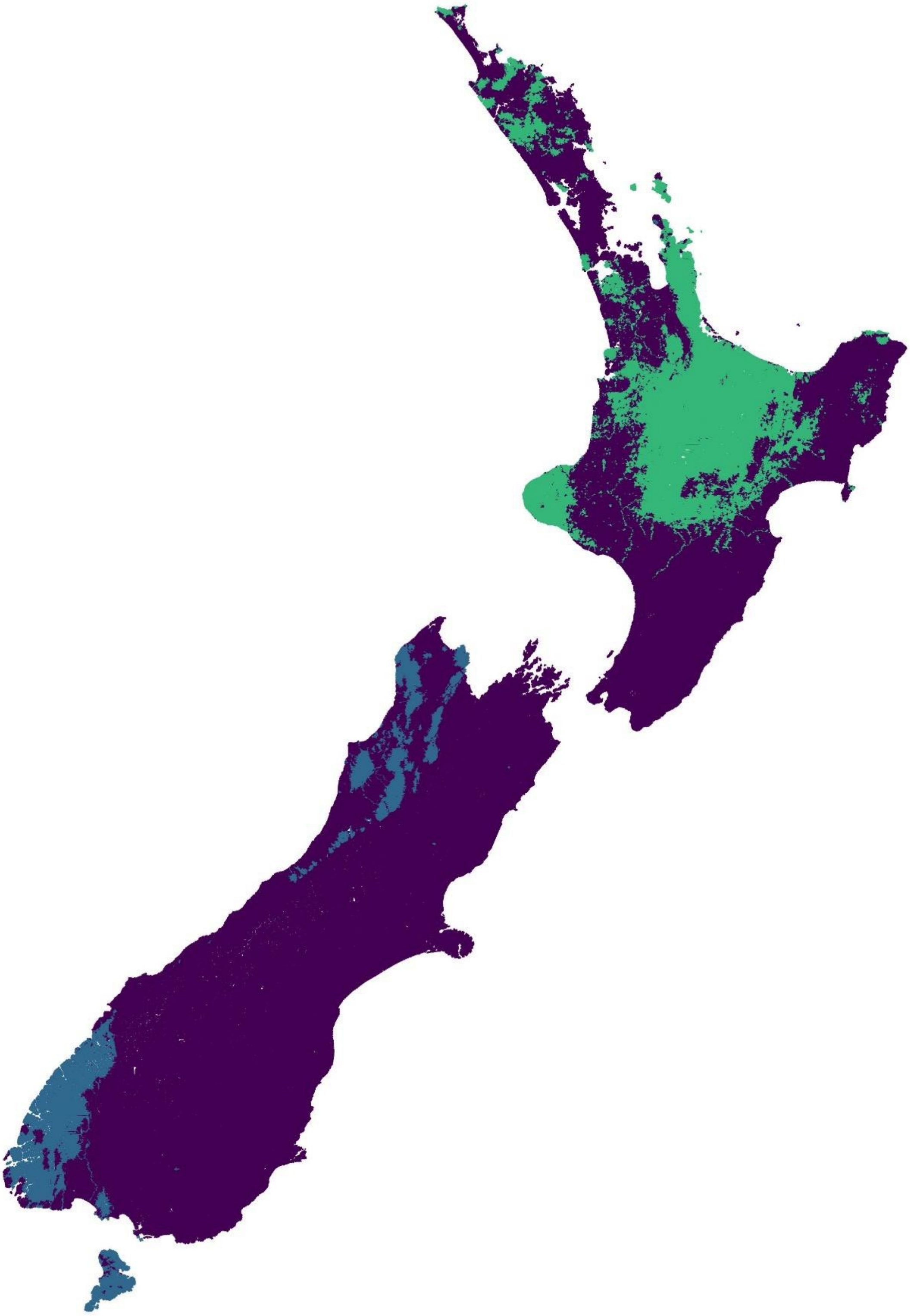


Fig. A3. Median DRP change-point based on the REC geology classes. Green=0.021 mg/L, mid-blue=0.019 mg/L and dark-blue=0.017 mg/L.

APPENDIX B: Boosted Regression Trees

Whilst Appendix L largely found significant correlations between the selected invertebrate-metric and nutrients for different regions, FENZ classes and REC climate classes, there was a lot of unexplained variation for the reasons outlined in that and other sections. Further, the results of Appendix A have shown how boosted regression trees can be used to account for natural environmental variation and have excellent performance in predicting the MCI.

One of the downsides of Appendix A is that only sites where both MCI and measured nutrient concentrations occur were used, substantially limiting the use of the entire, much larger, macroinvertebrate monitoring network. One way to circumvent this is to use the entire macroinvertebrate monitoring network, but use modelled nutrients rather than measured nutrients.

As an exploratory exercise, BRTs are used to model MCI, %EPT-abundance and the ASPM from the factors in Table B1 (sourced from the FENZ geodatabase (Leathwick *et al.*, 2010) and Larned, Snelder and Unwin (2017).

Multiple models were repeated using three macroinvertebrate datasets, being the national SOE dataset (2012-2016) compiled by Clapcott *et al* (2017) for all three metrics (n=1851), (2) Prof Death's dataset as described in Death *et al.*, (2015) for MCI only (n=963), and (3) the national SOE dataset and Prof Death's combined for MCI only. Whilst Prof Death's dataset is a collation of one-off surveys primarily throughout the lower North Island, it was included as it represents a large dataset that has been collected and identified in a consistent manner. BRTs were ran with a tree complexity of 10, learning rate of 0.001, and cross-validated using a bag fraction of 0.2, and carried out using the Dismo package (Hijmans *et al.*, 2012) in R (R Core Team, 2016).

Table B1. Predictors used in BRT explorations (Leathwick *et al.*, 2010; Larned, Snelder and Unwin, 2017)

DRP	USAvgSlope
NO3N	USCalcium
SegJanAirT	USHardness
SegLowFlow	USPhosporus
SegFlow4th	USPeat
SegFlowVariability	USLake
SegSlopeSqrt	USWetland
SegRipShade	USNative
SegRipNative	USGlacier
DSDist2Coast	ReachSed
USAvgTNorm	ReachHab

Predicting MCI

When using the Death *et al.*, (2015) data, MCI was well predicted with a cross-validated correlation of 0.799 (se=0.013). Nitrate-nitrogen had the greatest relative influence in predicting MCI, followed by flow variability and native riparian cover as per their relative in (Figure B1).

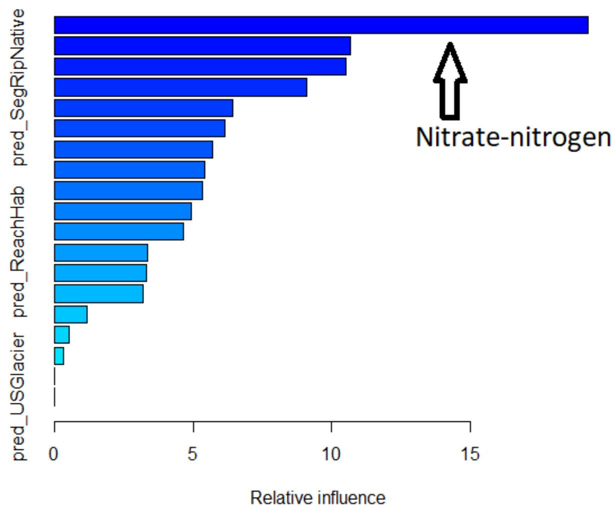


Fig. B1. The relative influence of factors used to predict MCI using the Death *et al.*, (2015) dataset.

When using the Clapcott *et al* (2017) dataset, MCI was well predicted with a cross-validated correlation of 0.742 (se=0.011). Nitrate-nitrogen had the greatest relative influence in predicting MCI, followed by flow variability, slope, January air temperature and native riparian cover as per their relative in (Figure B2).

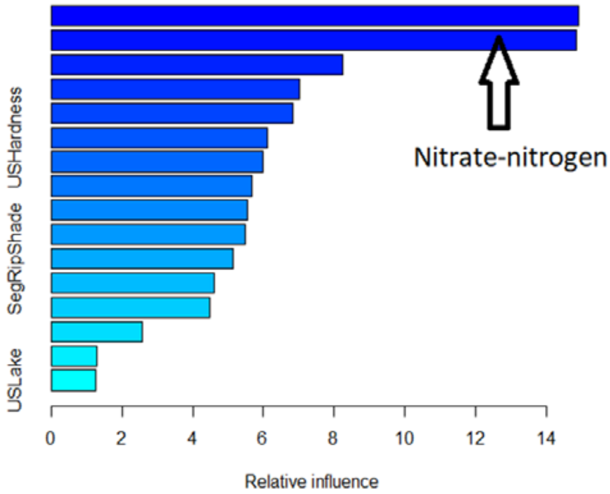


Fig. B2. The relative influence of factors used to predict MCI using the Clapcott *et al* (2017) dataset.

When using the combined dataset, MCI was well predicted with a cross-validated correlation of 0.797 (se=0.007). Nitrate-nitrogen had the greatest relative influence in predicting MCI, followed by flow variability, January air temperature, slope and native riparian cover as per their relative in (Figure B3).

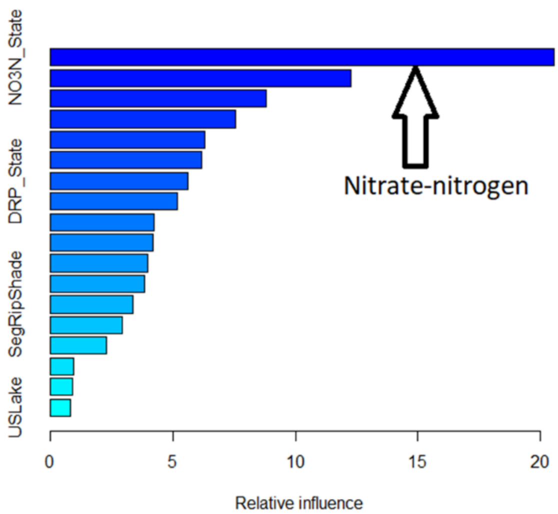


Fig. B3. The relative influence of factors used to predict MCI using the combined dataset.

Predicting ASPM

When using the Clapcott *et al* (2017) dataset, ASPM was well predicted with a cross-validated correlation of 0.74 (se=0.011). Nitrate-nitrogen had the greatest relative influence in predicting ASPM, followed by flow variability, January air temperature and native riparian cover as per their relative in (Figure B4).

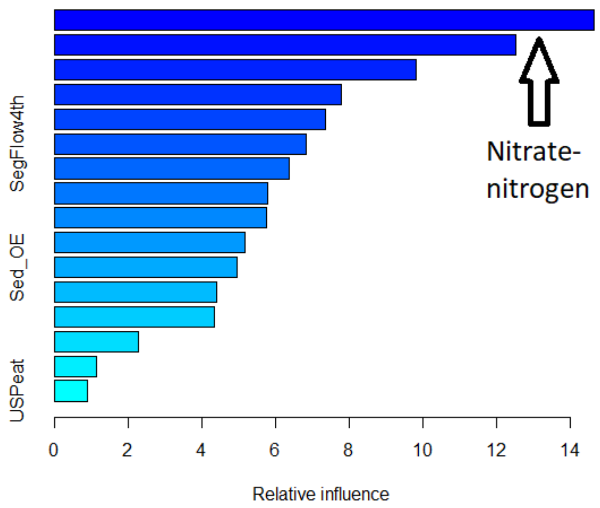


Fig. B4. The relative influence of factors used to predict ASPM using the Clapcott *et al* (2017) dataset.

Predicting %EPT-abundance

When using the Clapcott *et al* (2017) dataset, %-EPT-abundance was well predicted with a cross-validated correlation of 0.68 (se=0.012). Nitrate-nitrogen had the greatest relative influence in predicting MCI, followed by January air temperature, flow variability and DRP as per their relative in (Figure B5).

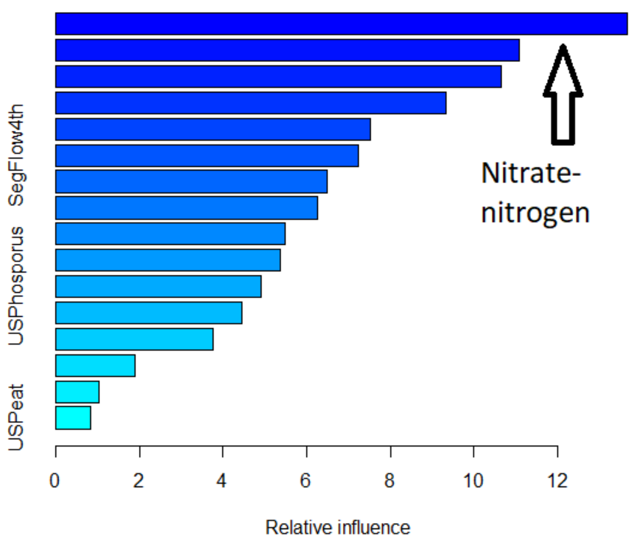


Fig. B5. The relative influence of factors used to predict %EPT-abundance using the Clapcott *et al* (2017) dataset.

Conclusion: In all explorations, the macroinvertebrate metrics were well predicted by the BRT models, with nitrate-nitrogen being the most influential factor in predicting all metrics. Flow variability, native riparian cover and summer temperature were also consistently among the next most influential predictors.

APPENDIX C: Quantile regressions

One of the difficulties with understanding relationships in ecological data is that stressor-response relationships are often fuzzy as there are multiple limiting and/or interacting factors that could be determining the response variable. Often there is insufficient data accurately delineate all these relationships – particularly when predicting metrics intended to summarize an entire community, such as the MCI.

Cade & Noon (2003) provide a helpful introduction into quantile regressions and explain how regression of the upper quantile can be used to identify relationships between responses when a driver is limiting and assuming there are no other interacting factors. Whilst quantile regression does require a large dataset, it allows relationships between responses and limiting factors (as per Liebig's law of the minimum) to be clearly elucidated when there are many other limiting factors that have not been measured and accounted for. This is best explained graphically, see Figure C1. However, quantile regressions do not elucidate responses when there are interactions between drivers that may cause nutrient limitation to come into effect at lower doses than observed in the upper quantile regression. The upper quantile provides the relationship of the most permissive scenario. Therefore, any nutrient criteria to provide for a metric that can be nutrient limited should not be set higher than the upper quantile for the desired score.

Here linear regressions are plotted of the upper quantiles (90th, 85th, 80th, and 75th percentiles), to relate MCI, QMCI and ASPM to DIN and DRP concentrations (Figures C2-3). This is carried out for two datasets: (1) macroinvertebrate and measured nutrient concentrations (where they occur in the same location; n=293-388) (Figure C2), and (2) the entire Clapcott *et al* (2017) SOE dataset (n=1851) and modelled nutrient concentrations from (Larned, Snelder and Unwin, 2017) (Figure C3). Median macroinvertebrate scores from five annual surveys between 2012-16 (inclusive) were used. Measured nutrients were the medians over the same five-year period (beginning 12 months prior to the first invertebrate survey to ensure nutrient measurements from after the last invertebrate survey are not included. Consistent with the EU Best Practice guidelines (Phillips *et al.*, 2018), Tables C1-2 (using datasets (1) and (2) respectively), show the DIN and DRP concentrations derived from the 75th percentile regression that correspond to the bottom of the proposed A, B and C bands for MCI, QMCI and ASPM, being A=130/6.5/0.6, B=110/5.5/0.4 and C=90/4.5/0.3 respectively. Using the smaller dataset with measured nutrients, the DIN criteria for bottom of C-band ranged from 1.80-2.46 mg/L, and DRP criteria for C-band were >=0.036 mg/L. Using the larger dataset with modelled nutrients, the DIN criteria for bottom of C-band ranged from 1.32-1.38 mg/L, and DRP criteria for C-band were >=0.039 mg/L.

Table C1. Nutrient concentrations (mg/L) from quantile regressions between measured DIN & DRP and three macroinvertebrate metrics at the 75th percentile.

Metric	Nutrient	A	B	C
MCI	DIN	N/A	0.69	1.82
QMCI		N/A	0.84	1.8
ASPM		N/A	1.52	2.46
MCI	DRP	N/A	0.023	>0.05
QMCI		0.001	0.018	0.036
ASPM		N/A	0.041	>0.05

Table C2. Nutrient concentrations (mg/L) from quantile regressions between modelled DIN & DRP and three macroinvertebrate metrics at the 75th percentile.

Metric	Nutrient	A	B	C
MCI	DIN	N/A	0.58	1.32
QMCI		0.3	0.84	1.38
ASPM		N/A	0.90	1.37
MCI	DRP	N/A	0.026	>0.05
QMCI		0.012	0.025	0.039
ASPM		N/A	0.036	>0.05

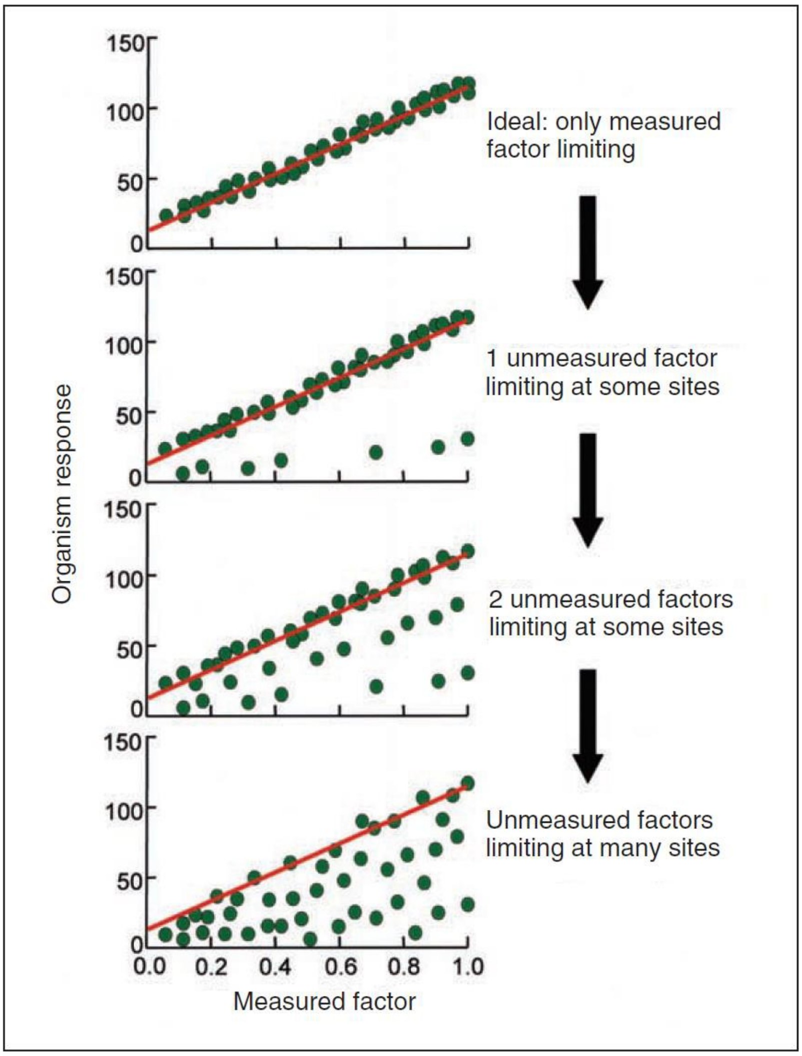


Figure 2. *The top graph represents the ideal statistical situation where an organism response is driven primarily by the measured factor(s) included in the linear regression model; ie all other potential limiting factors are at permissive levels. As we proceed from top to bottom, an increasing number of factors that were not measured become limiting at some sample locations and times, increasing the heterogeneity of organism response with respect to the measured factor(s) included in the regression model.*

Fig. C1. A graphical explanation of how quantile regression can be used to elucidate relationships between limiting factors and response variables. Reproduced from Cade & Noon (2003).

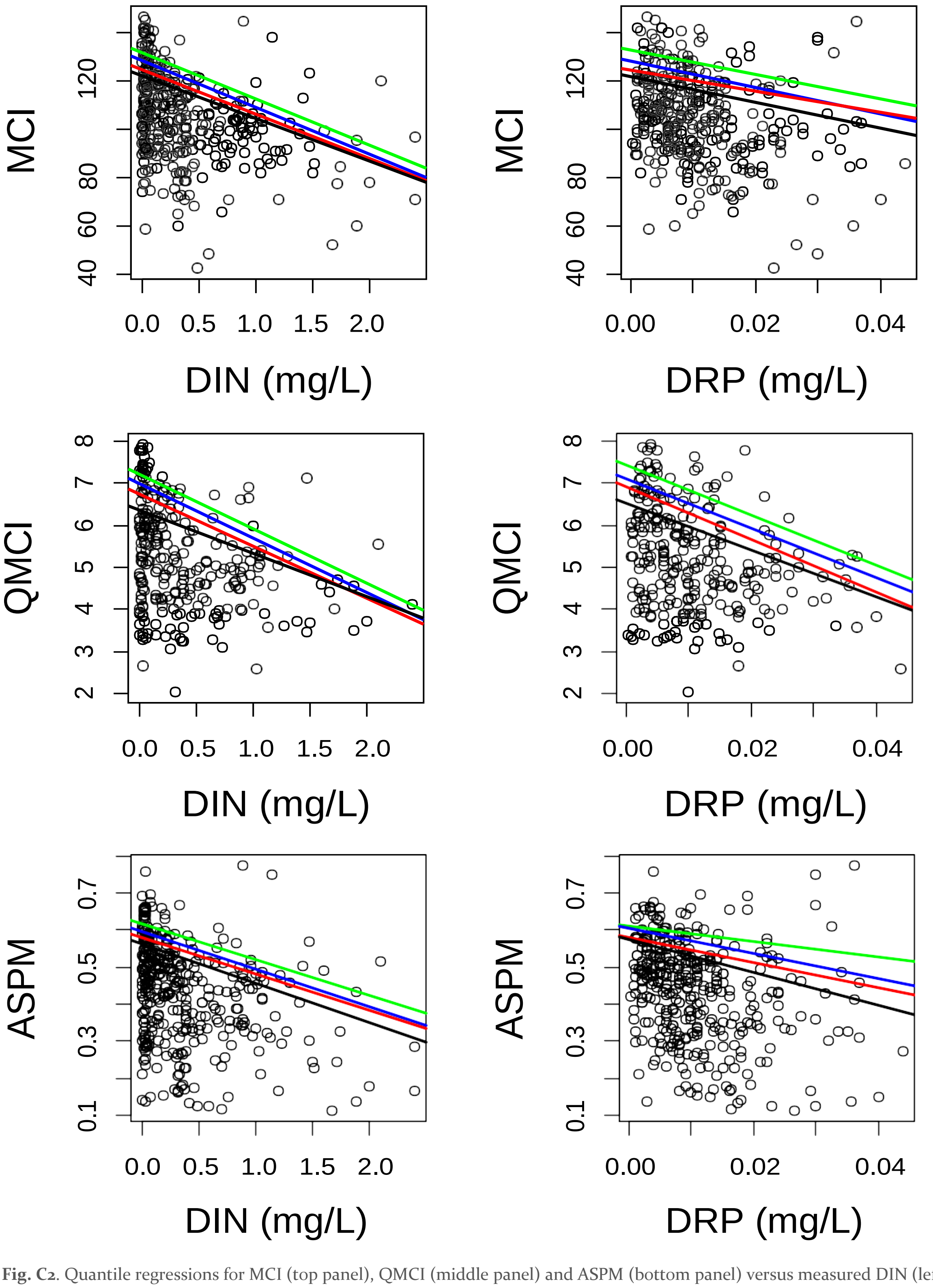


Fig. C2. Quantile regressions for MCI (top panel), QMCI (middle panel) and ASPM (bottom panel) versus measured DIN (left) and DRP (right) at 293-388 sites. The green, blue, red and black lines represent the 90th, 85th, 80th, and 75th percentiles respectively.

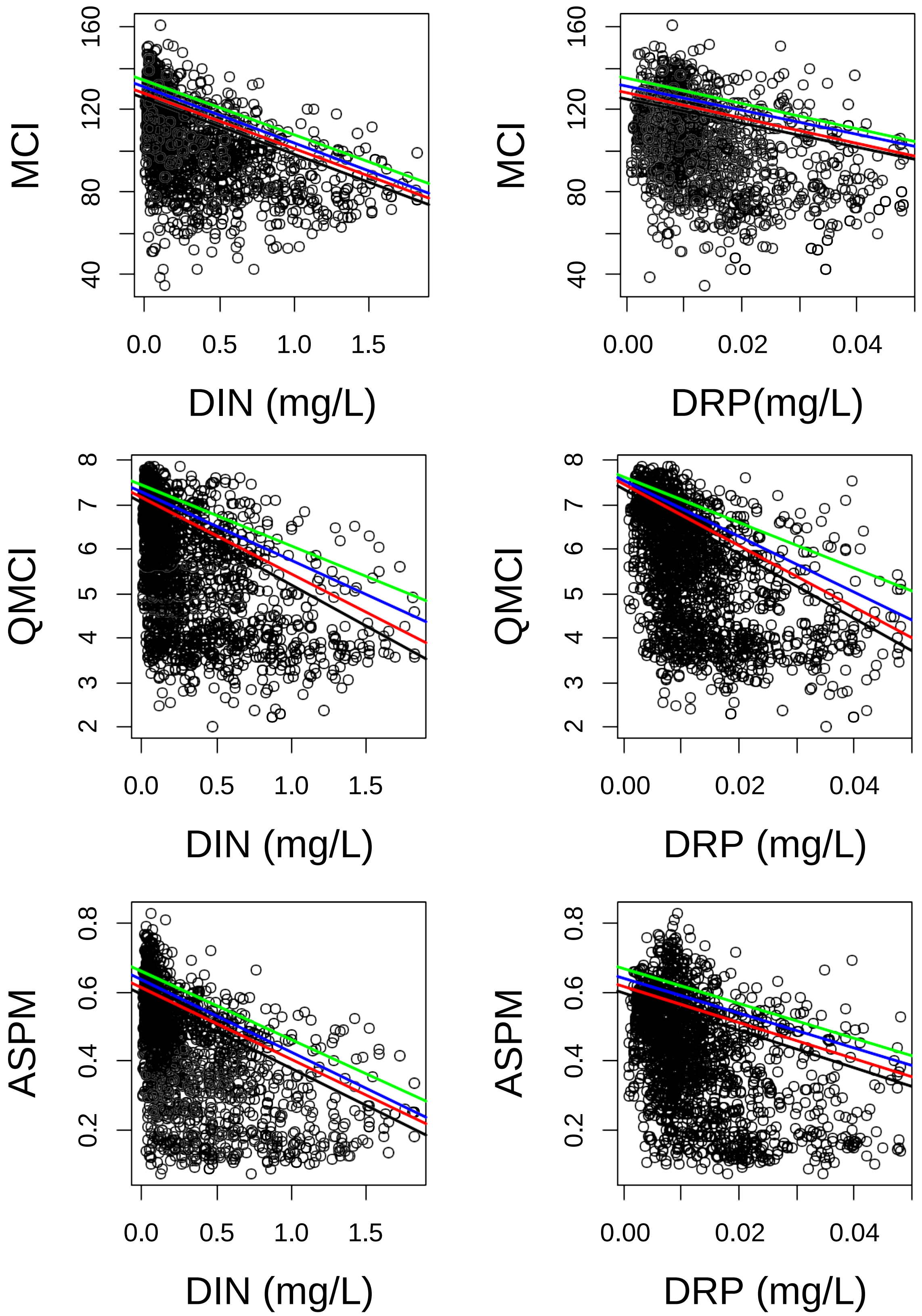


Fig. C3. Quantile regressions for MCI (top panel), QMCI (middle panel) and ASPM (bottom panel) versus modelled DIN (left) and DRP (right) (n=1851 sites). The green, blue, red and black lines represent the 90th, 85th, 80th, and 75th percentiles respectively.

APPENDIX D: Gradient forests

This analysis aims to use gradient forests to explore the extent to which macroinvertebrate assemblage turnover relates to DIN and DRP concentrations, accounting for the influence of other determining factors or stressors.

Annual benthic macroinvertebrate community data used in this study was sourced from New Zealand's regional monitoring network between 1990 and 2016, and then standardized to a consistent taxonomic resolution, by Clapcott *et al.*, (2017). Whilst most sites have been surveyed consistently for multiple, consecutive years, sites have periodically changed, and number increased, over the last 20 years. Macroinvertebrates are typically surveyed in riffles either using kick nets or Surber nets, then stored either in ethanol or formalin, then identified using common keys (Winterbourn, Gregson and Dolphin, 1989). Given the inevitable differences that may occur from different survey techniques, collectors and observers, relative abundance (percentages) was used instead absolute abundance. A total of 396 species were identified from 15,508 surveys collected from 1966 sites nationwide. Only surveys that had monthly concentrations of dissolved inorganic nitrogen (DIN) and Dissolved Reactive Phosphorus (DRP) for the preceding 12 months were used. This yielded a collection of 1784 surveys across 312 sites. Species were only included in the study if they were present in at least 20 surveys with at least five unique variables (n=179, gradient forest requirement).

Most sites do not have site-specific physical habitat surveys, and where they do, they are often surveyed inconsistently. Instead, most environmental variables were extracted from the Freshwater Environments New Zealand (FENZ) geodatabase (Leathwick *et al.*, 2010), except the hydrological characteristics which were sourced from Booker & Woods (2014) and predicted fine sediment cover from (Clapcott *et al.*, 2011) (28 variables, Table D1). FENZ is a geodatabase that contains modelled habitat characteristics, and calculated riparian and catchment cover, for every river reach in New Zealand. Booker & Woods (2014) modelled hydrological statistics of flow volume, flow variability and stream width for all river reaches.

Table D1. Environmental predictors from FENZ (Leathwick *et al.*, 2010) used in BRT explorations.	
SegJanAirT	Average January Air Temperature (°C)
SegMinTNorm	Average minimum daily air temperature (°C) normalised with respect to SegJanAirT
SegSlope	Slope of segment (°)
SegRipShade	The likely proportion of stream shaded from riparian
SegRipNative	Proportion of native riparian vegetation within a 100 m buffer of the river
USCalcium	Average calcium concentration of rocks in the catchment, 1 = very low to 4 = very high
USHardness	Average hardness of rocks in the catchment, 1 = very low to 5 = very high
USPhosporus	Average phosphorus concentration of rocks in the catchment, 1 = very low to 5 = very high
USPeat	Proportion of upstream catchment covered by peat
USWetland	Proportion of upstream catchment covered by wetland
USGlacier	Proportion of upstream catchment covered by glacier
ReachSed	Weighted average of proportional cover of bed sediment using categories of: 1 = mud; 2 = sand; 3 = fine gravel; 4 = coarse gravel; 5 = cobble; 6 = boulder; 7 = bedrock
ReachHab	Weighted average of proportional cover of local habitat using categories of: 1 = still; 2 = backwater; 3 = pool; 4 = run; 5 = riffle; 6 = rapid; 7 = cascade
DSDist2Coast	Distance to coast (km)
USDaysRain	Days/year with rainfall greater than 25 mm in the upstream catchment
USLake	Proportion of upstream catchment covered by lake
USWetland	Proportion of upstream catchment covered by wetland
USIndigFor	Proportion of upstream catchment covered by indigenous forest
USNative	Proportion of upstream catchment covered by native vegetation
USPasture	Proportion of upstream catchment covered by pasture
USGlacier	Proportion of upstream catchment covered by glaciers
MALF	Mean annual 7-day low flow (m^3/sec)
MeanF	Mean of all daily flows (m^3/sec)
Feb	Mean daily February flow divided by the overall mean daily flow
WidthMALF	Predicted wetted width (m) at MALF
FRE3	Predicted annual frequency of flows exceeding three times the annual median flow
SedAdded	The difference between the current and human-absent predicted fine sediment cover (%)
SedExpected	Predicted fine sediment cover (%) expected in human-absent conditions

Gradient forests are an extension of the random forest approach and assess how the compositional turnover of ecological assemblages change over gradients (Ellis, Smith and Pitcher, 2012). Random forests are a collective of regression trees, whereby each tree is fitted to a bootstrapped sample (with replacement) and then validated on the out-of-bag sample (Breiman, 2001). Random forest predictions are the average of the predictions of each tree.

Regression trees, and consequently random forests, work by partitioning observations at splits of predictors that minimise the sum of squares error. They have a high level of flexibility and can handle non-linear relationships and complex interactions (Cutler *et al.*, 2007; Hastie, Tibshirani and Friedman, 2009; Ellis, Smith and Pitcher, 2012). Furthermore, the R package *extendedForest* (Ellis, 2019a) not only computes the Breiman and Cutler's random forests, but also records the raw importance, and can calculate the conditional (instead of marginal) permutation importance of correlated predictors for a given correlation, following Strobl et al (2008) and explained in Appendix A of Ellis (2012).

In the gradient forest approach, the *extendedForest* package (Ellis, 2019a) is used to grow univariate random forests that predict each species abundances from environmental predictors. The quantum of species turnover is indicated by the amount of change across neighbouring partitions of a given split. The *gradientForest* package (Ellis, 2019b) calculates assemblage compositional turnover by aggregating these quanta across the entire species assemblage (when cross-validated $R^2>0$), weighted by the goodness of fit of each species and predictor importance (see Ellis (2012) for more details).

Overall, there was very little collinearity between predictors (Table D2); however, the MeanF was highly correlated with MALF ($r^2=0.95$), and USNative was well correlated with USPasture ($r^2=0.77$).

Of the 180 macroinvertebrate species included, 83 had positive R^2 from the random forest models, the average R^2 was 0.18, though the range was wide (0.6-0.004). The six best-fitting taxa represented a range of groups, feeding strategies, river type preferences and pollution tolerance levels (Table D3). The gradient forest analysis ranked DRP, DIN, Year, SegJanAirT, USDaysRain and Feb flows as the six most important predictors of macroinvertebrate assemblages across the country (Fig. D1.).

Figure D2 shows the cumulative importance of splits in predicting the cumulative change in species relative abundance (most important species listed in the legend) and overall community composition across the DIN and DRP gradients. The 5th, 10th and 20th percentiles of the overall change in community composition range occurred at 0.029, 0.055, 0.111 mg/L DRP respectively and 0.36, 0.69, 1.37 mg/L DIN respectively.

Table D2. Correlation coefficients (R^2) between predictors in the gradient forest analysis of macroinvertebrate communities

	Year	DRP	DIN	SegJanAirT	SegMinTNorm	SegSlope	SegRipShade	SegRipNative	DSDist2Coast	USDaysRain	USCalcium	USHardness	USPhosporus	USPeat	USLake	USWetland	USIndigFor	USNative	USPasture	USGlacier	ReachSed	ReachHab	MALF	MeanF	Feb	WidthMALF	FRE3	SedExpected	SedAdded
Site	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.01	0.00	0.00
Year		0.00	0.00	0.02	0.00	0.00	0.03	0.00	0.01	0.01	0.00	0.03	0.00	0.00	0.01	0.00	0.02	0.04	0.04	0.01	0.03	0.02	0.02	0.03	0.02	0.06	0.00	0.00	0.03
DRP			0.06	0.01	0.00	0.00	0.00	0.01	0.04	0.00	0.03	0.07	0.01	0.00	0.00	0.00	0.03	0.08	0.02	0.00	0.05	0.04	0.01	0.01	0.00	0.02	0.06	0.00	0.12
DIN				0.00	0.00	0.01	0.01	0.08	0.01	0.05	0.14	0.03	0.08	0.00	0.01	0.00	0.11	0.23	0.23	0.00	0.15	0.01	0.02	0.02	0.01	0.05	0.10	0.00	0.16
SegJanAirT					0.00	0.00	0.01	0.00	0.00	0.07	0.01	0.04	0.27	0.00	0.01	0.07	0.01	0.02	0.01	0.00	0.04	0.03	0.00	0.00	0.08	0.00	0.02	0.14	0.07
SegMinTNorm						0.01	0.00	0.00	0.19	0.00	0.00	0.01	0.02	0.03	0.01	0.00	0.02	0.03	0.02	0.01	0.00	0.00	0.01	0.01	0.05	0.01	0.20	0.01	0.00
SegSlope							0.28	0.06	0.01	0.00	0.06	0.07	0.00	0.00	0.00	0.00	0.02	0.03	0.10	0.00	0.04	0.01	0.01	0.01	0.00	0.05	0.03	0.00	0.01
SegRipShade								0.27	0.03	0.02	0.01	0.02	0.00	0.00	0.00	0.01	0.01	0.00	0.03	0.00	0.04	0.02	0.03	0.04	0.00	0.18	0.00	0.01	0.08
SegRipNative									0.00	0.01	0.09	0.09	0.02	0.00	0.04	0.01	0.13	0.23	0.22	0.00	0.04	0.02	0.00	0.00	0.00	0.00	0.02	0.00	0.06
DSDist2Coast										0.02	0.04	0.01	0.05	0.00	0.00	0.01	0.00	0.03	0.02	0.00	0.01	0.00	0.00	0.00	0.02	0.01	0.04	0.00	0.01
USDaysRain											0.08	0.01	0.09	0.01	0.02	0.02	0.24	0.14	0.13	0.01	0.11	0.06	0.04	0.06	0.05	0.10	0.31	0.01	0.05
USCalcium												0.12	0.13	0.01	0.00	0.00	0.19	0.31	0.31	0.01	0.14	0.04	0.01	0.01	0.02	0.05	0.08	0.00	0.13
USHardness													0.01	0.05	0.01	0.00	0.11	0.27	0.29	0.01	0.12	0.14	0.02	0.02	0.06	0.03	0.02	0.02	0.14
USPhosporus														0.01	0.00	0.01	0.09	0.02	0.05	0.00	0.02	0.00	0.00	0.00	0.07	0.01	0.13	0.02	0.01
USPeat															0.00	0.01	0.01	0.01	0.00	0.00	0.03	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.01
USLake																0.00	0.00	0.02	0.04	0.06	0.00	0.00	0.18	0.17	0.07	0.12	0.01	0.02	0.01
USWetland																	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
USIndigFor																		0.47	0.38	0.00	0.18	0.06	0.00	0.00	0.02	0.02	0.24	0.00	0.16
USNative																			0.77	0.02	0.29	0.12	0.06	0.06	0.07	0.13	0.09	0.01	0.41
USPasture																				0.02	0.26	0.09	0.05	0.05	0.08	0.09	0.07	0.00	0.30
USGlacier																					0.00	0.00	0.44	0.28	0.07	0.15	0.00	0.02	0.00
ReachSed																						0.34	0.00	0.01	0.04	0.08	0.10	0.20	0.60
ReachHab																							0.04	0.03	0.04	0.00	0.04	0.19	0.27
MALF																								0.95	0.10	0.60	0.00	0.07	0.02
MeanF																									0.08	0.68	0.00	0.04	0.03
Feb																										0.10	0.06	0.02	0.02
WidthMALF																											0.00	0.01	0.16
FRE3																												0.00	0.07
SedExpected																													0.08

Table D3. The six best fitting taxa used in the gradient forests, along with their fits (R^2), functional feeding groups, common habitat and tolerance to organic pollution.

Taxa	**R^2**	**Group**	**Functional feeding group**	**Common habitat**	**Tolerance to organic pollution**
Paracalliope spp	0.60	Amphipod	Collector-gatherer	Lowland, slow-flowing weedy streams	Moderately tolerant
Deleatidium spp	0.50	Mayfly	Scraper	Most abundant macroinvertebrate. Cool, stony-bottom rivers and streams.	Low tolerance
Confluens spp	0.45	Caddisfly	Scraper	Stony, bush-covered streams	Moderately tolerant
Pycnocentrodes spp	0.42	Caddisfly	Collector-gatherer and scraper	Stony and fine gravel bed streams	Moderately tolerant
Potamopyrgus antipodarum	0.42	Mud snail	Scraper	Widespread on any submerged surface	High tolerance
Glossiphoniidae	0.41	Leech	Predator	Lowland, slow-flowing weedy streams	High tolerance

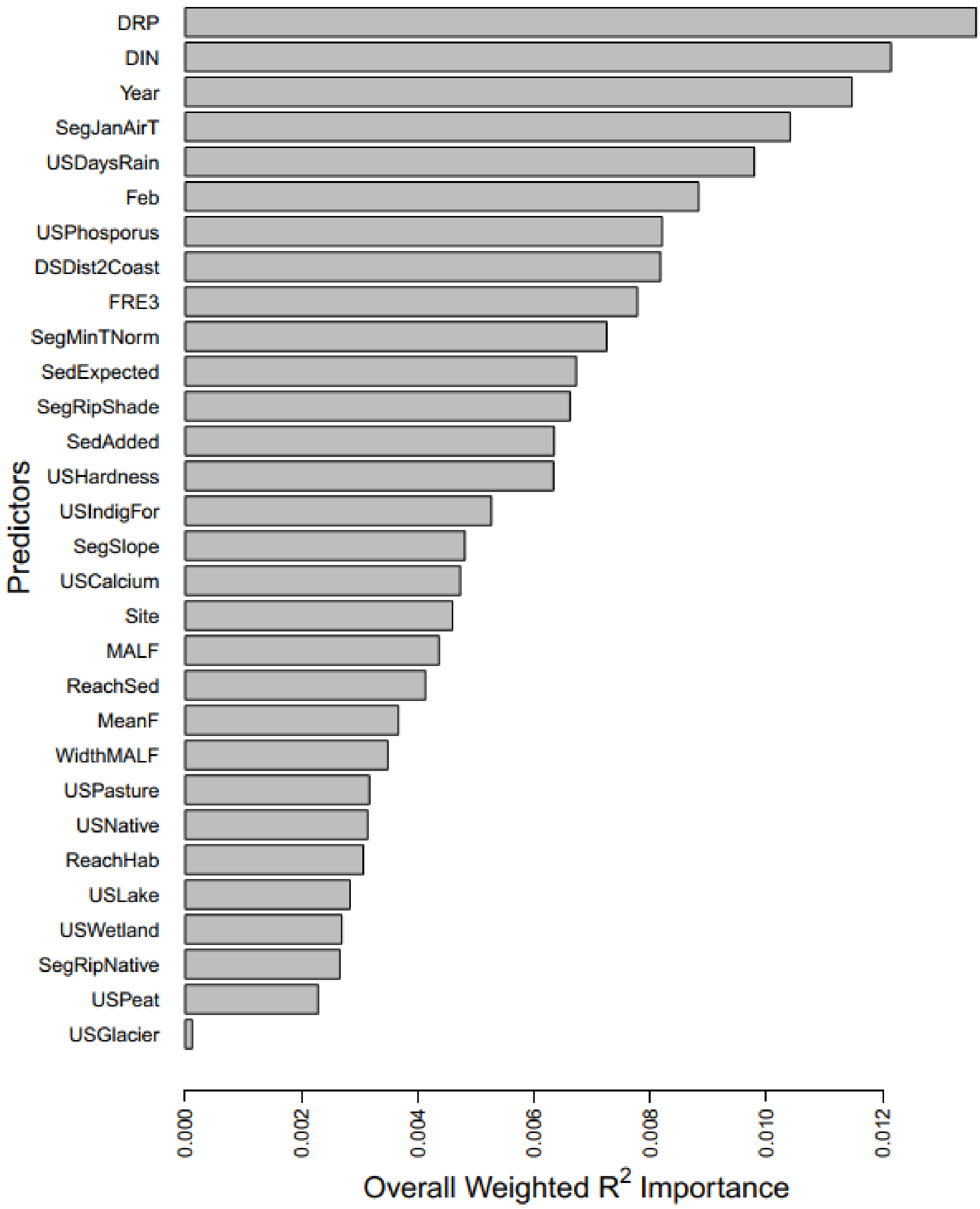


Fig. D1. The overall weighted R^2 importance of predictors in determining macroinvertebrate assemblage changes.

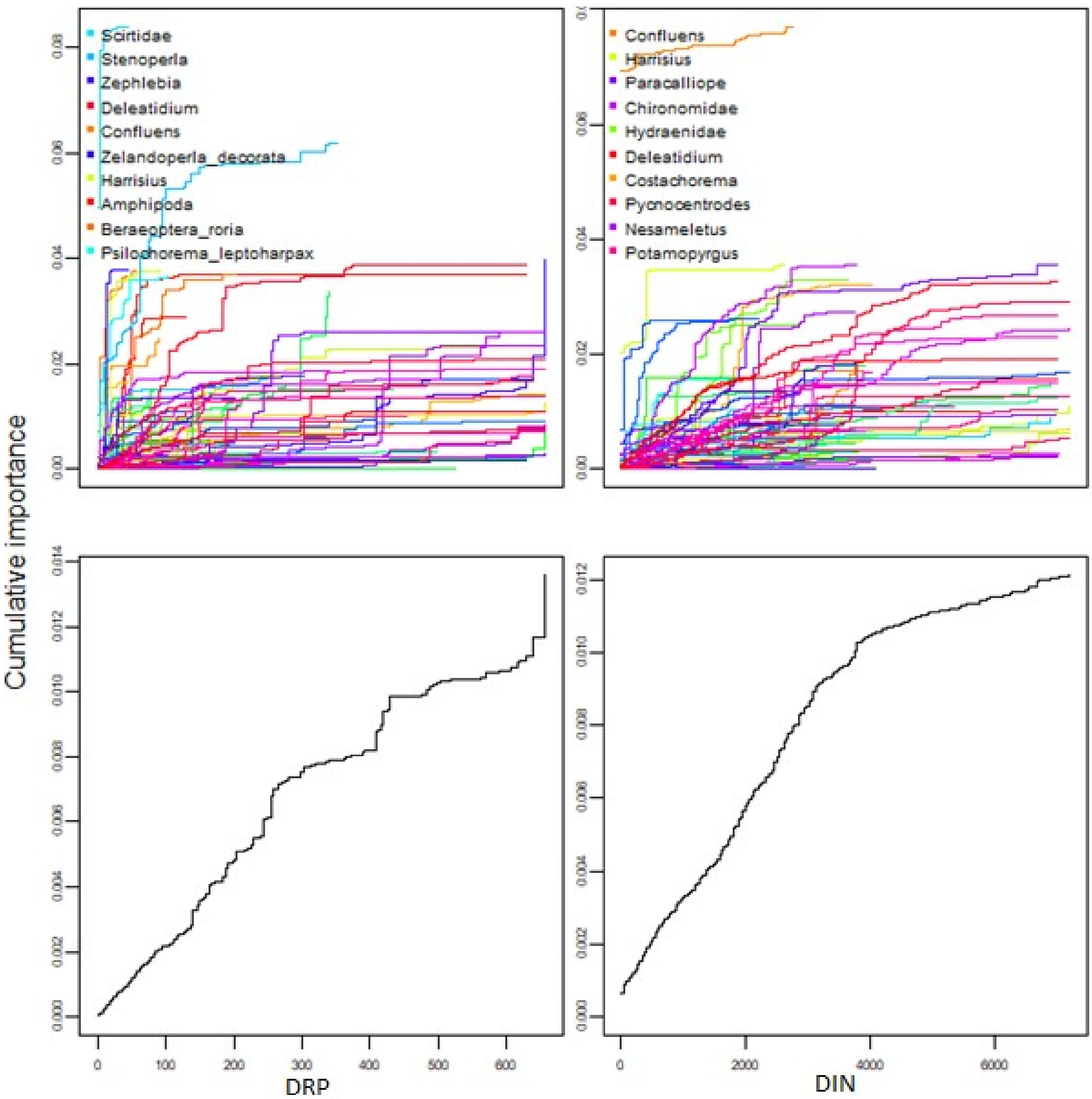


Fig. D2. The cumulative importance distributions of splits improvement scaled by R^2 weighted importance and standardized by density of observations. The top plots indicate cumulative change in abundance of individual species along DRP (left) and DIN (right) gradients. The bottom plots indicate cumulative change in overall composition of the community along DRP (left) and DIN (right) gradients. Nutrient concentrations are in mg/m³. Legends list the most important species in each relationship.

APPENDIX E: Fish assemblage O/E

This analysis aims to explore the relationship between the fish observed/expected ratios and nutrient concentrations, after accounting for other potential stressors.

Canning (2018) used the NZ Freshwater Fish Database and Boosted Regression Trees (BRTs) to predict the distribution of 24 native fish across New Zealand in both contemporary and reference condition. For every river reach, the number of fish species predicted to currently be present (observed) was divided by the number of fish species expected to occur in reference conditions (expected). The observed/expected (or O/E) ratio can provide an indication of ecological integrity of fish assemblages.

The fish O/E for every reach was then modelled (also using BRTs) only using the following stressors: presence/absence of a downstream dam; predicted nitrate-nitrogen and DRP concentrations; predicted E. Coli concentrations; the O/E riparian vegetation cover; the O/E fine sediment cover; and the predicted presence of exotic *Oncorhynchus mykiss* (Rainbow Trout), *Salmo trutta* (Brown Trout), *Perca fluviatilis* (Perch), *Scardinius erythrophthalmus* (Rudd), *Carassius auratus* (Goldfish), *Gambusia affinis* (Gambusia), *Oncorhynchus tshawytscha* (Chinook salmon) and *Ameiurus nebulosus* (Catfish). Both DRP and nitrate-nitrogen had the largest relative influence in predicting fish O/E, collectively accounting for approximately 48% of the models explanatory power, followed by downstream dams (20.7%), O/E riparian cover (16.1%) and O/E sediment cover (8.0%). Full details of the modelling is explained and discussed in Canning (2018).

Figure E1 shows the fitted functions plotted across the DIN and DRP gradients derived from the model of O/E as predicted by stressors. For DIN, the fitted function declines until a plateau between 1-1.3 mg/L and then continues declines. For DRP, there is an initial increase in fitted function at very low concentrations (those within the range typically expected to naturally occur (McDowell *et al.*, 2013), followed by a brief plateau between 0.008-0.015 mg/L, followed by a sharp decline between 0.015-0.02 mg/L, followed by a shallow decline between 0.02-0.025 mg/L and then a plateau.

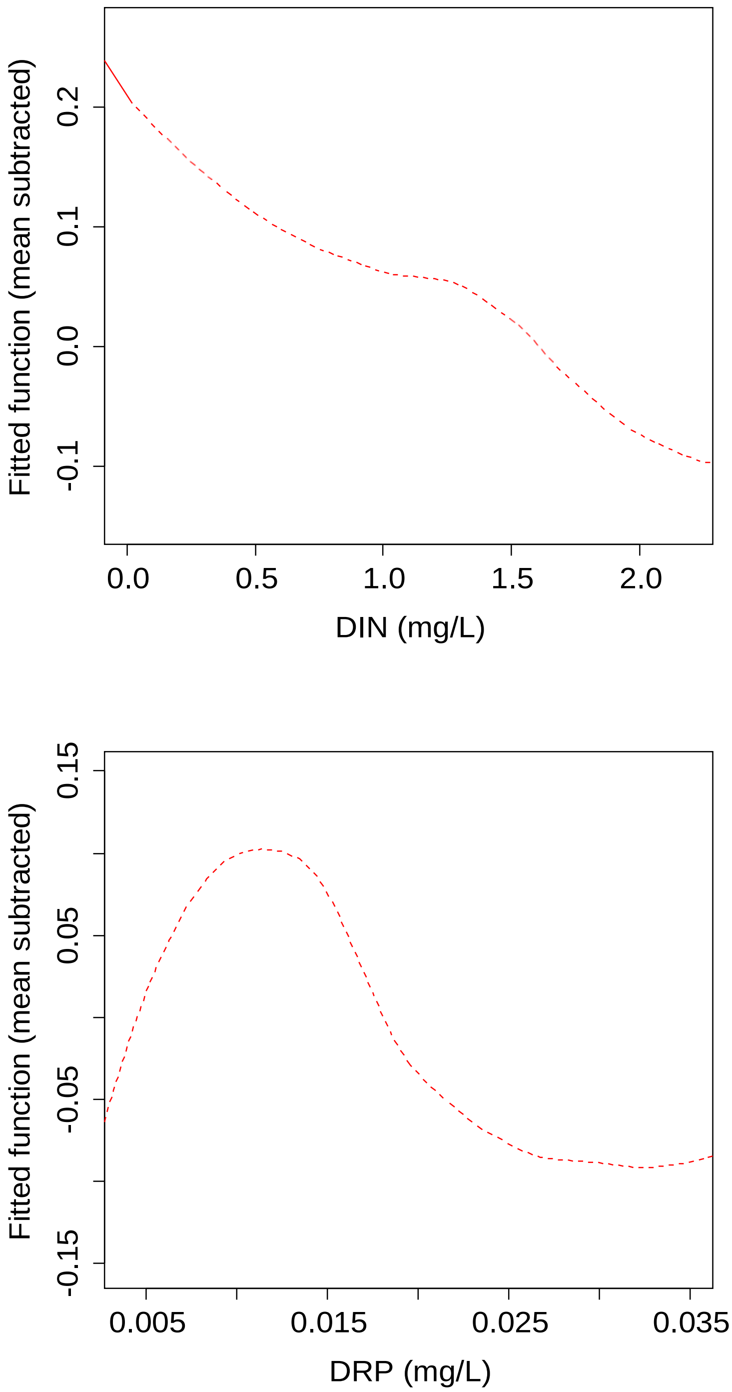


Fig. E1. Fitted functions of Fish O/E for DIN (left) and DRP (right) derived from BRT modelling of stressors, described in Canning (2018).

APPENDIX F: Invertebrate Change Analysis

This analysis aims to examine how the probability of taxa occurrence changes across DIN and DRP gradients.

New Zealand's rivers are managed by 15 regional authorities, all of which regularly monitor a range of sites across their river networks. Monitoring sites are not necessarily representative of the river network (i.e., they are non-random) and may be chosen for a range of regions, for example, a council may wish monitor nutrients near a flow recorder to calculate nutrient loads, or they be interested in monitoring the effects of a particular activity, or gathering data on reference conditions in conservation land. None the less, nutrients are currently monitored monthly at well over 1000 sites nationwide.

Annual benthic macroinvertebrate community data used in this study was sourced from New Zealand's regional monitoring network between 1990 and 2016, and then standardized to a consistent taxonomic resolution, by Clapcott *et al.*, (2017). Whilst most sites have been surveyed consistently for multiple, consecutive years, sites have periodically changed, and number increased, over the last 20 years. Macroinvertebrates are typically surveyed in riffles either using kick nets or Surber nets, then stored either in ethanol or formalin, then identified using common keys (Winterbourn, Gregson and Dolphin, 1989). A total of 396 species were identified from 15,508 surveys collected from 1966 sites nationwide. Only surveys that had monthly concentrations of dissolved inorganic nitrogen (DIN) and Dissolved Reactive Phosphorus (DRP) for the preceding 12 months were used. Of the 396 taxa observed, only those that occurred on at least 300 surveys were included in the analysis (n=60). There were 1784 surveys that contained at least one 60 common taxa included as well as having paired DIN and DRP monitoring.

Each site was assigned to one of 50 bins for DIN (increasing sequentially from 0 to 2.5 mg/L by 0.05 mg/L) and one of 30 bins for DRP (increasing sequentially from 0 to 0.06 mg/L by 0.002 mg/L), depending on the median concentrations of the previous 12 months.

For each taxa and nutrient, a binomial regression was fitted across all surveys to predict the probability of occurrence for each taxon in a given bin. Binomial regressions were calculated using R (R Core Team, 2016). All taxa assessed had highly significant binomial regressions ($p<0.001$). Each taxon was considered to be substantially impacted by nutrients when their probability of occurrence changed by more than 20% relative to that predicted for very low nutrient concentration bins. Absolute change is used rather than directional change as increases in some taxa can adversely affect ecosystem health along with reductions in taxa. Whether the relationships change with environment type has not been explored – exploration is at the national scale. Furthermore, this analysis assessed the probability of occurrence, not abundance or density, it is unknown whether the two correlate.

Figure F1 shows the proportion of all taxa (%) that are likely to experiencing a less than substantial impact on their probability of occurrence for both DIN and DRP. Curves are LOESS with a 0.50 span. For DIN, 95% taxa protection occurred at 0.2 mg/L, 80% at 0.44 mg/L, 60% at 0.67 mg/L and 40% at 1.14 mg/L. For DRP, 95% taxa protection occurred at 0.008 mg/L, 80% at 0.011 mg/L, 60% at 0.018 mg/L and 40% at 0.029 mg/L.

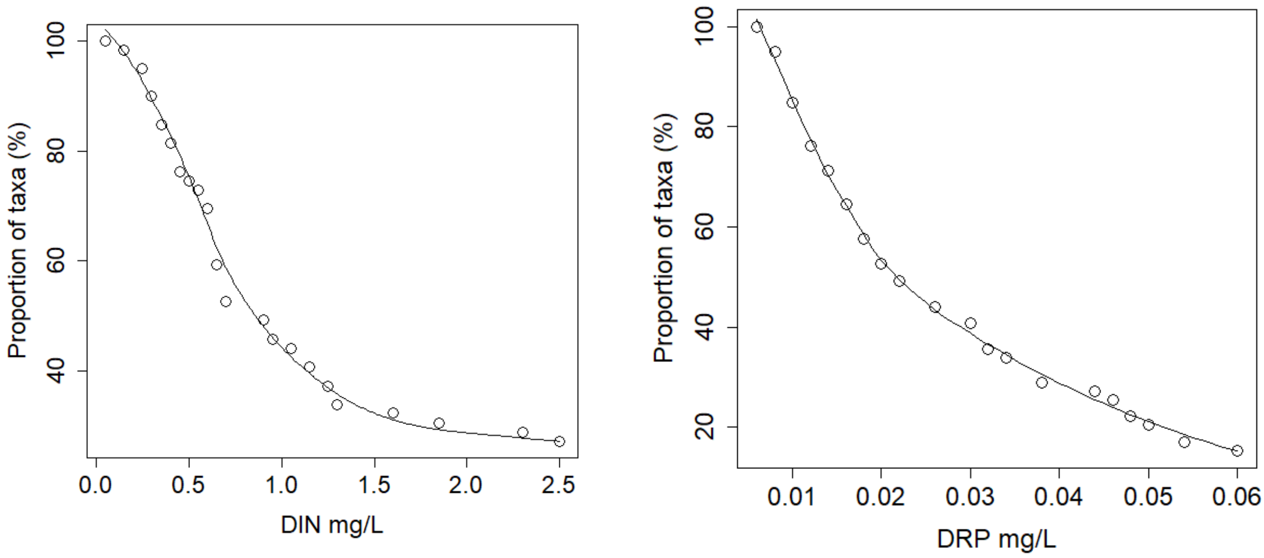


Fig. F1. The proportion of taxa that are predicted to have >20% change (positive or negative) in probability of occurrence than when DIN (left) or DRP (right) are increased from a 0 mg/L baseline.

APPENDIX G: Proposed Attribute Tables

Value	Ecosystem health		
Freshwater Body Type	Rivers [1]		
Attribute	Dissolved inorganic nitrogen		
Attribute Unit	DIN mg/L		
Attribute State	**Numeric Attribute State[2]**		**Narrative Attribute State**
	Median	**95th percentile**	**Description**
A	≤ 0.24	≤ 0.56	Ecological communities and ecosystem processes are similar to those of natural reference conditions. No adverse effects attributable to DIN enrichment are expected.
B	> 0.24 and ≤0.50	> 0.56 and ≤01.10	Ecological communities are slightly impacted by minor DIN elevation above natural reference conditions. If other conditions also favour eutrophication, sensitive ecosystems may experience additional algal and plant growth, loss of sensitive macroinvertebrate taxa, and higher respiration and decay rates.
C	> 0.5 and ≤ 1.0	> 1.10 and ≤ 2.05	Ecological communities are impacted by moderate DIN elevation above natural reference conditions, but sensitive species are not experiencing nitrate toxicity. If other conditions also favour eutrophication, DIN enrichment may cause increased algal and plant growth, loss of sensitive macroinvertebrate & fish taxa, and high rates of respiration and decay.
National Bottom Line	1.0	2.05	
D	>1.0	>2.05	Ecological communities impacted by substantial DIN elevation above natural reference conditions. In combination with other conditions favouring eutrophication, DIN enrichment drives excessive primary production and significant changes in macroinvertebrate and fish communities, as taxa sensitive to hypoxia and nitrate toxicity are lost.

1. Groundwater concentrations also need to be managed to ensure resurgence via springs and seepage does not degrade rivers through DIN enrichment.
2. Must be derived from the median of monthly monitoring over the most recent five years.

DRP option one:

Value	Ecosystem health		
Freshwater Body Type	Rivers		
Attribute	Dissolved reactive phosphorus		
Attribute Unit	DRP mg/L		
Attribute State	Numeric Attribute State[1]		Narrative Attribute State
	Median	95^{th} percentile	Description
A	≤ 0.006	≤ 0.021	Ecological communities and ecosystem processes are similar to those of natural reference conditions. No adverse effects attributable to DRP enrichment are expected.
B	> 0.006 and ≤0.010	> 0.021 and ≤0.030	Ecological communities are slightly impacted by minor DRP elevation above natural reference conditions. If other conditions also favour eutrophication, sensitive ecosystems may experience additional algal and plant growth, loss of sensitive macroinvertebrate taxa, and higher respiration and decay rates.
C	> 0.010 and ≤ 0.018	> 0.030 and ≤ 0.054	Ecological communities are impacted by moderate DRP elevation above natural reference conditions. If other conditions also favour eutrophication, DRP enrichment may cause increased algal and plant growth, loss of sensitive macro-invertebrate & fish taxa, and high rates of respiration and decay.
National Bottom Line	0.018	0.054	
D	>0.018	>0.054	Ecological communities impacted by substantial DRP elevation above natural reference conditions. In combination with other conditions favouring eutrophication, DRP enrichment drives excessive primary production and significant changes in macroinvertebrate and fish communities, as taxa sensitive to hypoxia are lost.

1. Must be derived from the median of monthly monitoring over the most recent five years.

DRP option two:

Value	Ecosystem health				
Freshwater Body Type	Rivers				
Attribute	Dissolved reactive phosphorus				
Attribute Unit	DRP mg/L				
Attribute State	**Numeric Attribute State[1]**				**Narrative Attribute State**
	Median		**95th percentile**		
	Default	**VA**	**Default**	**VA**	**Description**
A	≤ 0.006	≤ 0.013	≤ 0.021	≤ 0.028	Ecological communities and ecosystem processes are similar to those of natural reference conditions. No adverse effects attributable to DRP enrichment are expected.
B	> 0.006 and ≤0.010	> 0.013 and ≤0.017	> 0.021 and ≤0.030	> 0.028 and ≤0.037	Ecological communities are slightly impacted by minor DRP elevation above natural reference conditions. If other conditions also favour eutrophication, sensitive ecosystems may experience additional algal and plant growth, loss of sensitive macroinvertebrate taxa, and higher respiration and decay rates.
C	> 0.010 and ≤ 0.018	> 0.017 and ≤ 0.025	> 0.030 and ≤ 0.054	> 0.037 and ≤ 0.061	Ecological communities are impacted by moderate DRP elevation above natural reference conditions. If other conditions also favour eutrophication, DRP enrichment may cause increased algal and plant growth, loss of sensitive macro-invertebrate & fish taxa, and high rates of respiration and decay.
National Bottom Line	0.018	0.025	0.054	0.061	
D	>0.018	>0.025	>0.054	>0.061	Ecological communities impacted by substantial DRP elevation above natural reference conditions. In combination with other conditions favouring eutrophication, DRP enrichment drives excessive primary production and significant changes in macroinvertebrate and fish communities, as taxa sensitive to hypoxia are lost.

1. Must be derived from the median of monthly monitoring over the most recent five years.
2. VA refers to rivers in the volcanic acidic (VA) class in River Environment Classification

APPENDIX H: Comparison with Death et al

The derivation of nutrient criteria in this report is considerably different from that in (Death *et al.*, 2018, no date). Differences do not necessarily represent the authors' views but have arisen from STAG discussions and determinations. Table H1 summarizes a brief list of differences.

Table H1. Broad comparison between the nutrient criteria derived here and that by Death et al (2018)

	This analysis	Death et al (2018)
Weighting	No weightings applied. Each line of evidence and each trophic level were treated equally.	Each line of evidence was weighted depending on whether the response was deemed directly or indirectly affected by nutrients, regardless of the trophic level.
Spatial representation	Only nationally representative data used	Local, regional and national datasets used
Use of modelled data	No response was modelled. Modelled nutrients were only used when measured nutrients were unavailable, even if it meant using a substantially smaller and noisier dataset.	Modelled MCI, modelled O/E MCI and modelled nutrients used more liberally. This maximizes the amount of data available and reduces white noise but relies heavily on models being reliable.
Fish IBI	Quantile regressions were used to determine the relationships with nutrients as fish IBI is also substantially impacted by other drivers.	Mean regression used. Produces more precautionary criteria as there are likely interaction effects undetected by quantile regression. Fish IBI received lower weighting to reflect the indirect nature of the relationship.
Macroinvertebrate metrics	Metrics include MCI, QMCI and ASPM.	Metrics include MCI, QMCI and EPT.
Percentile-based criteria	No criteria based on distribution percentiles were included.	Criteria based on national nutrient concentration percentiles included.

APPENDIX I: Percentile-based exploration of nutrient criteria

This exploration aims to explore the distribution of nutrient concentrations across sites that meet (or are better than) a given band of a given metric, and then examine what nutrient criteria would look like if this approach was adopted in deriving national nutrient criteria instead of the proposed approach.

For each line of evidence used in the principal derivation (see main report body), the 80th and 90th percentiles are calculated for DIN and DRP for all sites that are better than or complaint with a given band. The same data and bands used for principle derivation were used in this analysis.

For example with MCI and DIN, across all sites where MCI is B band or better (>=110), then the 80th percentile is 0.32 mg/L, indicating that 20% of sites exceed 0.32 mg/L, and the 90th percentile is 0.50 mg/L, indicating that 10% of sites exceed 0.50 mg/L.

Each line of evidence was then averaged to derive the overall criteria in the same way described in Figure 4.

Overall, the DIN bottom-line derived using a percentile-based approach would be 0.68 mg/L or 0.96 mg/L, depending if the 80th or 90th percentile were adopted respectively. Whilst the DRP bottom-line derived using a percentile-based approach would be 0.015 mg/L or 0.020 mg/L, depending if the 80th or 90th percentile were adopted respectively. In both cases, the 90th percentiles align closely with the proposed nutrient bottom-lines presented in the report body of 1.0 mg/L and 0.018 mg/L for DIN and DRP respectively.

Table I1. The 80th and 90th percentiles for all DIN and DRP concentrations across all sites where a metric meets or exceeds a given band. Included are also the average for each trophic group and the overall averages (using same averaging as in Fig. 4.) The same averaging without microbes and the minimum line are also provided.

	Band	Peri	MCI	QMCI	ASPM	IBI	R	GPP	Cotton	Peri - mean	Inverts - mean	Fish - mean	Microbes	**AVG w/ microbes**	AVG w/o microbes	Min
80th percentile DIN	A	0.41	0.08	0.24	0.12	0.29	0.47	0.73	0.52	0.41	0.14	0.12	0.57	**0.31**	0.23	0.08
	B	0.52	0.32	0.34	0.53	0.37	0.56	0.73	0.56	0.52	0.43	0.53	0.62	**0.53**	0.49	0.32
	C	0.59	0.68	0.80	0.71	0.42	0.67	0.74	0.71	0.59	0.72	0.71	0.71	**0.68**	0.68	0.42
90th percentile DIN	A	0.67	0.20	0.51	0.20	0.46	0.61	0.67	0.71	0.67	0.27	0.20	0.66	**0.45**	0.38	0.20
	B	0.82	0.50	0.66	0.95	0.61	0.70	0.72	0.76	0.82	0.77	0.95	0.73	**0.82**	0.85	0.50
	C	0.86	0.98	1.05	1.00	0.71	0.96	0.82	1.12	0.86	1.01	1.00	0.96	**0.96**	0.96	0.71
80th percentile DRP	A	0.016	0.016	0.011	0.012	0.013	0.013	0.013	0.011	0.016	0.013	0.012	0.012	**0.013**	0.014	0.011
	B	0.018	0.013	0.012	0.014	0.015	0.013	0.014	0.013	0.018	0.013	0.014	0.013	**0.015**	0.015	0.012
	C	0.018	0.015	0.015	0.015	0.016	0.013	0.014	0.014	0.018	0.015	0.015	0.014	**0.015**	0.016	0.013
90th percentile DRP	A	0.020	0.030	0.014	0.019	0.017	0.015	0.015	0.015	0.020	0.021	0.019	0.015	**0.019**	0.020	0.014
	B	0.021	0.013	0.015	0.021	0.019	0.015	0.016	0.016	0.021	0.018	0.021	0.016	**0.019**	0.020	0.013
	C	0.021	0.015	0.022	0.023	0.020	0.015	0.016	0.017	0.021	0.021	0.023	0.016	**0.020**	0.022	0.015

APPENDIX J: Potential exclusion criteria

As there is always natural variability, in some rare situations, it may be possible for high nutrient enrichment to not result in the significant degradation of the aquatic life and ecosystem functioning components of ecosystem health. Demonstrating the lack of an impact in a single river adequately across all components of ecosystem (NOT simply with MCI alone) will be very difficult and will likely still be questionable. However, the following is a suggested criteria (all points should be met) for delineating potential exceptional sites where nutrient criteria below the proposed national bottom-lines may be considered:

a) The 92nd percentile for Chlorophyll a of consecutive monthly samples over a seven year period must not exceed 200 mg/m² or the community's target (where more stringent);
b) The median MCI must not be less than 90 or the community's expectations (where more stringent);
c) The median QMCI must not be less than 4.5 or the community's expectations (where more stringent);
d) The median ASPM must not be less than 0.3 or the community's expectations (where more stringent);
e) Anyone of the mean biomass density, body size, production and turnover ratio (production:biomass) for any of the following macroinvertebrate groups ephemeroptera, plecoptera, trichoptera, odonata, chironomidae, muscidae, culicidae oligochaeta, platyhelminthes, nemertea, gastropoda must not deviate more than 20% from any of the five reference sites;
f) The invertebrate metrics calculated for criteria b-e, must be derived using seven-years of consecutive annual samples, taken between January to March (inclusive) and no earlier than 20 days after a gravel moving flushing event. Samples are to be collected using seven Surber samplers in random riffles, using full invertebrate counts (no sub-sampling). Taxonomic resolution and sensitivity scores to be use is that from Table A1.1 from:

   Clapcott, J., Wagenhoff, A., Neale, M., Storey, R., Smith, B., Death, R., … Young, R. (2017). Macroinvertebrate metrics for the National Policy Statement for Freshwater Management. Cawthron: Nelson, New Zealand;

g) Dissolved Oxygen must not have a mean minimum less than 5.0 mg/L and a 1-day minimum of 4.0 mg/L or the community's target (where more stringent);

h) Gross primary production must not exceed 8.0 g O_2 m^{-2} d^{-1} in non-wadeable rivers or 8.0 g O_2 m^{-2} d^{-1} in wadeable rivers or the community's target (where more stringent);

i) Ecosystem respiration must not be less than 0.6 or greater than 13.0 g O_2 m^{-2} d^{-1} in non-wadeable rivers, and must be less than 0.8 or greater than 9.5 g O_2 m^{-2} d^{-1} in wadeable rivers or the community's target (where more stringent;

j) For conditions f-h, these must be derived using continuous dissolved oxygen monitoring for a minimum the entirety of December-March (inclusive) for seven consecutive years – not a single 7-day period;

k) The instantaneous flow must not be more than 20% lower than that estimated to occur in abstraction-free conditions or the community's target (where more stringent;

l) The fish IBI must not be lower than 18 or the community's target (where more stringent);

m) There must be a native vegetation riparian buffer of at least the width necessary to improve water quality, modulate stream temperatures, provide food and resources and improve in-stream biodiversity dependent on land use and type, defined as the recommendations in Table 3 of

   Hansen, B., Reich, P., Lake, P. S., & Cavagnaro, T. (2010). Minimum width requirements for riparian zones to protect flowing waters and to conserve biodiversity: a review and recommendations with application to the State of Victoria. Monash University, Melbourne.

n) The average rate of cotton decay must not deviate by more than 20% in any of the four seasons from measurements at any of a minimum five comparable sites in reference conditions;

   a. Cotton decay must be measured as loss tensile-strength, following:
      Tiegs, S. D., Clapcott, J. E., Griffiths, N. A., & Boulton, A. J. (2013). A standardized cotton-strip assay for measuring organic-matter decomposition in streams. Ecological indicators, 32, 131-139;

   b. Each survey must use at least ten cotton strips at each site per season, immersed in-stream for 30 days and be placed in locations deemed to be representative of that stream with regards to (but not limited to) depth, flow velocity, turbulence, leaf litter deposition, riparian cover;

o) The seasonal mean of 5-day biochemical oxygen demand collected monthly (three months per season) over a consecutive seven-year period must not be 20% higher any of a minimum five comparable sites in reference conditions;

p) Desired health of downstream ecosystems is achieved;

q) All reference condition sites must meet the following criteria:

    i. At least 90% native vegetation cover,
    ii. Less than 10% low intensity farming,
    iii. No horticulture,
    iv. No water abstraction,
    v. Native vegetation riparian buffers at least 100m wide,
    vi. No point source discharges,
    vii. No in-stream engineering works, and
    viii. No toxicant or nutrient greater than the relevant default guideline defined in:

    ANZECC & ARMCANZ 2000, Australian and New Zealand Guidelines for Fresh and Marine Water Quality, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra.

APPENDIX K: Comparison with reference condition

Any freshwater management targets set need to be ecologically achievable. In the case of setting nutrient targets, the targets should not be more stringent than that estimated to occur in reference state.

This section assesses the proposed nutrient criteria with river-class specific reference state estimates using two approaches:

1. Figure K1 compares the proposed DIN and DRP bands respectively with the predicted reference condition estimated by McDowell et al (2013) for the 3rd level river classification (climate, topography, and geology). McDowell et al (2013) used mixed-effects models to relate nutrient concentrations with river environment classes and land use. Median concentrations at reference state were then estimated by extrapolating relationships to no human influence. This analysis assumes that there is sufficient data across a pressure gradient to develop a reliable model that encapsulates the true relationship. If data are skewed, show heteroskedasticity, influenced by outliers or have low power across a considerable portion of the gradient, then predictions may be unreliable. Nonetheless, it still uses the best available data at the time. Many river classes in this analysis had very sites (in cases as little as two!), whilst using mixed effects modelling recruits the entire dataset to improve reliability, this can never completely account for a lack of data.

2. Figures K2-4 compare the proposed DIN and DRP bands with boxplots of measured 5-years medians (2012-2016) at sites considered reference reference condition, based on the REC's (Snelder, Biggs and Weatherhead, 2010) climate and geology classes (Figures K2 & K3), and the FENZ (Leathwick et al., 2010) level one alpha class (Figure K4). Sites in reference condition were defined as draining catchments with more than 80% native cover and less than 15% pasture. Measured data is that collected by SOE monitoring and made available on LAWA. This approach assumes that there are no other factors that substantially affect measured water quality, such as large point source discharges into native forest, or that the nutrient contribution from the small portion of pasture is low. Like approach (1), this method is also limited by data as some river classes have little data, it is for this reason that finer scale REC classes were not used.

Across both procedures, the measured and estimated DIN reference conditions almost always fell within the proposed A-band and never exceeded the proposed bottom-line. DRP, however, showed more scatter across the bands, though largely in the A and B bands. River classes with Volcanic Acidic (VA) geology appear to have considerably higher concentrations of DRP in reference state, with some sites or estimates near or exceeding the proposed DRP bottom-line. Unless a higher bottom-line is permitted for the VA geology class, then between 5-15% of sites could require exemptions for natural exceedances.

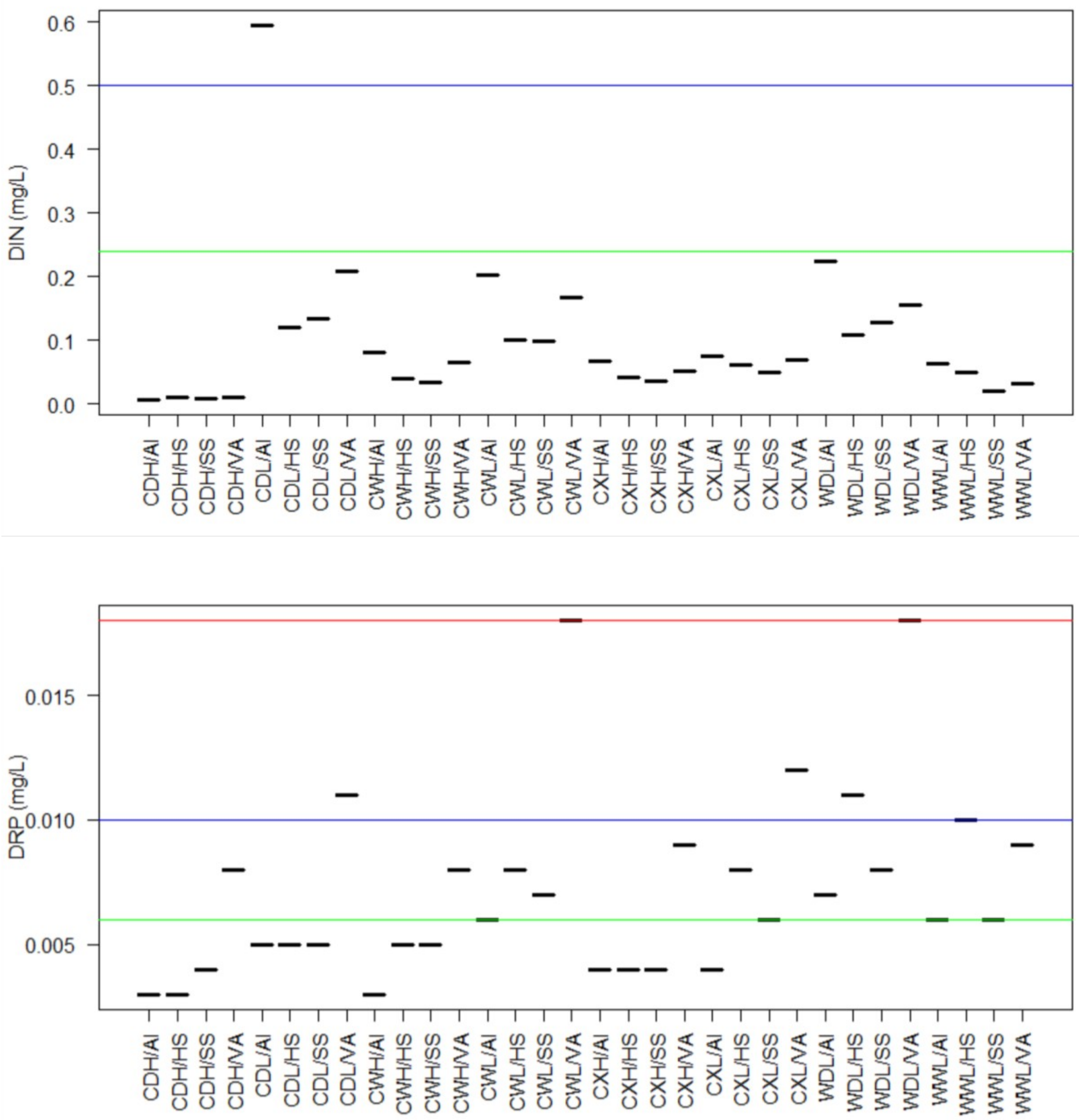
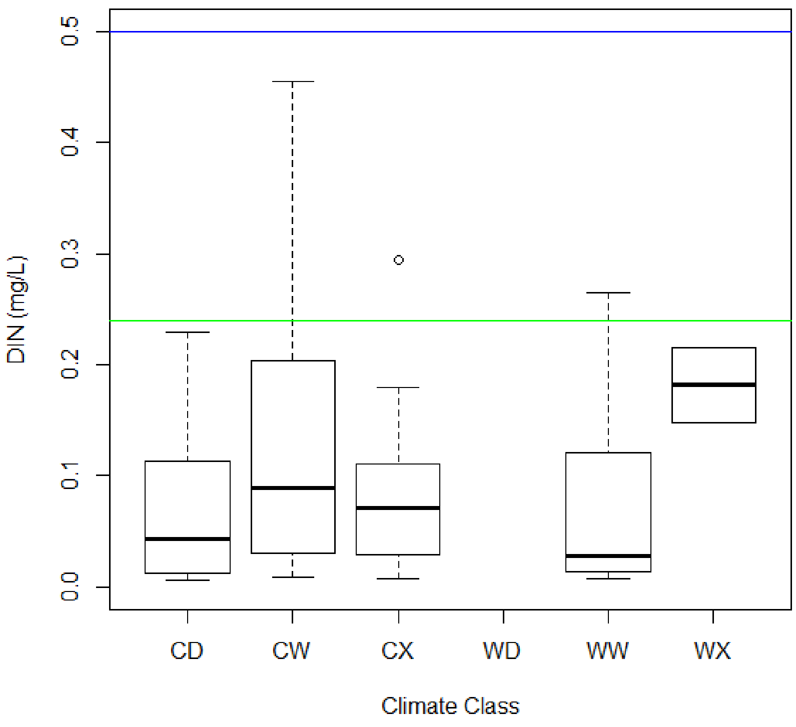


Fig. K1. The predicted median reference DIN (top) and DRP (bottom) concentrations estimated by McDowell et al (2013) for the 3rd level river classification (climate, topography, and geology). Lines indicate bottom of bands, red=C, blue=B, and green=A.



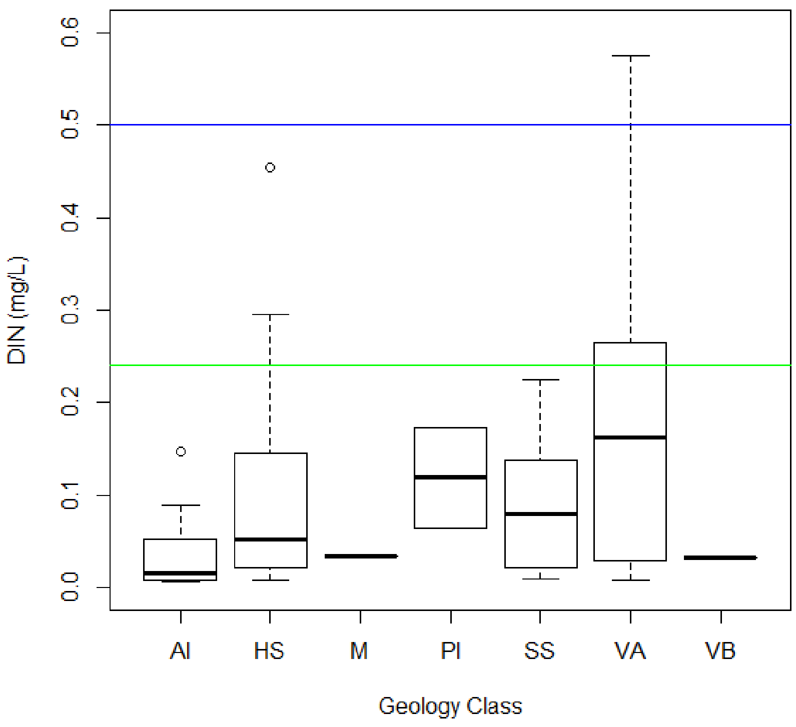


Fig. K2. The measured 5-year median DIN concentrations (mg/L) at sites considered reference condition across the REC climate classes (top) and geology classes (bottom). Lines indicate bottom of bands, red=C, blue=B, and green=A.

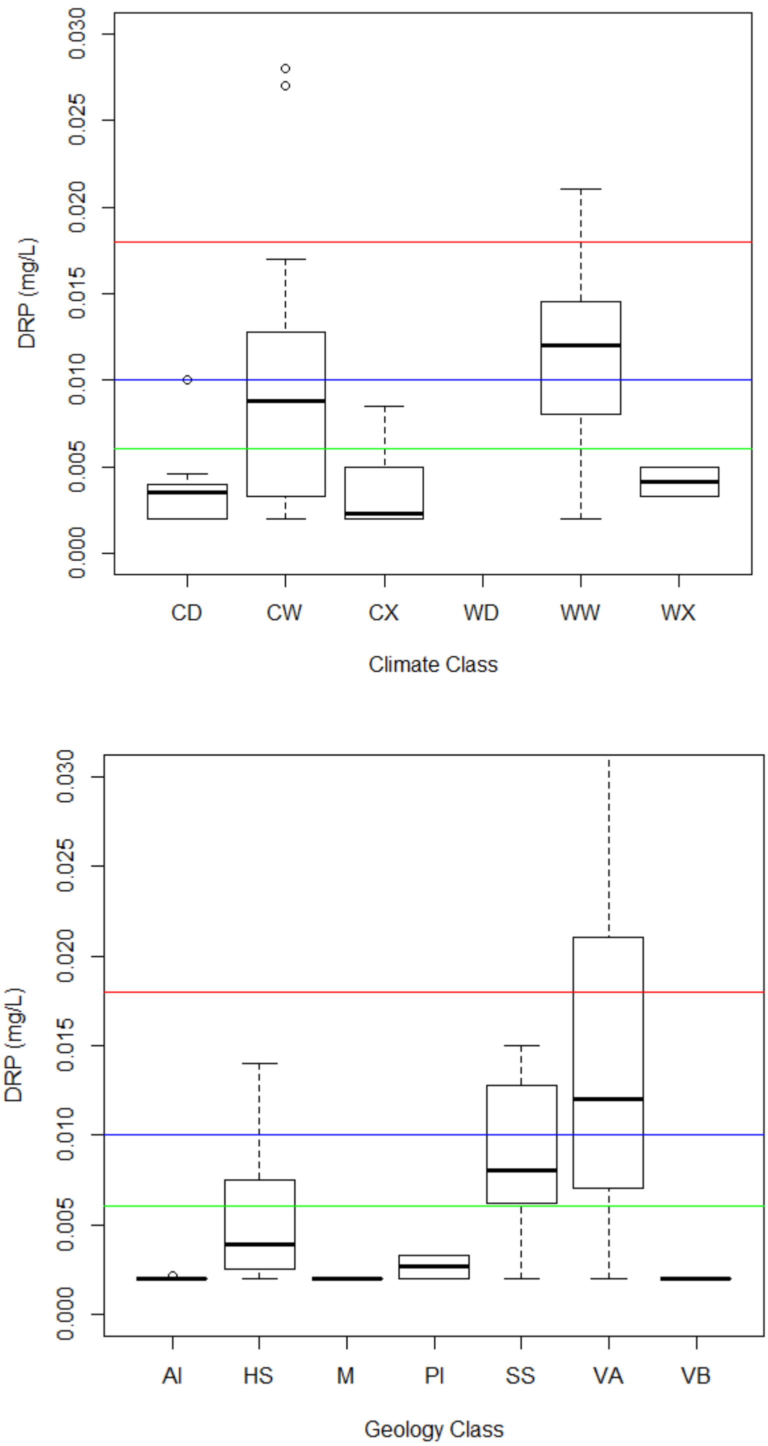
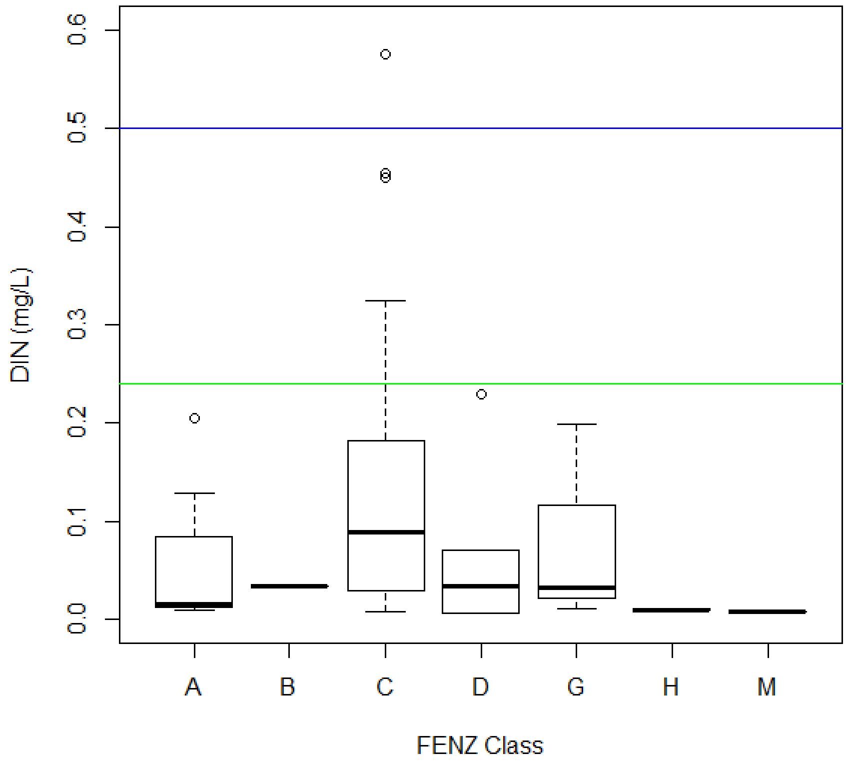


Fig. K3. The measured 5-year median DRP concentrations (mg/L) at sites considered reference condition across the REC climate classes (top) and geology classes (bottom). Lines indicate bottom of bands, red=C, blue=B, and green=A.



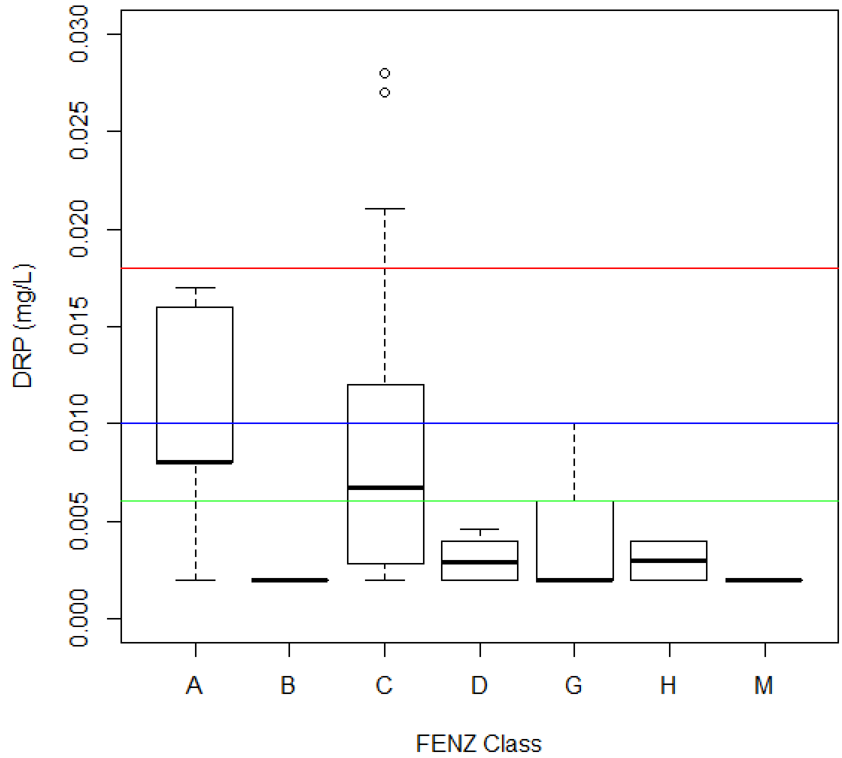


Fig. K4. The measured 5-year median DIN (top) and DRP (bottom) concentrations (mg/L) at sites considered reference condition across the FENZ LevelOneAlpha classes. Lines indicate bottom of bands, red=C, blue=B, and green=A.

APPENDIX L: Regional & river class invertebrate-nutrient relationships

Previous work (e.g., Dolédec *et al.*, 2006; Clapcott *et al.*, 2012; Death *et al.*, 2015, 2018) has found relationships between key invertebrate-based indicators of ecological health, such as the macroinvertebrate community index (MCI; Stark and Maxted, 2007), and nitrogen and/or phosphorus. However, there is uncertainty whether these relationships are limited to particular river types or regions, or whether they are widespread.

Here we explore whether three invertebrate-based metrics of ecosystem health (MCI, quantitative MCI (QMCI) and average score per metric (ASPM; Collier, 2008)) correlated with dissolved inorganic nitrogen (DIN) and dissolved reactive phosphorus (DRP) in each region, river environment classification's (REC) climate class (Snelder, Biggs and Weatherhead, 2010) and the freshwater environments (FENZ) level one river classes.

Macroinvertebrate data used was sourced from annual state of environment monitoring between 2012-2016. Whilst regional councils differ markedly in their spatial coverage; representativeness; collection technique; invertebrate identification intensity, taxonomic resolution and quality; and use of sensitivity scores, the dataset provides the most widespread coverage of regular invertebrate surveys in NZ. To reduce the impact of differences in regionalized scoring and error, the scores for each sample derived Clapcott et al (2017) were used as they ensured the scores for MCI, QMCI and ASPM were calculated consistently.

Where monthly nutrient monitoring (sourced from lawa.org.nz) occurred at the same location as the annual macroinvertebrate surveys, then average scores for each of the three metrics were correlated with the median DIN and DRP concentrations over the same five-year period (n_{MCI}& n_{ASPM} =385; n_{QMCI}=290).

Given that most macroinvertebrate monitoring sites did not have a paired nutrient monitoring site available on LAWA, correlations were also examined using the entire macroinvertebrate dataset (n=1851) and modelled DIN and DRP concentrations (Larned, Snelder and Unwin, 2017).

When examining linear regressions, it's important keep at the forefront seven key assumptions:

(1) the relationship is indeed linear,

(2) residuals are normally distributed,

(3) No or little multicollinearity,

(4) no auto-correlation,

(5) there is homogeneity of variance across the gradient (homoscedasticity),

(6) the sample size is sufficiently large (ie 20+), and

(7) there is sufficient data across the spectrum to reduce extrapolation.

Many of the regressions DO NOT meet these assumptions and should be interpreted with caution. Many of the relationships display clear cases of heteroscedasticity, insufficient data or inadequate spread across the nutrient gradients and either do not reach significance or have fits that were informative across a reasonable portion of the gradient, as a result, many region and classes specific relationships are unsuitable for deriving region or class specific nutrient criteria. Though the relationships improve substantially when the entire dataset is used with modelled nutrients as the coverage of data across the gradients are more complete and models can smooth otherwise messy data. Among the relationships derived from the complete macroinvertebrate dataset, there appeared to be little substantial differences among relationships between regions across the country, between REC classes and between FENZ classes.

In this exploration, the variation observed is also not unforeseen. Variation arises for multiple reasons, including: other limiting or interacting factors not being included; substantial differences in sampling techniques; the inevitable loss of information when an entire assemblage is summarised in a single number (such as the MCI); differences in invertebrate re-establishment rates post-floods; differences in the seral stage sampled (early seral stages tend to have higher MCI scores); inconsistencies between labs and lab technicians in correctly identifying invertebrates; inconsistencies in the intensification of enumerating samples (some do presence absence right through to others that to do full counts); inherent indeterminism in ecological systems; differences arising from predator-prey cycles; gross variability in river nutrient concentration at the time sampled (can change substantially within hours yet only monthly grab samples are taken); variability in community stoichiometry as influenced by weather or management dependent nutrient fluctuations; and differences among labs in testing nutrients.

Tables L1-37 provide the sub-classification, R^2, p-value and number of sites for each relationship. Figures L1-37 provide the corresponding scatter plots with regression lines where significant.

Table L1. Correlation coefficients and p-values from regional regressions between MCI and measured DIN

Region	R2	p-value	sites
Northland	0.48	0.01	14
Auckland	0.33	0.03	14
Waikato	0.04	0.62	9
Taranaki	0.10	0.31	12
Hawkes Bay	0.00	0.98	50
Manawatu	0.26	0.00	60
Wellington	0.33	0.00	42
Tasman	0.35	0.00	30
Marlborough	0.23	0.02	24
Canterbury	0.18	0.00	42
West Coast	0.26	0.01	26
Otago	0.21	0.01	30
Southland	0.60	0.00	32

Table L2. Correlation coefficients and p-values from regional regressions between MCI and measured DRP

Region	R2	p-value	sites
Northland	0.04	0.47	14
Auckland	0.19	0.12	14
Waikato	0.01	0.78	9
Taranaki	0.13	0.26	12
Hawkes Bay	0.33	0.00	50
Manawatu	0.10	0.01	60
Wellington	0.43	0.00	42
Tasman	0.08	0.12	30
Marlborough	0.10	0.14	24
Canterbury	0.25	0.00	42
West Coast	0.31	0.00	26
Otago	0.11	0.07	30
Southland	0.32	0.00	32

Table L3. Correlation coefficients and p-values from regional regressions between QMCI and measured DIN

Region	R2	p-value	sites
Northland	0.16	0.15	14
Waikato	0.34	0.10	9
Taranaki	1.00	NA	2
Hawkes Bay	0.03	0.19	50
Manawatu	0.21	0.00	60
Wellington	0.46	0.00	42
Tasman	0.63	0.41	3
Marlborough	1.00	NA	2
Canterbury	0.26	0.00	42
West Coast	0.69	0.08	5
Otago	0.07	0.17	29
Southland	0.53	0.00	32

Table L4. Correlation coefficients and p-values from regional regressions between QMCI and measured DRP

Region	R2	p-value	sites
Northland	0.00	0.86	14
Waikato	0.33	0.10	9
Taranaki	1.00	NA	2
Hawkes Bay	0.25	0.00	50
Manawatu	0.05	0.08	60
Wellington	0.46	0.00	42
Tasman	0.99	0.07	3
Marlborough	1.00	NA	2
Canterbury	0.26	0.00	42
West Coast	0.69	0.08	5
Otago	0.02	0.42	29
Southland	0.31	0.00	32

Table L5. Correlation coefficients and p-values from regional regressions between ASPM and measured DIN

Region	R2	p-value	Sites
Northland	0.30	0.04	14
Auckland	0.36	0.02	14
Waikato	0.07	0.49	9
Taranaki	0.05	0.49	12
Hawkes Bay	0.02	0.33	50
Manawatu	0.29	0.00	60
Wellington	0.37	0.00	42
Tasman	0.27	0.00	30
Marlborough	0.27	0.01	24
Canterbury	0.26	0.00	42
West Coast	0.06	0.24	26
Otago	0.17	0.02	30
Southland	0.50	0.00	32

Table L6. Correlation coefficients and p-values from regional regressions between ASPM and measured DRP

Region	R2	p-value	sites
Northland	0.03	0.53	14
Auckland	0.15	0.16	14
Waikato	0.01	0.79	9
Taranaki	0.12	0.26	12
Hawkes Bay	0.33	0.00	50
Manawatu	0.06	0.06	60
Wellington	0.46	0.00	42
Tasman	0.14	0.04	30
Marlborough	0.21	0.02	24
Canterbury	0.31	0.00	42
West Coast	0.69	0.00	26
Otago	0.15	0.04	30
Southland	0.23	0.01	32

Table L8. Correlation coefficients and p-values from FENZ class regressions between MCI and measured DIN			
FENZ class	R2	p-value	Sites
A	0.01	0.57	50
C	0.09	0.00	289
G	0.29	0.00	38

Table L9. Correlation coefficients and p-values from FENZ class regressions between MCI and measured DRP			
FENZ class	R2	p-value	Sites
A	0.17	0.00	50
C	0.07	0.00	289
G	0.19	0.01	38

Table L10. Correlation coefficients and p-values from FENZ class regressions between QMCI and measured DIN			
FENZ class	R2	p-value	sites
A	0	0.978224	40
C	0.06	0.000282	203
G	0.12	0.032507	37

Table L11. Correlation coefficients and p-values from FENZ class regressions between QMCI and measured DRP			
FENZ class	R2	p-value	sites
A	0.01	0.48	40
C	0.06	0.00	204
G	0.22	0.00	37

Table L12. Correlation coefficients and p-values from FENZ class regressions between ASPM and measured DIN

FENZ class	R2	p-value	Sites
A	0.003902	0.666522	50
C	0.056861	4.21E-05	289
G	0.169976	0.010108	38

Table L13. Correlation coefficients and p-values from FENZ class regressions between ASPM and measured DRP

	R2	p-value	Sites
A	0.161014	0.003876	50
C	0.074094	2.65E-06	289
G	0.150875	0.015956	38

Table L14. Correlation coefficients and p-values from REC climate regressions between MCI and measured DIN

REC climate class	R2	p-value	Sites
WD	0.01	0.65	22
WW	0.09	0.04	47
CW	0.09	0.00	191
CD	0.20	0.00	80

Table L15. Correlation coefficients and p-values from REC climate regressions between MCI and measured DRP

REC climate class	R2	p-value	Sites
WD	0.17	0.05	22
WW	0.01	0.57	47
CW	0.02	0.05	191
CD	0.19	0.00	80

Table L16. Correlation coefficients and p-values from REC climate regressions between QMCI and measured DIN

REC climate class	R2	p-value	Sites
WD	0.03	0.53	18
WW	0.05	0.19	34
CW	0.11	0.00	146
CD	0.12	0.00	74

Table L17. Correlation coefficients and p-values from REC climate regressions between QMCI and measured DRP

REC climate class	R2	p-value	Sites
WD	0.16	0.10	18
WW	0.05	0.19	34
CW	0.05	0.01	146
CD	0.17	0.00	74

Table L18. Correlation coefficients and p-values from REC climate regressions between ASPM and measured DIN

REC climate class	R2	p-value	Sites
WD	0.01	0.73	22
WW	0.10	0.03	47
CW	0.08	0.00	191
CD	0.18	0.00	80

Table L19. Correlation coefficients and p-values from REC climate regressions between ASPM and measured DRP

REC climate class	R2	p-value	Sites
WD	0.10	0.16	22
WW	0.00	0.70	47
CW	0.05	0.00	191
CD	0.20	0.00	80

Table L20. Correlation coefficients and p-values from regional regressions between MCI and modelled DIN			
Region	R2	p-value	sites
Northland	0.333	0.000	36
Auckland	0.569	0.000	107
Waikato	0.403	0.000	621
BOP	0.064	0.002	142
Gisborne	0.028	0.197	60
Taranaki	0.041	0.130	57
Hawkes Bay	0.144	0.000	92
Manawatu	0.248	0.000	139
Wellington	0.357	0.000	62
Tasman	0.496	0.000	59
Marlborough	0.218	0.006	33
Canterbury	0.292	0.000	203
West Coast	0.248	0.001	44
Otago	0.281	0.000	63
Southland	0.375	0.000	133

Table L21. Correlation coefficients and p-values from regional regressions between MCI and modelled DRP			
Region	R2	p-value	sites
Northland	0.180	0.010	36
Auckland	0.385	0.000	107
Waikato	0.196	0.000	621
BOP	0.005	0.588	57
Gisborne	0.007	0.322	142
Taranaki	0.143	0.003	60
Hawkes Bay	0.260	0.000	92
Manawatu	0.096	0.000	139
Wellington	0.469	0.000	62
Tasman	0.373	0.000	59
Marlborough	0.266	0.002	33
Canterbury	0.136	0.000	203
West Coast	0.183	0.004	44
Otago	0.149	0.002	63
Southland	0.374	0.000	133

Table L22. Correlation coefficients and p-values from regional regressions between QMCI and modelled DIN			
Region	R2	p-value	sites
Northland	0.230	0.004	35
Auckland	1.000	NA	2
Waikato	0.354	0.000	617
BOP	0.129	0.000	142
Gisborne	0.007	0.526	60
Taranaki	0.668	0.391	3
Hawkes Bay	0.052	0.031	90
Manawatu	0.213	0.000	139
Wellington	0.324	0.000	62
Tasman	1.000	NA	2
Marlborough	1.000	NA	2
Canterbury	0.399	0.000	203
West Coast	0.641	0.104	5
Otago	0.201	0.001	52
Southland	0.230	0.000	133

Table L23. Correlation coefficients and p-values from regional regressions between QMCI and modelled DRP			
Region	R2	p-value	sites
Northland	0.22	0.00	35
Auckland	1.00	NA	2
Waikato	0.19	0.00	617
BOP	0.07	0.83	3
Gisborne	0.03	0.04	142
Taranaki	0.10	0.01	60
Hawkes Bay	0.26	0.00	90
Manawatu	0.11	0.00	139
Wellington	0.51	0.00	62
Tasman	1.00	NA	2
Marlborough	1.00	NA	2
Canterbury	0.33	0.00	203
West Coast	0.61	0.12	5
Otago	0.24	0.00	52
Southland	0.33	0.00	133

Table L24. Correlation coefficients and p-values from regional regressions between ASPM and modelled DIN

Region	R2	p-value	sites
Northland	0.251	0.002	36
Auckland	0.524	0.000	107
Waikato	0.412	0.000	621
BOP	0.134	0.000	142
Gisborne	0.011	0.436	60
Taranaki	0.028	0.213	57
Hawkes Bay	0.067	0.013	92
Manawatu	0.260	0.000	139
Wellington	0.374	0.000	62
Tasman	0.341	0.000	59
Marlborough	0.277	0.002	33
Canterbury	0.314	0.000	203
West Coast	0.120	0.021	44
Otago	0.221	0.000	63
Southland	0.285	0.000	133

Table L25. Correlation coefficients and p-values from regional regressions between ASPM and modelled DRP

Region	R2	p-value	sites
Northland	0.127	0.033	36
Auckland	0.404	0.000	107
Waikato	0.190	0.000	621
BOP	0.001	0.820	57
Gisborne	0.050	0.008	142
Taranaki	0.141	0.003	60
Hawkes Bay	0.240	0.000	92
Manawatu	0.107	0.000	139
Wellington	0.491	0.000	62
Tasman	0.449	0.000	59
Marlborough	0.361	0.000	33
Canterbury	0.213	0.000	203
West Coast	0.138	0.013	44
Otago	0.143	0.002	63
Southland	0.263	0.000	133

Table L26. Correlation coefficients and p-values from FENZ class regressions between MCI and modelled DIN			
FENZ class	R2	p-value	Sites
A	0.03	0.00	379
C	0.16	0.00	1315
G	0.31	0.00	108

Table L27. Correlation coefficients and p-values from FENZ class regressions between MCI and modelled DRP			
FENZ class	R2	p-value	Sites
A	0.09	0.00	379
C	0.04	0.00	1315
G	0.33	0.00	108

Table L28. Correlation coefficients and p-values from FENZ class regressions between QMCI and modelled DIN			
FENZ class	R2	p-value	Sites
A	0.00	0.23	301
C	0.14	0.00	1099
G	0.10	0.00	104

Table L29. Correlation coefficients and p-values from FENZ class regressions between QMCI and modelled DRP			
FENZ class	R2	p-value	sites
A	0.11	0.00	301
C	0.13	0.00	1099
G	0.29	0.00	104

Table L30. Correlation coefficients and p-values from FENZ class regressions between ASPM and modelled DIN			
FENZ class	R2	p-value	sites
A	0.01	0.02	379
C	0.18	0.00	1315
G	0.20	0.00	108

Table L31. Correlation coefficients and p-values from FENZ class regressions between ASPM and modelled DRP			
FENZ class	R2	p-value	sites
A	0.11	0.00	379
C	0.07	0.00	1315
G	0.23	0.00	108

Table L32. Correlation coefficients and p-values from REC climate class regressions between MCI and modelled DIN			
REC climate class	R2	p-value	sites
WD	0.26	0.00	124
WW	0.18	0.00	245
CW	0.27	0.00	861
CD	0.22	0.00	403

Table L33. Correlation coefficients and p-values from REC climate class regressions between MCI and modelled DRP			
REC climate class	R2	p-value	sites
WD	0.17	0.00	124
WW	0.12	0.00	245
CW	0.14	0.00	861
CD	0.15	0.00	403

Table L34. Correlation coefficients and p-values from REC climate class regressions between QMCI and modelled DIN

REC climate class	R2	p-value	sites
WD	0.28	0.00	106
WW	0.18	0.00	224
CW	0.24	0.00	738
CD	0.22	0.00	301

Table L35. Correlation coefficients and p-values from REC climate class regressions between QMCI and modelled DRP

REC climate class	R2	p-value	sites
WD	0.22	0.00	106
WW	0.24	0.00	224
CW	0.26	0.00	738
CD	0.21	0.00	301

Table L36. Correlation coefficients and p-values from REC climate class regressions between ASPM and modelled DIN

REC climate class	R2	p-value	sites
WD	0.25	0.00	124
WW	0.22	0.00	245
CW	0.28	0.00	861
CD	0.20	0.00	403

Table L37. Correlation coefficients and p-values from REC climate class regressions between ASPM and modelled DRP

REC climate class	R2	p-value	sites
WD	0.18	0.00	124
WW	0.20	0.00	245
CW	0.17	0.00	861
CD	0.16	0.00	403

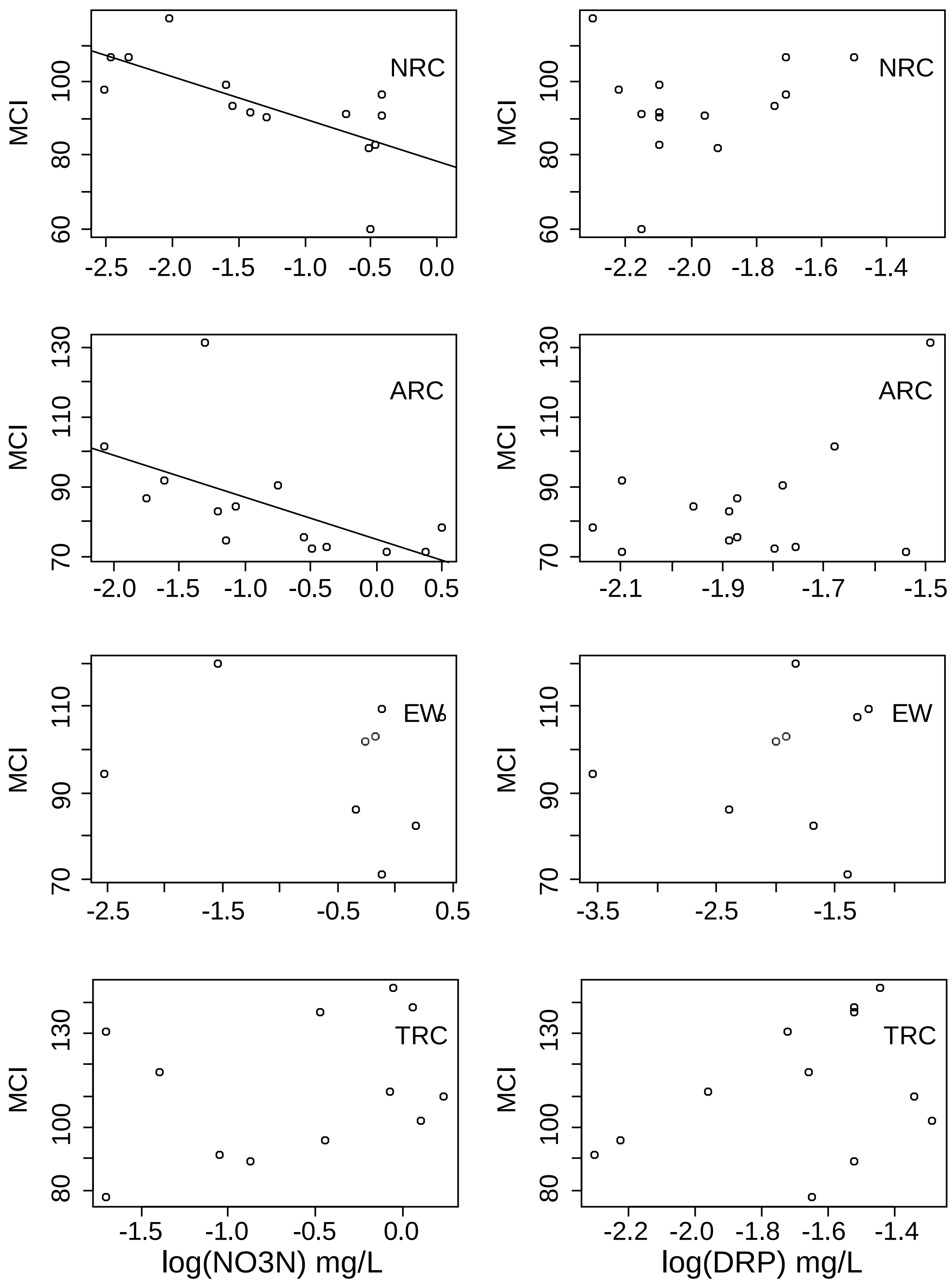


Fig. L1. Regional 5-year average MCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where MCI was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. NRC=Northland, ARC=Auckland, EW=Waikato and TRC=Taranaki.

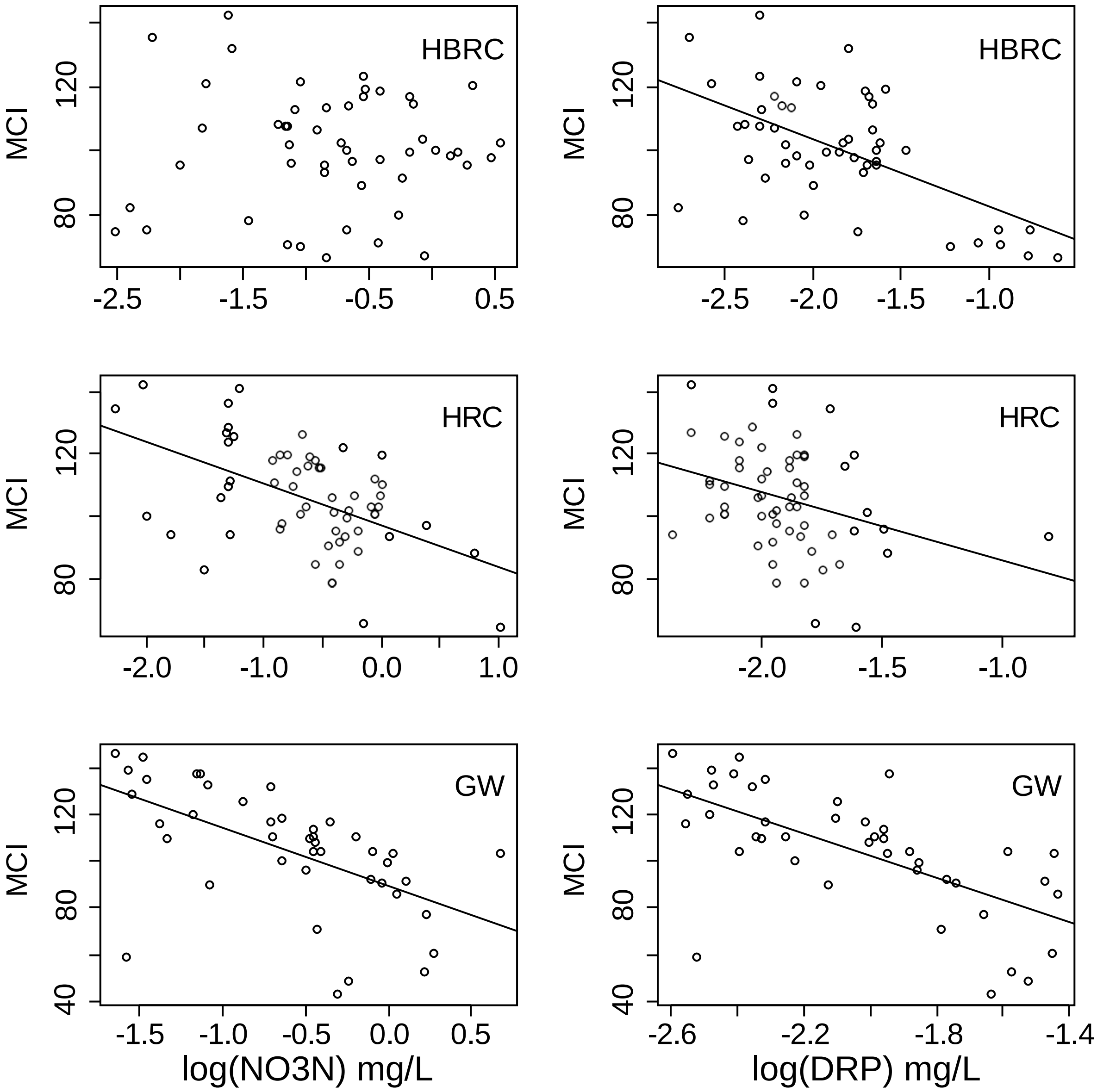


Fig. L2. Regional 5-year average MCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where MCI was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. HBRC=Hawkes Bay, HRC=Manawatu-Whanganui and GW=Wellington.

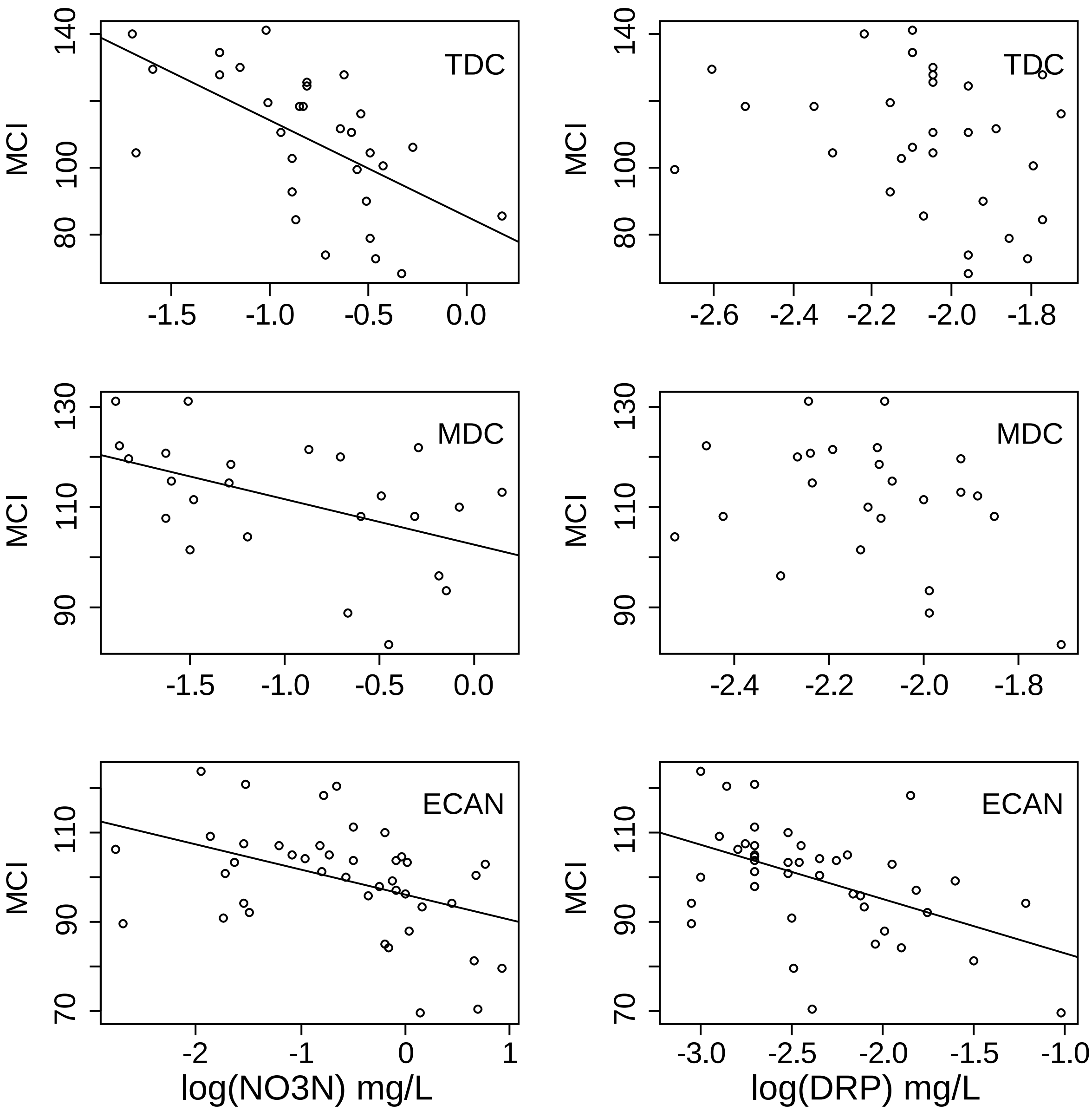


Fig. L3. Regional 5-year average MCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where MCI was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. TDC=Tasman, MDC=Marlborough, and ECAN=Canterbury.

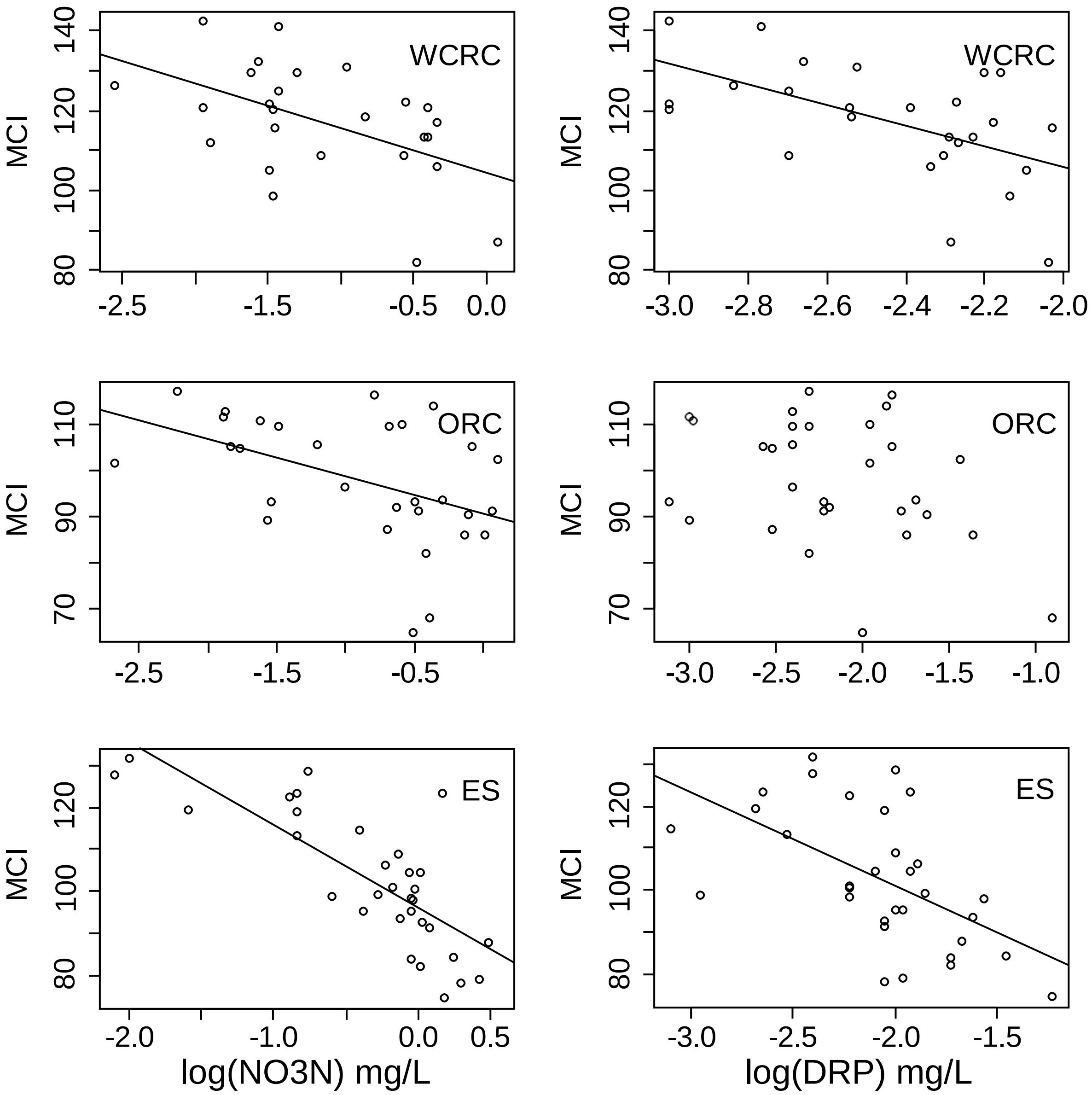


Fig. L4. Regional 5-year average MCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where MCI was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. WCRC=West Coast, ORC=Otago, and ES=Southland.

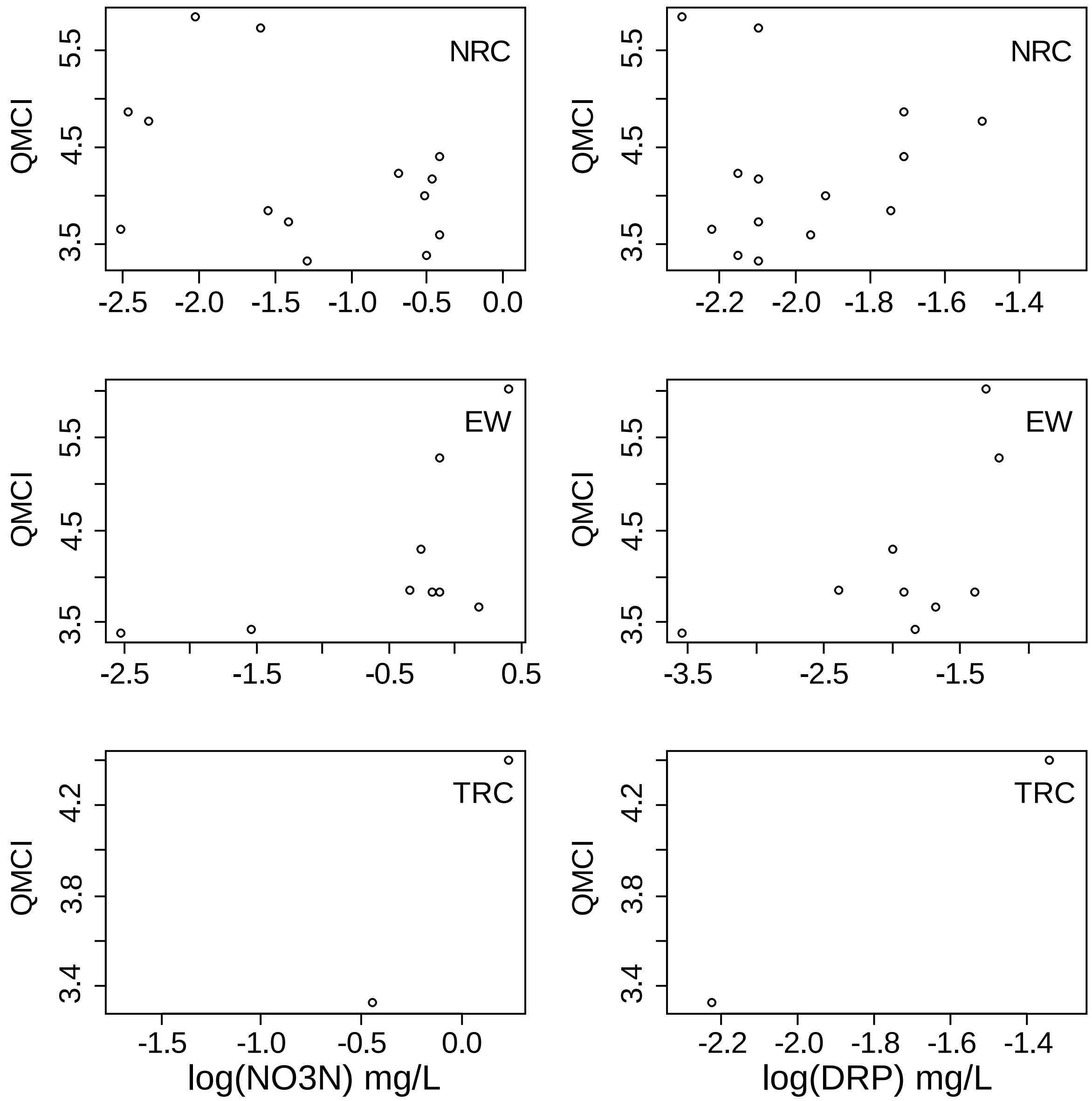


Fig. L5. Regional 5-year average QMCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where QMCI was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. NRC=Northland, EW=Waikato and TRC=Taranaki.

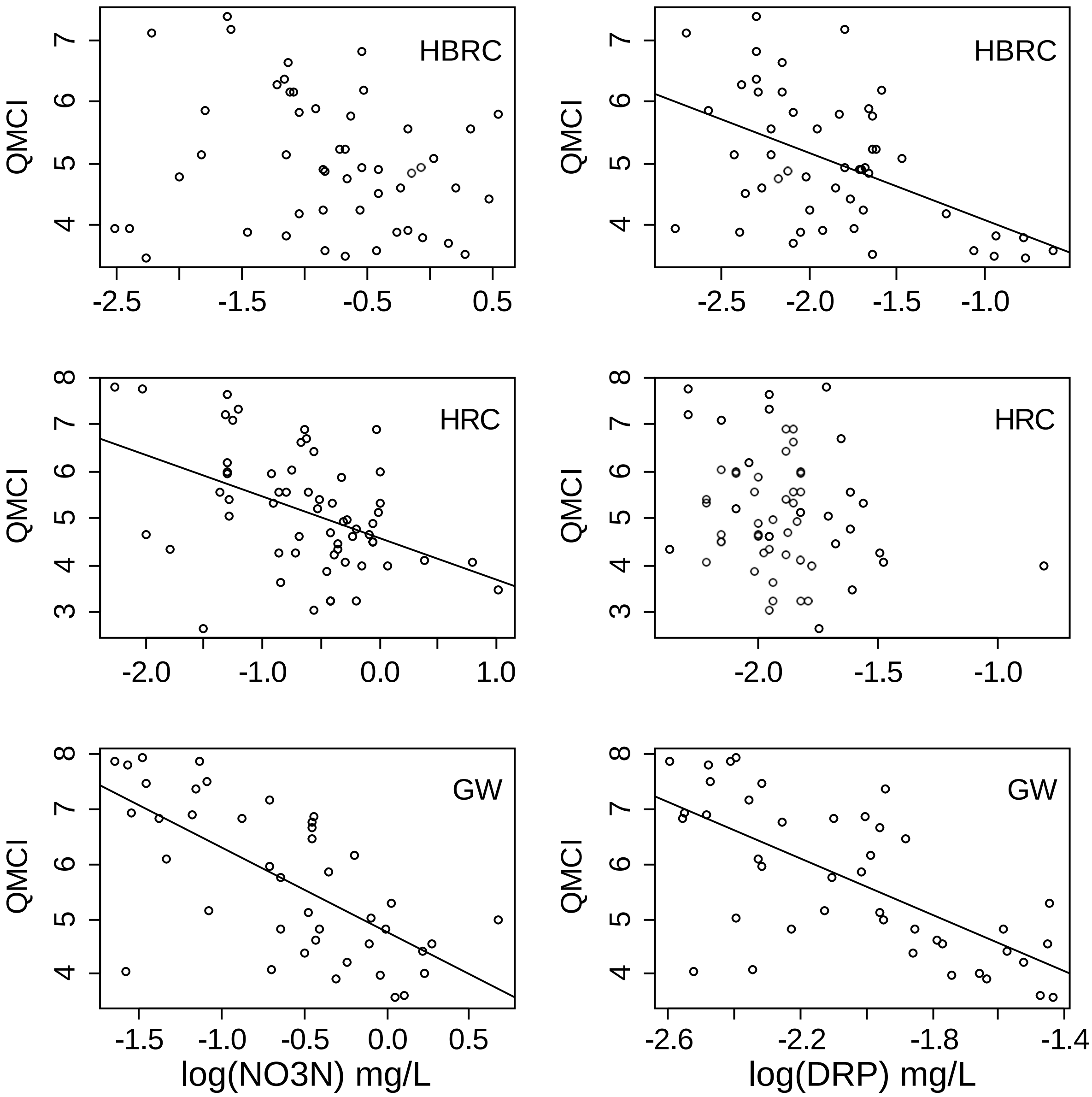


Fig. L6. Regional 5-year average QMCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where QMCI was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. HBRC=Hawkes Bay, HRC=Manawatu-Whanganui and GW=Wellington.

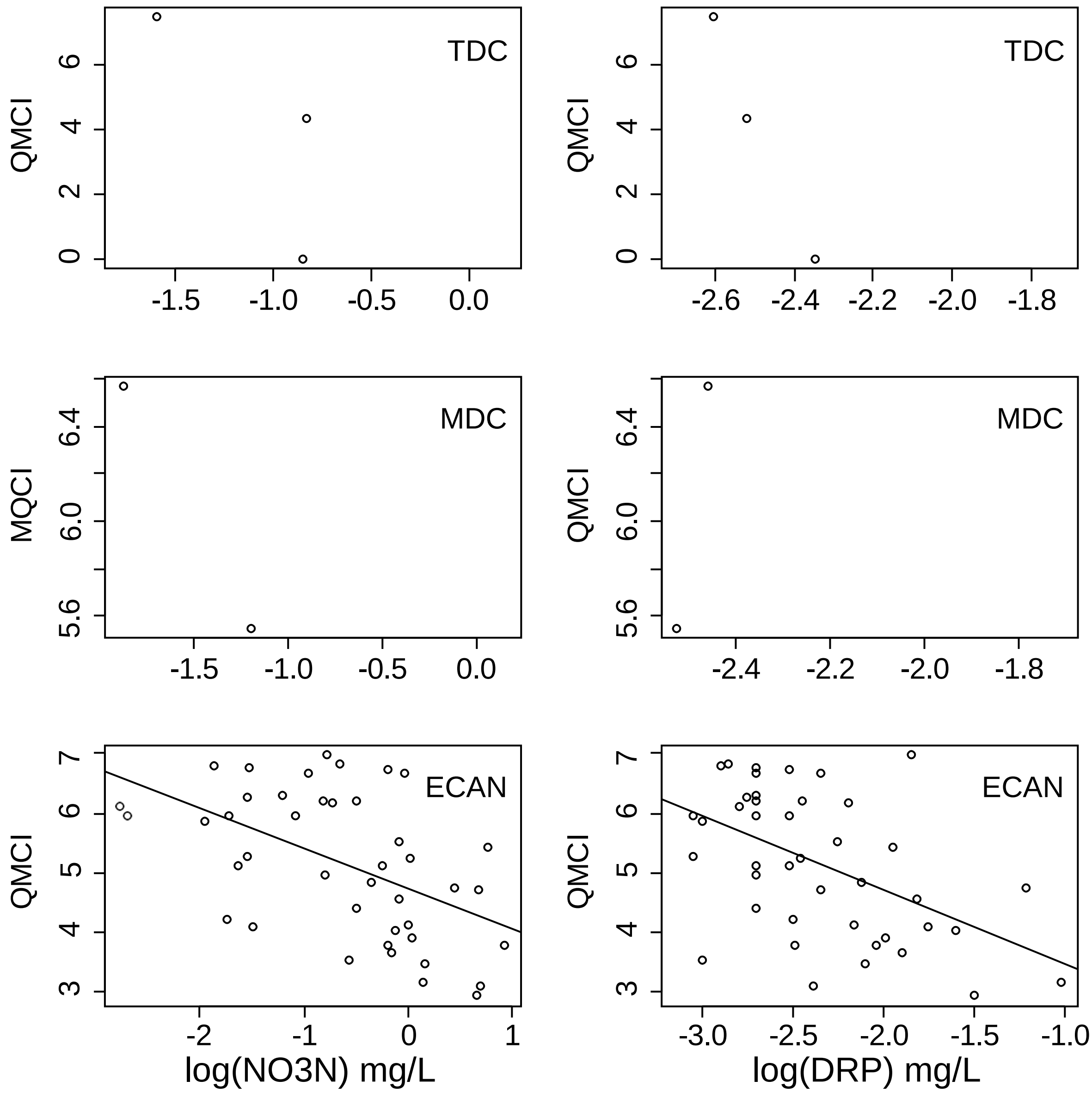


Fig. L7. Regional 5-year average QMCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where QMCI was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. TDC=Tasman, MDC=Marlborough, ECAN=Canterbury

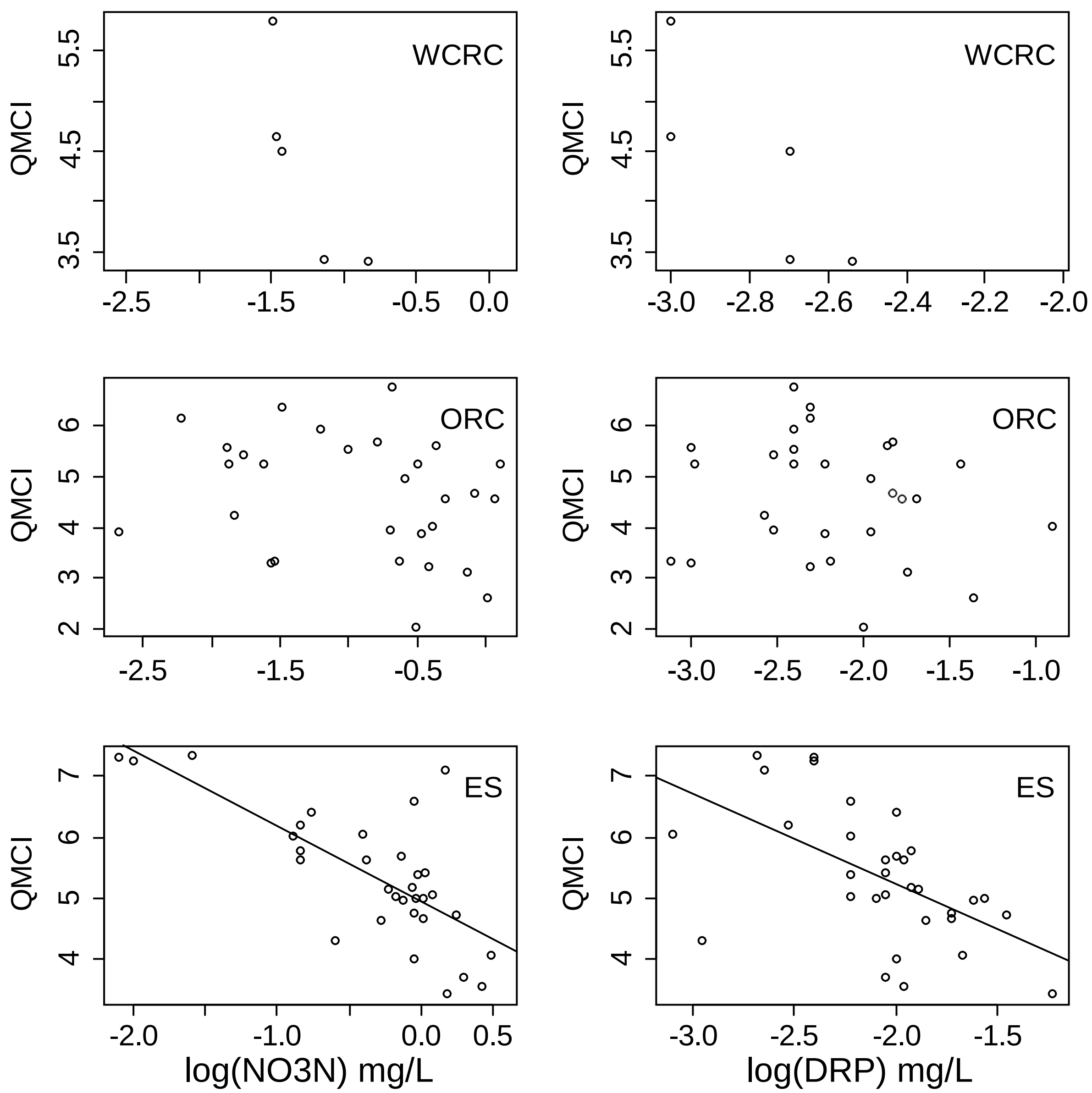


Fig. L8. Regional 5-year average QMCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where QMCI was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. WCRC=West Coast, ORC=Otago and ES=Southland.

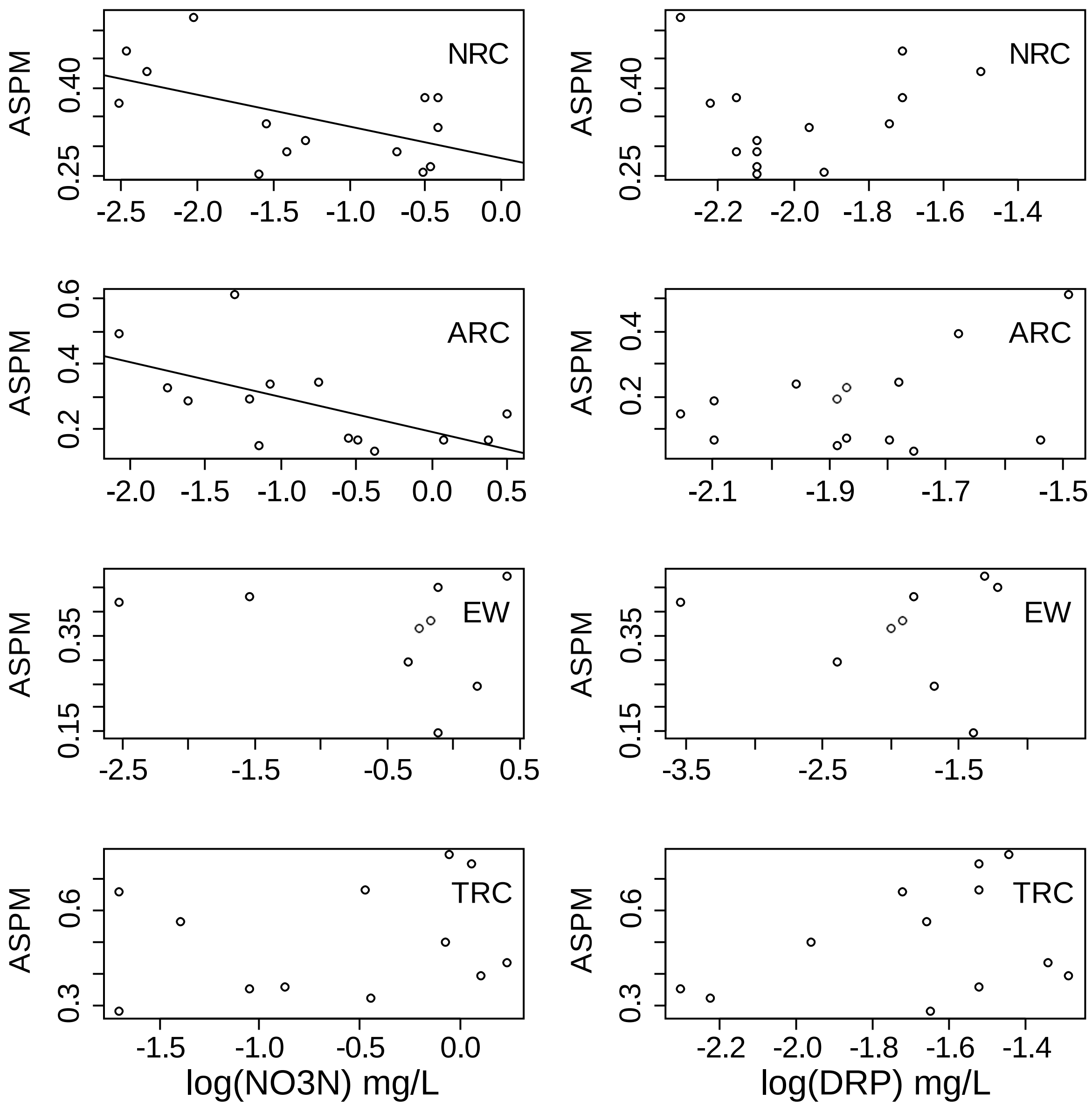


Fig. L9. Regional 5-year average ASPM versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where ASPM was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. NRC=Northland, ARC=Auckland, EW=Waikato and TRC=Taranaki.

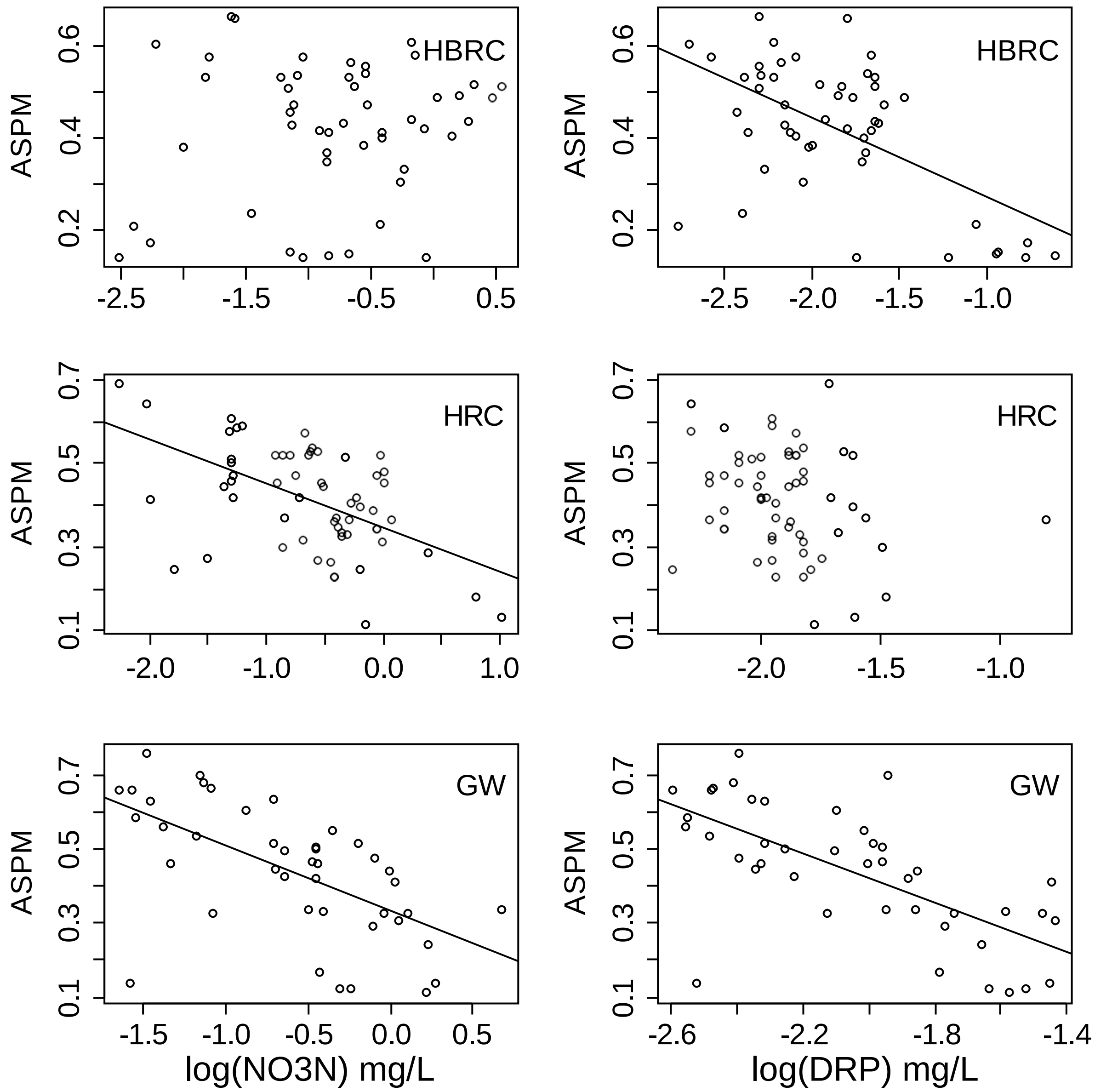


Fig. L10. Regional 5-year average ASPM versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where ASPM was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. HBRC=Hawkes Bay, HRC=Manawatu-Whanganui and GW=Wellington.

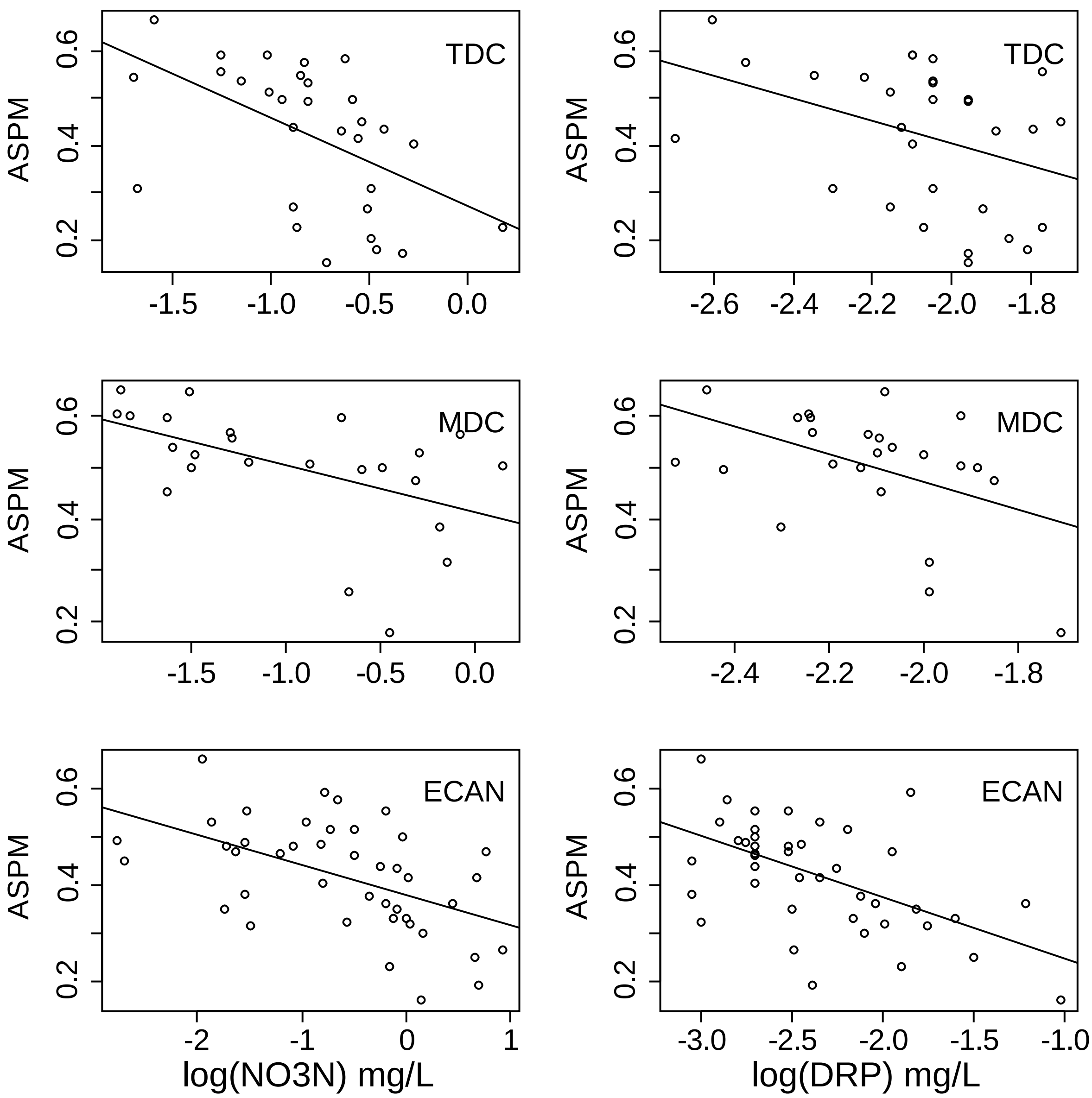


Fig. L11. Regional 5-year average ASPM versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where ASPM was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. TDC=Tasman, MDC=Marlborough and ECAN=Canterbury.

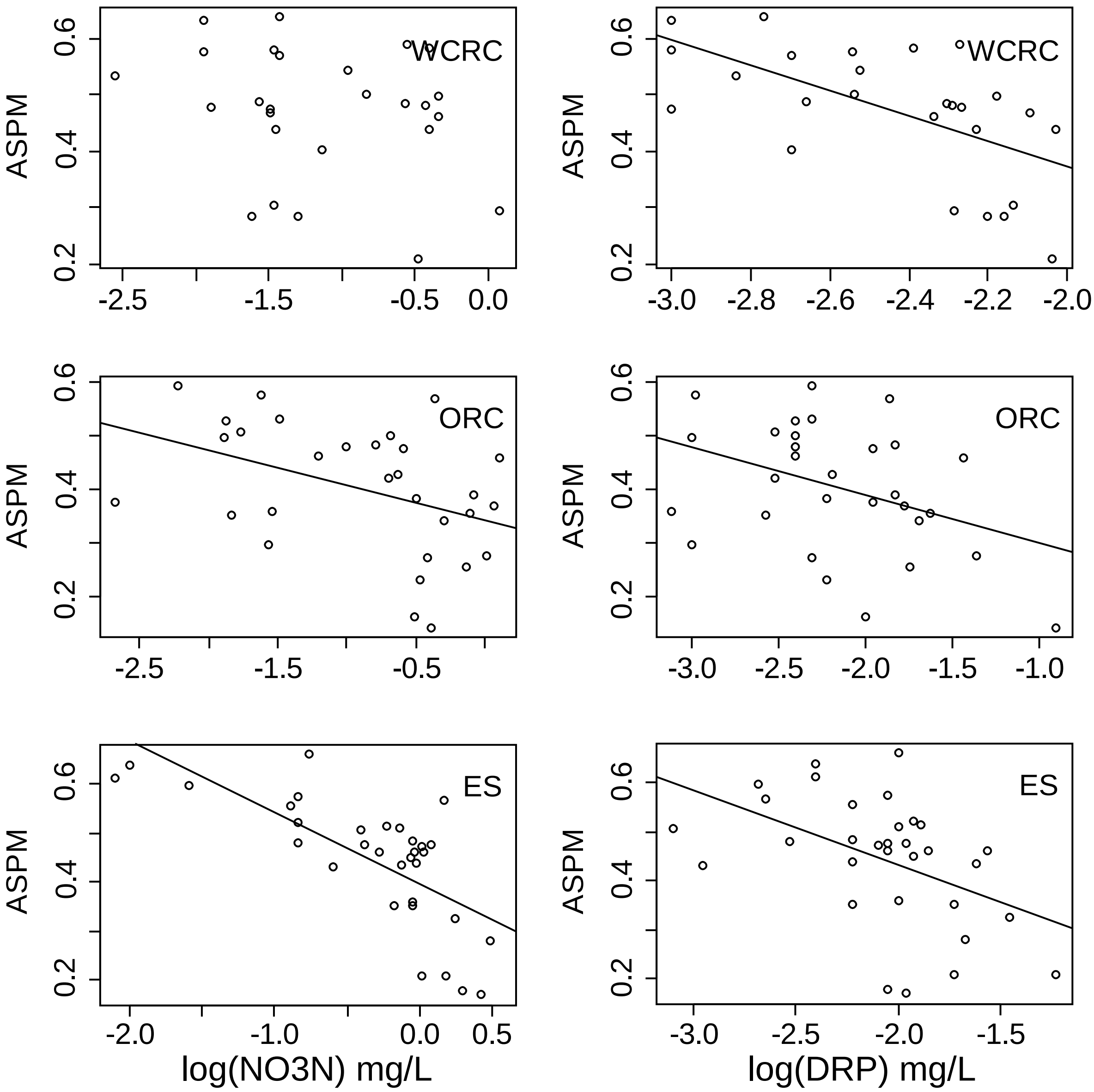


Fig. L12. Regional 5-year average ASPM versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where ASPM was measured at the same location as nutrients monitored monnthly (only those logged on LAWA) for 2012-2016. WCRC=West Coast, ORC=Otago and ES=Southland.

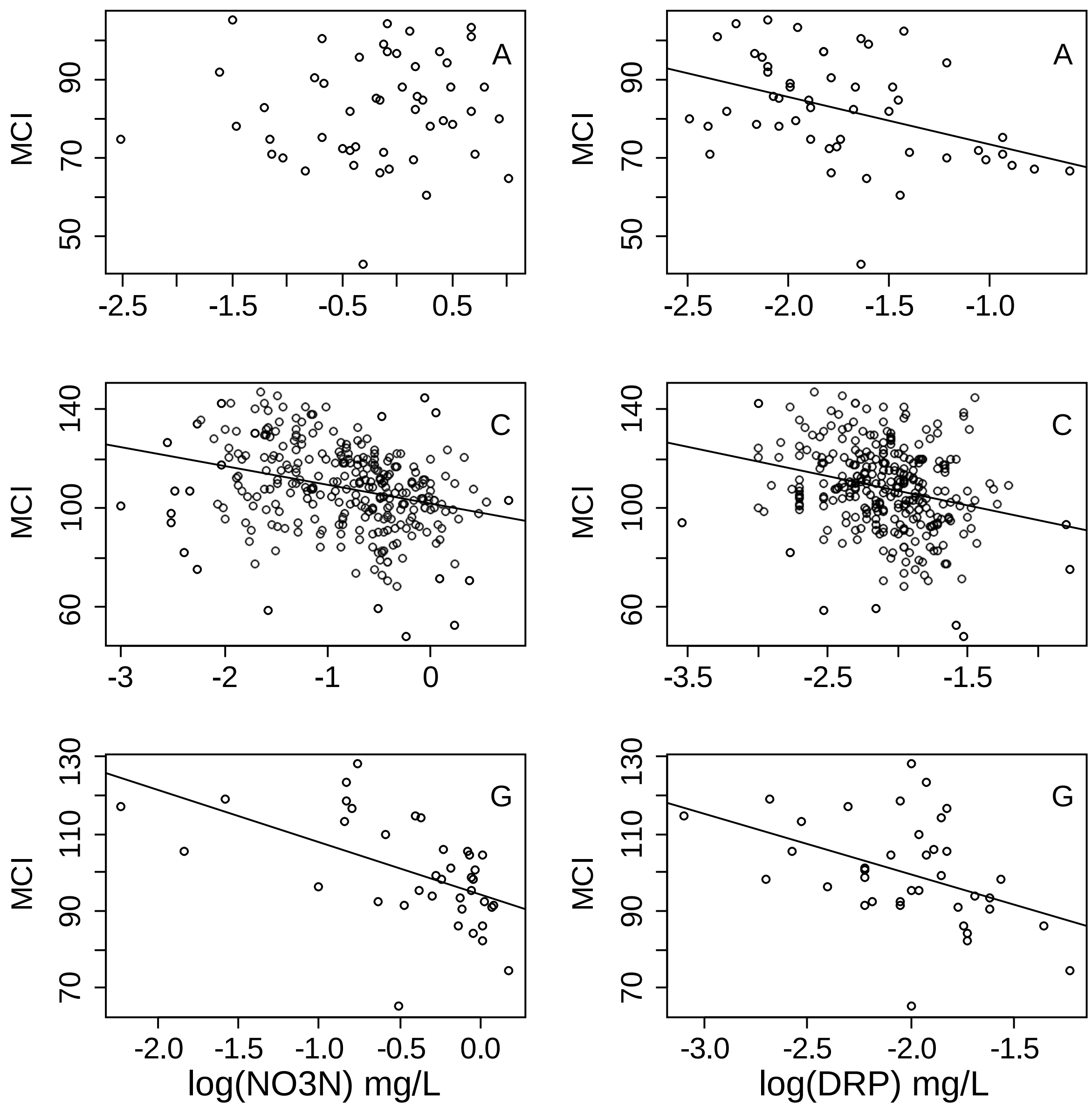


Fig. L13. Five year average MCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where MCI was measured at the same location as nutrients monitored monthly (only those logged on LAWA) for 2012-2016. Plots A, C and G are the three most data rich FENZ LevelOneAlpha classes.

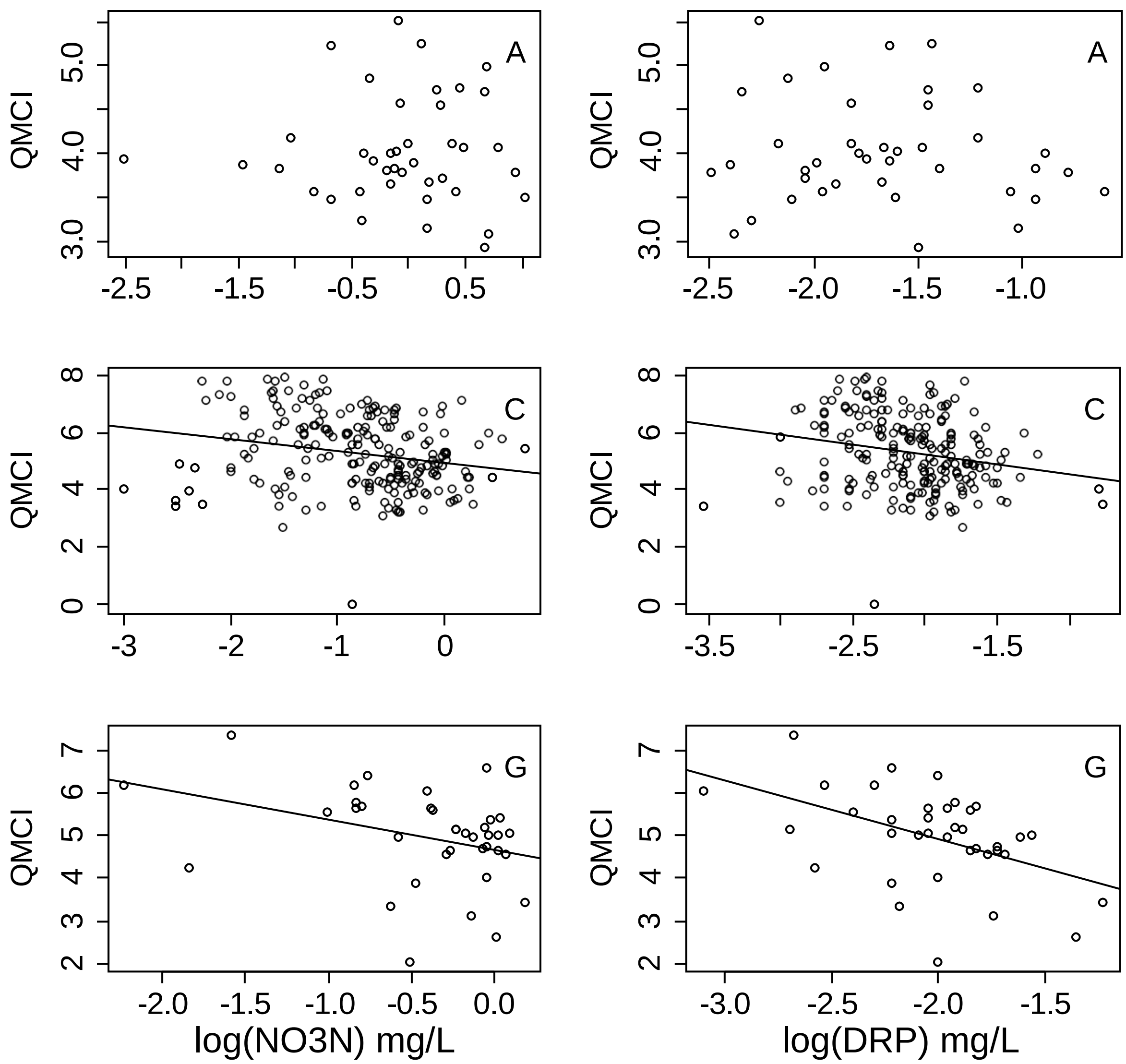


Fig. L14. Five year average QMCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where QMCI was measured at the same location as nutrients monitored monthly (only those logged on LAWA) for 2012-2016. Plots A, C and G are the three most data rich FENZ LevelOneAlpha classes.

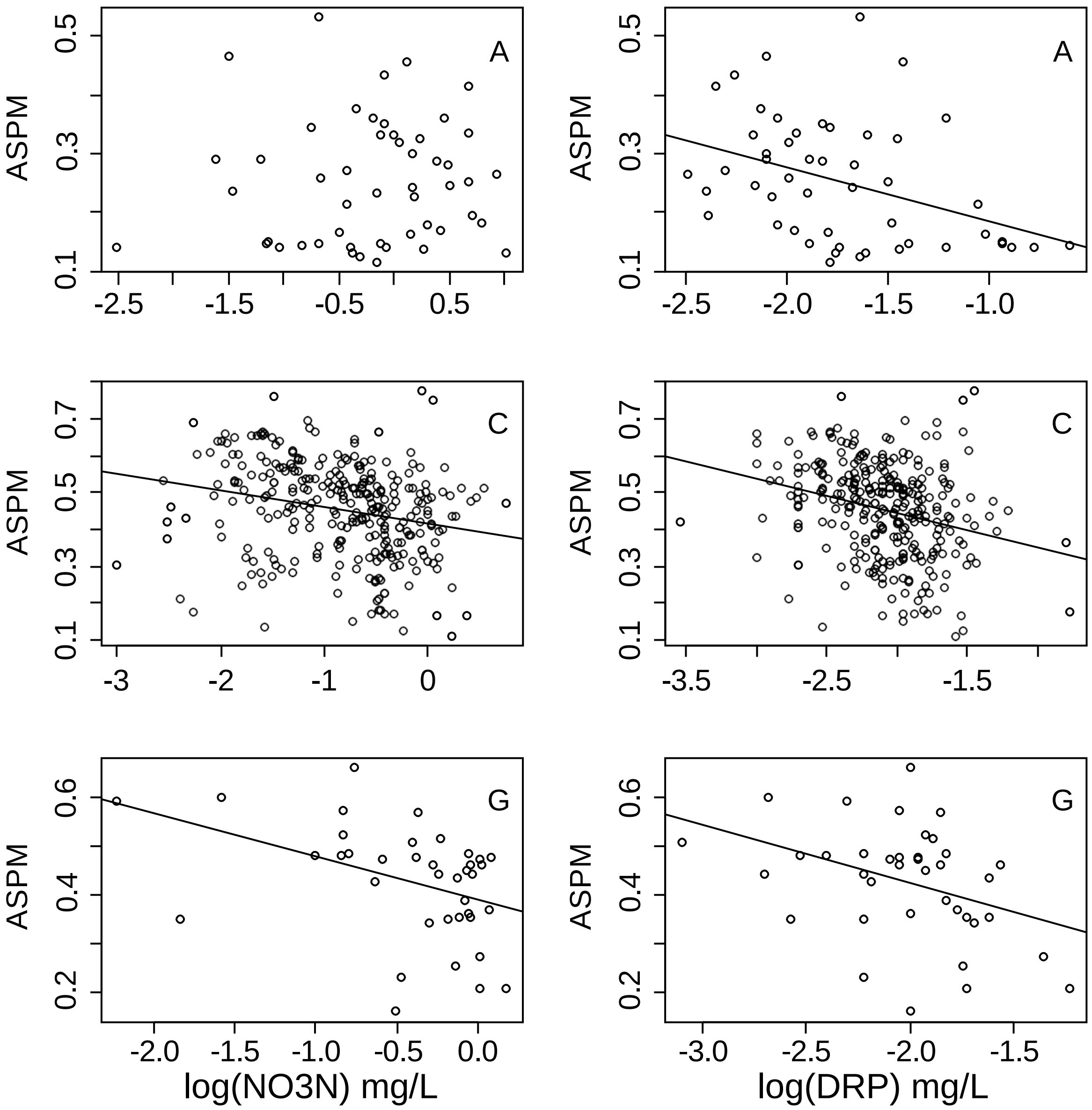


Fig. L16. Five year average ASPM versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where ASPM was measured at the same location as nutrients monitored monthly (only those logged on LAWA) for 2012-2016. Plots A, C and G are the three most data rich FENZ LevelOneAlpha classes.

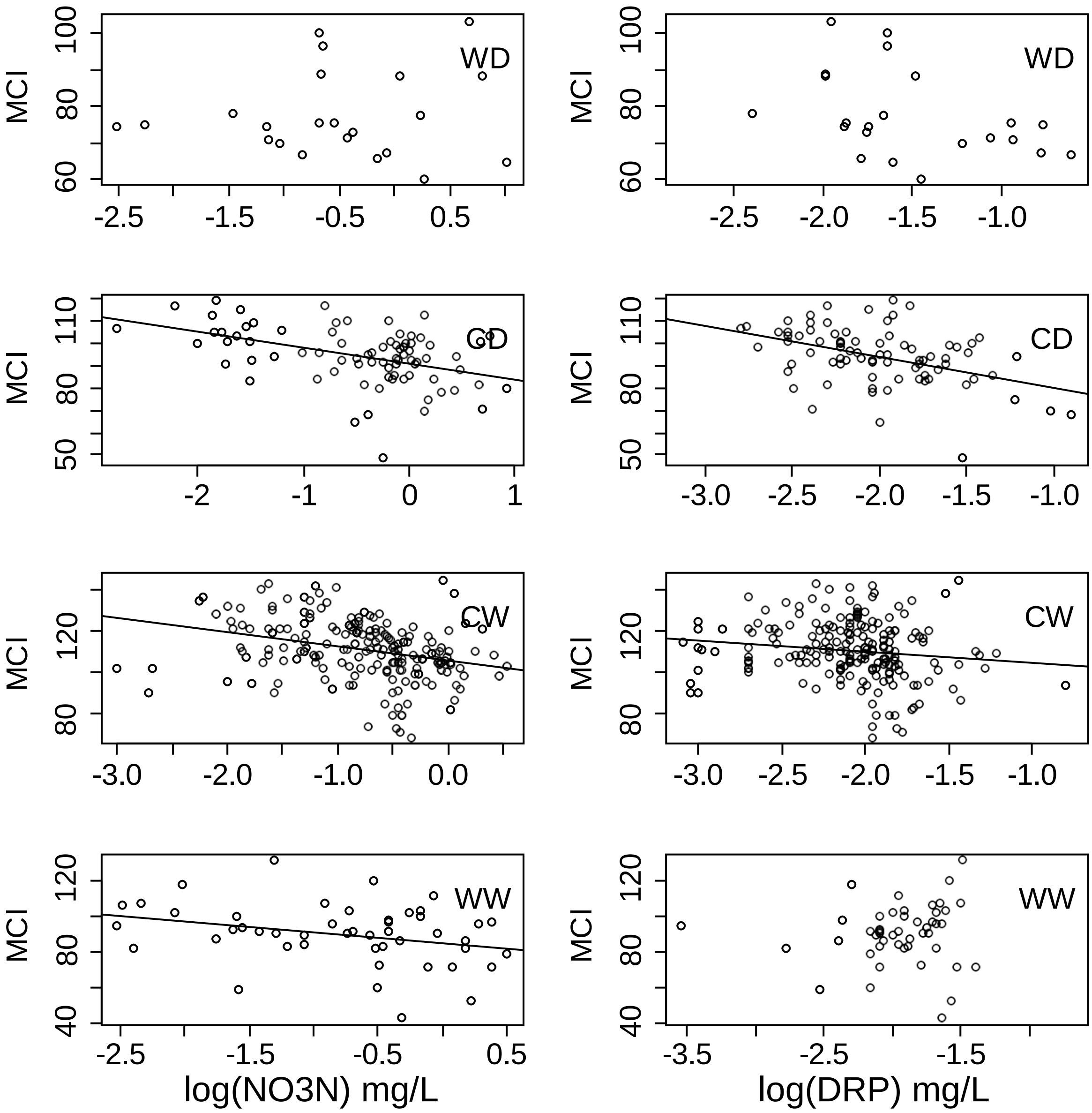


Fig. L17. Five year average MCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where MCI was measured at the same location as nutrients monitored monthly (only those logged on LAWA) for 2012-2016, based on the REC climate class. WD=warm dry, CD=cool dry, CW=cool wet, and WW=warm wet.

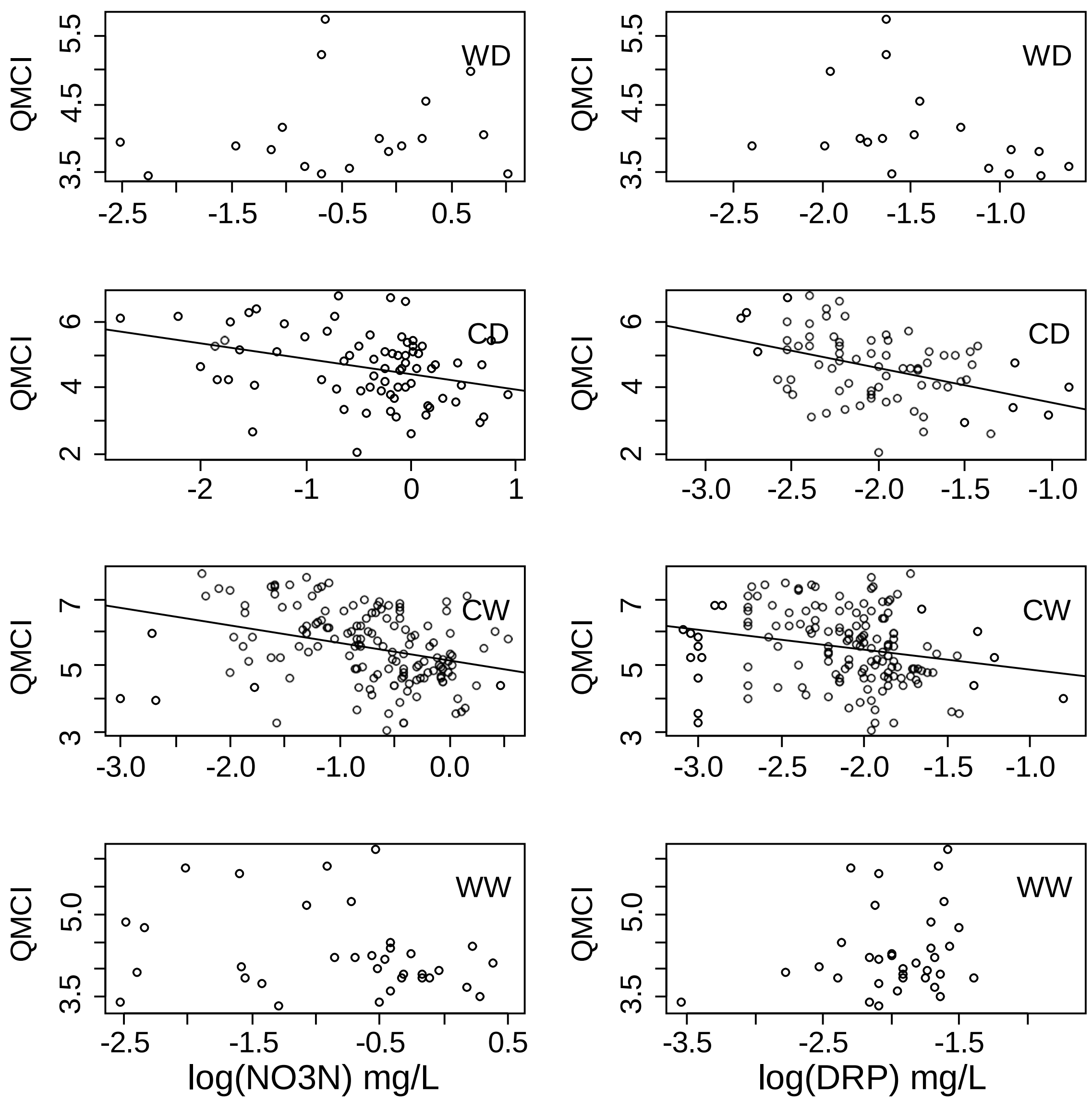


Fig. L18. Five year average QMCI versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where QMCI was measured at the same location as nutrients monitored monthly (only those logged on LAWA) for 2012-2016, based on the REC climate class. WD=warm dry, CD=cool dry, CW=cool wet, and WW=warm wet.

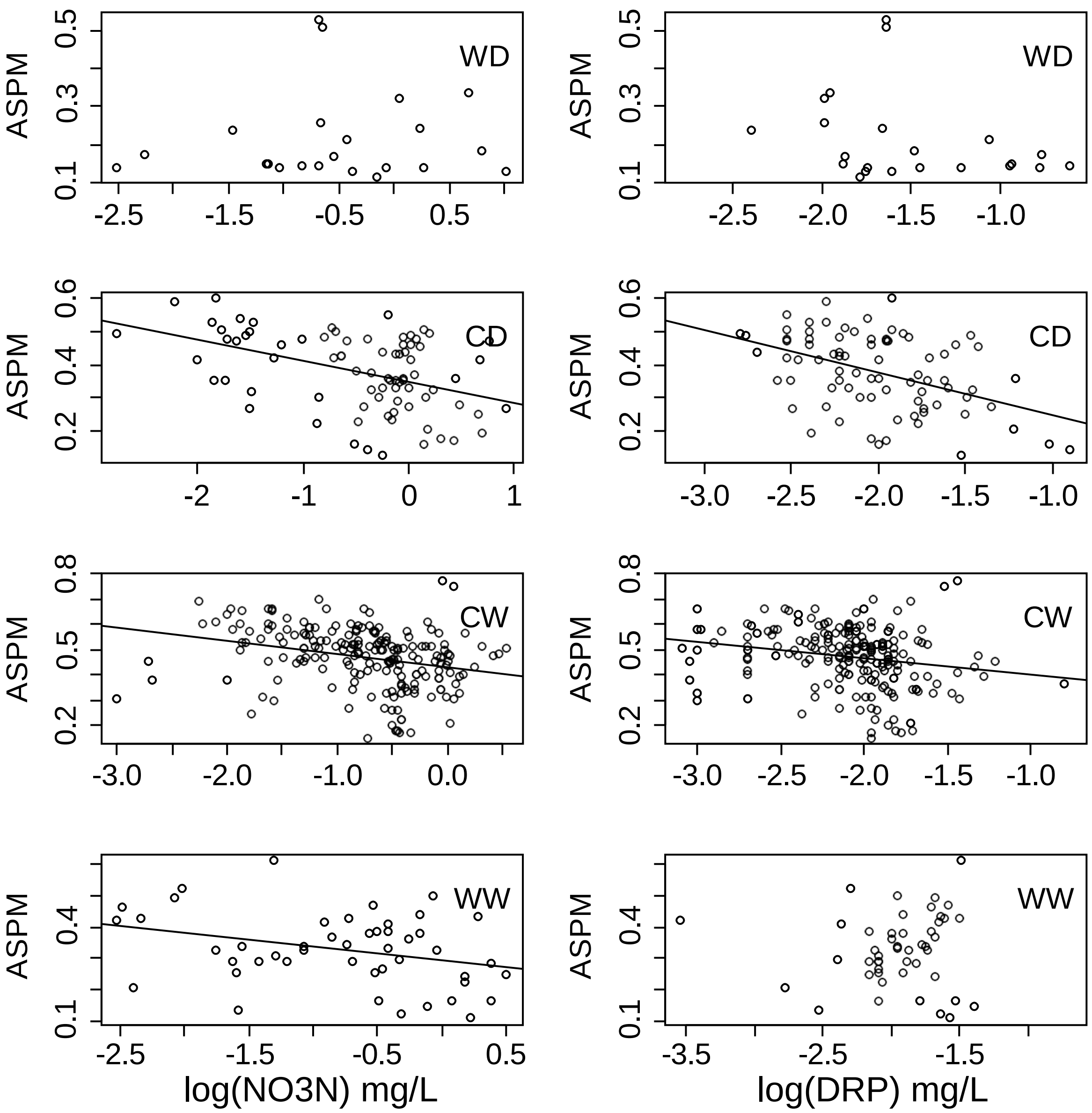


Fig. L19. Five year average ASPM versus median log(NO3-N) (left panel) or median log(DRP) (right panel) at sites where ASPM was measured at the same location as nutrients monitored monthly (only those logged on LAWA) for 2012-2016, based on the REC climate class. WD=warm dry, CD=cool dry, CW=cool wet, and WW=warm wet.

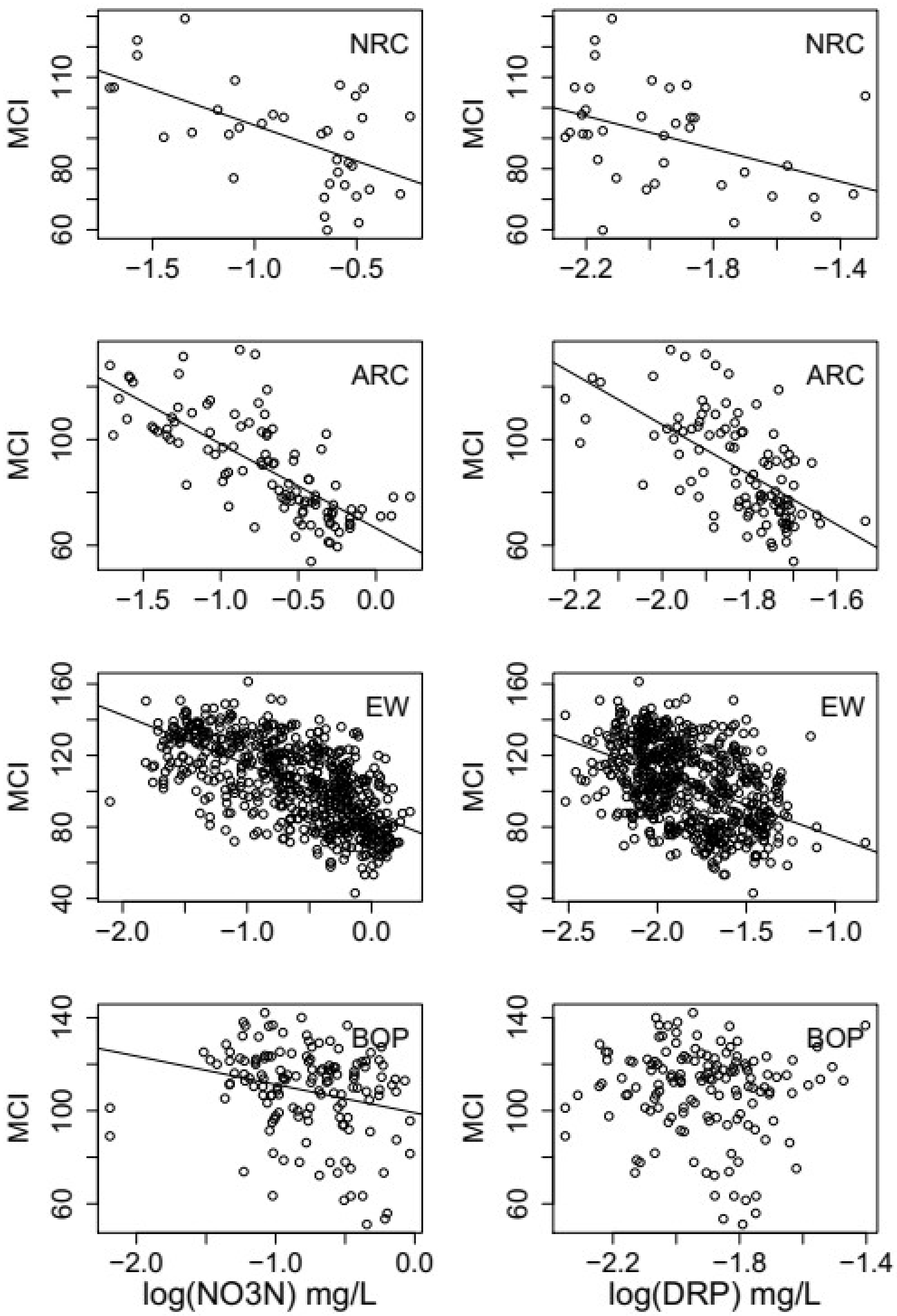


Fig. L20. Regional 5-year (2012-2016) average MCI versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). NRC=Northland, ARC=Auckland, EW=Waikato and BOP=Bay of Plenty.

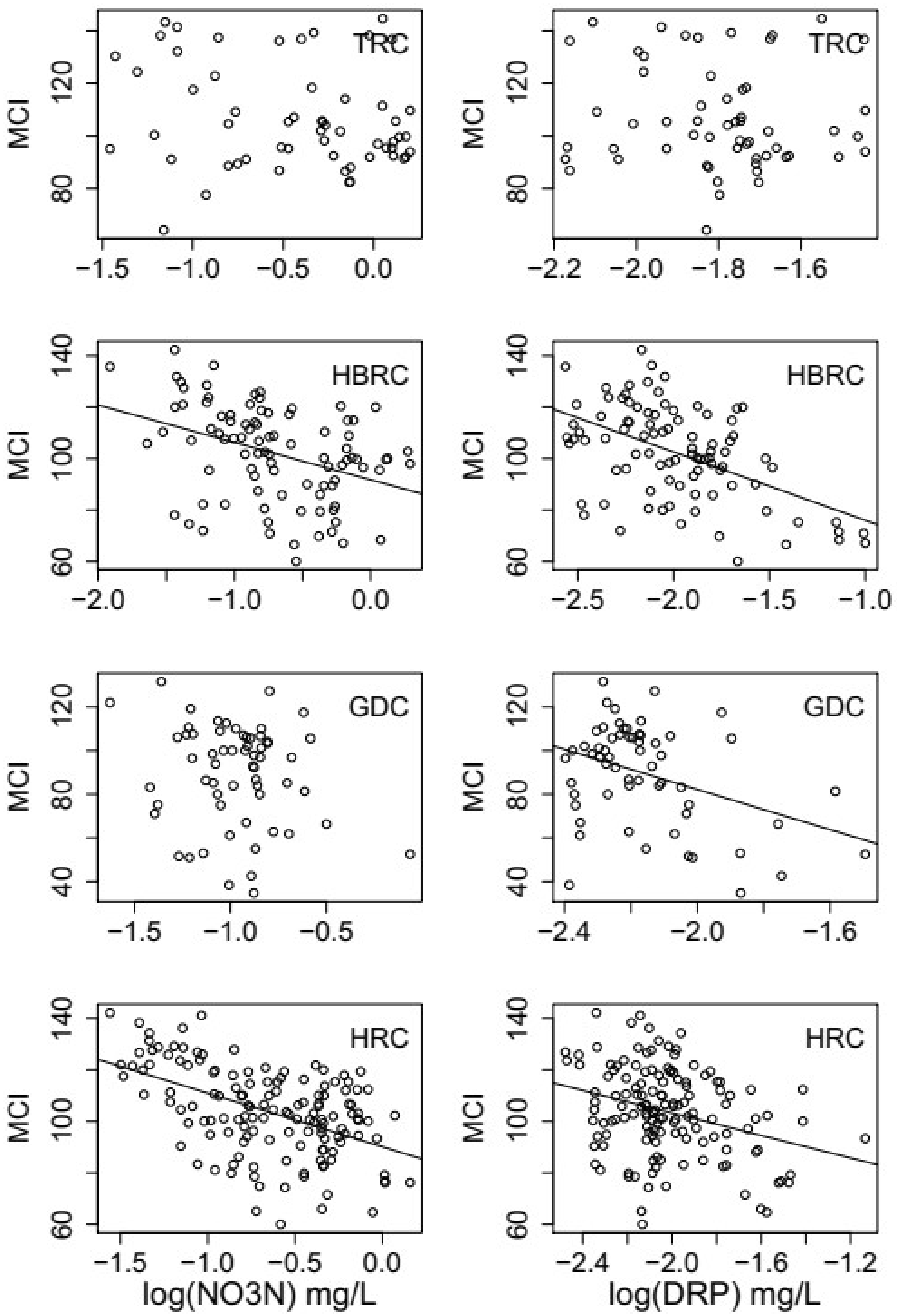


Fig. L21. Regional 5-year (2012-2016) average MCI versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). TRC=Taranaki, HBRC=Hawkes Bay, GDC=Gisborne and HRC=Manawatu-Whanganui.

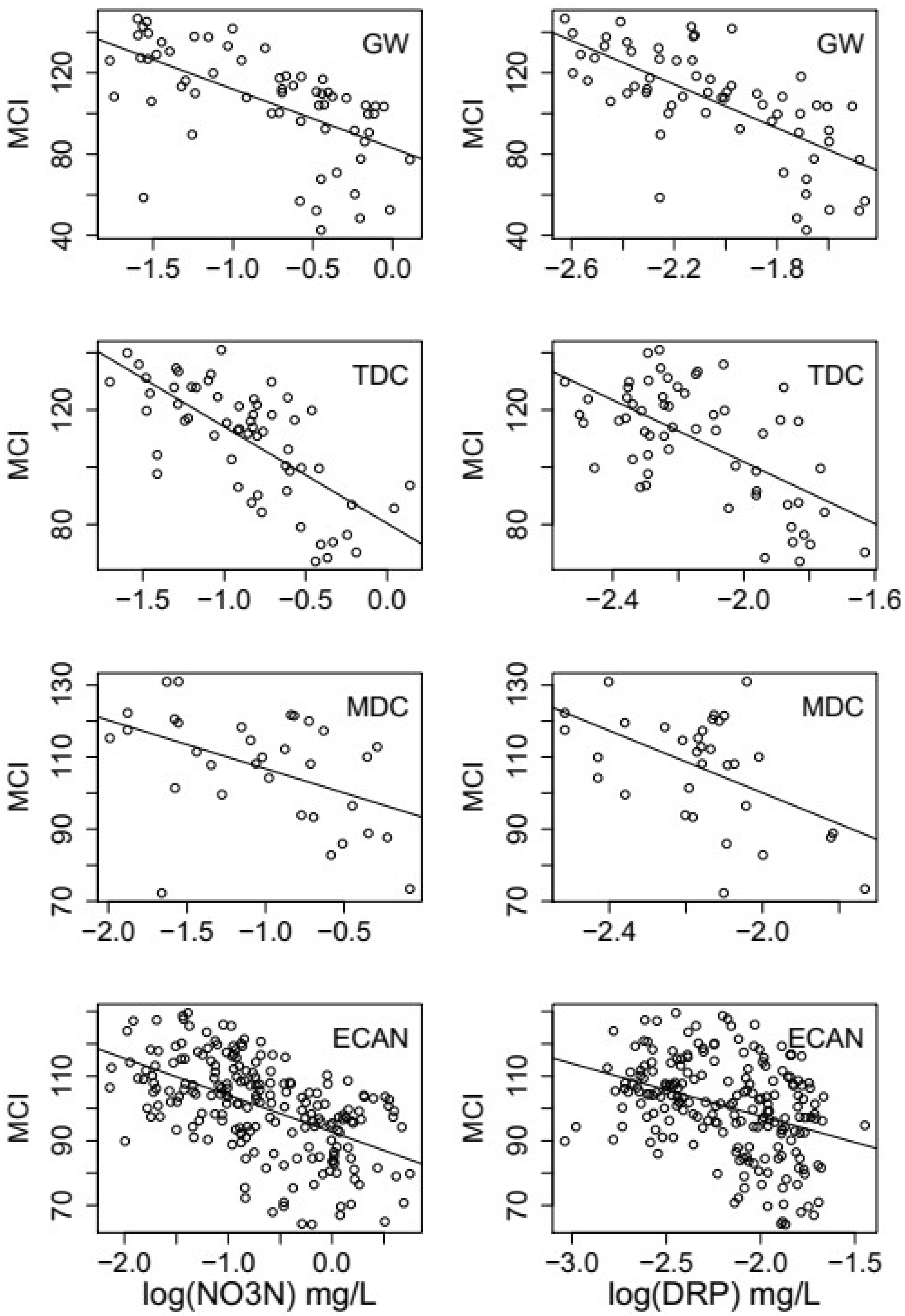


Fig. L22. Regional 5-year (2012-2016) average MCI versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). GW=Wellington, TDC=Tasman, MDC=Marlborough and ECAN=Canterbury.

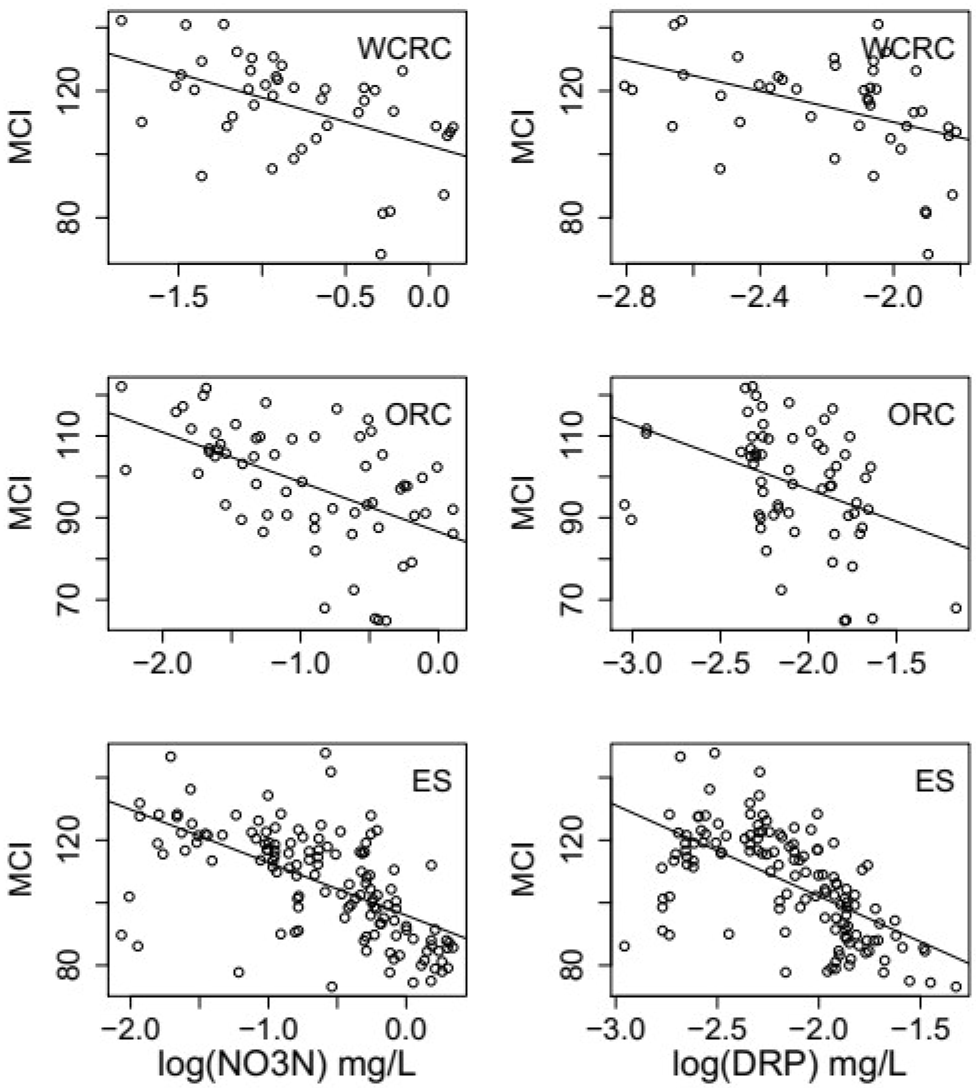


Fig. L23. Regional 5-year (2012-2016) average MCI versus modelled median log(NO_3-N) (left panel) or median log(DRP) (right panel). WCRC=West Coast, ORC=Otago and ES=Southland.

QMCI vs modelled nutrients: Regional

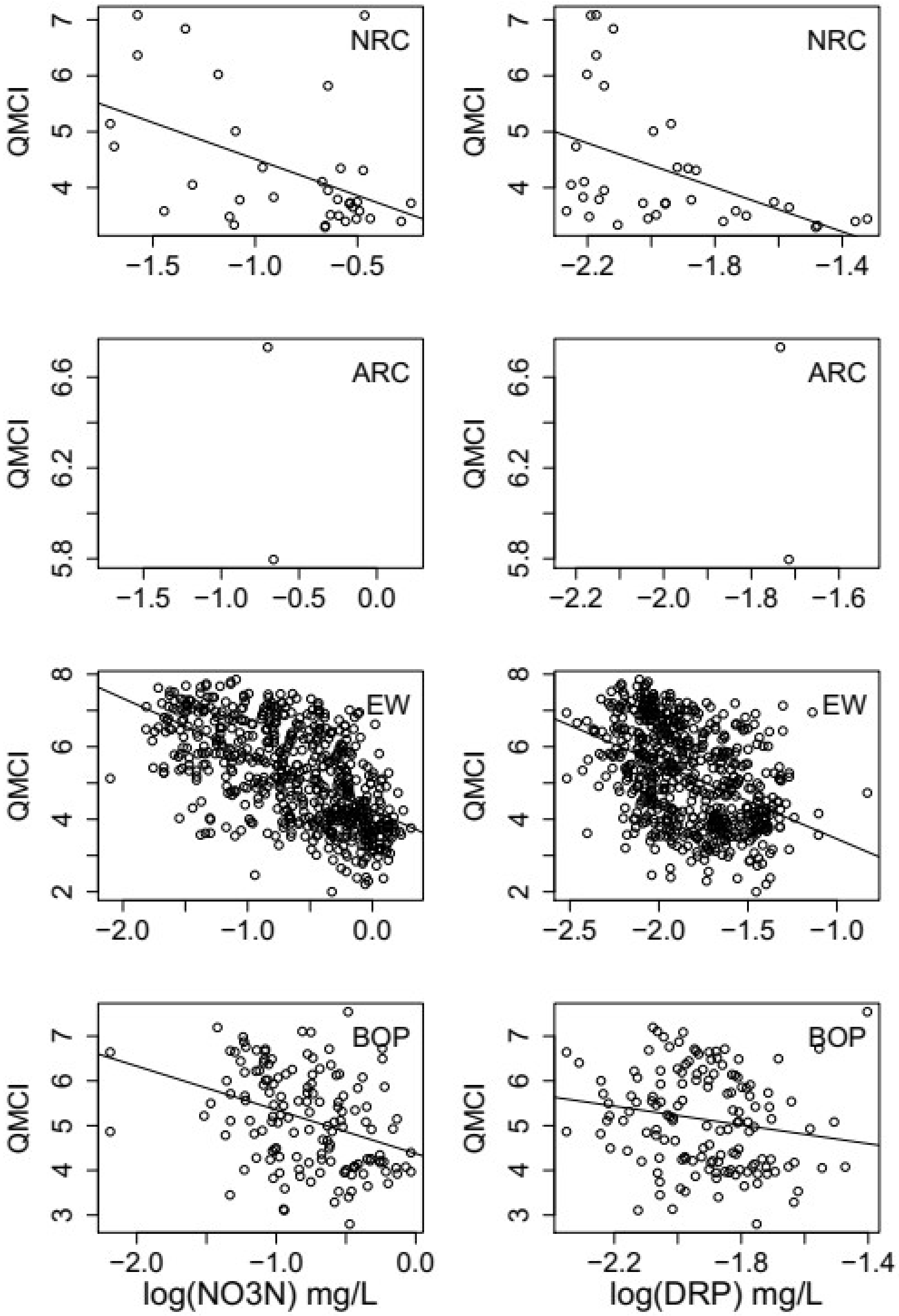


Fig. L24. Regional 5-year (2012-2016) average QMCI versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). NRC=Northland, ARC=Auckland, EW=Waikato and BOP=Bay of Plenty.

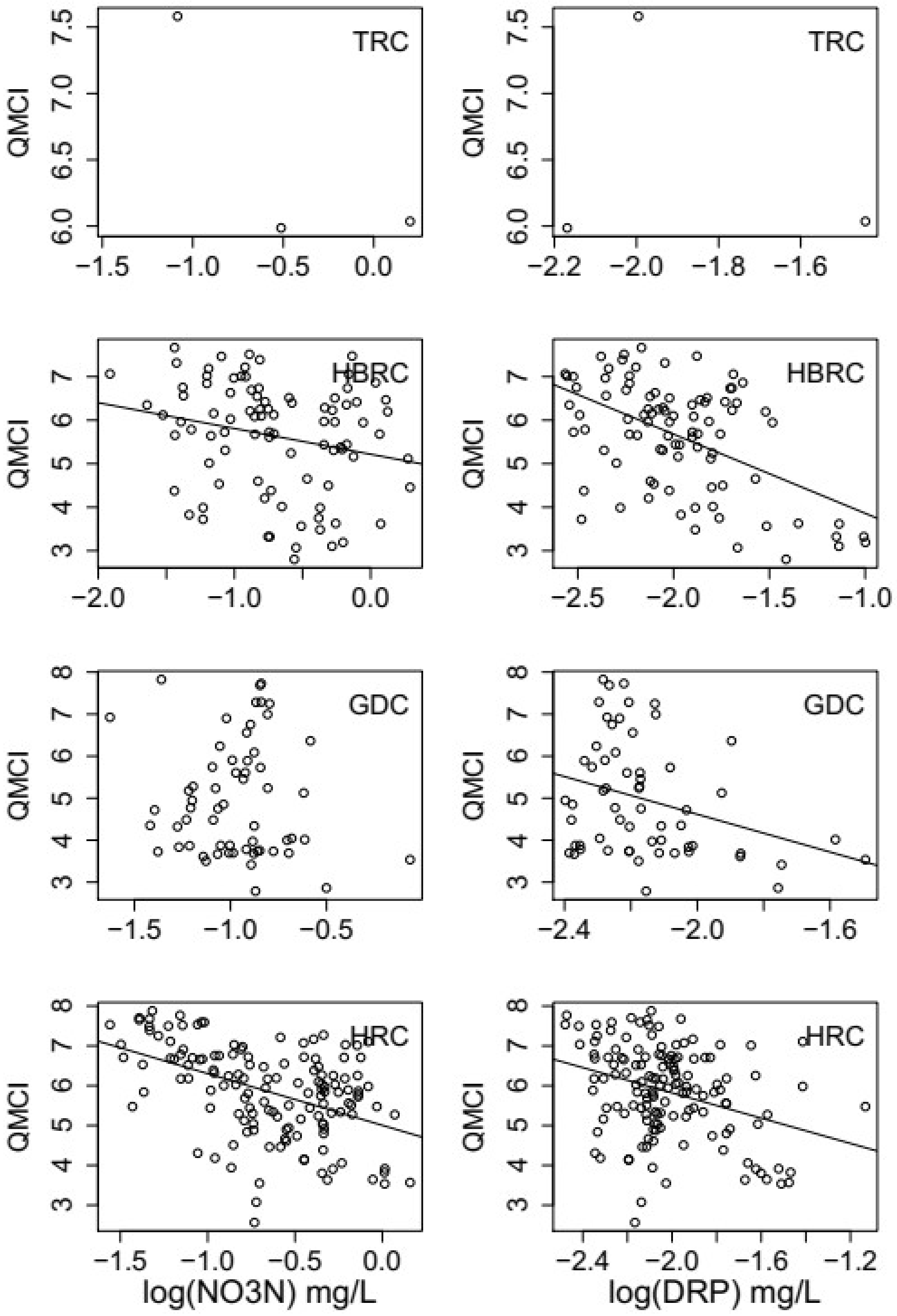


Fig. L25. Regional 5-year (2012-2016) average QMCI versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). TRC=Taranaki, HBRC=Hawkes Bay, GDC=Gisborne, and HRC=Manawatu-Whanganui.

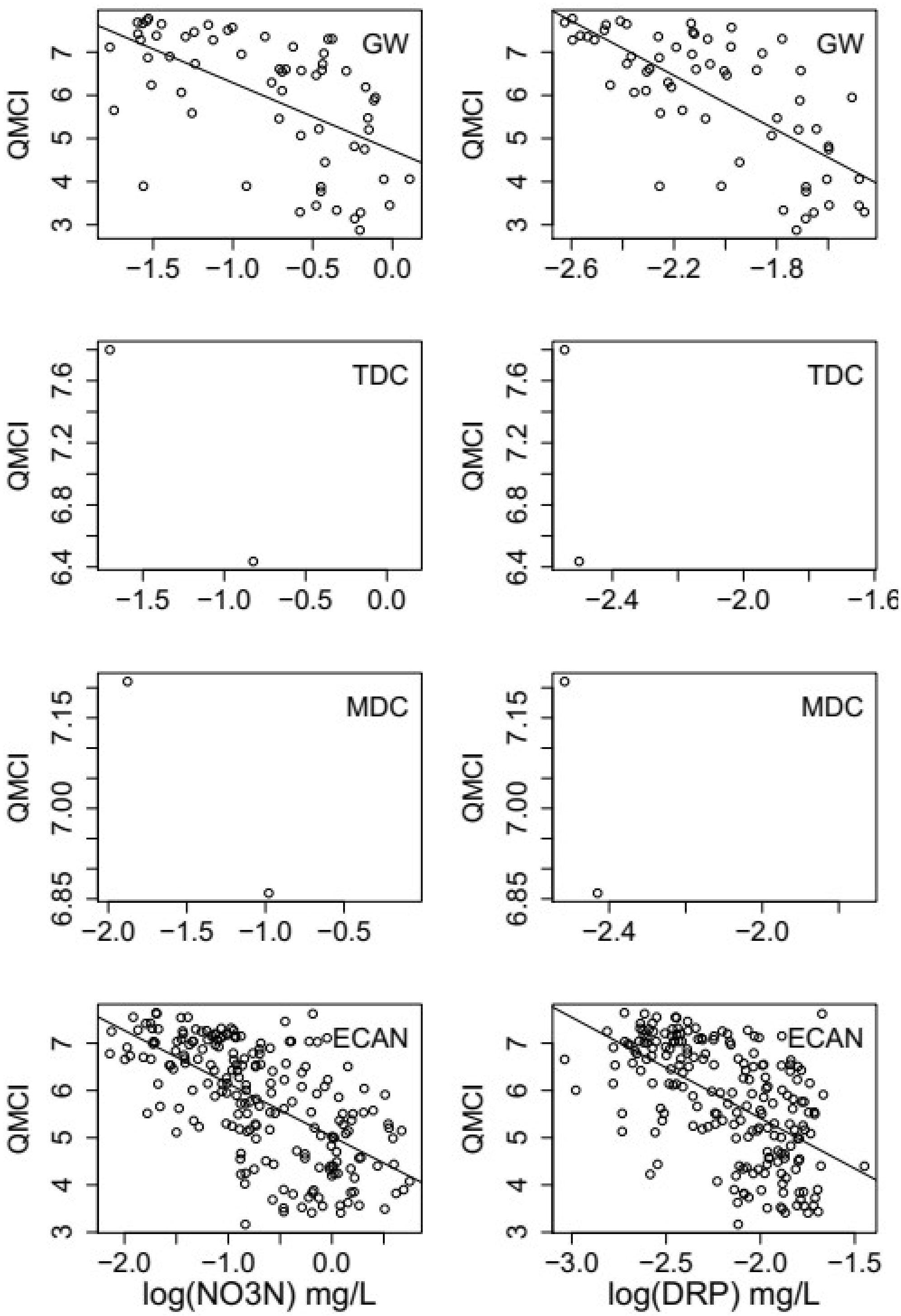


Fig. L26. Regional 5-year (2012-2016) average QMCI versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). GW=Wellington, TDC=Tasman, MDC=Marlborough and ECAN=Canterbury

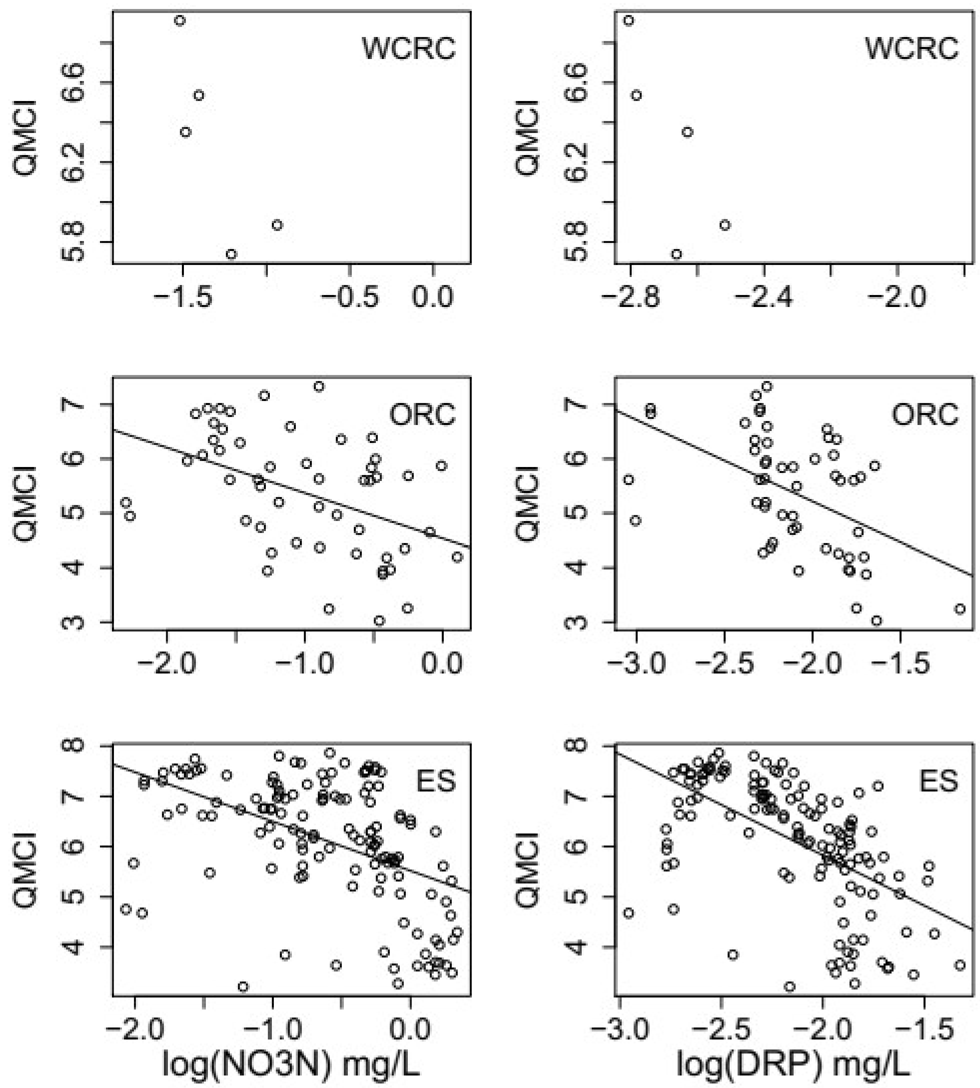


Fig. L27. Regional 5-year (2012-2016) average QMCI versus modelled median log(NO_3-N) (left panel) or median log(DRP) (right panel). WCRC=West Coast, ORC=Otago and ES=Southland.

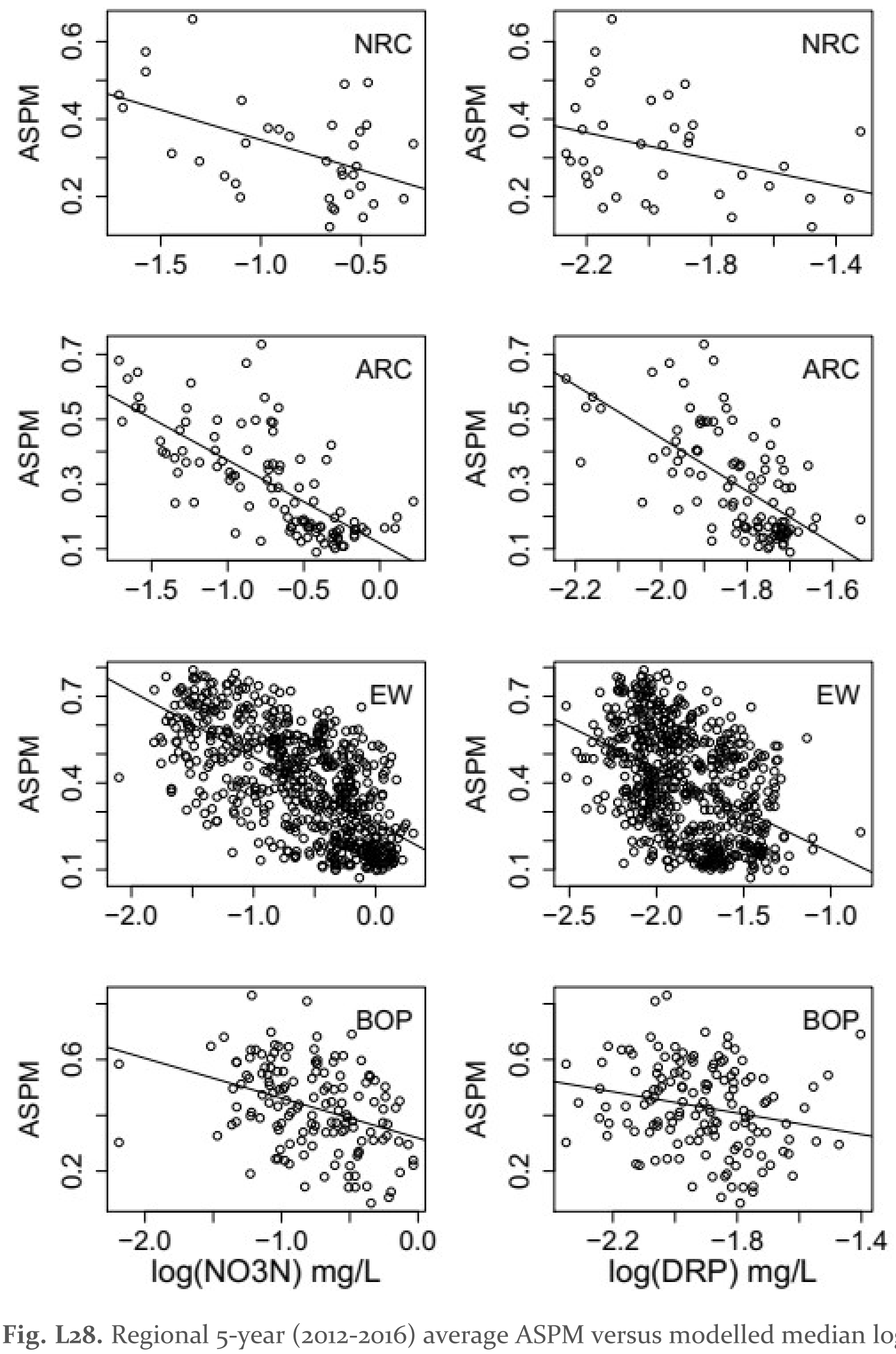


Fig. L28. Regional 5-year (2012-2016) average ASPM versus modelled median log(NO_3-N) (left panel) or median log(DRP) (right panel). NRC=Northland, ARC=Auckland, EW=Waikato and BOP=Bay of Plenty.

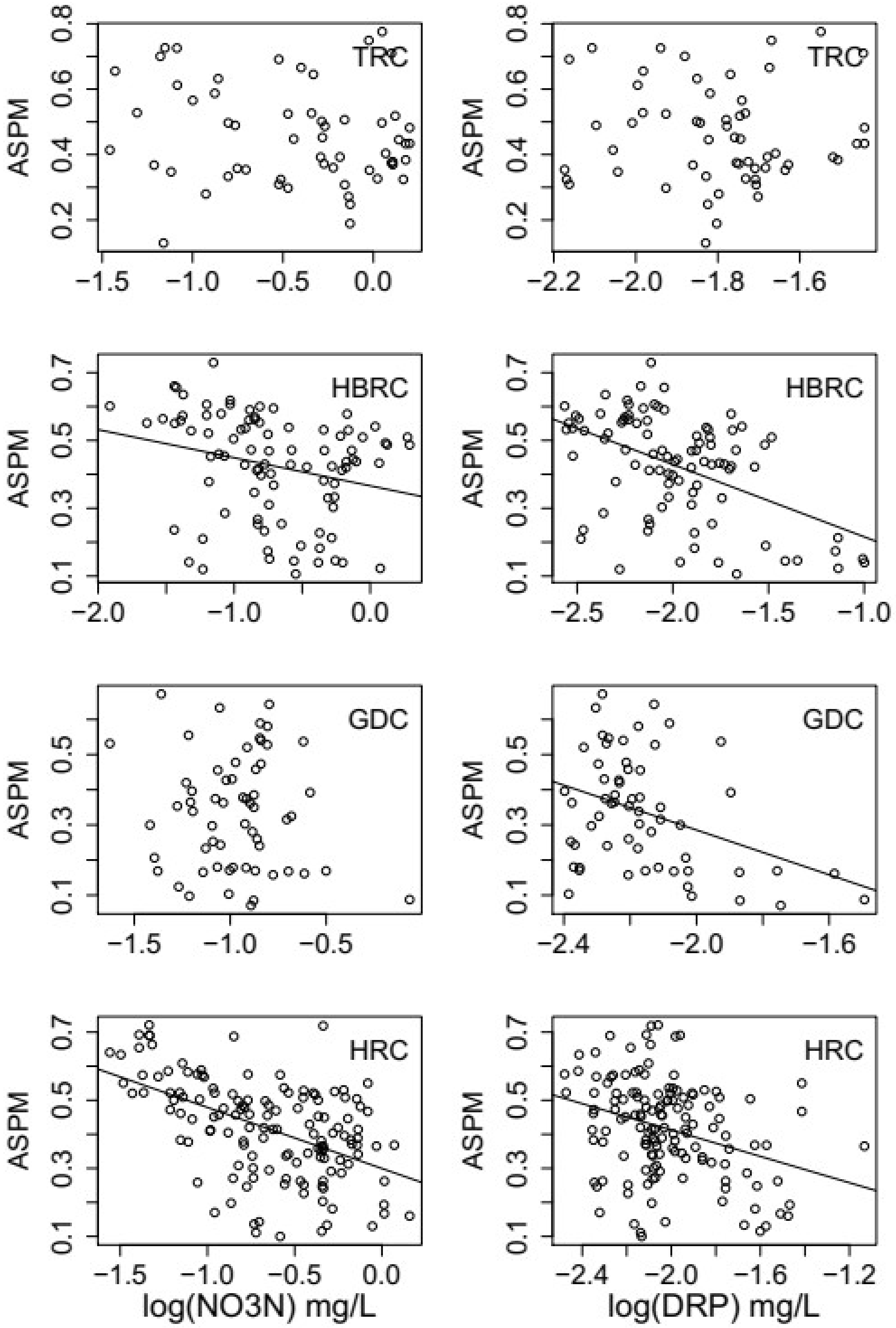


Fig. L29. Regional 5-year (2012-2016) average ASPM versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). TRC=Taranaki, HBRC=Hawkes Bay, GDC=Gisborne and HRC=Manawatu-Whanaganui.

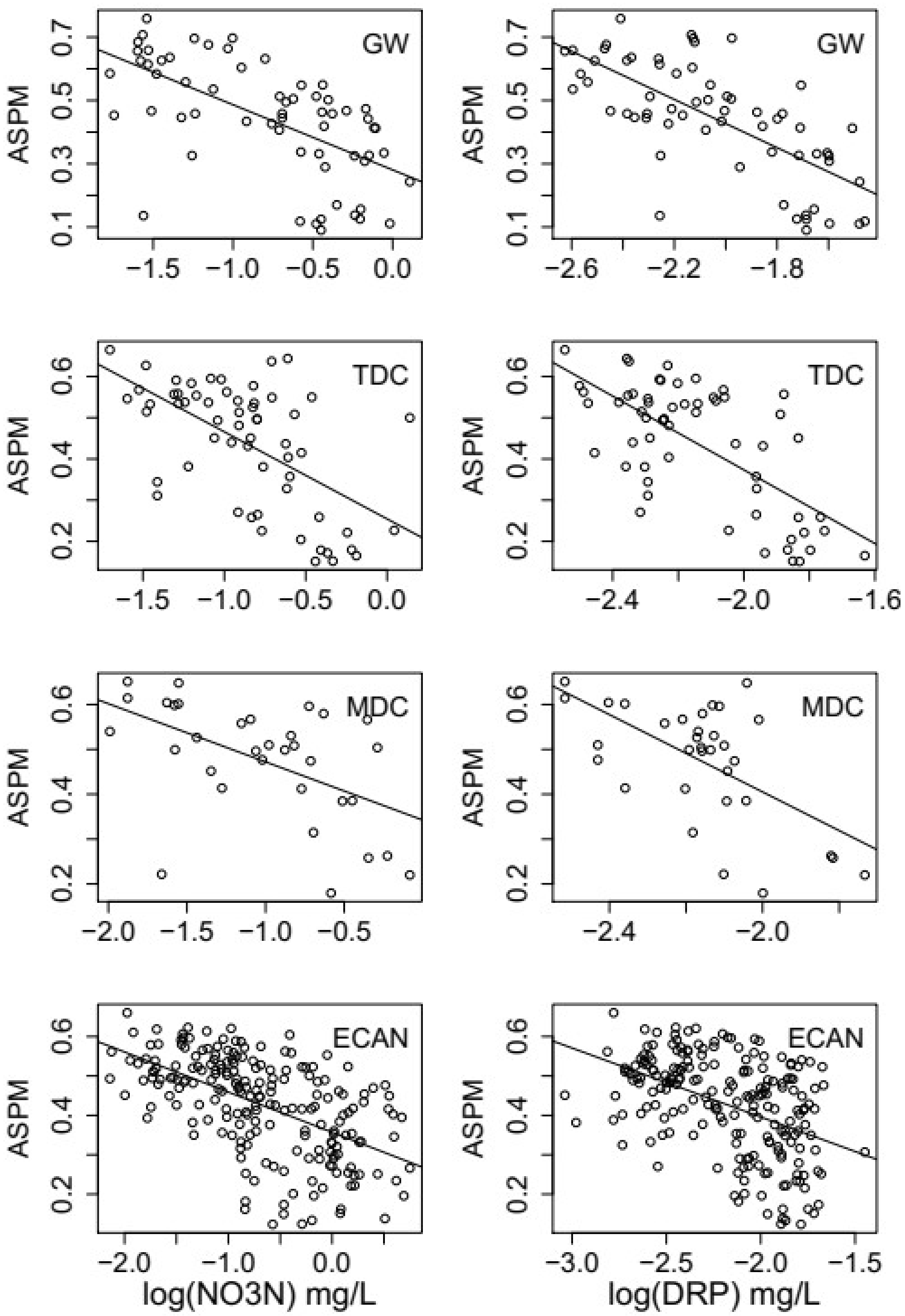


Fig. L30. Regional 5-year (2012-2016) average ASPM versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). GW=Wellington, TDC=Tasman, MDC=Marlborough and ECAN=Canterbury.

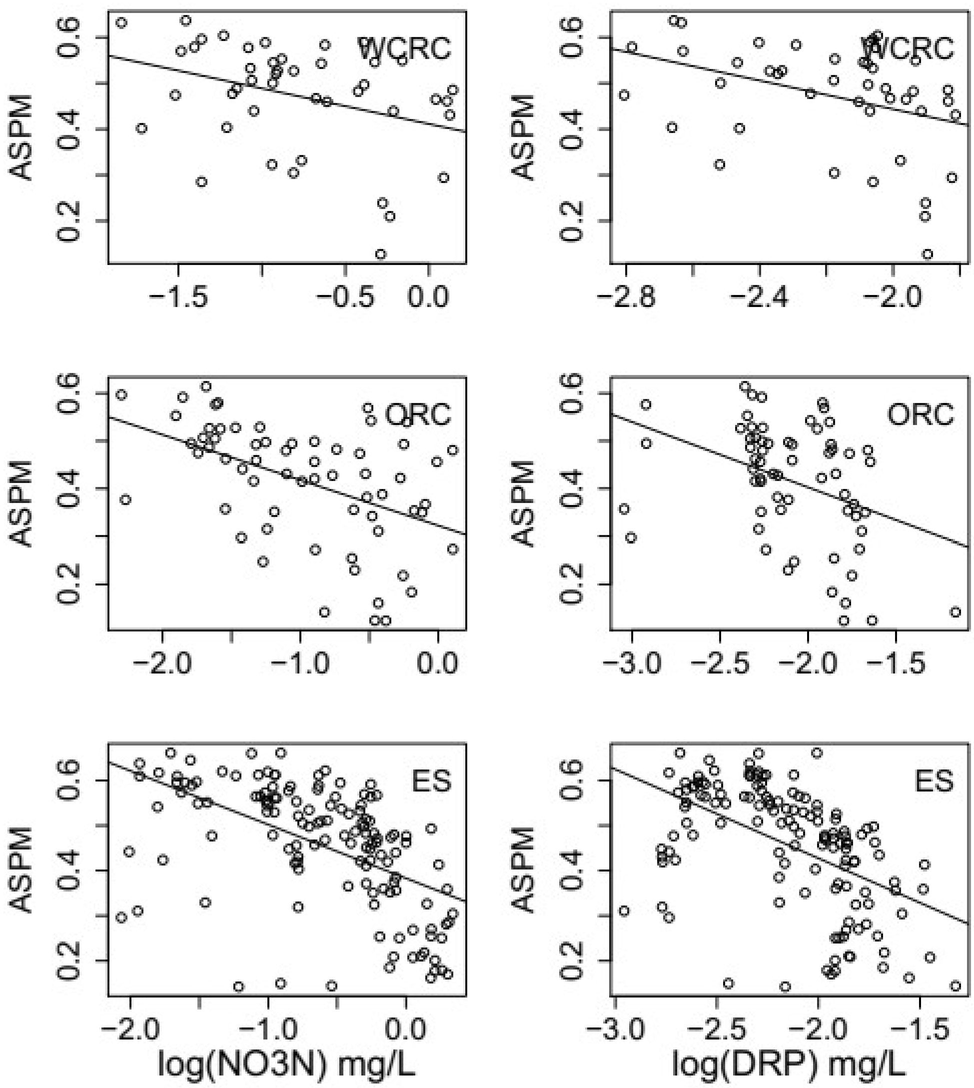


Fig. L31. Regional 5-year (2012-2016) average ASPM versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). WCRC=West Coast, ORC=Otago and ES=Southland.

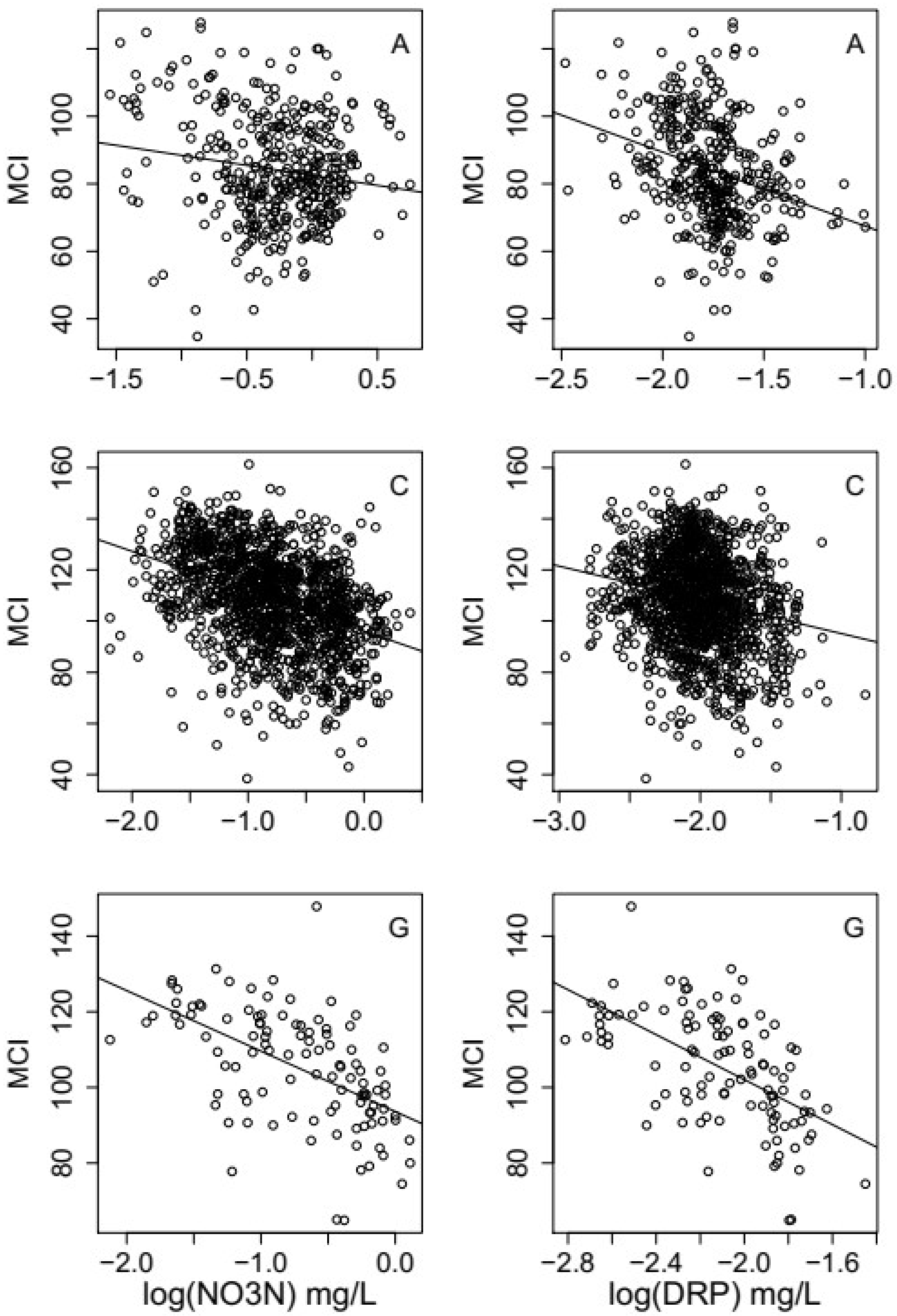


Fig. L32. Five year average (2012-2016) MCI versus modelled median log(NO_3-N) (left panel) or median log(DRP) (right panel). Plots A, C and G are the three most data rich FENZ LevelOneAlpha classes.

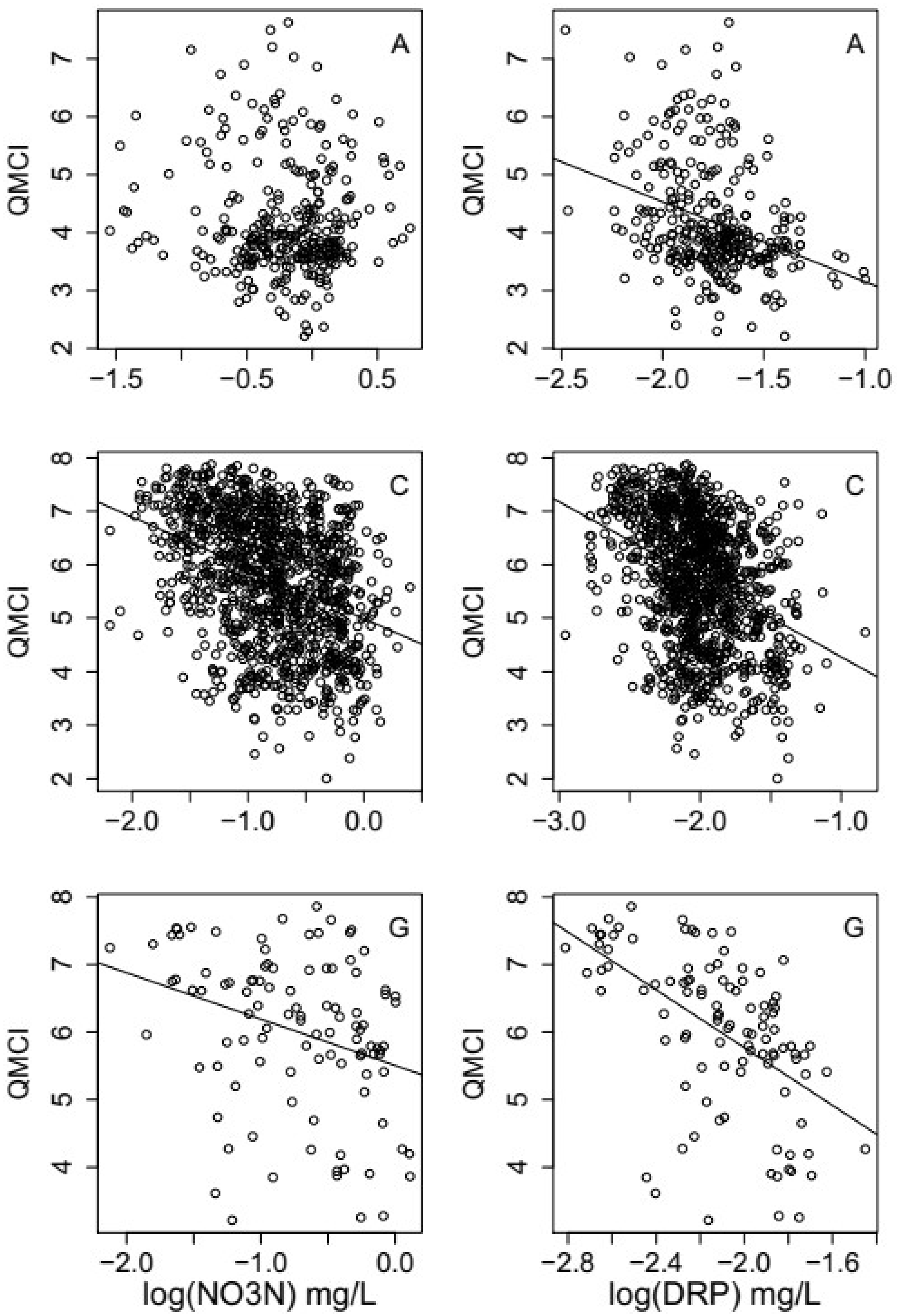


Fig. L33. Five year average (2012-2016) QMCI versus median modelled log(NO3-N) (left panel) or median log(DRP) (right panel). Plots A, C and G are the three most data rich FENZ LevelOneAlpha classes.

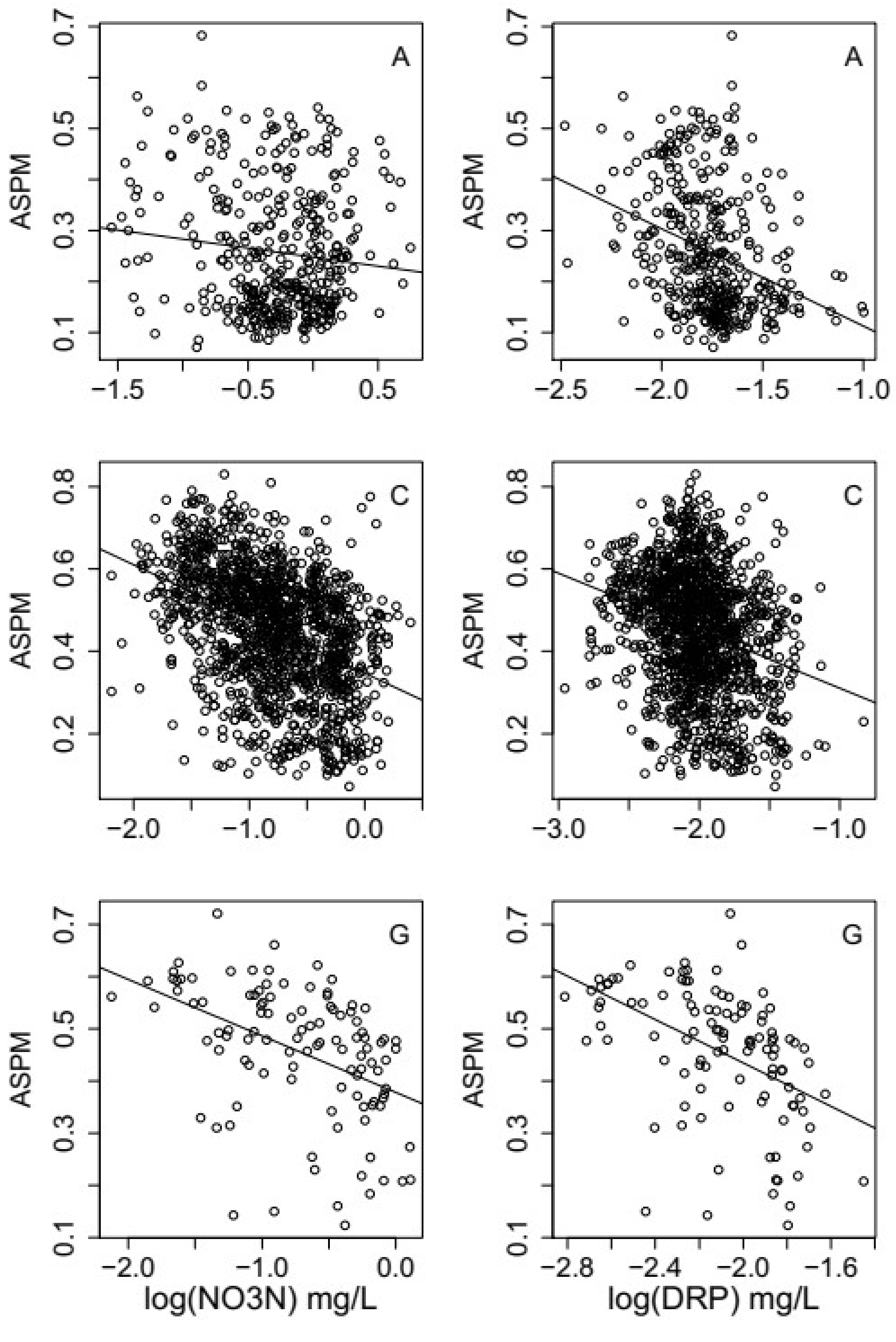


Fig. L34. Five year average (2012-2016) ASPM versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel). Plots A, C and G are the three most data rich FENZ LevelOneAlpha classes.

metrics vs modelled nutrients: rec climate

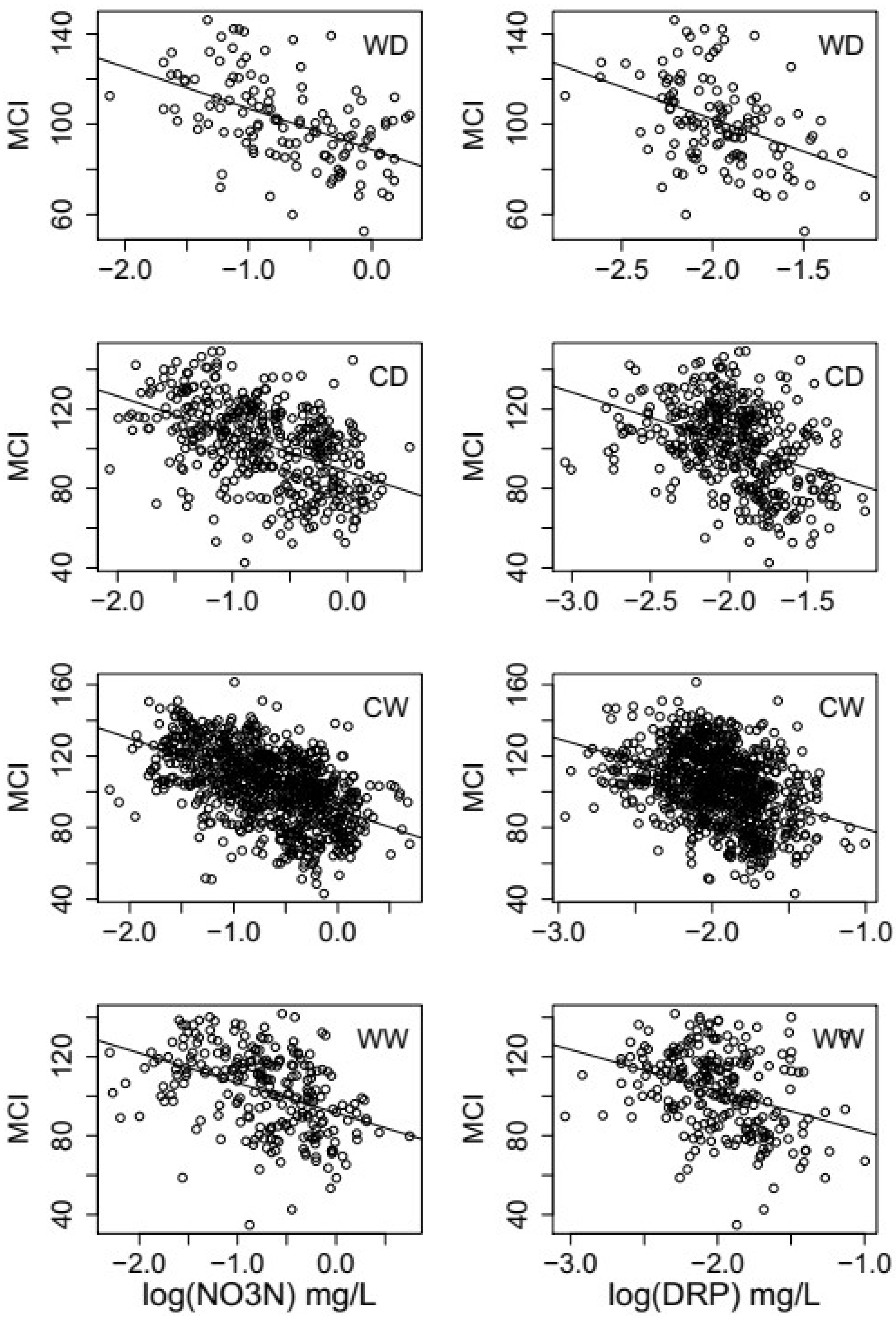


Fig. L35. Five year average (2012-2016) MCI versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel), based on the REC climate class. WD=warm dry, CD=cool dry, CW=cool wet, and WW=warm wet.

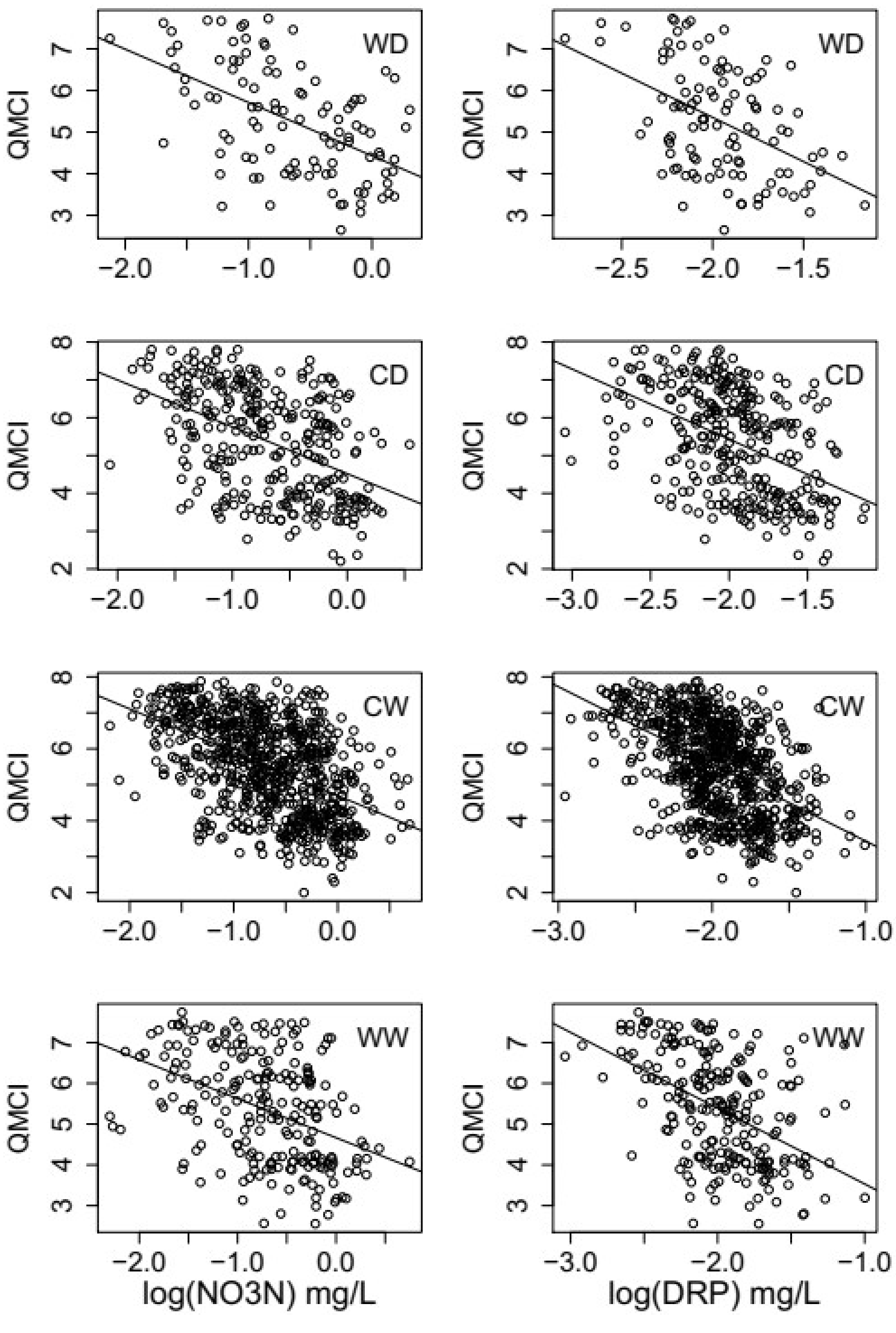


Fig. L36. Five year average (2012-2016) QMCI versus modelled median log(NO3-N) (left panel) or median log(DRP) (right panel), based on the REC climate class. WD=warm dry, CD=cool dry, CW=cool wet, and WW=warm wet.

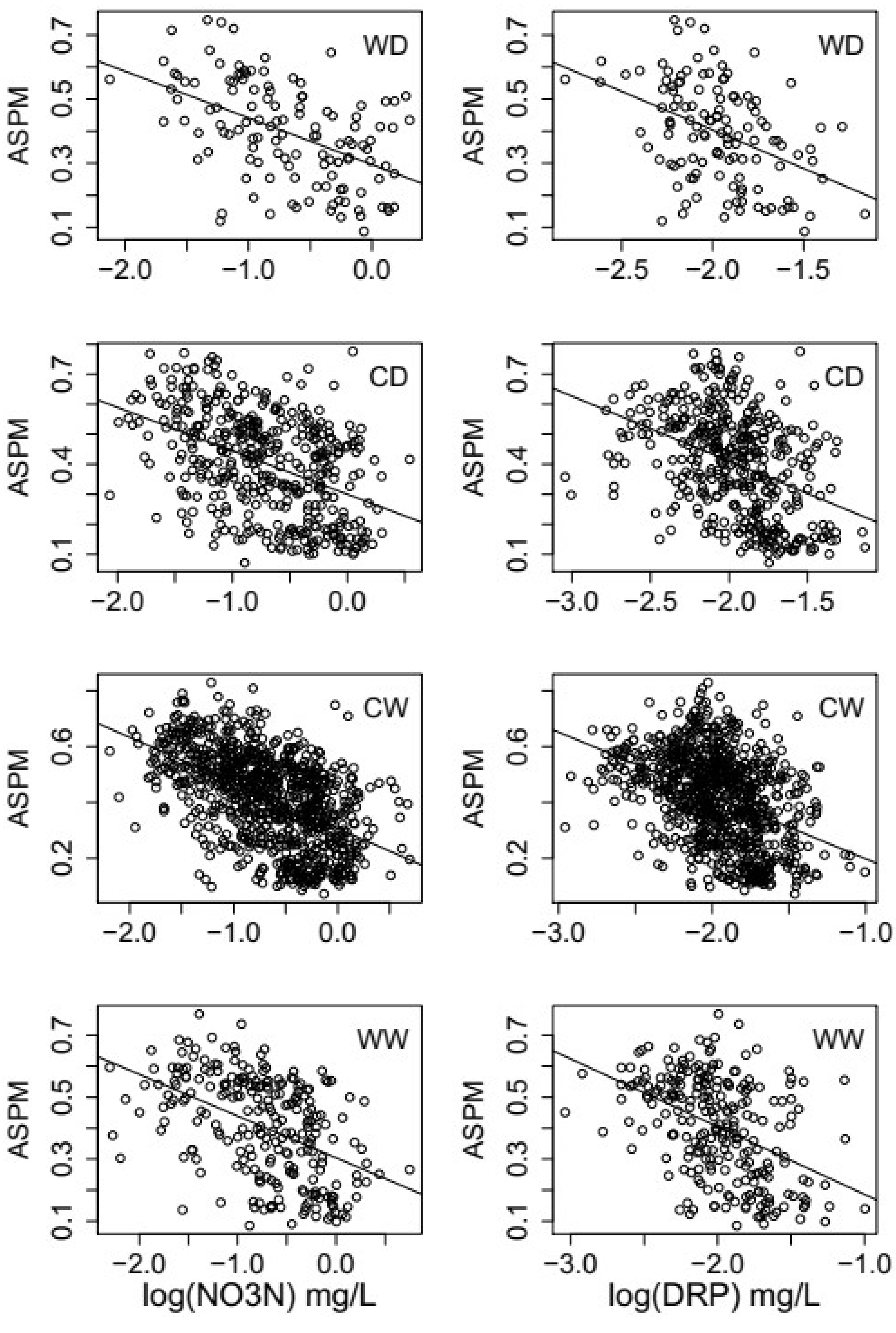


Fig. L37. Five year average (2012-2016) ASPM versus modelled median log(NO_3-N) (left panel) or median log(DRP) (right panel), based on the REC climate class. WD=warm dry, CD=cool dry, CW=cool wet, and WW=warm wet.

References

Abrams, P. A. and Roth, J. D. (1994) 'The Effects of Enrichment of Three-Species Food Chains with Nonlinear Functional Responses', *Ecology*. John Wiley & Sons, Ltd, 75(4), pp. 1118–1130. doi: 10.2307/1939435.

Anderson, N. H. and Sedell, J. R. (1979) 'Detritus Processing by Macroinvertebrates in Stream Ecosystems', *Annual Review of Entomology*. Annual Reviews, 24(1), pp. 351–377. doi: 10.1146/annurev.en.24.010179.002031.

ANZECC (2000) 'Australian and New Zealand guidelines for fresh and marine water quality', *Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra*, pp. 1–103.

ANZG (2018) *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*.

Arsuffi, T. L. and Suberkropp, K. (1989) 'Selective feeding by shredders on leaf-colonizing stream fungi: comparison of macroinvertebrate taxa', *Oecologia*, 79(1), pp. 30–37. doi: 10.1007/BF00378236.

Attayde, J. L. and Ripa, J. (2008) 'The Coupling Between Grazing and Detritus Food Chains and the Strength of Trophic Cascades Across a Gradient of Nutrient Enrichment', *Ecosystems*, 11(6), pp. 980–990. doi: 10.1007/s10021-008-9174-8.

Back, J. A. and King, R. S. (2013) 'Sex and size matter: ontogenetic patterns of nutrient content of aquatic insects', *Freshwater Science*. The University of Chicago Press, 32(3), pp. 837–848. doi: 10.1899/12-181.1.

Bärlocher, F. and Kendrick, B. (1975) 'Leaf-conditioning by microorganisms', *Oecologia*, 20(4), pp. 359–362. doi: 10.1007/BF00345526.

Bdmstedt, U. and Skjoldal, H. R. (1980) 'RNA concentration of zooplankton: Relationship with size and growth1', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 25(2), pp. 304–316. doi: 10.4319/lo.1980.25.2.0304.

Belgrano, A. *et al.* (2004) 'Aquatic food webs : An ecosystem approach'. Oxford University Press: Oxford, pp. x, 262.

Bellingham, P. J. *et al.* (2013) 'Litterfall, nutrient concentrations and decomposability of litter in a New Zealand temperate montane rain forest', *New Zealand Journal of Ecology*. New Zealand Ecological Society, 37(2), p. 162.

Benke, A. C. and Huryn, A. D. (2010) 'Benthic invertebrate production—facilitating answers to ecological riddles in freshwater ecosystems', *Journal of the North American Benthological Society*, 29(1), pp. 264–285. doi: 10.1899/08-075.1.

Benstead, J. P. *et al.* (2009) 'Nutrient enrichment alters storage and fluxes of detritus in a headwater stream ecosystem', *Ecology*. John Wiley & Sons, Ltd, 90(9), pp. 2556–2566. doi: 10.1890/08-0862.1.

Biggs, B. J. F. (2000) 'Eutrophication of streams and rivers: dissolved nutrient-chlorophyll relationships for benthic algae', *Journal of the North American Benthological Society*, 19(1), pp. 17–31.

Booker, D. J. and Woods, R. A. (2014) 'Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments', *Journal of Hydrology*, 508, pp. 227–239. doi: https://doi.org/10.1016/j.jhydrol.2013.11.007.

Breiman, L. (2001) 'Random forests', *Machine learning*. Springer, 45(1), pp. 5–32.

Brennan, A., Cross, P. C. and Creel, S. (2015) 'Managing more than the mean: using quantile regression to identify factors related to large elk groups', *The Journal of applied ecology*. 2015/08/27. John Wiley and Sons Inc., 52(6), pp. 1656–1664. doi: 10.1111/1365-2664.12514.

Cade, B. S. and Noon, B. R. (2003) 'A gentle introduction to quantile regression for ecologists', *Frontiers in Ecology and the Environment*, 1(8), pp. 412–420. doi: doi:10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2.

Camargo, J. A. and Alonso, Á. (2006) 'Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: A global assessment', *Environment International*, 32(6), pp. 831–849. doi: https://doi.org/10.1016/j.envint.2006.05.002.

Camargo, J. A., Alonso, A. and Salamanca, A. (2005) 'Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates', *Chemosphere*, 58(9), pp. 1255–1267. doi: http://dx.doi.org/10.1016/j.chemosphere.2004.10.044.

Canning, A. D. (2018) 'Predicting New Zealand riverine fish reference assemblages', *PeerJ*, 2018(5). doi: 10.7717/peerj.4890.

Canning, A. D. and Death, R. G. (2017) 'Trophic cascade direction and flow determine network flow stability', *Ecological Modelling*, 355. doi: 10.1016/j.ecolmodel.2017.03.020.

Canning, A. D. and Death, R. G. (2018) 'Relative ascendency predicts food web robustness', *Ecological Research*, 33(5). doi: 10.1007/s11284-018-1585-1.

Chamaillé-Jammes, S., Fritz, H. and Murindagomo, F. (2007) 'Detecting climate changes of concern in highly variable environments: Quantile regressions reveal that droughts worsen in Hwange National Park, Zimbabwe', *Journal of Arid Environments*, 71(3), pp. 321–326. doi: https://doi.org/10.1016/j.jaridenv.2007.05.005.

Civitello, D. J. *et al.* (2018) 'Assessing the direct and indirect effects of food provisioning and nutrient enrichment on wildlife infectious disease dynamics', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society, 373(1745), p. 20170101. doi: 10.1098/rstb.2017.0101.

Clapcott, Joanne *et al.* (2017) *Macroinvertebrate metrics for the National Policy Statement for Freshwater Management*. Nelson, New Zealand.

Clapcott, J *et al.* (2017) *Macroinvertebrate metrics for the National Policy Statement for Freshwater Management*. Nelson, New Zealand.

Clapcott, J. *et al.* (2018) *Freshwater biophysical ecosystem health framework. Prepared for Ministry for the Environment*. Cawthron Report.

Clapcott, J. E. *et al.* (2010) 'APPLIED ISSUES: Exploring the response of functional indicators of stream health to land-use gradients', *Freshwater Biology*. Blackwell Publishing Ltd, 55(10), pp. 2181–2199. doi: 10.1111/j.1365-2427.2010.02463.x.

Clapcott, J. E. *et al.* (2011) 'Sediment Assessment Methods: Protocols and guidelines for assessing the effects of deposited fine sediment on in-stream values', *Cawthron Institute,* *Nelson, New Zealand*, 108.

Clapcott, J. E. *et al.* (2012) 'Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity', *Freshwater Biology*. Blackwell Publishing Ltd, 57(1), pp. 74–90. doi: 10.1111/j.1365-2427.2011.02696.x.

Clapcott, J. E. *et al.* (2016) 'Land use affects temporal variation in stream metabolism', *Freshwater Science*. The University of Chicago Press, 35(4), pp. 1164–1175. doi: 10.1086/688872.

Collier, K. J. (2008) 'Average score per metric: An alternative metric aggregation method for assessing wadeable stream health', *New Zealand Journal of Marine and Freshwater Research*. Taylor & Francis, 42(4), pp. 367–378. doi: 10.1080/00288330809509965.

Crawley, M. J. (2013) *The R book*. Second edition. Chichester, West Sussex, United Kingdom : Wiley, 2013. Available at: https://search.library.wisc.edu/catalog/9910152827402121.

Cross, W. F. *et al.* (2003) 'Consumer-resource stoichiometry in detritus-based streams', *Ecology Letters*. John Wiley & Sons, Ltd (10.1111), 6(8), pp. 721–732. doi: 10.1046/j.1461-0248.2003.00481.x.

Cross, W. F. *et al.* (2005) 'Contrasting response of stream detritivores to long-term nutrient enrichment', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 50(6), pp. 1730–1739. doi: 10.4319/lo.2005.50.6.1730.

Cross, W. F. *et al.* (2006) 'Whole-system nutrient enrichment increases secondary production in a detritus-based ecosystem', *Ecology*, 87(6), pp. 1556–1565.

Cross, W. F., Wallace, J. B. and Rosemond, A. D. (2007) 'Nutrient enrichment reduces constrains on material flows in a detritus-based food web', *Ecology*. John Wiley & Sons, Ltd, 88(10), pp. 2563–2575. doi: 10.1890/06-1348.1.

Cutler, D. R. *et al.* (2007) 'Random forests for classification in ecology', *Ecology*. Wiley Online Library, 88(11), pp. 2783–2792.

Danger, M. *et al.* (2013) 'Phosphorus content in detritus controls life-history traits of a detritivore', *Functional Ecology*. John Wiley & Sons, Ltd (10.1111), 27(3), pp. 807–815. doi: 10.1111/1365-2435.12079.

Davis, J. M. *et al.* (2010) 'Long-term nutrient enrichment decouples predator and prey production', *Proceedings of the National Academy of Sciences*, 107(1), pp. 121 LP – 126. doi: 10.1073/pnas.0908497107.

Death, R. G. *et al.* (2015) 'How good are Bayesian belief networks for environmental management? A test with data from an agricultural river catchment', *Freshwater Biology*, 60(11), pp. 2297–2309. doi: doi:10.1111/fwb.12655.

Death, R. G. *et al.* (2018) *Why aren't we managing water quality to protect ecological health?, Farm Environmental Planning—Science, Policy and Practice*. Edited by L. D. Currie and C. L. Christensen. Palmerston North, New Zealand: Fertilizer and Lime Research Centre, Massey University.

Death, R. G. *et al.* (no date) 'Clean but not green: a weight-of-evidence approach for setting nutrient criteria in New Zealand rivers', *Marine and Freshwater Research*.

Demi, L. M. *et al.* (2018) 'Litter P content drives consumer production in detritus-based streams spanning an experimental N:P gradient', *Ecology*. John Wiley & Sons, Ltd, 99(2), pp. 347–359. doi: 10.1002/ecy.2118.

Demi, L. M. *et al.* (2019) 'Experimental N and P additions alter stream macroinvertebrate community composition via taxon-level responses to shifts in detrital resource stoichiometry', *Functional Ecology*. John Wiley & Sons, Ltd (10.1111), 33(5), pp. 855–867. doi: 10.1111/1365-2435.13289.

Demi, L. M. *et al.* (2020) 'Experimental N and P additions relieve stoichiometric constraints on organic matter flows through five stream food webs', *Journal of Animal Ecology*. John Wiley & Sons, Ltd, n/a(n/a). doi: 10.1111/1365-2656.13197.

Dodds, W. K. (2006) 'Eutrophication and trophic state in rivers and streams', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 51(1part2), pp. 671–680. doi: 10.4319/lo.2006.51.1_part_2.0671.

Dodds, W. K. (2007) 'Trophic state, eutrophication and nutrient criteria in streams', *Trends in Ecology & Evolution*, 22(12), pp. 669–676. doi: http://dx.doi.org/10.1016/j.tree.2007.07.010.

Dolédec, S. *et al.* (2006) 'Comparison of structural and functional approaches to determining landuse effects on grassland stream invertebrate communities', *Journal of the North American Benthological Society*. The University of Chicago Press, 25(1), pp. 44–60. doi: 10.1899/0887-3593(2006)25[44:COSAFA]2.0.CO;2.

Duggan, I. C., Boothroyd, I. K. G. and Speirs, D. A. (2007) 'Factors affecting the distribution of stream macroinvertebrates in geothermal areas: Taupo Volcanic Zone, New Zealand', *Hydrobiologia*, 592(1), pp. 235–247. doi: 10.1007/s10750-007-0748-9.

Elith, J., Leathwick, J. R. and Hastie, T. (2008) 'A working guide to boosted regression trees', *Journal of Animal Ecology*, 77(4), pp. 802–813.

Ellis, N. (2019a) 'extendedForest'.

Ellis, N. (2019b) 'gradientForest'.

Ellis, N., Smith, S. J. and Pitcher, C. R. (2012) 'Gradient forests: calculating importance gradients on physical predictors', *Ecology*. John Wiley & Sons, Ltd, 93(1), pp. 156–168. doi: 10.1890/11-0252.1.

Elser, J. J. *et al.* (1996) 'Organism Size, Life History, and N:P Stoichiometry: Toward a unified view of cellular and ecosystem processes', *BioScience*, 46(9), pp. 674–684. doi: 10.2307/1312897.

Elser, J. J. *et al.* (2000) 'Biological stoichiometry from genes to ecosystems', *Ecology Letters*. John Wiley & Sons, Ltd (10.1111), 3(6), pp. 540–550. doi: 10.1111/j.1461-0248.2000.00185.x.

Evans-White, M. A. *et al.* (2009) 'Thresholds in macroinvertebrate biodiversity and stoichiometry across water-quality gradients in Central Plains (USA) streams', *Journal of the North American Benthological Society*. The University of Chicago Press, 28(4), pp. 855–868. doi: 10.1899/08-113.1.

Fath, B. D., Patten, B. C. and Choi, J. S. (2001) 'Complementarity of ecological goal functions', *Journal of Theoretical Biology*, 208(4), pp. 493–506. doi: http://dx.doi.org/10.1006/jtbi.2000.2234.

Feijoó, C. *et al.* (2014) 'Stoichiometric homeostasis in the food web of a chronically nutrient-rich stream', *Freshwater Science*. The University of Chicago Press, 33(3), pp. 820–831. doi: 10.1086/677056.

Ferreira, V. *et al.* (2015) 'A meta-analysis of the effects of nutrient enrichment on litter decomposition in streams', *Biological Reviews*. John Wiley & Sons, Ltd (10.1111), 90(3), pp. 669–688. doi: 10.1111/brv.12125.

Fink, P. and Von Elert, E. (2006) 'Physiological responses to stoichiometric constraints: nutrient limitation and compensatory feeding in a freshwater snail', *Oikos*. John Wiley & Sons, Ltd (10.1111), 115(3), pp. 484–494. doi: 10.1111/j.2006.0030-1299.14951.x.

Frainer, A. *et al.* (2016) 'Stoichiometric imbalances between detritus and detritivores are related to shifts in ecosystem functioning', *Oikos*. John Wiley & Sons, Ltd (10.1111), 125(6), pp. 861–871. doi: 10.1111/oik.02687.

Frost, P. C. *et al.* (2006) 'Threshold elemental ratios of carbon and phosphorus in aquatic consumers', *Ecology Letters*. John Wiley & Sons, Ltd (10.1111), 9(7), pp. 774–779. doi: 10.1111/j.1461-0248.2006.00919.x.

Frost, P. C., Ebert, D. and Smith, V. H. (2008) 'Responses of a bacterial pathogen to phosphorus limitation of its aquatic invertebrate host', *Ecology*. John Wiley & Sons, Ltd, 89(2), pp. 313–318. doi: 10.1890/07-0389.1.

Gillooly, J. F. *et al.* (2005) 'The metabolic basis of whole-organism RNA and phosphorus content', *Proceedings of the National Academy of Sciences of the United States of America*, 102(33), pp. 11923 LP – 11927. doi: 10.1073/pnas.0504756102.

Goedkoop, W., Demandt, M. and Ahlgren, G. (2007) 'Interactions between food quantity and quality (long-chain polyunsaturated fatty acid concentrations) effects on growth and development of Chironomus riparius', *Canadian Journal of Fisheries and Aquatic Sciences*. NRC Research Press, 64(3), pp. 425–436. doi: 10.1139/f07-016.

Gómez, G. D. and Balcázar, J. L. (2008) 'A review on the interactions between gut microbiota and innate immunity of fish', *FEMS Immunology & Medical Microbiology*, 52(2), pp. 145–154. doi: 10.1111/j.1574-695X.2007.00343.x.

González, A. L., Romero, G. Q. and Srivastava, D. S. (2014) 'Detrital nutrient content determines growth rate and elemental composition of bromeliad-dwelling insects', *Freshwater Biology*. John Wiley & Sons, Ltd (10.1111), 59(4), pp. 737–747. doi: 10.1111/fwb.12300.

Graça, M. A. S. (2001) 'The Role of Invertebrates on Leaf Litter Decomposition in Streams – a Review', *International Review of Hydrobiology*. John Wiley & Sons, Ltd, 86(4-5), pp. 383–393. doi: 10.1002/1522-2632(200107)86:4/5<383::AID-IROH383>3.0.CO;2-D.

Gulis, V., Ferreira, V. and Graca, M. A. S. (2006) 'Stimulation of leaf litter decomposition and associated fungi and invertebrates by moderate eutrophication: implications for stream assessment', *Freshwater Biology*. John Wiley & Sons, Ltd (10.1111), 51(9), pp. 1655–1669. doi: 10.1111/j.1365-2427.2006.01615.x.

Guo, F., Kainz, M. J., Sheldon, F., *et al.* (2016) 'Effects of light and nutrients on periphyton and the fatty acid composition and somatic growth of invertebrate grazers in subtropical streams', *Oecologia*, 181(2), pp. 449–462. doi: 10.1007/s00442-016-3573-x.

Guo, F., Kainz, M. J., Valdez, D., *et al.* (2016a) 'High-quality algae attached to leaf litter boost invertebrate shredder growth', *Freshwater Science*. The University of Chicago Press, 35(4), pp.

1213–1221. doi: 10.1086/688667.

Guo, F., Kainz, M. J., Valdez, D., *et al.* (2016b) 'The effect of light and nutrients on algal food quality and their consequent effect on grazer growth in subtropical streams', *Freshwater Science*. The University of Chicago Press, 35(4), pp. 1202–1212. doi: 10.1086/688092.

Hall, S. R. (2008) 'Stoichiometrically Explicit Food Webs: Feedbacks between Resource Supply, Elemental Constraints, and Species Diversity', *Annual Review of Ecology, Evolution, and Systematics*. Annual Reviews, 40(1), pp. 503–528. doi: 10.1146/annurev.ecolsys.39.110707.173518.

Harris, J. M. (1993) 'The presence, nature, and role of gut microflora in aquatic invertebrates: A synthesis', *Microbial Ecology*, 25(3), pp. 195–231. doi: 10.1007/BF00171889.

Hastie, T., Tibshirani, R. and Friedman, J. (2009) 'Random forests', in *The elements of statistical learning*. Springer, pp. 587–604.

Hessen, D. O. *et al.* (2010) 'Genome streamlining and the elemental costs of growth', *Trends in Ecology & Evolution*, 25(2), pp. 75–80. doi: https://doi.org/10.1016/j.tree.2009.08.004.

Hessen, D. O. *et al.* (2013) 'Ecological stoichiometry: An elementary approach using basic principles', *Limnology and Oceanography*. John Wiley & Sons, Ltd, 58(6), pp. 2219–2236. doi: 10.4319/lo.2013.58.6.2219.

Hessen, D. O., Ventura, M. and Elser, J. J. (2008) 'Do phosphorus requirements for RNA limit genome size in crustacean zooplankton?', *Genome*. NRC Research Press, 51(9), pp. 685–691. doi: 10.1139/G08-053.

Hessen, D. and Persson, J. (2009) 'Genome size as a determinant of growth and life-history traits in crustaceans', *Biological Journal of the Linnean Society*, 98(2), pp. 393–399. doi: 10.1111/j.1095-8312.2009.01285.x.

Hickey, C. W. and Martin, M. L. (2009) *A Review of Nitrate Toxicity to Freshwater Aquatic Species*. Environment Canterbury.

Hijmans, R. J. *et al.* (2012) 'dismo: Species distribution modeling', *R package version 0.7-17*.

Huxel, G. R., McCann, K. and Polis, G. A. (2002) 'Effects of partitioning allochthonous and autochthonous resources on food web stability', *Ecological Research*, 17(4), pp. 419–432.

Jabiol, J. *et al.* (2019) 'Litter Quality Modulates Effects of Dissolved Nitrogen on Leaf Decomposition by Stream Microbial Communities', *Microbial Ecology*, 77(4), pp. 959–966. doi: 10.1007/s00248-019-01353-3.

Johnson, P. T. J. *et al.* (2010) 'Linking environmental nutrient enrichment and disease emergence in humans and wildlife', *Ecological Applications*. John Wiley & Sons, Ltd, 20(1), pp. 16–29. doi: 10.1890/08-0633.1.

Joy, M., David, B. and Lake, M. (2013) *New Zealand Freshwater Fish Sampling Protocols (Part 1): Wadeable rivers and streams*. Palmerston North, New Zealand: Massey University.

Joy, M. K. and Death, R. G. (2004) 'Application of the Index of Biotic Integrity Methodology to New Zealand Freshwater Fish Communities', *Environmental Management*. Springer-Verlag, 34(3), pp. 415–428. doi: 10.1007/s00267-004-0083-0.

King, R. S. and Richardson, C. J. (2003) 'Integrating Bioassessment and Ecological Risk Assessment: An Approach to Developing Numerical Water-Quality Criteria', *Environmental*

Management, 31(6), pp. 795–809. doi: 10.1007/s00267-002-0036-4.

Kovalenko, K. E. (2019) 'Interactions among anthropogenic effects on aquatic food webs', *Hydrobiologia*. doi: 10.1007/s10750-019-04018-x.

Kunzewicz, Z. W. *et al.* (2008) 'The implications of projected climate change for freshwater resources and their management', *Hydrological Sciences Journal*. Taylor & Francis, 53(1), pp. 3–10. doi: 10.1623/hysj.53.1.3.

Larned, S. T., Snelder, T. and Unwin, M. (2017) *Water quality in New Zealand rivers: Modelled water quality state*. Christchurch, New Zealand.

Leathwick, J. R. *et al.* (2010) *Freshwater Ecosystems of New Zealand (FENZ) Geodatabase* . Wellington: Department of Conservation.

Leflaive, J. *et al.* (2008) 'Nutrient effects on the genetic and functional diversity of aquatic bacterial communities', *FEMS Microbiology Ecology*, 66(2), pp. 379–390. doi: 10.1111/j.1574-6941.2008.00593.x.

Liao, H. *et al.* (1995) 'Relative Weight (Wr) as a Field Assessment Tool: Relationships with Growth, Prey Biomass, and Environmental Conditions', *Transactions of the American Fisheries Society*. John Wiley & Sons, Ltd, 124(3), pp. 387–400. doi: 10.1577/1548-8659(1995)124<0387:RWWRAA>2.3.CO;2.

Macreadie, P. I. *et al.* (2017) 'Can we manage coastal ecosystems to sequester more blue carbon?', *Frontiers in Ecology and the Environment*. John Wiley & Sons, Ltd, 15(4), pp. 206–213. doi: 10.1002/fee.1484.

Manning, D. W. P. *et al.* (2015) 'Detrital stoichiometry as a critical nexus for the effects of streamwater nutrients on leaf litter breakdown rates', *Ecology*. John Wiley & Sons, Ltd, 96(8), pp. 2214–2224. doi: 10.1890/14-1582.1.

Martin-Creuzburg, D., Beck, B. and Freese, H. M. (2011) 'Food quality of heterotrophic bacteria for Daphnia magna: evidence for a limitation by sterols', *FEMS Microbiology Ecology*, 76(3), pp. 592–601. doi: 10.1111/j.1574-6941.2011.01076.x.

Martin-Creuzburg, D. and Elert, E. von (2009) 'Ecological significance of sterols in aquatic food webs BT - Lipids in Aquatic Ecosystems', in Kainz, M., Brett, M. T., and Arts, M. T. (eds). New York, NY: Springer New York, pp. 43–64. doi: 10.1007/978-0-387-89366-2_3.

Matheson, F. *et al.* (2016) *Instream plant and nutrient guidelines: Review and development of an extended decision-making framework Phase 3*. Hamilton, New Zealand: National Institute of Water and Atmospheric Research.

McCarthy, S. D. S., Rafferty, S. P. and Frost, P. C. (2010) 'Responses of alkaline phosphatase activity to phosphorus stress in Daphnia magna', *The Journal of Experimental Biology*, 213(2), pp. 256 LP – 261. doi: 10.1242/jeb.037788.

McDowell, R. W. *et al.* (2013) 'Establishment of reference or baseline conditions of chemical indicators in New Zealand streams and rivers relative to present conditions', *Marine and Freshwater Research*, 64(5), pp. 387–400. doi: https://doi.org/10.1071/MF12153.

Ministry for the Environment and Statistics New Zealand (2015) *New Zealand's Environmental Reporting Series: Environment Aotearoa 2015*, *New Zealand's Environmental Reporting Series*. Wellington, New Zealand.

Ministry for the Environment and Statistics New Zealand (2019) *New Zealand's Environmental Reporting Series: Environment Aotearoa 2019*, *New Zealand's Environmental Reporting Series*. Wellington, New Zealand. Available at: www.mfe.govt.nz.

Moss, B. *et al.* (2011) 'Allied attack: climate change and eutrophication', *Inland Waters*. Taylor & Francis, 1(2), pp. 101–105. doi: 10.5268/IW-1.2.359.

Mougi, A. and Kondoh, M. (2016) 'Food-web complexity, meta-community complexity and community stability', *Scientific Reports*. The Author(s), 6, p. 24478. doi: 10.1038/srep24478http://www.nature.com/articles/srep24478#supplementary-information.

Mueller-Navarra, D. (1995) 'Evidence that a highly unsaturated fatty acid limits Daphnia growth in nature', *Archiv fur Hydrobiologie*. SCHWEIZERBART'SCHE VERLAGSBUCHHANDLUNG, 132, p. 297.

Nayak, S. K. (2010) 'Role of gastrointestinal microbiota in fish', *Aquaculture Research*. John Wiley & Sons, Ltd (10.1111), 41(11), pp. 1553–1573. doi: 10.1111/j.1365-2109.2010.02546.x.

Neutel, A.-M., Heesterbeek, J. A. P. and de Ruiter, P. C. (2002) 'Stability in real food webs: weak links in long loops', *Science*, 296(5570), pp. 1120–1123.

NIWA (2018) *A draft technical guide to the Periphyton Attribute Note Under the National Policy Statement for Freshwater Management 2014 (as amended 2017)*. Wellington, New Zealand.

O'Gorman, E. J., Fitch, J. E. and Crowe, T. P. (2012) 'Multiple anthropogenic stressors and the structural properties of food webs', *Ecology*. John Wiley & Sons, Ltd, 93(3), pp. 441–448. doi: 10.1890/11-0982.1.

Pascoal, C. *et al.* (2003) 'Assessing structural and functional ecosystem condition using leaf breakdown: studies on a polluted river', *Freshwater Biology*. John Wiley & Sons, Ltd (10.1111), 48(11), pp. 2033–2044. doi: 10.1046/j.1365-2427.2003.01130.x.

Patrício, J. and Marques, J. C. (2006) 'Mass balanced models of the food web in three areas along a gradient of eutrophication symptoms in the south arm of the Mondego estuary (Portugal)', *Ecological Modelling*, 197(1–2), pp. 21–34. doi: http://dx.doi.org/10.1016/j.ecolmodel.2006.03.008.

Persson, A. *et al.* (2001) 'Effects of Enrichment on Simple Aquatic Food Webs.', *The American Naturalist*. The University of Chicago Press, 157(6), pp. 654–669. doi: 10.1086/320620.

Persson, J. *et al.* (2010) 'To be or not to be what you eat: regulation of stoichiometric homeostasis among autotrophs and heterotrophs', *Oikos*. John Wiley & Sons, Ltd (10.1111), 119(5), pp. 741–751. doi: 10.1111/j.1600-0706.2009.18545.x.

Phillips, G. *et al.* (2018) 'Best practice for establishing nutrient concentrations to support good ecological status', in *Technical Report EUR 29329 EN*. Publications Office of the European Union Luxembourg, p. 142.

Pinckney, J. L. *et al.* (2001) 'The role of nutrient loading and eutrophication in estuarine ecology.', *Environmental Health Perspectives*. Environmental Health Perspectives, 109(suppl 5), pp. 699–706. doi: 10.1289/ehp.01109s5699.

Post, D. M. (2002) 'The long and short of food-chain length', *Trends in Ecology & Evolution*, 17(6), pp. 269–277. doi: http://dx.doi.org/10.1016/S0169-5347(02)02455-2.

R Core Team (2016) 'R: A language and environment for statistical computing.' Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.r-project.org/.

Rijal, B. (2018) 'Quantile regression: an alternative approach to modelling forest area burned by individual fires', *International Journal of Wildland Fire*, 27(8), pp. 538–549. Available at: https://doi.org/10.1071/WF17120.

Romero, J., Ringø, E. and Merrifield, D. L. (2014) 'The gut microbiota of fish', *Aquaculture nutrition: Gut health, probiotics and prebiotics*. Hoboken, NY: John Wiley and Sons, Ltd, pp. 75–100.

Rooney, N. and McCann, K. S. (2012) 'Integrating food web diversity, structure and stability', *Trends in Ecology & Evolution*, 27(1), pp. 40–46.

Rosemond, A. D. *et al.* (2015) 'Experimental nutrient additions accelerate terrestrial carbon loss from stream ecosystems', *Science*, 347(6226), pp. 1142 LP – 1145. doi: 10.1126/science.aaa1958.

Saint-Béat, B. *et al.* (2015) 'Trophic networks: How do theories link ecosystem structure and functioning to stability properties? A review', *Ecological Indicators*, 52, pp. 458–471. doi: http://dx.doi.org/10.1016/j.ecolind.2014.12.017.

Sanpera-Calbet, I. *et al.* (2017) 'Biochemical quality of basal resources in a forested stream: effects of nutrient enrichment', *Aquatic Sciences*, 79(1), pp. 99–112. doi: 10.1007/s00027-016-0482-3.

Scharf, F. S., Juanes, F. and Sutherland, M. (1998) 'Inferring ecological relationships from the edges of scatter diagrams: Comparison of regression techniques', *Ecology*. John Wiley & Sons, Ltd, 79(2), pp. 448–460. doi: 10.1890/0012-9658(1998)079[0448:IERFTE]2.0.CO;2.

Schindler, D. E. and Eby, L. A. (1997) 'Stoichiometry of fishes and their prey: Implications for nutrient recycling', *Ecology*. John Wiley & Sons, Ltd, 78(6), pp. 1816–1831. doi: 10.1890/0012-9658(1997)078[1816:SOFATP]2.0.CO;2.

Schmidt, T. S., Clements, W. H. and Cade, B. S. (2012) 'Estimating risks to aquatic life using quantile regression', *Freshwater Science*. [The University of Chicago Press, Society for Freshwater Science], 31(3), pp. 709–723. doi: 10.1899/11-133.1.

Shearer, K. A. and Hayes, J. (2019) *Assessing aquatic invertebrate prey indices for fish: trout. Prepared for Ministry for the Environment*. Nelson, New Zealand.

Singer, G. A. and Battin, T. J. (2007) 'Anthropogenic subsidies alter stream consumer-resource stoichiometry, biodiversity and food chains', *Ecological Applications*. John Wiley & Sons, Ltd, 17(2), pp. 376–389. doi: 10.1890/06-0229.

Smith, V. H., Tilman, G. D. and Nekola, J. C. (1999) 'Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems', *Environmental Pollution*, 100(1), pp. 179–196. doi: https://doi.org/10.1016/S0269-7491(99)00091-3.

Snelder, T., Biggs, B. J. and Weatherhead, M. (2010) *New Zealand river environment classification user guide*. Ministry for the Environment.

Snelder, T. H., Larned, S. T. and McDowell, R. W. (2018) 'Anthropogenic increases of catchment nitrogen and phosphorus loads in New Zealand', *New Zealand Journal of Marine and Freshwater Research*. Taylor & Francis, 52(3), pp. 336–361. doi:

10.1080/00288330.2017.1393758.

Stark, J. D. and Maxted, J. R. (2007) 'A user guide for the Macroinvertebrate Community Index', *Prepared for the Ministry for the Environment*, 58.

Steiner, C. F. *et al.* (2005) 'Temporal stability of aquatic food webs: partitioning the effects of species diversity, species composition and enrichment', *Ecology Letters*. John Wiley & Sons, Ltd (10.1111), 8(8), pp. 819–828. doi: 10.1111/j.1461-0248.2005.00785.x.

Stelzer, R. S. and Lamberti, G. A. (2002) 'Ecological stoichiometry in running waters: Periphyton chemical composition and snail growth', *Ecology*. John Wiley & Sons, Ltd, 83(4), pp. 1039–1051. doi: 10.1890/0012-9658(2002)083[1039:ESIRWP]2.0.CO;2.

Strobl, C. *et al.* (2008) 'Conditional variable importance for random forests', *BMC bioinformatics*. Springer, 9(1), p. 307.

Sullam, K. *et al.* (2012) 'Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis', *Molecular Ecology*. John Wiley & Sons, Ltd (10.1111), 21(13), pp. 3363–3378. doi: 10.1111/j.1365-294X.2012.05552.x.

Sutcliffe Jr., W. H. (1970) 'Relationship Between Growth Rate and Ribonucleic Acid Concentration in Some Invertebrates', *Journal of the Fisheries Research Board of Canada*. NRC Research Press, 27(3), pp. 606–609. doi: 10.1139/f70-065.

Tant, C. J. *et al.* (2015) 'Nutrient enrichment alters the magnitude and timing of fungal, bacterial, and detritivore contributions to litter breakdown', *Freshwater Science*. The University of Chicago Press, 34(4), pp. 1259–1271. doi: 10.1086/683255.

Tuomi, P. *et al.* (1995) 'Nutritional enrichment of a microbial community: The effects on activity, elemental composition, community structure and virus production', *FEMS Microbiology Ecology*, 16(2), pp. 123–134. doi: 10.1111/j.1574-6941.1995.tb00276.x.

Turner, A. M. and Chislock, M. F. (2010) 'Blinded by the stink: Nutrient enrichment impairs the perception of predation risk by freshwater snails', *Ecological Applications*. John Wiley & Sons, Ltd, 20(8), pp. 2089–2095. doi: 10.1890/10-0208.1.

Ulanowicz, R. E. (1997) *Ecology, The Ascendent Perspective*. Columbia University Press.

Ulanowicz, R. E. *et al.* (2009) 'Quantifying sustainability: Resilience, efficiency and the return of information theory', *Ecological Complexity*, 6(1), pp. 27–36. doi: http://dx.doi.org/10.1016/j.ecocom.2008.10.005.

Ulanowicz, R. E. (2019) 'Quantifying sustainable balance in ecosystem configurations', *Current Research in Environmental Sustainability*. doi: https://doi.org/10.1016/j.crsust.2019.09.001.

USEPA (2019) *Ecoregional Nutrient Criteria for Rivers and Streams*. Available at: https://www.epa.gov/nutrient-policy-data/ecoregional-nutrient-criteria-rivers-and-streams (Accessed: 2 February 2020).

Vinson, M. R. and Baker, M. A. (2008) 'Poor Growth of Rainbow Trout Fed New Zealand Mud Snails Potamopyrgus antipodarum', *North American Journal of Fisheries Management*. Taylor & Francis, 28(3), pp. 701–709. doi: 10.1577/M06-039.1.

Wacker, A. and Martin-Creuzburg, D. (2012) 'Biochemical nutrient requirements of the rotifer Brachionus calyciflorus: co-limitation by sterols and amino acids', *Functional Ecology*. John

Wiley & Sons, Ltd (10.1111), 26(5), pp. 1135–1143. doi: 10.1111/j.1365-2435.2012.02047.x.

Wagenhoff, A., Clapcott, J. E., *et al.* (2017) 'Identifying congruence in stream assemblage thresholds in response to nutrient and sediment gradients for limit setting', *Ecological Applications*, 27(2), pp. 469–484. doi: doi:10.1002/eap.1457.

Wagenhoff, A., Liess, A., *et al.* (2017) 'Thresholds in ecosystem structural and functional responses to agricultural stressors can inform limit setting in streams', *Freshwater Science*. The University of Chicago Press, 36(1), pp. 178–194. doi: 10.1086/690233.

Wardle, D. A., Bonner, K. I. and Barker, G. M. (2002) 'Linkages between plant litter decomposition, litter quality, and vegetation responses to herbivores', *Functional Ecology*. John Wiley & Sons, Ltd (10.1111), 16(5), pp. 585–595. doi: 10.1046/j.1365-2435.2002.00659.x.

Weber, N., Bouwes, N. and Jordan, C. E. (2014) 'Estimation of salmonid habitat growth potential through measurements of invertebrate food abundance and temperature', *Canadian Journal of Fisheries and Aquatic Sciences*. NRC Research Press, 71(8), pp. 1158–1170. doi: 10.1139/cjfas-2013-0390.

Whitehead, P. G. *et al.* (2009) 'A review of the potential impacts of climate change on surface water quality', *Hydrological Sciences Journal*. Taylor & Francis, 54(1), pp. 101–123. doi: 10.1623/hysj.54.1.101.

Winterbourn, M. J. (2000) 'Feeding ecology', in Collier, K. J. and Winterbourn, M. J. (eds) *New Zealand stream invertebrates: Ecology and implications for management*. Christchurch: The New Zealand Limnological Society, pp. 100–124.

Winterbourn, M. J., Gregson, K. L. D. and Dolphin, C. H. (1989) *Guide to the Aquatic Insects of New Zealand*. Entomological Society of New Zealand Auckland.

Winterbourn, M. J., Rounick, J. S. and Cowie, B. (1981) 'Are New Zealand stream ecosystems really different?', *New Zealand Journal of Marine and Freshwater Research*. Taylor & Francis, 15(3), pp. 321–328. doi: 10.1080/00288330.1981.9515927.

Wojewodzic, M. W. *et al.* (2011) 'Joint effect of phosphorus limitation and temperature on alkaline phosphatase activity and somatic growth in Daphnia magna', *Oecologia*, 165(4), pp. 837–846. doi: 10.1007/s00442-010-1863-2.

Woodward, G. *et al.* (2012) 'Continental-Scale Effects of Nutrient Pollution on Stream Ecosystem Functioning', *Science*, 336(6087), pp. 1438 LP – 1440. doi: 10.1126/science.1219534.

Young, R. G., Matthaei, C. D. and Townsend, C. R. (2008) 'Organic matter breakdown and ecosystem metabolism: functional indicators for assessing river ecosystem health', *Journal of the North American Benthological Society*. The Society for Freshwater Science, 27(3), pp. 605–625. doi: 10.1899/07-121.1.

Zhao, L. *et al.* (2016) 'Weighting and indirect effects identify keystone species in food webs', *Ecology Letters*, 19(9), pp. 1032–1040. doi: 10.1111/ele.12638.

Zimmerman, J. K. H. and Vondracek, B. (2006) 'Effects of stream enclosures on drifting invertebrates and fish growth', *Journal of the North American Benthological Society*. The University of Chicago Press, 25(2), pp. 453–464. doi: 10.1899/0887-3593(2006)25[453:EOSEOD]2.0.CO;2.

Appendix 7:

Summary of perspectives of a sub-group of STAG members who do not support Recommendation 13 of the primary report

Purpose

This paper sets out the key steps in the process that led a sub-group of STAG members - Bryce Cooper, Chris Daughney, Ian Hawes, Clive Howard-Williams and Jon Roygard - to conclude they could not support recommendation 13 of the primary report,

- explains the key concerns of those members, and
- suggests alternative ways to strengthen controls in the National Policy Statement for Freshwater Management (NPS-FM) relating to the effects of nutrients on ecosystem health.

Summary

The sub-group agrees that DIN and DRP affect ecosystem health and that the controls in the current NPS-FM on the effects of nutrients in rivers should be strengthened.

However, the sub-group has concluded that the evidence does not support the creation of a single set of nationally applicable bottom lines and thresholds for DIN and DRP (recommendation 13 of the primary STAG report).

The sub-group reached this conclusion because:

- the evidence provided to establish nationally applicable bands and bottom lines is insufficient to provide confidence that a given DIN or DRP concentration will achieve desired improvement in ecosystem health or ensure that the target of a specific ecosystem health metric will be met. The supplementary technical report (Appendix 6) on the development of DIN and DRP attributes accepts this point.

- they hold concerns about the reliability and effectiveness of nationally applied nutrient criteria in managing for ecosystem health, given they have been derived from weak relationships that vary spatially. This could have the effect of not triggering a management response in rivers where this is necessary to protect ecosystem health and vice versa.

The sub-group recommends that controls in the current NPS-FM on the effects of nutrients in rivers should be strengthened by:

- Giving effect to recommendation 8 of the primary report and strengthening the periphyton attribute in the current NPS-FM by providing a default nutrient table with spatially variable bottom lines and band thresholds, via guidance, as recommended in the primary report of the STAG. Sub-group members note that these default bottom lines are nearly all more stringent than those proposed in recommendation 13 and would be applicable to at least 72% of national river length.
- Increasing the level of protection from toxicity by making the current bottom of the ‘B band’ the national bottom line for ammonia and nitrate The current national bottom line provides for 80% species protection from chronic toxicity and the sub-group’s recommendation is to

raise this to 95% species protection from chronic toxicity which is more consistent with other ecosystem health protection measures recommended by the STAG.

- Introducing national monitoring requirements for DIN and DRP in rivers that, where increasing trends are detected in a freshwater management unit, would trigger the requirement to develop a management plan for reducing nutrient concentrations.

Process

In the lead up to finalising its primary report in June 2019, a group of members within the STAG raised concerns about proposals to introduce national bottom lines for DIN and DRP for rivers. Those members formed a sub-group and provided the STAG with a paper in June 2019 explaining why they considered the science was unresolved in this area and held concerns about introducing national bottom lines and management thresholds based on the information available.

The STAG discussed these concerns at length and, in its June 2019 report:

- included the overarching statement that '… these proposed changes will be subject to a public submission process. This process will bring public and practitioner experience to bear as well as enable the contribution of scientists employed in the various sectors of the economy impacted by our recommendations. While a public submission process is essential many of our recommendations are based on scientific judgements and should be subject to peer review.'[1]
- noted that 'almost all members supported the introduction of attribute limits for nitrogen and phosphorus for ecosystem health protection as outlined above [in the report]'[2]
- expressed an alternative approach for managing the impacts of nutrients on river health within the NPS-FM.[3]

Following the public release of a government discussion document accompanied by a series of draft national regulations, standards and policies a meeting of the STAG requested a supplemental technical report be prepared to clarify the processes taken in relation to the introduction of numeric biophysical tables for DIN and DRP to the National Objectives Framework (NOF) within the NPS-FM.

Officials asked the STAG to consider whether there was enough information and justification provided in the supplementary technical report to resolve questions and issues raised previously by STAG members and whether additional peer review was required.

An update on progress with the supplementary technical report, along with several graphs and tables, was presented to the STAG at its meeting on 22/23 January 2020. On 3 February, a draft of the supplementary technical report on the development of DIN and DRP attributes was circulated to the STAG for review and comments provided. At that time the subgroup who previously identified concerns developed a table detailing remaining concerns as well as additional commentary and shared that with the report author. A final version of the supplementary technical report (now Appendix 6) was provided to the STAG on 16 April 2020 and this subgroup have accordingly reviewed and modified their paper to form this document (now Appendix 7).

[1] p.12
[2] P.42
[3] ibid

Key concerns

The sub-group of STAG members who do not support Recommendation 13 of the primary report all agree that supply of nutrients can impact on the health of freshwater ecosystems and need to be managed.

The salient question for the sub-group is whether the controls currently in the NPS-FM are sufficient to manage those effects and, if not, what is the best approach to strengthening the NPS-FM to achieve that outcome?

Sub-group concerns pertaining to the effectiveness of nutrient attributes proposed in recommendation 13

Sub-group members agree that nutrients assimilated into the food web through primary production (or microbial processes) can pass through and potentially influence higher trophic levels. These relationships are indirect, however, and are influenced by so many other factors as to potentially negate the derivation of a single, nationally applicable, set of nutrient criteria that could be used reliably and effectively in a management framework.

It is of significant concern to sub-group members, having reviewed the draft of the supplementary technical report, that the national bottom lines and thresholds proposed for DIN and DRP have been derived based on weak relationships that vary substantially from river to river. Sub-group members note that STAG has recommended spatially variable bottom lines and thresholds based on river classes for other 'stressor' attributes (suspended sediment, deposited sediment, and nutrients for periphyton control).

Sub-group members are also concerned that the proposed DIN and DRP bottom lines are 'blunt tools' that will result in a significant number of 'unders and overs' – meaning that the levels of DIN and DRP may not trigger a management response in rivers where this is necessary to protect ecosystem health and vice versa.

Similarly, although not being philosophically opposed to the concept of introducing limits on DIN & DRP, the members of the sub-group are of the opinion that the available evidence does not show a high probability that reducing DIN or DRP to the suggested levels will lead to improvement in ecosystem health. The supplementary technical report (Appendix 6) on the development of DIN and DRP attributes accepts this point. Sub-group members ask: if the proposed national bottom lines for DIN and DRP will not necessarily achieve desired outcomes at the local scale, are they scientifically defensible, necessary and helpful for achieving ecosystem health outcomes?

Sub-group members also consider that any proposal to introduce bottom lines for DIN and DRP needs to be made in the context of existing and other proposed controls. In this regard, the additional default nutrient tables proposed by STAG for periphyton will cover most rivers and in most cases introduce more stringent requirements. Further, the macroinvertebrate and fish attributes proposed by STAG would give protection to these components of ecosystems in all rivers and would require nutrient management where this is impacting on bottom lines.

While the STAG has proposed a default nutrient table to manage periphyton, it is recommended that a coordinated national programme be instigated to speed up the availability of more robust, locally relevant, nutrient-periphyton relationships for use in freshwater-related policy and decision-making. Consideration should be given to extending such a programme to other ecosystem health metrics.

Suggested alternatives

While sub-group members do not believe there is sufficient evidence to justify the introduction of the proposed national bottom line and thresholds for DIN and DRP, the sub-group does agree that these nutrients can have effects on ecosystem function and aquatic health that do need to be managed locally and, that in some instances, these effects will not be captured by the existing and other proposed controls.

Sub-group members consider that DIN and DRP can be effective as lead indicators – in some instances levels of these nutrients may change more rapidly and be more readily identifiable than changes in other ecosystem health attributes. Accordingly, they could play a useful role in a monitoring and management framework, especially when trying to safeguard aquatic ecosystems from degradation. The sub-group therefore proposes that:

- Recommendation 13 should be replaced with a recommendation that would introduce mandatory national monitoring requirements for rivers that would trigger the requirement to assess and, if appropriate, develop a management plan for reducing nutrient concentrations. If ecosystem health metrics do not meet community aspirations or national bottom lines, then managers should undertake targeted investigations at a suitable scale to determine the cause(s). Guidance should be developed as to the conditions under which elevated nutrients may be influential on such ecosystems, and managers should then derive DIN and DRP reduction targets that are likely to achieve the desired states. Where nutrients are not influential, or where ecosystem health metrics already meet community aspirations, then managers should ensure that DIN and DRP are maintained at the current state (as per recommendation 3 from the primary STAG report) or reduced to concentrations consistent with protecting downstream ecosystems (as per the footnote the current periphyton attribute in the NOF), whichever is the most stringent.

Sub-group members believe that it is necessary to keep a tight focus on what one is trying to achieve when considering the introduction of new management categories and metrics into the NPS-FM.

If the objective is to strengthen the management of the effects of nutrients on ecosystem health the sub-group are of the view that there are ways to achieve this that are more locally relevant and better supported by the currently available data, compared to what is currently proposed in recommendation 13. There are two ways this can be approached:

- Give effect to recommendation 8 from the STAG's primary report, which is to strengthen the periphyton attribute in the current NPS by providing a default nutrient table with spatially variable bottom lines and band thresholds (via national guidance as recommended in the primary report of the STAG). It is noted that these default bottom lines are nearly all more stringent than those proposed in recommendation 13 and would be applicable to *at least* 72% of national river length.
- While the STAG has proposed a default nutrient table for the periphyton attribute, it is recommended that a coordinated national programme be instigated to speed up the availability of more robust, locally relevant, nutrient-periphyton relationships for use in freshwater-related policy and decision-making. Consideration should be given to extending such a programme to other ecosystem health metrics in addition to periphyton.

- Address the fact that the toxicity bottom lines for nitrate and ammonia in the current NPS-FM can be inconsistent with healthy ecosystem outcomes and need to be strengthened. These bottom lines have been set to protect 80% of species from chronic toxicity effects. This sub-group considers that raising the toxicity bottom lines to the current B band thresholds (95% protection from chronic toxicity effects) would be more consistent with ensuring healthy ecosystems.

STAG members' final comments in response to the perspectives of the sub-group of STAG Appendix 7

We, the majority of the STAG membership who do support Recommendation 13 of the primary report, believe that one final comment in response to the arguments presented in Appendix 7, by the minority sub-group who do not support the recommendation, may help to clarify the situation for those unfamiliar with the deliberations and nature of this technical advisory group.

It is important to recognise the scientific consensus among members of the STAG in this area – all members agree that:

- *Elevated DIN and DRP concentrations adversely affect ecosystem health,*
- *The controls in the current NPS-FM on the effects of these nutrients in rivers on ecosystem health are insufficient and should be strengthened.*

Disagreement between members in this area is confined to one key matter – whether or not there is enough evidence currently to support the introduction of nationally applicable bottom lines and thresholds for DIN and DRP.

The minority sub-group believe more evidence is required before introducing nationally applicable bottom lines and thresholds for DIN and DRP into the National Objectives Framework. In the interim they favour limiting nutrient concentrations via a default nutrient limits table as part of the periphyton attributes and increasing the stringency of the current nitrate toxicity attribute.

We, the majority, do recognise the value of such a table in the periphyton attribute, but there will still be water environments that elude nutrient control. We believe that there is sufficient evidence available ***now,*** *as summarised in Appendix 6, to support the introduction of nationally applicable bottom lines and thresholds for DIN and DRP. We are mindful that successive state of the environment reports produced by the Ministry for the Environment, including the Our Freshwater 2020 report released in April, have concluded that water quality in New Zealand's rivers continues to degrade, threats to New Zealand's freshwater fish and ecosystems continue to grow and the health of these ecosystems continues to decline. We believe we cannot wait for every residual uncertainty in the evidence to be resolved before taking action. We note that there will be additional data generation and analysis as councils and other environmental regulators implement the Essential Freshwater actions, and therefore future opportunities for further refinement of national limits for DIN and DRP if warranted.*

We are very uncomfortable with the use of nitrate toxicity data (which is poor for New Zealand ecosystems and does not yield a relatable phosphate limit), as a basis for nutrient limits. As we understand it, this would make New Zealand the only country to try to manage the effects of nutrients on ecosystem health based on nitrate toxicity. However, the fact that the nitrate limits being proposed in Recommendation 13 are generally consistent with both ecosystem and emerging human health toxicity thresholds (including evidence for links between nitrate in drinking water and colo-rectal cancer), further increases our confidence in the value of these proposed attribute limits.