NIWA
Taihoro Nukurangi

Deriving potential fine sediment attribute thresholds for the National Objectives Framework

Prepared for Ministry for the Environment

June 2019

Prepared by:
Paul Franklin
Rick Stoffels
Joanne Clapcott (Cawthron Institute)
Doug Booker
Annika Wagenhoff (Cawthron Institute)
Chris Hickey

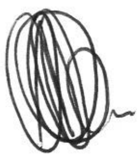
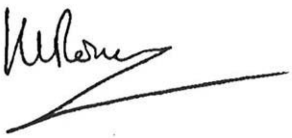
For any information regarding this report please contact:

Paul Franklin
Scientist
Freshwater Ecology
+64-7-859 1882
paul.franklin@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 11115
Hamilton 3251

Phone +64 7 856 7026

NIWA CLIENT REPORT No:	2019039HN
Report date:	June 2019
NIWA Project:	MFE19202

Quality Assurance Statement		
	Reviewed by:	Neale Hudson
	Formatting checked by:	Aarti Wadhwa
	Approved for release by:	Helen Rouse

Contents

Tables

Figures

Executive summary

This report sets out our framework for drawing together the various workstreams focused on characterising the relationships between fine sediment environmental state variables (ESVs) and indicators of ecosystem health. Based on the outcomes of these workstreams, we have derived numeric thresholds that could potentially form the basis of fine sediment attributes in the National Objectives Framework (NOF).

NOF attributes are measurable characteristics of fresh water that support particular values. They assist with the definition of numeric freshwater objectives and the development of associated limits and management actions. Ecosystem health is designated as a compulsory national value in the National Policy Statement for Freshwater Management (NPS-FM). Ecosystem health is a broad term generally used to describe the condition of an ecosystem. The NPS-FM states that in a healthy freshwater ecosystem "ecological processes are maintained, there is a range and diversity of indigenous flora and fauna, and there is resilience to change." For this project we have interpreted this to imply that a healthy ecosystem is one where both ecosystem structure and function are similar to that expected under minimally disturbed conditions.

Greater fine sediment (generally defined as organic and inorganic particles <2 mm in diameter) inputs to a river system are observed as increases in the suspended solids load (i.e., suspended sediment) and/or accumulation of fine sediments in/on the river bed (i.e., deposited sediment). Elevated fine sediment is widely documented as having an impact on ecosystem health. Consequently, there is a strong basis for the incorporation of fine sediment attributes in the NPS-FM.

Defining numeric thresholds for fine sediment attributes is complicated by numerous factors including:

- variation in sediment state across sites[1]
- variation in sediment state at a site over time[1]
- multiple modes of impact[2], and
- variation in effects across different species and life stages[2].

Our ability to characterise these variations and modes of impact are also challenged by the existence of:

- multiple potential fine sediment ESVs (e.g., suspended sediment, deposited sediment – with each variously composed of particulate organic or inorganic matter)[3]
- different ways of measuring individual ESVs (e.g., turbidity, total suspended solids, visual clarity)[3]
- multiple potential indicators of ecosystem health (e.g., fish, macroinvertebrates) and ways to measure them (e.g., presence/absence, absolute abundance, relative abundance, community composition)[3], and

[1] See Section 2.2.1
[2] See Section 2.2.2
[3] See Section 2.2.3 and Appendix M to Appendix O

- variations in the spatial and temporal resolution and overlap of available data on both sediment ESVs and ecological response variables[3].

The approach we took to deriving potential fine sediment attributes involved the following steps:

1. We used existing literature to characterise hypothesised interaction pathways between fine sediment stressors and ecosystem health indicators[4].
2. We identified and reviewed available data sets for their suitability to:
   - characterise natural spatial and temporal variations in sediment ESVs across New Zealand, and
   - quantify interaction pathways between sediment ESVs and ecosystem health indicators.
3. We developed a classification of natural spatial patterns in sediment ESV state[5].
4. We used a range of analytical methods to characterise quantitative relationships between sediment ESVs and ecosystem health indicators[6].
5. We combined the multiple lines of evidence to make recommendations on potential sediment ESV attribute thresholds[7].

In deriving potential fine sediment attribute thresholds we adhered to a number of guiding principles including basing 'bottom lines'[8] on the least acceptable state for ecosystem health, avoiding potentially significant adverse ecosystem effects, and accounting for spatial patterns in both ecological distributions and natural sediment state.

We chose to adopt a data-driven approach to deriving numerical criteria. This has the benefit of providing transparency and reproducibility in the derivation of potential attribute thresholds, but it also means we were constrained by the nature, extent, consistency and resolution of existing data. We used a formal weight of evidence assessment to score different lines of evidence based on their relevance, reliability and suitability for defining numeric criteria that are consistent with the guiding principles.

We defined potential fine sediment attributes for three separate fine sediment ESVs: % cover of deposited fine sediment; turbidity; and visual clarity (proposed attribute tables are presented at the end of this Executive Summary; Table 1-1 to Table 1-3). The attribute thresholds for deposited fine sediment were based on the response of macroinvertebrate communities, whereas the attribute thresholds for the suspended sediment ESVs (turbidity and visual clarity) were based on the response of fish communities. This reflects the respective sensitivity of these communities to the deposited and suspended sediment ESVs.

To account for natural patterns in sediment state across New Zealand, we have proposed attribute thresholds that are defined across classes in a sediment state classification (SSC). Separate SSCs have been derived for deposited sediment (Figure 1-1) and suspended sediment (Figure 1-2). The SSCs we

[4] See Appendix A to Appendix C
[5] See Section 3 and Appendix D
[6] See Section 4 and Appendix F to Appendix K
[7] See Sections 5 and 6 and Appendix L
[8] The minimum acceptable state for attributes defined in the NPS-FM (MfE 2017)

created are hierarchical, allowing aggregation of classes at different levels of the SSCs. Our analyses indicate that the level of aggregation has a significant influence on the bias of thresholds. Effectively, the thresholds result in more variable outcomes as the level of aggregation increases (i.e., number of classes decreases). This occurs because the proposed thresholds are derived based on the ecological effects of increased fine sediment levels as they depart from reference state (i.e., minimally disturbed compared to natural condition). When classes are aggregated their predicted reference states are effectively 'averaged'. This results in larger 'unders' and 'overs' within a class where the natural reference state of a reach is under or over the average reference state for the whole class. Because the same thresholds apply within a class, increasing the level of aggregation increases the number of reaches that risk having under protective ('unders') or over protective ('overs') thresholds. Consequently, we have recommended defining thresholds at the lowest level of aggregation (12 classes).

Our analyses focused on characterising community level responses in macroinvertebrates and fish to increasing deposited fine sediment. Community level responses are complex because they integrate the diverse responses of multiple species. However, in our view basing thresholds for ecosystem health on a community level response was more consistent with the guiding principles than using the sensitivity of individual sentinel species (i.e., a single species that may be particularly sensitive to a stressor of interest).

Because there are species and life-stage specific differences in sensitivity to elevated fine sediment we recognise that the proposed thresholds may not provide adequate protection for individual valued species or during critical life-stages. A good example is the high sensitivity of salmonid spawning habitats to elevated fine sediment cover. It is our view that where such sensitive species or life-stages are identified as being of value, then objectives can be set to achieve a higher attribute band, or value-specific attribute thresholds can be defined.

Sediment delivery and transport processes are highly dynamic and, therefore, ecosystem stressor exposure is a function of long-term averages, as well as the impact of short-term events. However, the data currently available dictate that only the impacts of long-term average conditions can be considered. Consequently, any thresholds we have derived can only be considered protective of changes to the long-term average condition. To establish the impacts of shorter-term sediment dynamics on ecosystem health, there is a need to collect data on both sediment ESVs and ecological response variables at a greater temporal resolution. The spatial coverage of existing data also limited our ability to effectively account for spatial variations in environmental gradients in some cases.

We have developed potential attributes for two different suspended sediment ESVs, namely turbidity and visual clarity. Turbidity and visual clarity are typically strongly correlated, but for the purposes of this project the turbidity and visual clarity attribute thresholds were derived independently using currently available data. Comparison of the results show that for some classes the thresholds for turbidity and visual clarity are not numerically similar when using published equations for converting between the two ESVs. This is likely the result of fewer data being available for visual clarity to derive exposure-response relationships. If there is a preference to implement a suspended sediment attribute using visual clarity as the ESV, some consideration should be given as to whether the data derived thresholds should be used, or whether the turbidity thresholds (which were derived using more data and therefore, should be more robust) should be used to derive a visual clarity attribute by applying a conversion factor.

There are mechanistic and pragmatic arguments that support the use of visual clarity rather than turbidity as the preferred suspended sediment attribute. Turbidity is considered a good proxy variable for several sediment-related variables (including visual clarity), but evidence showing that turbidity measurements are instrument dependent has led to recommendations that the use of nephelometric turbidity units (NTU) as an absolute quantity should be abandoned. Turbidity units (NTU/FNU) are relative and not standardised SI units whereas visual clarity (m) is – this means that evaluation of compliance is more robust for visual clarity. However, presently, turbidity is monitored routinely by all councils, whereas visual clarity is not. Consequently, more turbidity than visual clarity data are available, meaning that we were able to better characterise ecological responses to turbidity (ignoring potentially significant issues with the comparability of data collected using different instruments) than visual clarity. It is also currently more cost effective to monitor turbidity continuously and across a larger range than it is to measure visual clarity continuously (using a beam transmissometer). A number of councils are currently conducting continuous measurement of turbidity at a limited number of sites for sediment load calculation. These data could potentially be used for evaluating compliance with the proposed limits, but sensors are typically calibrated to higher sediment concentrations resulting in greater measurement uncertainty at concentrations in the range of the proposed limits.

We have not explicitly calculated uncertainty in our analyses. As far as practicable we have, however, indicated where uncertainties exist. Some normative decisions were integral to the determination of the proposed thresholds, particularly the magnitude of acceptable deviation from reference in the different community metrics. Ideally these results would be validated with independent data, but limitations on available data prevented this. We have endeavoured to be transparent about where normative decisions have been made, and where possible have demonstrated how thresholds would vary if those normative decisions were different.

Throughout these analyses the interactions between ESVs have not been explicitly considered. It is possible that management actions directed toward one ESV, will also influence the state of the other ESVs. However, it was outside the scope of this project to determine how the different ESV measures correlate with different management actions and how their responses may be interrelated. We have also not considered any interactions with other stressors (e.g., temperature, flows or nutrients).

Table 1-1: Potential attribute band thresholds for deposited fine sediment cover. Thresholds are defined for the 12 classes at Level 4 of the deposited sediment state classification.

Value	Ecosystem Health												
Freshwater Body Type	Rivers												
Attribute	Deposited fine sediment												
Attribute Unit	% fine sediment cover (percentage cover of the streambed in a run habitat determined by the instream visual method, SAM2)												
Attribute State	**SSC class[1]**												**Narrative Attribute State**
	1	2	3	4	5	6	7	8	9	10	11	12	
	Site median[2]												
A	<84	<9	<42	<12	<80	<30	<41	<22	<48	<15	<76	<27	Minimal impact of deposited fine sediment on instream biota. Ecological communities are similar to those observed in natural reference conditions.
B	<90	<15	<50	<17	<86	<38	<48	<33	<54	<22	<82	<36	Low to moderate impact of deposited fine sediment on instream biota. Abundance of sensitive macroinvertebrate species may be reduced.
C	≤97	≤21	≤60	≤23	≤92	≤46	≤56	≤45	≤61	≤29	≤89	≤45	Moderate to high impact of deposited fine sediment on instream biota. Sensitive macroinvertebrate species may be lost.
National Bottom Line[3]	97	21	60	23	92	46	56	45	61	29	89	45	
D	>97	>21	>60	>23	>92	>46	>56	>45	>61	>29	>89	>45	High impact of deposited fine sediment on instream biota. Ecological communities are significantly altered and sensitive fish and macroinvertebrate species are lost or at high risk of being lost.

[1] Classes are streams and rivers defined according to the fourth level of aggregation (L4) of the deposited sediment State Classification (SSC).

[2] The minimum record length for grading a site based on an instream visual assessment of % fine sediment cover (SAM2) is 2 years based on a monthly monitoring regime.

[3] Bottom-line thresholds are anticipated to provide a sufficient level of protection at an overall macroinvertebrate community level (i.e., will cause <20% decrease in the *macroinvertebrate community deviation metric*). Bottom-line thresholds may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations for certain biota. For example, salmonid spawning habitats may require sediment cover of <10%. Fine sediments with high organic enrichment may also result in higher levels of impacts on macroinvertebrate communities or sensitive fish life-stages.

Table 1-2: Potential attribute band thresholds for turbidity. Thresholds are defined for the 12 classes at Level 4 of the suspended sediment state classification.

Value	Ecosystem Health												
Freshwater Body Type	Rivers												
Attribute	Suspended fine sediment												
Attribute Unit	Turbidity (NTU/FNU)												
Attribute State	**SSC class[1]**												**Narrative Attribute State**
	1	2	3	4	5	6	7	8	9	10	11	12	
	Site median[2]												
A	<2.0	<6.2	<1.3	<3.3	<7.5	<4.8	<2.3	<4.3	<1.2	<1.1	<1.1	<2.4	Minimal impact of suspended sediment on instream biota. Ecological communities are similar to those observed in natural reference conditions.
B	<2.5	<7.9	<1.6	<3.9	<9.8	<6.3	<2.8	<5.2	<1.4	<1.3	<1.3	<2.7	Low to moderate impact of suspended sediment on instream biota. Abundance of sensitive fish species may be reduced.
C	≤3.2	≤10.5	≤2.0	≤4.8	≤13.1	≤8.3	≤3.3	≤6.4	≤1.6	≤1.5	≤1.6	≤3.1	Moderate to high impact of suspended sediment on instream biota. Sensitive fish species may be lost.
National Bottom Line[3]	3.2	10.5	2.0	4.8	13.1	8.3	3.3	6.4	1.6	1.5	1.6	3.1	
D	>3.2	>10.5	>2.0	>4.8	>13.1	>8.3	>3.3	>6.4	>1.6	>1.5	>1.6	>3.1	High impact of suspended sediment on instream biota. Ecological communities are significantly altered and sensitive fish and macroinvertebrate species are lost or at high risk of being lost.

[1] Classes are streams and rivers defined according to the fourth level of aggregation (L4) of the suspended sediment State Classification (SSC).

[2] The minimum record length for grading a site is 2 years of monthly samples. Continuous turbidity[9] data may be used to calculate 2-year median turbidity.

[3] Bottom-line thresholds are anticipated to provide a sufficient level of protection at an overall fish community level (i.e., will cause <20% decrease in the *fish community deviation metric*). Bottom-line thresholds may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations for certain biota.

[9] Turbidity sensors should be calibrated to the range of interest defined by the attribute band thresholds for the relevant SSC class

Table 1-3: Potential attribute band thresholds for visual clarity. Thresholds are defined for the 12 classes at Level 4 of the suspended sediment state classification.

<table>
<tr><td>Value</td><td colspan="13">Ecosystem Health</td></tr>
<tr><td>Freshwater Body Type</td><td colspan="13">Rivers</td></tr>
<tr><td>Attribute</td><td colspan="13">Suspended fine sediment</td></tr>
<tr><td>Attribute Unit</td><td colspan="13">Visual clarity (m)</td></tr>
<tr><td rowspan="3">Attribute State</td><td colspan="12">SSC class[1]</td><td rowspan="3">Narrative Attribute State</td></tr>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td></tr>
<tr><td colspan="12">Site median[2]</td></tr>
<tr><td>A</td><td>>2.25</td><td>>2.43</td><td>>1.45</td><td>>1.43</td><td>>0.66</td><td>>1.06</td><td>>1.78</td><td>>0.63</td><td>>3.10</td><td>>3.38</td><td>>2.84</td><td>>2.79</td><td>Minimal impact of suspended sediment on instream biota. Ecological communities are similar to those observed in natural reference conditions.</td></tr>
<tr><td>B</td><td>>1.88</td><td>>2.02</td><td>>1.21</td><td>>1.22</td><td>>0.53</td><td>>0.87</td><td>>1.53</td><td>>0.53</td><td>>2.71</td><td>>2.93</td><td>>2.43</td><td>>2.51</td><td>Low to moderate impact of suspended sediment on instream biota. Abundance of sensitive fish species may be reduced.</td></tr>
<tr><td>C</td><td>≥1.55</td><td>≥1.65</td><td>≥1.00</td><td>≥1.02</td><td>≥0.42</td><td>≥0.70</td><td>≥1.30</td><td>≥0.44</td><td>≥2.35</td><td>≥2.51</td><td>≥2.06</td><td>≥2.23</td><td>Moderate to high impact of suspended sediment on instream biota. Sensitive fish species may be lost.</td></tr>
<tr><td>National Bottom Line[3]</td><td>1.55</td><td>1.65</td><td>1.00</td><td>1.02</td><td>0.42</td><td>0.70</td><td>1.30</td><td>0.44</td><td>2.35</td><td>2.51</td><td>2.06</td><td>2.23</td><td></td></tr>
<tr><td>D</td><td><1.55</td><td><1.65</td><td><1.00</td><td><1.02</td><td><0.42</td><td><0.70</td><td><1.30</td><td><0.44</td><td><2.35</td><td><2.51</td><td><2.06</td><td><2.23</td><td>High impact of suspended sediment on instream biota. Ecological communities are significantly altered and sensitive fish and macroinvertebrate species are lost or at high risk of being lost.</td></tr>
<tr><td colspan="14">[1] Classes are streams and rivers defined according to the fourth level of aggregation (L4) of the suspended sediment State Classification (SSC).
[2] The minimum record length for grading a site is 2 years of monthly samples. Continuous visual clarity data may be used to calculate a 2-year median visual clarity.
[3] Bottom-line thresholds are anticipated to provide a sufficient level of protection at an overall fish community level (i.e., will cause <20% decrease in the fish community deviation metric). Bottom-line thresholds may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations for certain biota.</td></tr>
</table>

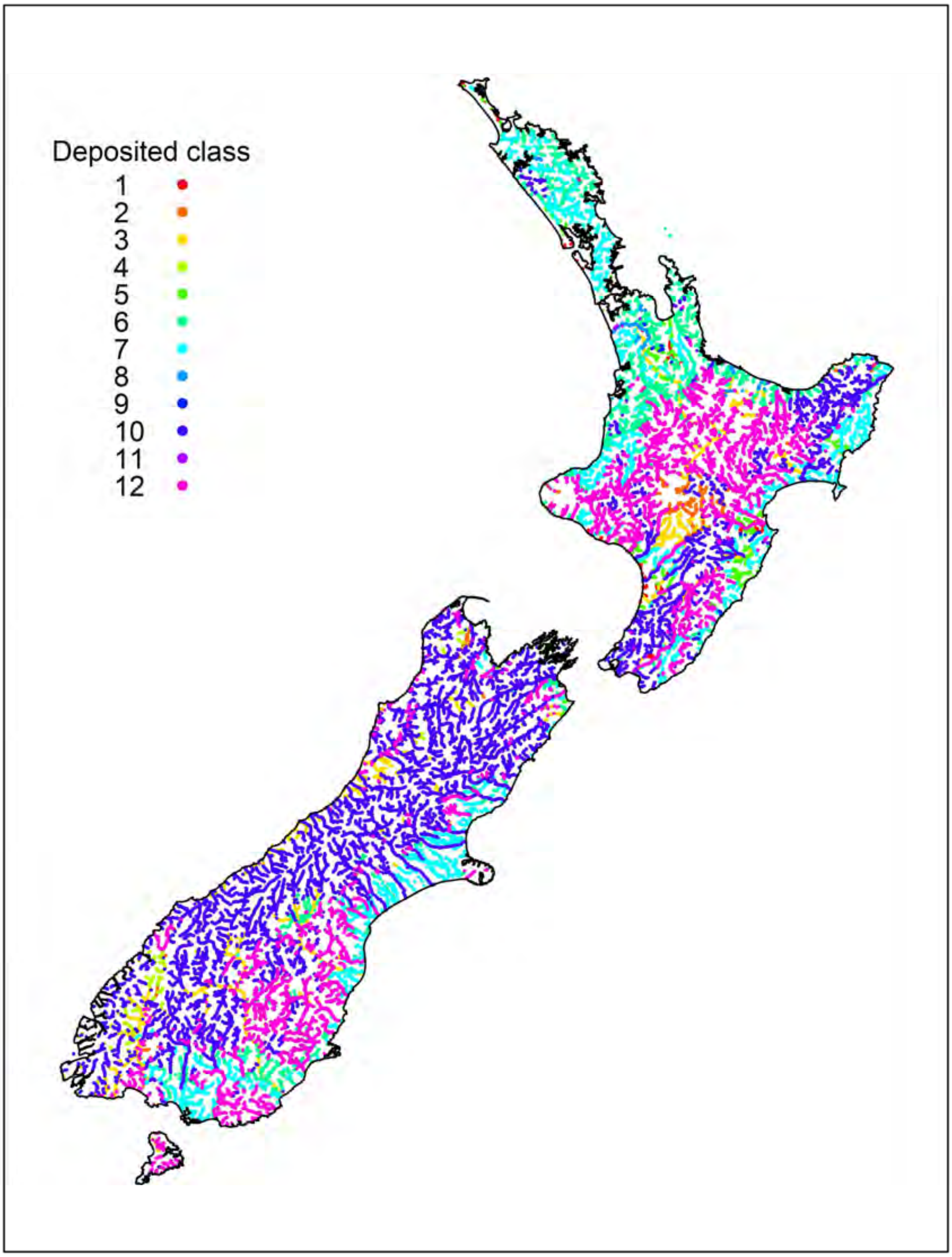


Figure 1-1: Map of the river network showing the 12 class SSC for deposited sediment. The map shows all streams order 4 and greater.

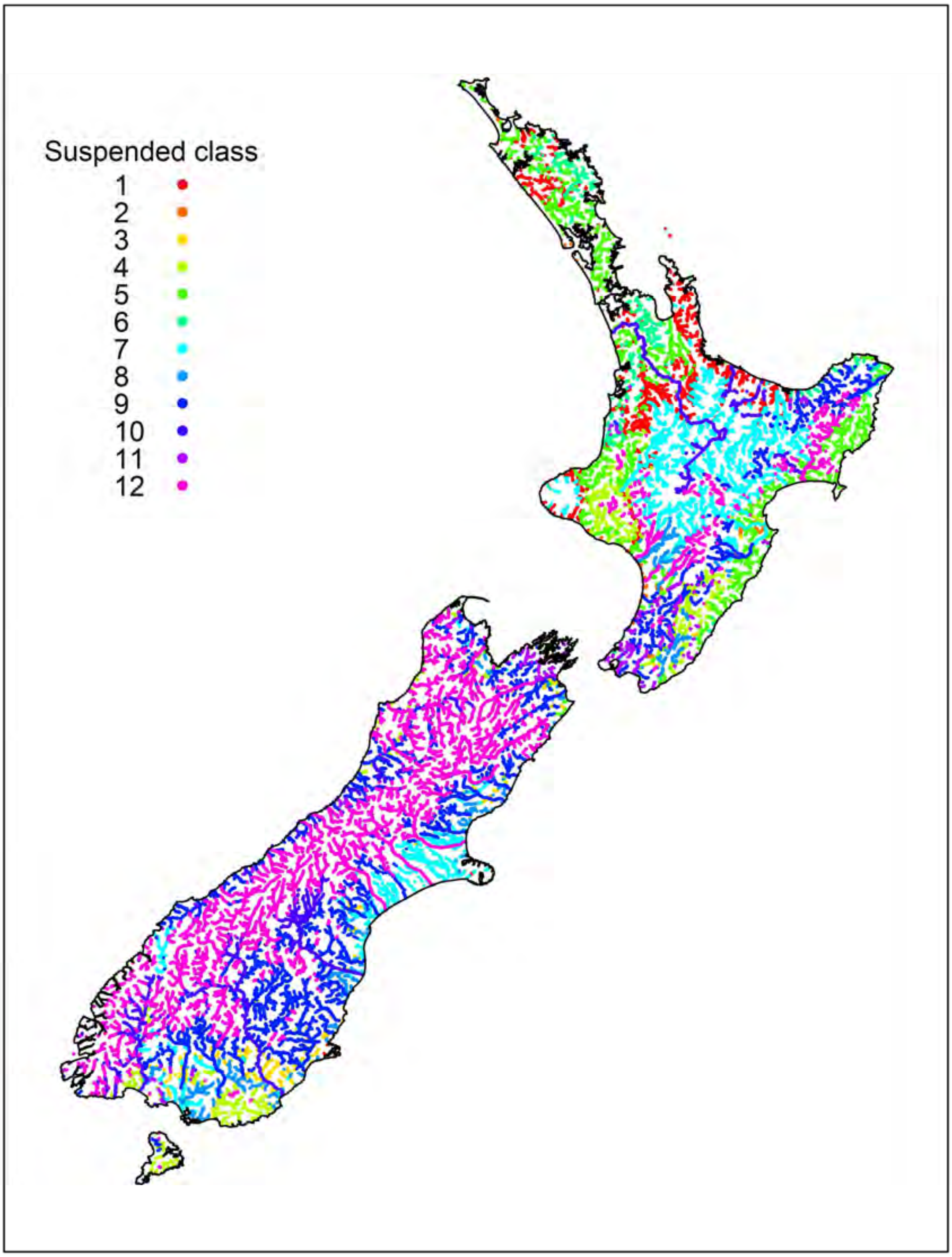


Figure 1-2: Map of the river network showing the 12 class SSC for suspended sediment. The map shows all streams order 4 and greater.

1 Introduction

1.1 Background

The National Policy Statement for Freshwater Management (NPS-FM) sets out requirements for regional councils, in consultation with their communities, to set objectives for the state of freshwater bodies, and to set limits on resource use to meet these objectives. A key component of the NPS-FM is the implementation of a National Objectives Framework (NOF), which provides an approach for establishing freshwater objectives for national (and other) values and recognises that certain physical, chemical and biological properties of fresh water environments (attributes) must be maintained or improved to sustain particular values.

Attributes are measurable characteristics of fresh water that support particular values. Appendix 2 of the NPS-FM (MfE 2017) sets out attribute tables for the compulsory values (ecosystem health and human health for recreation). Attribute tables define the attribute and set out numeric thresholds that define attribute state, including a national bottom line that specifies the minimum acceptable state required to sustain compulsory values.

Increased fine sediment inputs are widely acknowledged to impact negatively on freshwater ecosystems and a range of causal mechanisms are known to underpin the exhibited responses (Ryan 1991; Cantilli et al. 2006). However, the NOF does not currently define attributes for sediment despite the importance of this contaminant in freshwaters in New Zealand.

The Ministry for the Environment (MfE) recognises the important influence that fine sediment can have on the compulsory value of ecosystem health and seeks to establish attributes for both suspended and deposited fine sediment environmental state variables (ESVs) in freshwater bodies. A range of complimentary workstreams have been developed to inform the establishment of fine sediment attributes in the NOF. This has included a literature review summarising fine sediment effects on freshwaters (Davies-Colley et al. 2015), and work to understand links between catchment sediment loads and fine sediment ESVs (e.g. Hicks et al. 2016), to evaluate sediment ESV data availability (Depree et al. 2016), to investigate national sediment classification systems (Clapcott and Goodwin 2017; Depree 2017), and to characterise relationships between fine sediment ESVs and ecological responses (Depree et al. 2018).

This report refines, extends, and sets out a framework for drawing together the workstreams that have attempted to characterise the relationships between fine sediment ESVs and ecological response variables. Based on the outcomes of these workstreams, this report also defines potential numeric thresholds that could form the basis of fine sediment attributes in the NOF. The results and thresholds laid out in this report are the full and final version and take precedence over the information presented by Depree et al. (2018).

1.2 Scope

The scope of this project includes responding to the technical review comments on the draft report developed under contract 21511 (Depree et al. 2018), and addressing critical gaps identified by the project team in order to implement workable and defensible fine sediment attributes, including defining attribute bands. Specifically, this includes:

- resolving the use of different ESV measures across reference state models and analyses of ESV-ecological response relationships, and how these differences are accounted for in the establishment of attribute tables
- assessing alternative reference state models and potential consequences for establishing attribute tables
- addressing equivalence between analytical methods across different ESVs and ecological response variables
- evaluating and implementing analytical methods suitable for supporting definition of attribute bands, and
- instigating a more formal and transparent weight of evidence approach for combining the multiple lines of evidence to define thresholds for attribute states.

It is not within the scope of this project to provide detailed analyses of potential spatial and temporal uncertainties in the models or thresholds derived from the models. This project does not evaluate potential compliance with or management actions required to meet any proposed attribute thresholds arising from the project. It is also outside the scope of this project to consider the socio-economic impacts of implementing any proposed thresholds.

1.3 Structure of this report

This report lays out the framework and guiding principles that we have used to determine potential numerical thresholds for sediment attributes to protect ecosystem health. We first discuss the general framework for defining NOF attributes and some of the challenges associated with deriving numeric thresholds for fine sediment (Section 2). We then describe the sediment state classification (SSC) system that we have adopted for the derivation of reference ESV states (Section 3), which replaces Chapter 2 of Depree et al. (2018). Section 4 describes the analytical methods that are the basis of our data-driven approach to characterising stressor-response relationships and form the basis of deriving numerical thresholds. The analyses described here largely supersede those reported in Chapters 4, 5 and 6 in Depree et al. (2018). In Section 5 of this report we describe the weight of evidence approach we have adopted for evaluating and comparing the different lines of evidence that have been developed throughout this project. In combination with Section 6 of this report, which describes potential attribute tables for deposited sediment, turbidity and visual clarity ESVs, this replaces Chapter 7 of Depree et al. (2018). In Section 7, we discuss some of the outstanding issues that we have identified during the process of developing the proposed sediment attributes. Appendix A to Appendix O provide more detailed descriptions of the technical analyses and supporting information associated with development of the attributes.

2 How are NOF attributes defined?

2.1 General framework

Attributes are measurable characteristics of fresh water that support particular values. They assist with the definition of numeric freshwater objectives and the development of associated limits and management actions. Five broad principles have been set out for developing NOF attributes:

1. Link to the national value
   - Is the attribute required to support the value?
   - Does the attribute represent the value?
2. Measurement and band thresholds
   - Are there established protocols for measurement of the attribute?
   - Do experts agree on the summary statistic and associated time period?
   - Do experts agree on thresholds for the numerical bands and associated band descriptors?
3. Relationship to limits and management
   - Do we know what to do to manage this attribute?
   - Do we understand the drivers associated with the attribute?
   - Do quantitative relationships link the attribute state to resource use limits and/or management interventions?
4. Evaluation of current state of the attribute on a national scale
   - Can we adequately assess the current state of the attribute at a national scale, including the extent, magnitude and location of failures to meet the proposed bottom line for the attribute?
   - Are the data of sufficient quality, quantity, and representativeness to assess the current state of the attribute on a national scale?
5. Implications of including the attribute in the NOF
   - What are the socio-economic impacts of implementing the attribute at a national scale?

This project is concerned with addressing principles 1 and 2. Consideration has, therefore, been given to characterising relationships between sediment environmental state variables (ESVs) and characteristics of ecosystem health, and identifying appropriate ESV metrics and thresholds for numeric bands. Depree et al. (2018) reported the initial steps taken towards achieving these objectives. This report will summarise some of the key outcomes from that report and provide updates based on subsequent work and analyses completed in this phase of the project. Furthermore, it defines final attribute thresholds. All results and attribute thresholds reported here take precedence over those in Depree et al. (2018).

2.2 Challenges for defining numeric thresholds for a sediment attribute

Defining numeric thresholds for fine sediment ESVs is complicated by numerous factors including:

- variation in state across sites
- variation at a site over time
- multiple modes of impact, and
- variation in effects across different species and life-stages.

Our ability to characterise these variations and modes of impact are also challenged by the existence of:

- multiple potential fine sediment ESVs (e.g., suspended sediment, deposited sediment – with each comprising particulate organic or inorganic matter in varying proportions)
- different ways of measuring individual ESVs (e.g., turbidity, total suspended solids, visual clarity)
- multiple potential indicators of ecosystem health (e.g., fish, macroinvertebrates), and ways to measure them (e.g., presence/absence, absolute abundance, relative abundance, community composition), and
- variations in the spatial and temporal resolution and overlap of available data on both sediment ESVs and ecological response variables.

2.2.1 Variations in sediment state over space and time

Variation in natural environmental characteristics (e.g., geology, slope and rainfall) result in natural differences in the availability of fine sediments and vulnerability of the landscape to sediment loss. They also determine natural rates of sediment transport and deposition within rivers and streams. For example, areas with hard geology tend to have a lower supply of fine sediments than areas dominated by soft geology, and deposition of fine sediments tends to be higher in streams with a shallower slope compared to those with steeper slopes. Consequently, the state of both deposited and suspended fine sediments naturally varies across the country.

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.

Sediment delivery, transport and deposition also naturally vary over time at a site in a river. This results in the aquatic ecosystem being subject to different magnitudes, durations and frequencies of sediment exposure at different times at a site. Because multiple modes of impact exist, different aspects of ecosystem health may be impacted differently by the temporal variations in sediment state at a site. For example, acute impacts may be associated with short exposure to very high concentrations of sediment, while chronic impacts may be associated with long exposure to much lower sediment concentrations. It is important, therefore, to consider the temporal scale at which sediment state and ecological responses are being measured and characterised when deriving sediment attributes.

2.2.2 Multiple modes of impact and variations in effects across species and lifeSstages

Ecosystem health is a broad term generally used to describe the condition of an ecosystem. The NPS-FM states that in a healthy freshwater ecosystem "ecological processes are maintained, there is a range and diversity of indigenous flora and fauna, and there is resilience to change." For this study we have interpreted this to imply that a healthy ecosystem is one where both ecosystem structure and function are similar to that expected under minimally disturbed conditions.

Greater fine sediment (generally defined as organic and inorganic particles <2 mm in diameter) inputs to a river system are observed as increases in the suspended solids load (i.e., suspended sediment) and/or accumulation of fine sediments in/on the river bed (i.e., deposited sediment). Elevated fine sediment is widely documented as having an impact on ecosystem health, therefore, fulfilling criterion one for the development of NOF attributes (see Section 2.1). The effects on ecosystem structure and function can be pervasive, but not all aquatic communities are equally sensitive (Bilotta and Brazier 2008; Collins et al. 2011; Kemp et al. 2011; Wilkes et al. 2019). The functional traits of biota determine an individual species' ability to survive in different environmental conditions and their resistance and resilience to disturbances (Wilkes et al. 2019). Increased fine sediment also impacts ecosystem structure and function via multiple mechanisms (Ryan 1991; Collins et al. 2011). Consequently, the interaction between differing functional traits of individual biota and multiple impact pathways will determine the broader ecosystem response to elevated fine sediment (Wilkes et al. 2019).

The literature describing ecological responses to fine sediment was reviewed during an earlier phase of this project and the results reported in Depree et al. (2018). This review indicated that the effects of fine sediment on biotic communities are determined by several characteristics: the sediment concentration, the duration and frequency that aquatic environments are exposed to the elevated sediment levels, the particle-size distribution of the sediments, and the organic/inorganic composition. There was also evidence to support the existence of impacts on biota both directly through physical effects and indirectly through effects on habitat, food supply, migratory cues and behaviour (Ryan 1991; Bilotta and Brazier 2008; Kemp et al. 2011). The effects are most often chronic and sub-lethal, leading to a decline in growth and condition, curtailed migration, redistribution of populations and changes in population demographics. However, lethal and acute impacts may also occur in some circumstances. While many of the ecological responses were negative, in some cases the evidence was inconclusive, and for some species positive responses have been documented (Ryan 1991; Bilotta and Brazier 2008; Kemp et al. 2011; Grove et al. 2015). Overall, there was an indication that macroinvertebrate communities were generally more sensitive to deposited sediment than fish communities, and vice versa for suspended sediment. However, there is significant variation within those communities, with individual species with certain traits more susceptible to impacts than others (e.g., fish that lay their eggs in gravel are more sensitive to increases in deposited sediment). The key impact pathways for negative effects highlighted by the literature review are conceptualised in Figure 2-1. For further details please refer to the literature reviews in Appendix A (deposited sediment and macroinvertebrates), Appendix B (suspended sediment and macroinvertebrates) and Appendix C (suspended sediment and fish).

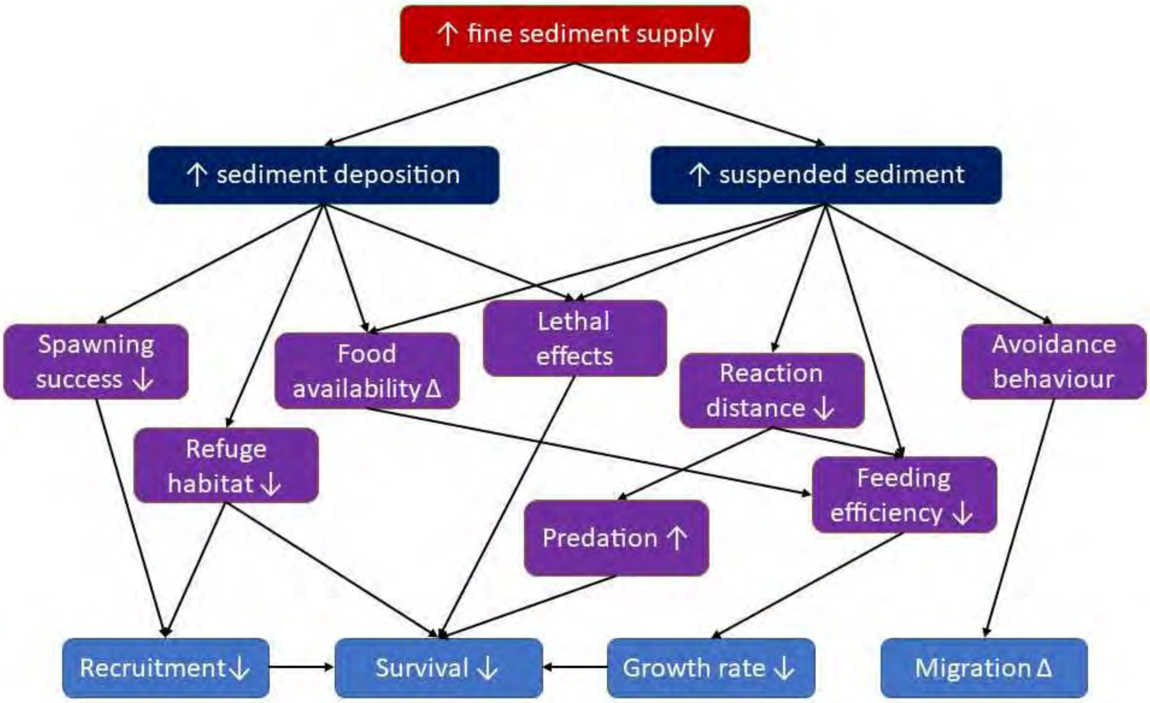


Figure 2-1: Conceptualisation of key effects pathways showing the negative impacts of increased fine sediments on aquatic organisms. ↑ = increases; ↓= decreases; Δ = changes (may be up or down).

Because there are multiple modes of impact it has been acknowledged that more than one sediment attribute would assist with managing the effects of elevated fine sediments (Collins et al. 2011). In the earlier phases of this project suspended sediment (measured as turbidity and/or visual clarity) and deposited sediment (measured as % cover of fine sediment) were identified as appropriate ESVs for development of sediment attributes (Depree et al. 2018). However, because there are multiple effects pathways (e.g., Figure 2-1) that vary between species and life-stages, determining appropriate sediment attribute thresholds requires evaluation of multiple lines of evidence.

2.2.3 Data availability

Depree et al. (2018) summarised a range of available data sets for both sediment ESVs and indicators of ecosystem health. A number of challenges for progressing analyses of spatial and temporal patterns in sediment ESV state and characterising ecological responses to changes in sediment ESV state were highlighted. These are summarised below.

Suspended sediment ESV data

The main sources of suspended sediment ESV data are regional council state of the environment (SOE) monitoring sites and the National River Water Quality Network (NRWQN) sites. Total suspended solids (TSS) is a direct measure of sediment concentration in the water column, whereas turbidity and visual clarity are surrogates for suspended sediment. The only suspended sediment ESV that is currently measured in all regions is turbidity. Visual clarity and TSS are currently measured routinely by 12 of 16 regional councils. Turbidity and visual clarity are routinely measured at the 77 NRWQN sites.

While continuous measurement of turbidity, TSS and visual clarity is technically feasible, data are not routinely collected this way for state of the environment monitoring. Some councils do now have continuous turbidity sensors deployed at a limited number of sites for sediment load estimation and these data may be useful in future for better understanding suspended sediment dynamics. Most available data are collected as monthly spot samples. There are presently around 830 sites where ≥10 years of monthly turbidity measurements are available, and around 720 sites where ≥10 years of monthly visual clarity measurements are available. The number of sites with long-term TSS measurements is much lower and the detection limit for TSS samples was often too high for characterising sediment state at 'clean' sites. Consequently, Depree et al. (2018) concluded that further analyses of suspended sediment should focus on using turbidity and visual clarity data.

The relatively high number of sites where long-term data are available means that spatial coverage of the suspended sediment ESV data is generally good at a national scale (Figure 2-2). However, the low number of continuous or event based suspended sediment ESV data means that analyses in this project are restricted to long-term site 'average' statistics (e.g., medians) of suspended sediment ESV state.

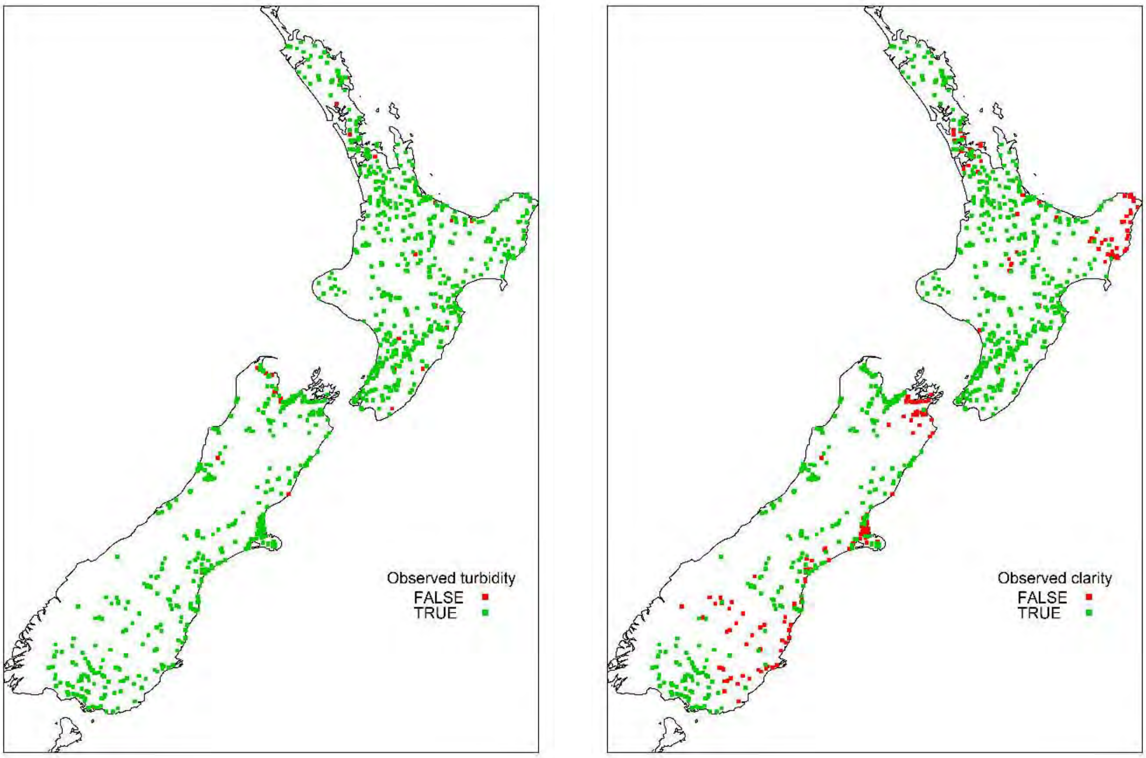


Figure 2-2: Spatial distribution of suspended sediment observations in the collated data set. Only sites where corresponding macroinvertebrate observations are available are shown. TRUE indicates sites where the sediment ESV is also measured and FALSE indicates sites where the sediment ESV is not measured.

Deposited sediment ESV data

Deposited sediment ESV data were collated from two main sources, regional councils and the New Zealand Freshwater Fish Database (NZFFD) as well as from disparate research projects (published and unpublished). Data for eight different ESV measures were identified. Deposited sediment data have generally not been routinely collected by most councils meaning that there are no sites with long-term regular monitoring data available. This means that it is very difficult to characterise temporal variability at a site in a generalisable way, and the analyses carried out in this project are effectively based on one-off observations of ESV state.

Spatial representativeness of the NZFFD % fines data is generally good due to the large number of sampling locations (Figure 2-3). However, for most of the other measures there are biases in the spatial representativeness of the data sets, particularly an under representation of high order (i.e., bigger) rivers. This reflects the relatively lower number of sites, the differences in sampling methodologies adopted by different councils where deposited sediment is measured, and the reliance on methods that require wading. Based on the analyses of Depree et al. (2018), it was concluded that the NZFFD % fines data and the SAM2 % cover instream data were the most suited for further analyses. However, the lack of repeated measures over time at a site for most methods means that our ability to understand natural temporal variability in deposited sediment is limited. There are also no concurrent measurements of NZFFD % fines and SAM2 % cover instream, raising difficulties for conversion between data types.

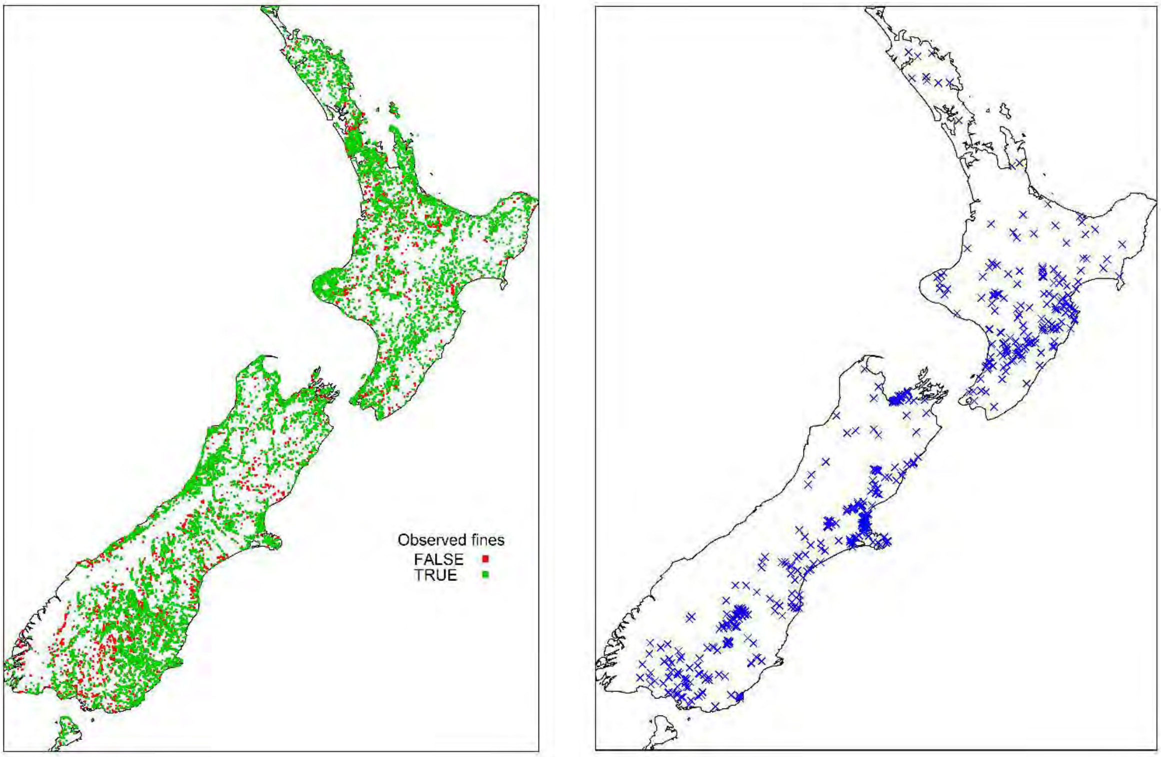


Figure 2-3: Spatial distribution of deposited sediment observations in the collated data sets. Left: NZFFD % fines (TRUE = % cover fines reported; FALSE = % cover fines not reported). Right: SAM2 % cover instream.

Ecological data

Depree et al. (2018) concluded that both fish and macroinvertebrate indicators should be considered when evaluating ecological responses to variation in sediment ESVs. Data on freshwater fish were retrieved from the NZFFD, with a total of 34,364 records available for analyses. While a proportion of NZFFD data records contain data on observed abundances, fish abundance was not used in the analyses for two reasons. First, abundance is strongly related to fishing effort and area fished, which are often not available or are imprecisely measured for many records. Also, fishing effort may not be transferable between sites due to differences in physical conditions (size of river, substrate size, presence of vegetation etc.). Second, the locations at which abundances have been observed are biased towards certain catchments and regions of the country. Fish distributions are strongly related to landscape setting such as distance from sea and altitude. These characteristics may also be related to sediment characteristics. Therefore, to best characterise the relationships between fish and sediment, this landscape-scale information must first be accounted for. This is best achieved by utilising fish observations spread across the entire range of observed conditions. Analyses were therefore, carried out using presence-absence data. At many sites only one record is available, making quantification of temporal variations in fish communities impossible.

Depree et al. (2018) collated three macroinvertebrate datasets for analysis of ecological response relationships. The NRWQN data include annual quantitative macroinvertebrate samples. Macroinvertebrate abundance data for the period 1990–2013 were collated. SOE monitoring data were also available from all regional or unitary councils. Most data were from the period 2000–2016 and samples were generally collected annually. However, a range of sampling and sample processing methods are used by the different regional and unitary councils. Also, macroinvertebrate data from published and unpublished research studies using a combination of sampling and sample processing methods were collated. To maintain equivalency between samples collected or processed using different methods, all data were converted to proportional relative abundance of species within a sample. Because there are multiple samples over time at the same sites, there is some scope for characterising temporal variations in macroinvertebrate communities at an annual time-step where matching temporal records for sediment ESVs exist.

Limitations in data matching

A critical challenge for this project was the limitations on data availability caused by the lack of coincidence and equivalency between the different data. Factors included:

- limited numbers of sediment ESV and ecological observations taken from the same site at the same time (one-off observations or repeated measures over time at the same site)
- differences in sampling frequency between sediment ESV and ecological observations (e.g., one-off v. monthly v. annual samples)
- mis-matches in the sediment ESV measures that could be paired with different ecological indicators (e.g., % cover instream and macroinvertebrates v NZFFD % fines and fish), and
- different response measures for the different ecological indicators (e.g., abundance v. relative abundance v. presence/absence).

Resolving these inconsistencies is critical to developing transparent and defensible sediment attributes.

2.3 Approach adopted for this project

A range of strategies were explored by Depree et al. (2018) to address the challenges described in the previous section, and to then develop sediment attributes. However, some of these challenges were not resolved within the constraints of that project and are the focus of the work continued here.

The broad approach taken by Depree et al. (2018) was to:

- Use existing literature to characterise hypothesised interaction pathways between fine sediment stressors and ecosystem health indicators.
- Identify and review available data sets for their suitability to:
  - characterise natural spatial and temporal variations in sediment ESVs across New Zealand, and
  - quantify interaction pathways between sediment ESVs and ecosystem health indicators.
- Develop a classification of natural spatial patterns in sediment ESV state.
- use a range of analytical methods to characterise quantitative relationships between sediment ESVs and ecosystem health indicators, and
- combine the multiple lines of evidence to make recommendations on appropriate sediment ESV attribute thresholds.

The principal technical review comments that remained unsatisfactorily resolved at the end of that workstream (and which are the focus of this report) are:

- the classification of spatial patterns in sediment ESV state and how that is linked to attribute development
- equivalency between the multiple lines of evidence, and
- the need for a transparent and reproducible method for integrating multiple lines of evidence to define attribute thresholds.

3 Accounting for natural patterns in fine sediment state

- Patterns of fine sediment state are expected to vary naturally across the country and this should be accounted for in setting limits.
- We set out five guiding principles for developing a sediment classification system and estimating reference ESV state for each class (Section 3.1).
- We developed separate Sediment State Classification (SSC) systems for deposited and suspended sediment using a hierarchical clustering method based on the combined climate-topography-geology classes from the River Environment Classification (Section 3.2).
- Separate classifications were derived for deposited and suspended sediment ESVs. The 12 class SSCs were recommended for adoption for both deposited and suspended sediment (Section 3.2).
- Reference ESV state was estimated for each of the 12 classes for both deposited and suspended sediment using a statistical model (Section 3.2).

The amount of instream deposited fine sediment and suspended sediment is expected to vary naturally across the New Zealand river network because of natural differences in geology, climate, topography, etc. Sediment management objectives must take this landscape-scale variability in natural state into consideration. This requires a method to determine reference states for instream deposited fine sediment and suspended sediment across segments of the New Zealand river network.

For the purposes of this investigation, the reference state of a segment was broadly defined as the average level of deposited and suspended sediment within that segment, through time, assuming minimal urban, agricultural and forestry development within the catchment upstream. The levels of deposited and suspended sediment in a segment in its reference state are dependent on factors such as climate, topography and geology, which interact to determine sediment supply and retention.

Reference states throughout New Zealand were required for three sediment ESVs: deposited fine sediment (proportion of streambed covered by sediment <2 mm diameter), turbidity (NTU), and visual clarity (m). This was achieved by:

1. developing a sediment state classification (SSC) for New Zealand rivers that sorts river segments into groups that have different sediment supply and retention characteristics, and
2. within each sediment class, estimating the reference state for each ESV.

This section summarises the key steps and outcomes of this process. Full details of the analyses are set out in Appendix D. Based on the outcomes of these analyses we recommend progressing the separate 12-class sediment classifications (i.e., aggregation level 4) for each of deposited and suspended sediment as shown in Table 3-1 to Table 3-3. These results replace Chapter 2 of Depree et al. (2018).

3.1 Guiding principles

Our approach to achieving these two objectives was guided by five principles:

1. The reference state classification should achieve a balance between simplicity, hence ease of use, and sensitivity to changes in the sediment status of streams.

2. The classification should build on existing river classification systems used in New Zealand, particularly those that have been used to inform catchment policy and management.

3. The classification should be (a) based on the key geomorphological and climatological variables that drive sediment supply and retention; and (b) also be based on observed deposited and suspended sediment data, hence capture real differences in the sediment characteristics of rivers.

4. The classification should group stream segments at a spatial resolution reflecting likely changes in the geomorphological and climatological variables driving sediment supply and retention.

5. Estimates of reference state within all regions of New Zealand should result in NOF management bands — hence management targets — that are achievable.

3.2 Sediment State Classification (SSC) development

Two SSCs were developed; one for deposited fine sediment (SSC_Dep) and one for suspended sediment (SSC_Sus). Separate SSCs for deposited and suspended sediments were deemed necessary since measures of suspended and deposited fine sediment are not well correlated within New Zealand river segments (Figure D-1). Turbidity was chosen as the basis for development of the SSC_Sus because sites at which turbidity were monitored were more numerous and had greater spatial coverage than those for visual clarity, and because turbidity and visual clarity are strongly correlated (Figure D-1).

The New Zealand River Environment Classification (REC; Snelder and Biggs 2002) climate, topography and geology variables were used as the basis of the SSCs. These REC variables were selected as being likely to drive supply and retention of fine sediment in New Zealand streams. The combined Climate-Topography-Geology (CTG) classes were used as the basis of grouping streams that should experience contrasting sediment supply and retention characteristics.

The sediment ESV characteristics of each CTG class were characterised using observed data collated by Whitehead (2019). Some CTG classes were aggregated where the different CTG classes were likely to experience similar sediment supply and retention characteristics to ensure enough data ($n \geq 20$) were available to effectively characterise sediment ESV characteristics. Sufficient data existed for the CTG classes compromising the majority of the river network. For those CTG classes where insufficient data existed for characterisation, we developed a method for mapping them to a class based on spatial proximity rules (see Appendix E for details). Table D-2 in Appendix D shows the unmapped CTG classes for each of the deposited and suspended SSCs, but for deposited sediment this most commonly occurred in cool dry mountain areas and for suspended sediment this was most common in cool hill and mountain areas and warm hill areas.

Although variation in sediment composition among CTG classes was evident, many classes exhibited similar ESV composition, justifying further aggregation. This was achieved using cluster analysis, a statistical method for grouping a set of objects such that the objects in a group are more similar to each other than to those in other groups. A hierarchical clustering method was used, meaning that the number of groups increased as the level of dissimilarity between groups decreased. SSCs were generated for both the deposited and suspended sediment ESVs at four different levels of dissimilarity (50%, 30%, 20% and 15%). This yielded 2, 4, 8 and 12 sediment classes at the four levels of aggregation for both deposited (Figure 3-1) and suspended (Figure 3-2) sediment. For both deposited and suspended fine sediment the cluster analysis yielded sediment classes that clearly had different climatic, topographical and geological characteristics.

For each class at the different levels of aggregation, we next estimated the ESV reference state. Due to the small number and restricted distribution of reference sites for deposited and suspended sediment, we used a model-based approach for estimating reference state. This involved selecting a statistical model that describes how ESV state changes with increasing anthropogenic disturbance (e.g., increasing conversion to pasture), and then using that model to estimate the predicted ESV state at zero anthropogenic disturbance (see Appendix D for details). The resulting reference state predictions for each class are presented for deposited sediment in Table 3-1, turbidity in Table 3-2 and visual clarity in Table 3-3.

Finally, we considered the extent to which the estimates of reference state may be biased by moving to a higher level of aggregation (i.e., going from Level 4 (12 classes) to Level 3 (8 classes)). The SSCs are hierarchical, so multiple reference states within a lower level of aggregation may correspond to a single reference state at the next highest level of aggregation. For all ESVs, the higher the level of aggregation the more biased the estimates of reference state. This means that management outcomes are likely to be more variable as the level of aggregation increases. Sites could end up with overly permissive thresholds, or overly restrictive thresholds, depending on whether the reference state at the lower level of aggregation is lower or higher, respectively, than the ‘average’ at the higher level of aggregation.

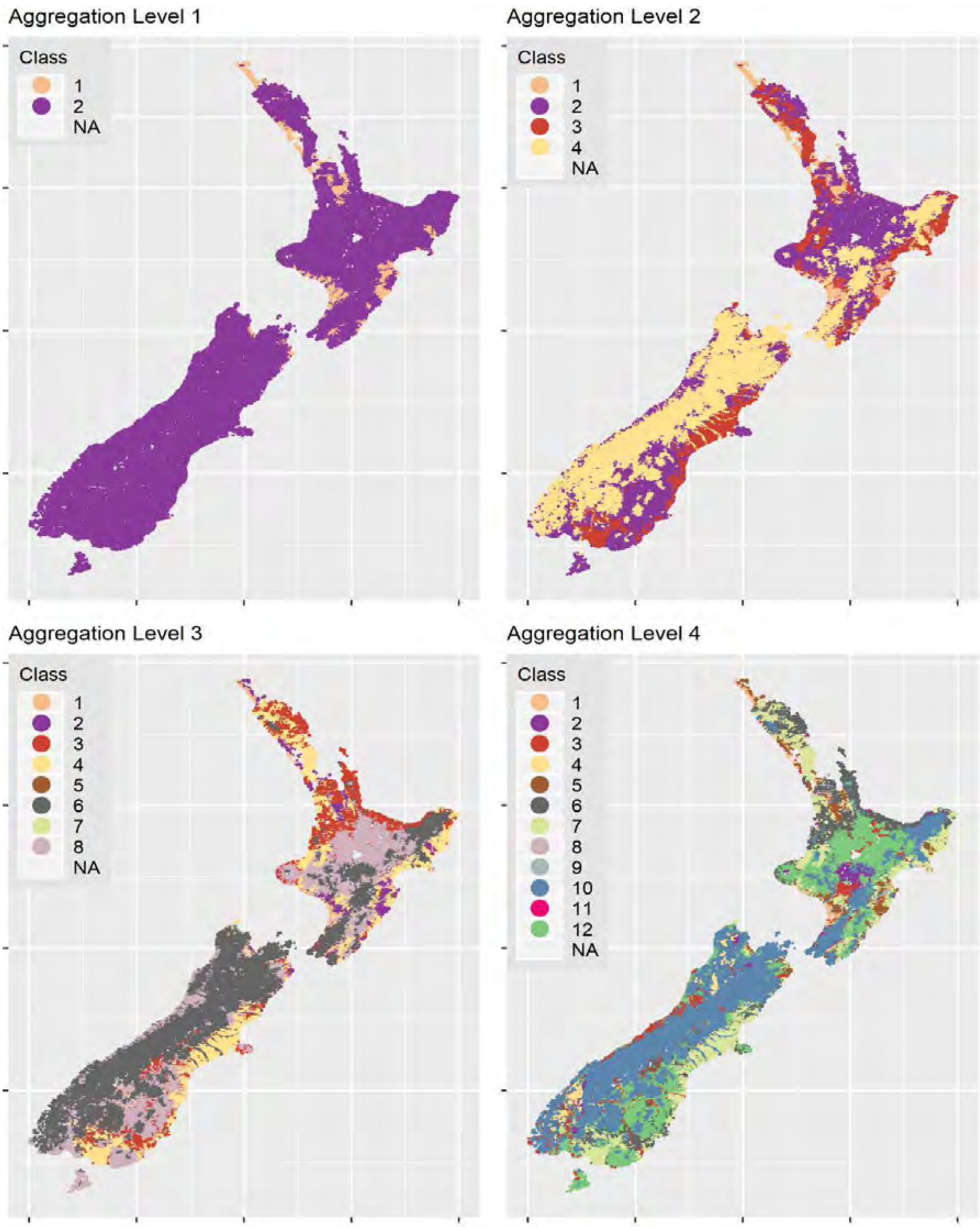


Figure 3-1: Spatial distribution of the deposited fine sediment classes under four different levels of aggregation of the REC CTG classes. See Table D-3 for a description of sediment classes. A small number of reaches could not be classified using this method due to a lack of data (Class = NA). For Aggregation Level 4, which we recommended for use, these unclassified reaches were subsequently allocated to classes using the spatial mapping process explained in Appendix E. The final classification for deposited fine sediment is shown in Figure 1-1.

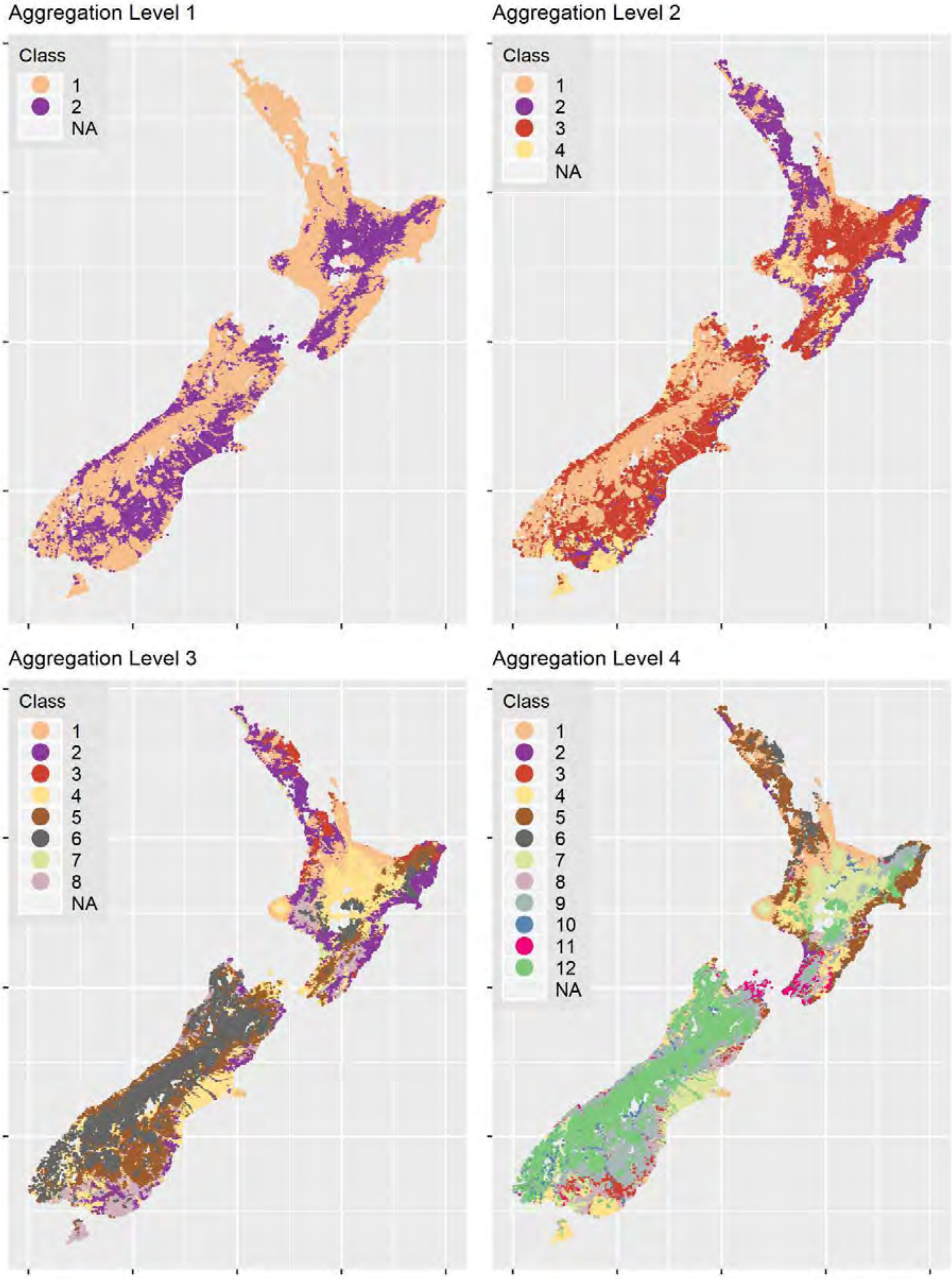


Figure 3-2: Spatial distribution of the suspended sediment (turbidity) classes under four different levels of aggregation of the REC CTG classes. See Table D-3 for a description of sediment classes. A small number of reaches could not be classified using this method due to a lack of data (Class = NA). For Aggregation Level 4, which we recommended for use, these unclassified reaches were subsequently allocated to classes using the spatial mapping process explained in Appendix E. The final classification for deposited fine sediment is shown in Figure 1-2.

Table 3-1: Reference values (Ref) for proportional cover of deposited fine sediment for each sediment class, at each level of aggregation. Agg. L1 = 50% dissimilarity. Agg. L2 = 30% dissimilarity. Agg. L3 = 20% dissimilarity. Agg. L4 = 15% dissimilarity. Also presented are the percentages of the New Zealand river network allocated to each class (% River Net.), at each level of aggregation. CTG = Climate-Topography-Geology classes from the REC.

<table>
<tr><th>Agg. L1</th><th>Ref</th><th>% River Net.</th><th>Agg. L2</th><th>Ref</th><th>% River Net.</th><th>Agg. L3</th><th>Ref</th><th>% River Net.</th><th>Agg. L4</th><th>Ref</th><th>% River Net.</th><th>CTG Classes</th></tr>
<tr><td rowspan="5">1</td><td rowspan="5">0.64</td><td rowspan="5">5.88</td><td rowspan="5">1</td><td rowspan="5">0.64</td><td rowspan="5">5.88</td><td>1</td><td>0.79</td><td>1.88</td><td>1</td><td>0.79</td><td>1.88</td><td>WD_Low_VA;
WD_Low_Al</td></tr>
<tr><td rowspan="2">2</td><td rowspan="2">0.68</td><td rowspan="2">3.42</td><td>5</td><td>0.74</td><td>3.05</td><td>WD_Low_SS</td></tr>
<tr><td>9</td><td>0.43</td><td>0.36</td><td>WD_Low_HS</td></tr>
<tr><td>5</td><td>0.13</td><td>0.14</td><td>8</td><td>0.13</td><td>0.14</td><td>WW_Lake_Any</td></tr>
<tr><td>7</td><td>0.69</td><td>0.45</td><td>11</td><td>0.69</td><td>0.45</td><td>WW_Low_Al</td></tr>
<tr><td rowspan="7">2</td><td rowspan="7">0.15</td><td rowspan="7">93.05</td><td rowspan="3">2</td><td rowspan="3">0.21</td><td rowspan="3">37.73</td><td>3</td><td>0.22</td><td>13.32</td><td>6</td><td>0.22</td><td>13.32</td><td>WW_Low_VA;
WW_Low_HS;
CD_Low_VA;
CD_Hill_Al;
CD_Low_HS</td></tr>
<tr><td rowspan="2">8</td><td rowspan="2">0.22</td><td rowspan="2">24.41</td><td>12</td><td>0.20</td><td>19.73</td><td>CW_Hill_VA;
CW_Low_VA;
CW_Low_SS;
CD_Hill_HS</td></tr>
<tr><td>3</td><td>0.33</td><td>4.68</td><td>CW_Lake_Any;
CW_Low_Al;
CD_Hill_SS</td></tr>
<tr><td>3</td><td>0.34</td><td>15.51</td><td>4</td><td>0.34</td><td>15.51</td><td>7</td><td>0.34</td><td>15.51</td><td>WW_Low_SS;
CD_Low_SS;
CD_Low_Al</td></tr>
<tr><td rowspan="3">4</td><td rowspan="3">0.09</td><td rowspan="3">39.82</td><td rowspan="3">6</td><td rowspan="3">0.09</td><td rowspan="3">39.82</td><td>10</td><td>0.09</td><td>36.41</td><td>WW_Hill_VA;
CW_Hill_HS;
CW_Low_HS;
CW_Mount_HS;
CW_Hill_SS;
CW_Hill_Al;
CD_Mount_HS;
CW_Mount_Al</td></tr>
<tr><td>2</td><td>0.04</td><td>1.46</td><td>WW_Hill_HS;
CW_Mount_VA</td></tr>
<tr><td>4</td><td>0.07</td><td>1.95</td><td>CW_Mount_SS</td></tr>
</table>

Table 3-2: Reference values (Ref) for turbidity (NTUs) for each sediment class, at each level of aggregation. Agg. L1 = 50% dissimilarity. Agg. L2 = 30% dissimilarity. Agg. L3 = 20% dissimilarity. Agg. L4 = 15% dissimilarity. Also presented are the percentages of the New Zealand river network allocated to each class (% River Net.), at each level of aggregation. CTG = Climate-Topography-Geology classes from the REC.

<table>
<tr><th>Agg. L1</th><th>Ref</th><th>% River Net.</th><th>Agg. L2</th><th>Ref</th><th>% River Net.</th><th>Agg. L3</th><th>Ref</th><th>% River Net.</th><th>Agg. L4</th><th>Ref</th><th>% River Net.</th><th>CTG Classes</th></tr>
<tr><td rowspan="8">1</td><td rowspan="8">2.4</td><td rowspan="8">56.82</td><td rowspan="3">1</td><td rowspan="3">2.1</td><td rowspan="3">30.83</td><td>1</td><td>1.6</td><td>7.05</td><td>1</td><td>1.6</td><td>7.05</td><td>WW_Low_VA;
CW_Low_VA</td></tr>
<tr><td>6</td><td>2.1</td><td>22.37</td><td>12</td><td>2.2</td><td>22.37</td><td>CW_Mount_HS;
CW_Hill_SS</td></tr>
<tr><td>7</td><td>4.9</td><td>1.42</td><td>2</td><td>4.9</td><td>1.42</td><td>WD_Low_Al</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">5.2</td><td rowspan="3">17.26</td><td rowspan="2">2</td><td rowspan="2">5.8</td><td rowspan="2">14.42</td><td>5</td><td>5.9</td><td>10.81</td><td>WW_Low_SS;
WD_Low_SS</td></tr>
<tr><td>8</td><td>3.6</td><td>3.61</td><td>CD_Low_SS</td></tr>
<tr><td>3</td><td>3.8</td><td>2.84</td><td>6</td><td>3.8</td><td>2.84</td><td>WW_Low_HS</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2">2.5</td><td rowspan="2">8.72</td><td rowspan="2">8</td><td rowspan="2">2.5</td><td rowspan="2">8.72</td><td>3</td><td>1.1</td><td>2.72</td><td>CD_Low_HS</td></tr>
<tr><td>4</td><td>2.7</td><td>6.01</td><td>CW_Low_SS</td></tr>
<tr><td rowspan="4">2</td><td rowspan="4">1.1</td><td rowspan="4">31.70</td><td rowspan="4">3</td><td rowspan="4">1.2</td><td rowspan="4">31.70</td><td rowspan="3">4</td><td rowspan="3">1.5</td><td rowspan="3">14.58</td><td>7</td><td>2</td><td>10.92</td><td>CD_Low_Al;
CW_Hill_VA</td></tr>
<tr><td>10</td><td>0.9</td><td>1.63</td><td>CW_Lake_Any</td></tr>
<tr><td>11</td><td>0.9</td><td>2.03</td><td>CW_Low_HS</td></tr>
<tr><td>5</td><td>1.0</td><td>17.12</td><td>9</td><td>1.0</td><td>17.12</td><td>CW_Hill_HS;
CD_Hill_HS;
CW_Low_Al</td></tr>
</table>

Table 3-3: Reference values (Ref) for visual clarity (m) for each sediment class, at each level of aggregation. Agg. L1 = 50% dissimilarity. Agg. L2 = 30% dissimilarity. Agg. L3 = 20% dissimilarity. Agg. L4 = 15% dissimilarity. Also presented are the percentages of the New Zealand river network allocated to each class (% River Net.), at each level of aggregation. Classes whose reference state estimates were denoted by an Asterix (*) were assigned the reference state of their parent class, due to insufficient data within that class, at that level, for implementation of the model-based estimation (see Appendix D) . CTG = Climate-Topography-Geology classes from the REC.

<table>
<tr><th>Agg. L1</th><th>Ref</th><th>% River Net.</th><th>Agg. L2</th><th>Ref</th><th>% River Net.</th><th>Agg. L3</th><th>Ref</th><th>% River Net.</th><th>Agg. L4</th><th>Ref</th><th>% River Net.</th><th>CTG Classes</th></tr>
<tr><td rowspan="8">1</td><td rowspan="8">2.0</td><td rowspan="8">56.82</td><td rowspan="3">1</td><td rowspan="3">2.9</td><td rowspan="3">30.83</td><td>1</td><td>2.7</td><td>7.05</td><td>1</td><td>2.7</td><td>7.05</td><td>WW_Low_VA;
CW_Low_VA</td></tr>
<tr><td>6</td><td>3.0</td><td>22.37</td><td>12</td><td>3.1</td><td>22.37</td><td>CW_Mount_HS;
CW_Hill_SS</td></tr>
<tr><td>7</td><td>2.9*</td><td>1.42</td><td>2</td><td>2.9*</td><td>1.42</td><td>WD_Low_Al</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">1.0</td><td rowspan="3">17.26</td><td rowspan="2">2</td><td rowspan="2">0.9</td><td rowspan="2">14.42</td><td>5</td><td>0.8</td><td>10.81</td><td>WW_Low_SS;
WD_Low_SS</td></tr>
<tr><td>8</td><td>0.7</td><td>3.61</td><td>CD_Low_SS</td></tr>
<tr><td>3</td><td>1.6</td><td>2.84</td><td>6</td><td>1.3</td><td>2.84</td><td>WW_Low_HS</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2">1.6</td><td rowspan="2">8.72</td><td rowspan="2">8</td><td rowspan="2">1.7</td><td rowspan="2">8.72</td><td>3</td><td>1.7*</td><td>2.72</td><td>CD_Low_HS</td></tr>
<tr><td>4</td><td>1.7</td><td>6.01</td><td>CW_Low_SS</td></tr>
<tr><td rowspan="4">2</td><td rowspan="4">3.1</td><td rowspan="4">31.70</td><td rowspan="4">3</td><td rowspan="4">3.0</td><td rowspan="4">31.70</td><td rowspan="3">4</td><td rowspan="3">2.7</td><td rowspan="3">14.58</td><td>7</td><td>2.1</td><td>10.92</td><td>CD_Low_Al;
CW_Hill_VA</td></tr>
<tr><td>10</td><td>3.9</td><td>1.63</td><td>CW_Lake_Any</td></tr>
<tr><td>11</td><td>3.3</td><td>2.03</td><td>CW_Low_HS</td></tr>
<tr><td>5</td><td>3.1</td><td>17.12</td><td>9</td><td>3.5</td><td>17.12</td><td>CW_Hill_HS;
CD_Hill_HS;
CW_Low_Al</td></tr>
</table>

4 Characterising the responses of ecosystem health to increasing fine sediment

- A range of methods were used to characterise and delimit potential effects-based numeric thresholds for fine sediment impacts on ecosystem health (see Section 4.2).
- Separate stressor-response relationships were developed for each sediment ESV (i.e., % cover of deposited fine sediment, turbidity and visual clarity) (see Sections 4.3 and 4.4).
- We found macroinvertebrate communities were the most sensitive ecological value to changes in deposited fine sediment (see Section 4.3). Consequently, macroinvertebrate responses were used as the basis for defining potential attribute bands for the deposited sediment ESV.
- We found fish communities were the most sensitive ecological value to changes in suspended fine sediment (see Section 4.4). Consequently, fish responses were used as the basis for defining potential attribute bands for the suspended sediment ESVs.
- Following the weight of evidence assessment (see Section 5) we recommended using the community deviation method (see Section 4.2.5) as the preferred method for deriving potential attribute bands for both the deposited (see Section 4.3.2) and suspended (see Section 4.4.1) sediment ESVs.

Sediment NOF attributes are defined to inform the process resulting in establishment of management objectives that protect ecosystem health. To define effects-based numeric thresholds, we need to characterise how ecosystem health responds across a gradient of increasing fine sediment and identify thresholds or 'tipping points' and 'safe levels'. Numeric thresholds are best determined by dose-response studies that describe the change in effect on an organism caused by differing levels of exposure to a stressor. In the absence of such studies, an alternative is to develop exposure-response relationships based on field data collected across a stressor gradient.

Exposure to a stressor can elicit a range of responses (Larned,Schallenberg (2018); Figure 4-1). The shape of the response will be determined by factors such as the traits of the organism, the magnitude, duration and frequency of exposure to the stressor, and the influence of confounding factors, such as additional stressors. To account for the potential for differing response shapes, a range of different analytical methods are available to explore stressor-response relationships and define resource management thresholds (Figure 4-2).

Our decisions on appropriate analytical methods for this study were guided by several requirements:

- Analyses must be based on existing data.
- Separate stressor-response relationships should be developed for each sediment ESV.
- Response variables should be representative indicators of ecosystem health.
- Analytical methods must meet at least one of three criteria:

- They can be used to characterise the shape of the stressor-response relationship.
- 'Bottom line' thresholds can be identified.
- Attribute band thresholds can be determined.

- Methods should ideally account for natural spatial variations in sediment ESV state.
-

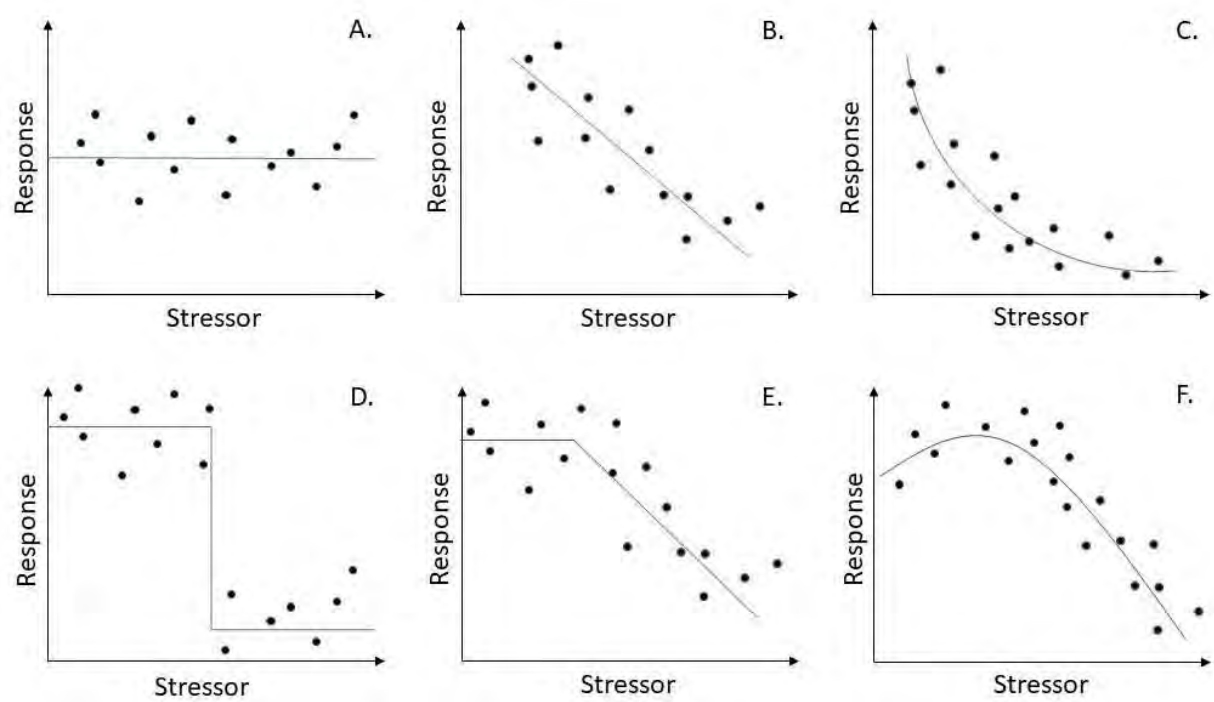


Figure 4-1: Illustration of different shape stressor-response relationships. A. No response; B. Linear decline; C. Logarithmic decline; D. Step response; E. Threshold response; F. Subsidy-stress response.

A range of analytical methods were ultimately used to characterise stressor-response relationships between sediment ESVs and ecosystem health indicators. The requirement to use existing data constrained the choice of methods owing to differences in the spatial coverage and temporal resolution of data, inconsistencies in how different data were collected in different places, and differences between endpoints (i.e., attributes of an ecological entity that can be used to measure effects) in the ecological data.

This section summarises the core analytical methods adopted for this project, documents fundamental decisions in the analytical process, and highlights key findings from those analyses. Further details on all analytical methods and the results arising from them can be found in the technical appendices (Appendix F to Appendix J).

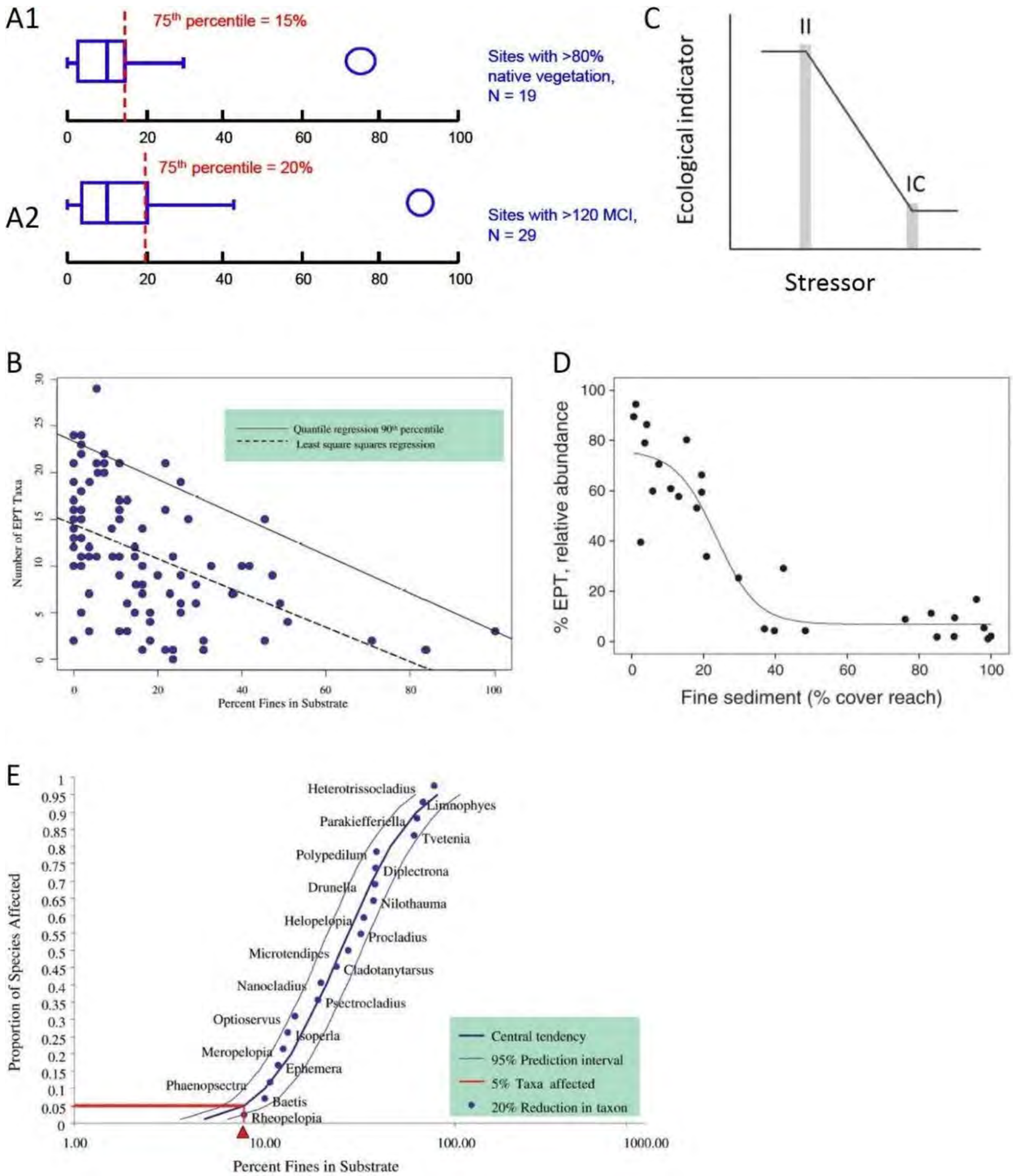


Figure 4-2: Examples of statistical approaches to characterising stressor-response relationships and identifying thresholds for resource management. A. The 75th percentile of the distribution of percent deposited fine sediment values observed at reference sites was used to inform sediment criteria by Clapcott et al. (2011); reference sites were either selected based on percent native vegetation in the catchment (A1) or on MCI values of above 120 (A2). B. Example of use of a least squares simple regression and a quantile regression model to quantify the relationship between the number of EPT taxa and percent deposited fine sediment presented by Cormier et al. (2008). C. Conceptual diagram illustrating impact initiation (II) and impact cessation (IC) thresholds presented by Wagenhoff et al. (2017). D. Example of a four-parameter sigmoidal regression model to quantify the relationship between %EPT and percent deposited fine sediment cover presented by Burdon et al. (2013). E. Species sensitivity distribution plot used to identify the deposited sediment threshold (8%) at which the abundance of 5% of the species is reduced by 20% presented by Cormier et al. (2008).

4.1 General analytical approach

The review and collation of data by Depree et al. (2018) indicated that all analyses in this project would take the general form of field-based gradient analyses using a space-for-time substitution approach. Effectively this means that field data collected from many different places with a range of different sediment ESV states are used to characterise how indicators of ecosystem health vary between places with different sediment ESV states (Figure 4-3).

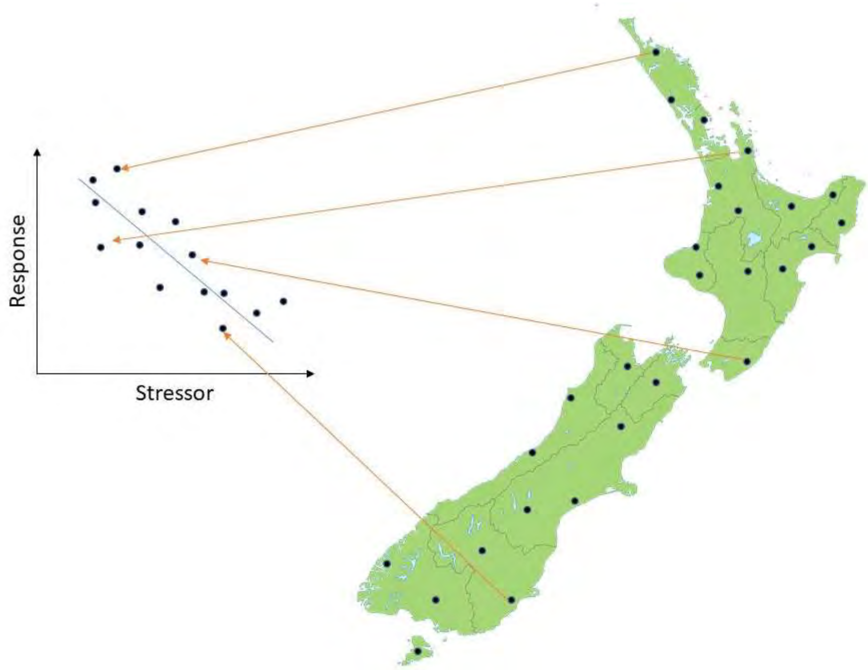


Figure 4-3: Illustration of the concept of space-for-time substitution for stressor-response analyses. Each dot on the map represents a place where the stressor (e.g., fine sediment) and response variables (e.g., macroinvertebrates) have been measured. The amount of fine sediment varies across the country and so is different in each place. When you bring the samples together, they represent a gradient of the stressor (e.g., from places with high fine sediment to places with low fine sediment), indicating how a response indicator changes across the gradient.

Structural indicators of ecosystem health in the form of measures of fish and macroinvertebrate communities were selected for all analyses. Structural characteristics (e.g., taxonomic composition, biological diversity or physical characteristics of habitats) of ecosystem health are more easily measured than functional characteristics (i.e., processes such as flows of energy and materials) and are more routinely surveyed. This means that more data are available for these characteristics, making evaluation of the effectiveness of potential thresholds and limits easier.

Available ecological data dictated that ecological endpoints had to vary between different analyses. For example, few data on fish abundance (i.e., counts of how many of a fish species are there) that were collected in a consistent and comparable manner are available at a national level. All fish analyses were, therefore, based on presence/absence (i.e., is a fish species there or not) data. However, macroinvertebrate abundance data are collected routinely at the National River Water Quality Network monitoring sites allowing analyses of changes in species abundance. Logically, a response in terms of species abundance (i.e., decline in numbers) would be anticipated prior to these species becoming absent at a site. It was, therefore, important to consider these differences in ecological endpoints of individual indicators when compiling the multiple lines of evidence collated in this project (see Section 5 for more information on combining multiple lines of evidence).

The suspended sediment ESVs used for all analyses were annual medians of turbidity and visual clarity calculated from monthly samples over the 12 months preceding the macroinvertebrate sampling date, or modelled long-term medians of turbidity and visual clarity (see Section 4.4.1) for the fish analyses. Deposited sediment has not been routinely and regularly monitored and so most deposited sediment ESVs that were used for the analyses were generally one-off observations paired with ecological observations.

4.2 Methodologies

A range of different methods were used to characterise stressor-response relationships for macroinvertebrates and fish:

1. Boosted regression tree (BRT) analyses to explore the shape of stressor-response relationships in macroinvertebrate metrics.
2. Quantile regression analyses of macroinvertebrate metrics.
3. Generalised linear modelling to characterise the variation in macroinvertebrate metrics with increasing deposited fine sediment.
4. Extirpation analyses of individual macroinvertebrate species combined with species sensitivity distributions (SSDs) to identify species protection levels.
5. A community deviation method to model community changes for fish and macroinvertebrates.

The various methods reflect differences in ecological data types and our attempts to understand equivalence between results derived for fish and macroinvertebrates. A brief description of each method follows. Full technical descriptions are provided in the technical appendices (Appendix F to Appendix J).

4.2.1 Boosted regression tree analyses

Boosted regression tree (BRT) analysis is a flexible modelling approach that allows modelling of complex response shapes. We used this tool to better understand the shape of the stressor-response relationships between benthic macroinvertebrate metrics and the sediment ESVs. A technical description of the BRT methodology is provided in Appendix F.

For these analyses, sediment ESV data were paired with macroinvertebrate metric data sampled from the same site. Four different macroinvertebrate metrics were used as indicators of ecosystem health:

- Macroinvertebrate Community Index (MCI).
- The number of taxa from the orders of Ephemeroptera, Plecoptera and Tricoptera (EPT taxon richness).
- Sediment sensitive Macroinvertebrate Community Index (sediment MCI).

- The number of taxa that decline with increasing deposited sediment (No. of decreases).

The deposited sediment ESV data used for these analyses were collected using the SAM2 instream visual assessment protocol (Clapcott et al. 2011) and are referred to as '% cover instream' in this report. For suspended sediment analyses, median turbidity (NTU) was used as the ESV measure.

BRT analyses were carried out at national scale (i.e., using all data from across the country) and within classes for those classes where sufficient data (>100 observations) were available to robustly build BRT models. Separate BRT models were built for each of the four macroinvertebrate metrics, for each of the sediment ESVs (i.e., % cover instream and turbidity), using 16 predictor variables. The 16 predictor variables included the sediment ESV of interest, plus a range of environmental descriptors chosen based on their high relative importance in exploratory analyses. The environmental descriptors were kept consistent across the deposited and suspended sediment analyses. Sample observations were equally weighted within the same NZReach of the RECv2 to account for the fact that some sites had many observations over time and others had few[10].

BRT model outputs included the percentage total deviance explained (%TDE) and a mean cross-validation (CV) coefficient. The %TDE is a measure of the goodness of fit of the model (i.e., how well it fits to the data) whereas the CV coefficient is a measure of the predictive performance of the model. The relative contribution of the different predictors and the predictors' partial dependence plots are also produced. The fitted functions in the partial dependence plots depict the modelled response shape across each of the predictors when all other predictors are held constant, typically at the mean value. These fitted functions were used for visual threshold identification. Inclusion of predictors other than the sediment ESV measures in the model improves our confidence that the fitted function describing the response shape to sediment reflects that stressor, rather than the response to another predictor that is correlated with increasing sediment.

4.2.2 Quantile regression

Quantile regression is a form of regression analysis that can be used to determine different measures of central tendency and statistical dispersion and, thus, obtain a more comprehensive analysis of the relationships between variables. It is particularly suited to characterising ecological responses where typically not all the factors that affect ecological processes are measured and so cannot be included in predictive models (Cade and Noon 2003). The primary stressor/response quantile regression analysis was undertaken on the NRWQN dataset for the period 1990 to 2013 collated by Depree et al. (2018). A global analysis (i.e., using all NRWQN data) for a range of macroinvertebrate metrics and selected species was undertaken for the initial assessment reported in Depree et al. (2018). The SSC was not available at that time to characterise the response relationships for specific classes.

All quantile regression analyses were performed using the 'quantreg' package (Koenker 2013) in R. Quantiles were fitted using either a linear model or non-linear Ricker model (Cade and Guo 2000; Grace et al. 2014) depending on the shape of the biotic response (i.e., wedge shaped v subsidy-stress shape).

[10] For example, if four observations are present in the same reach they would each have a weighting of 0.25, whereas if there is a single observation in a reach it would have a weighting of 1.

Effects-based thresholds corresponding to a C/D band threshold for turbidity and visual clarity were derived based on a 30% reduction from either a nominal reference ESV state (defined as either 0.5 NTU for turbidity and 6 m for visual clarity) or the ESV state at the maxima of the biotic response. All thresholds were calculated from the 95th percentile quantile regression relationships. Other deviation thresholds could be used to derive equivalent thresholds for different levels of impairment.

The quantile regression thresholds for macroinvertebrate metrics are retained in this analysis as a line of evidence for the weight of evidence (WOE) assessment. The extirpation analysis (Section 4.2.4) was used as an alternative to the quantile regression in this phase of the project because it was better suited to deriving species-specific (taxon identified at the MCI-level) thresholds for the various classes. The species-specific thresholds are then combined to generate a species sensitivity distribution (SSD), which is fitted with a mathematical model to derive numeric threshold values for differing levels of protection for the macroinvertebrate community (see Section 4.2.4).

4.2.3 Generalised linear modelling

Generalised linear modelling is a form of regression analysis that can be used for estimating relationships among variables. We used generalised linear models (GLMs) to characterise the relationship between several macroinvertebrate metrics and the three sediment ESVs: % cover of deposited fine sediment, turbidity and visual clarity. In contrast to the quantile regression approach, GLMs are used to characterise the 'average' response and typically incorporate further predictor variables to account for the potential influence of factors other than the stressor of interest on the response variable. The general approach is summarised below, with a more detailed technical overview of this method included in Appendix I.

For each macroinvertebrate metric, a GLM was fitted to observed data in a form that accounted for landscape scale environmental influences on the response variable (i.e., climate, topography and river size) for each ESV. Within each climate/topography setting the state of the macroinvertebrate metric was estimated at the reference ESV state defined in the relevant SSC class.

Deviations from the metric reference state representing the A/B, B/C and C/D thresholds were calculated over a range spanning from the reference state to the theoretical worst state of the metric. Percent deviations from the reference state were arbitrarily set at 6.67%, 13.3%, 20%. The predicted ESV state corresponding with the metric state at each deviation threshold was then determined from the GLM model for each climate/topography setting.

Results were then amalgamated across different climate/topography settings to provide single threshold estimates for each class (at the 12 class level) of the appropriate ESV SSC. This was achieved by weighting the thresholds derived for each climate/topography setting by the proportion of reaches with that climate/topography setting contained within each SSC class. For example, in SSC class L4.1 there could be 80% of reaches in warm-dry/hill settings and 20% of reaches in cool-dry /mountain settings and so the threshold for the class would be derived by calculating a weighted average of the thresholds for warm-dry/hill (weighting of 0.8) and cool-dry/mountain (weighting of 0.2) settings.

4.2.4 Extirpation analyses

The objective of this analysis was to determine the proportion of macroinvertebrate taxa that may be locally extirpated (i.e., disappear) as sediment ESV state 'worsens'. The analysis we undertook is referred to as a biological extirpation analysis (BEA; Cormier et al. 2018), and is a well-established analytical technique in ecotoxicology (Posthuma et al. 2002; US EPA 2016a). Fish species richness is too low in New Zealand for this approach to be applied and so extirpation analyses were limited to macroinvertebrate communities. SSDs were determined for each class with sufficient data at Level 3 of the suspended SSC. Insufficient deposited sediment data exist that can be paired with macroinvertebrate data to create SSDs for deposited sediment. The SSDs were then used to estimate extirpation thresholds for 1%, 2.5%, 5%, 10%, 25%, 50% and 75% of species within each class (e.g., Figure 4-4). Broadly, the analysis involves two steps:

1. We determine the value of the sediment ESV that will likely result in the local extirpation of individual species. In our case, local extirpation means the species has disappeared from the 'reach' (NZReach), as defined in the New Zealand REC. Following Cormier et al. (2018), a species is deemed locally extirpated when the probability of its occurrence declines to 5% due to a worsening ESV state. The ESV value that corresponds to the 5% probability of occurrence is referred to as that species' XC95. This first step is completed for all species in the assemblage where sufficient data exist (see Appendix H for details).

2. The XC95 values of species in the assemblage are then ranked from lowest (most sensitive; least resistant) to highest (least sensitive; most resistant), and that ranked list of XC95 values is then transformed to yield a species sensitivity distribution (SSD; Posthuma et al. 2002). The SSD is essentially a cumulative probability distribution, representing the proportion of the species in the analysis that are locally extirpated at different sediment ESV states.

For a technical description of this method and the results see Appendix H.

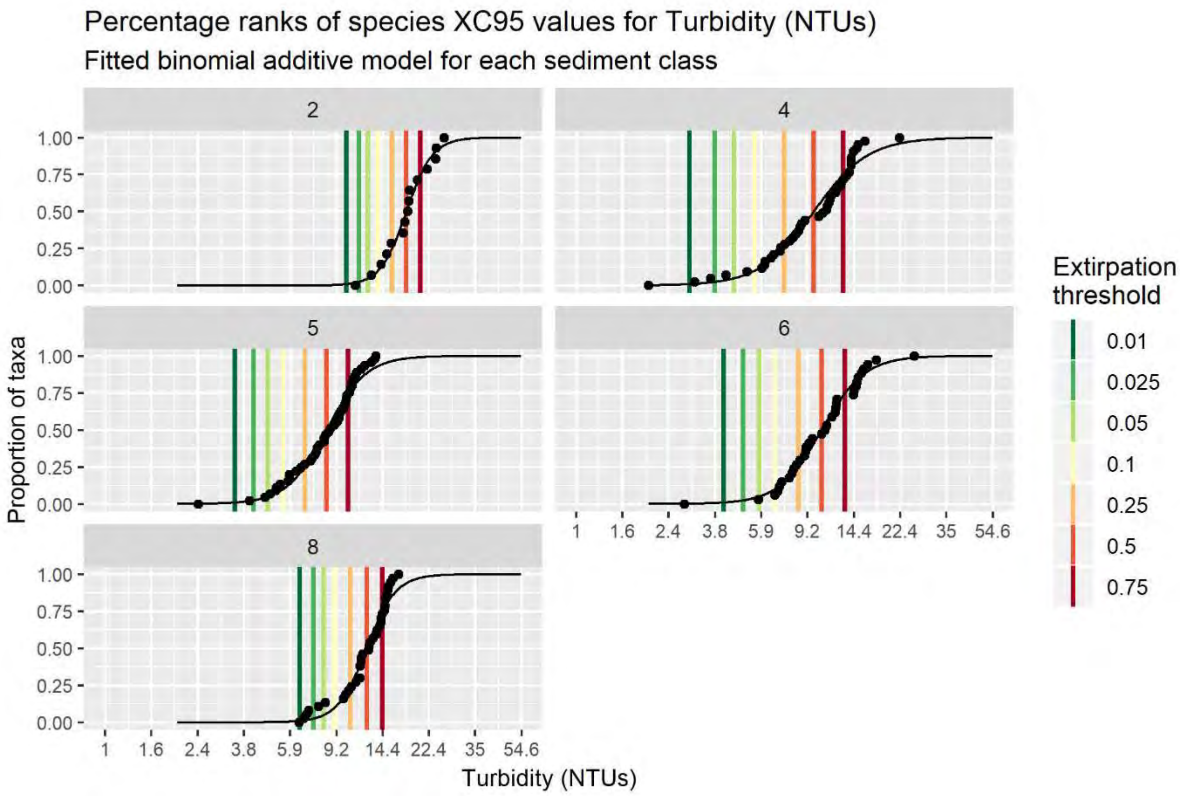


Figure 4-4: An example of SSDs derived at Level 3 of the suspended SSC for turbidity. Numbers correspond to the Level 3 classes.

We recommend the following correspondence between extirpation thresholds and sediment NOF management bands:

1. The threshold between the A and B management bands: 1% extirpation thresholds from BEA.
2. The threshold between the B and C management bands: 2.5% extirpation thresholds from BEA.
3. The threshold between the C and D management bands: 5% extirpation thresholds from BEA.

4.2.5 Community deviation method

The community deviation method was developed specifically for this project. It was designed to characterise changes in community composition resulting from declines in sediment ESV state from predicted reference conditions. A technical description of the community deviation method is provided in Appendix J.

The community deviation method was applied separately to the fish community and to the macroinvertebrate community for each of visual clarity, turbidity, and deposited fine sediment. The fish community comprised 10 native, non-geographically restricted species, plus brown trout, that were present in at least 5% of sites (Figure 4-5). Brown trout were included because they are a highly valued species known to respond to the sediment ESVs and because their habitat is protected under the Resource Management Act. The macroinvertebrate community comprised 25 of 31 species (taxon identified at the MCI-level) present at more than 5% of observed sites (e.g., Figure 4-6).

Data from the New Zealand Freshwater Fish Database (NZFFD) were used to characterise potential thresholds with respect to fish communities. The NZFFD contains many presence/absence observations for various fish species paired with reach-averaged observations of total fine sediment cover. However, no measurements of turbidity or visual clarity taken in the same place and at the same time as the fish observations exist in the NZFFD. Consequently, predicted site median visual clarity or turbidity for each NZFFD observation were in-filled using recently developed statistical models trained using available SOE monitoring data. These statistical models were random forest models with a suit of landscape-scale predictors. The model took the same training data and predictors as used by Whitehead (2019), but also included sediment yield estimated by Hicks et al. (2019) as a predictor. One outcome of the inclusion of sediment yield as a predictor was a decrease in predicted clarity and an increase in predicted turbidity for some rivers located in the Southern Alps and the West Coast of the South Island.

The macroinvertebrate analyses were based on the dataset compiled by Depree et al. (2018), which has paired observations of macroinvertebrates and annual median visual clarity, annual median turbidity and/or deposited fine sediment measured using the SAM2 methodology (i.e., % cover instream). The macroinvertebrate taxa within this dataset are identified to the MCI level. Hereafter these taxa will be referred to as “species”. All macroinvertebrate observations were converted to species presence-absence for the purposes of this analysis to improve equivalence with the fish data.

Observations within the same NZReach of the RECv2 were down-weighted to avoid pseudo-replication (i.e., the effects of having different numbers of observations at a site) within both the fish and macroinvertebrate datasets. Species predicted to prefer greater turbidity, less visual clarity, or greater coverage of deposited fine sediment, e.g., shortfin eels, were removed from the analysis of the appropriate sediment ESV.

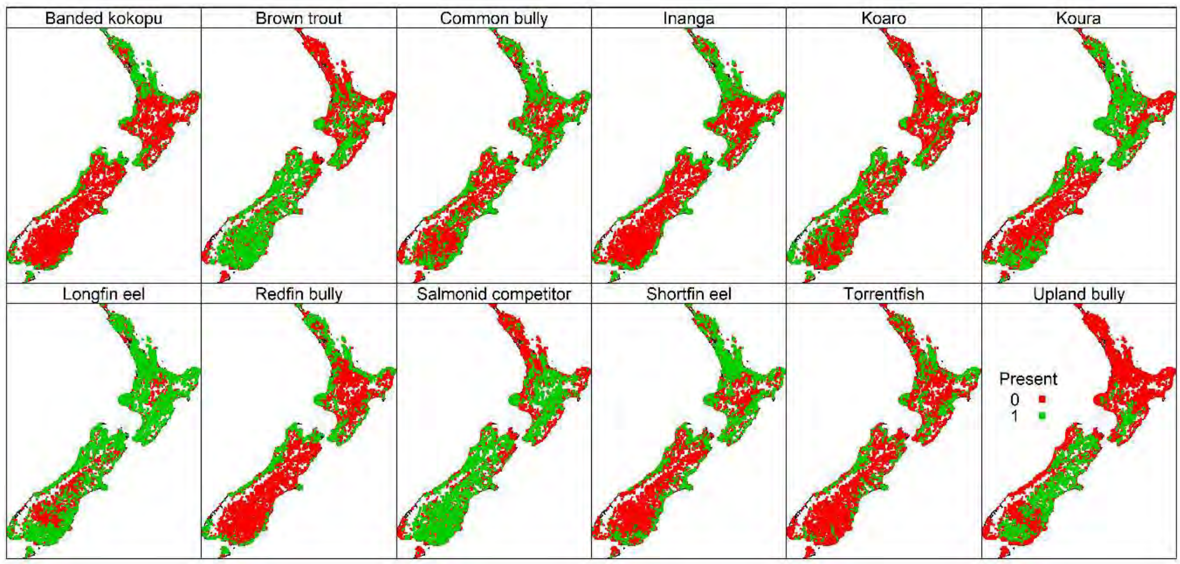


Figure 4-5: Distribution maps for the fish species included in the analyses. Data are from the New Zealand Freshwater Fish Database. Green indicates the species was present at the survey site at the time of the survey. Red indicates the species was not captured at the survey site at the time of the survey.

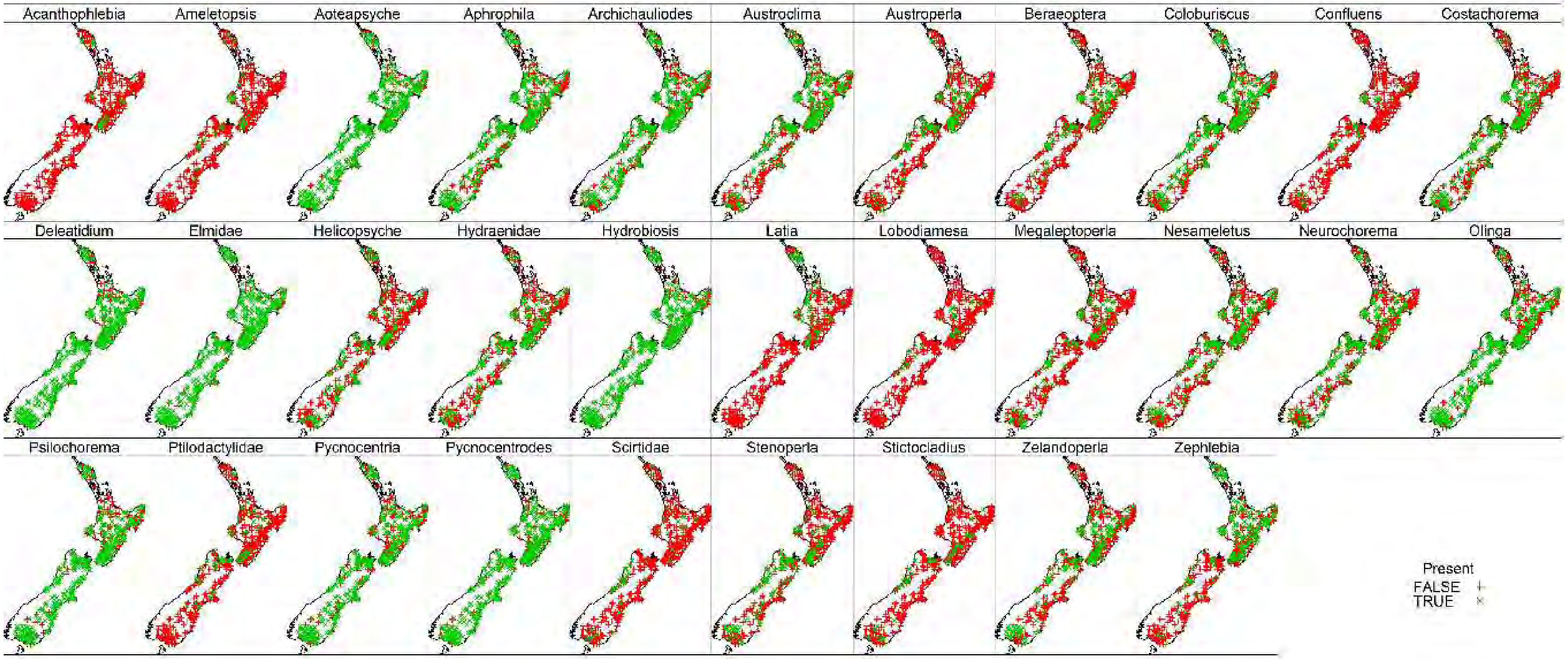


Figure 4-6: Distribution maps for the macroinvertebrate species included in the analyses for the turbidity ESV. Data are from SOE monitoring sites. Green indicates the species was present at the survey site at the time of the survey. Red indicates the species was not captured at the survey site at the time of the survey.

Applying the community deviation method involved three fundamental steps:

1. Modelling species probability of capture as a function of sediment ESVs within landscape settings.
2. Evaluating fish or macroinvertebrate community change in response to deviation of the sediment ESV state from reference conditions.
3. Deriving potential sediment ESV thresholds.

Probability of capture for each species was statistically modelled using a regression model as a function of the sediment ESV (e.g., Figure 4-7). All species were modelled individually, for each sediment ESV, within different landscape settings (i.e., climate, topography, network position, distance inland combination). These landscape settings incorporated variables considered important for describing expected species distributions due to influences on factors such as habitat types, flow regime, and migration ability (for fish). The probabilities of capture were translated to expected presence/absence using a threshold probability (Manel et al. 2001) and then used to inform interpretation of the expected consequences of changing sediment ESV state for fish and macroinvertebrate community composition.

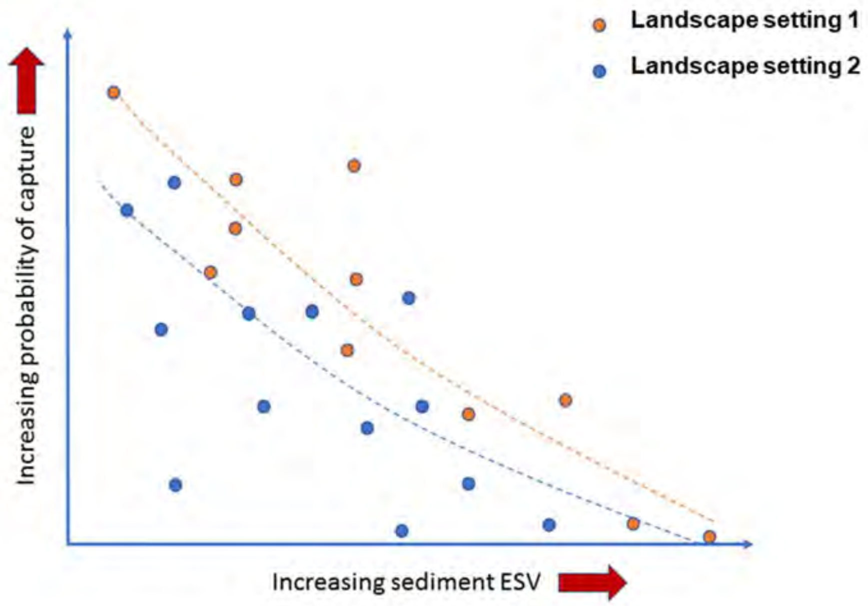


Figure 4-7: Simplified example of how variations in fish probability of capture with increasing sediment ESV are modelled across different landscape settings. The fitted curves derived from this step of the analysis are subsequently used to develop the metrics of community change.

Several steps were required to translate the predicted probability of capture versus sediment ESV responses for individual species into a metric of expected community change at different deposited or suspended sediment states (see Appendix J for details on the methods). In simple terms this first involved determining the probability of capture at reference sediment ESV state (see Section 3.2) and an array of different sediment ESV states for each individual species in each landscape setting. These values were then combined into a metric (ΔC) (Figure 4-8) describing the overall expected change in community relative to the community that could be expected at the reference state condition for deposited sediment for that landscape setting (Figure 4-9).

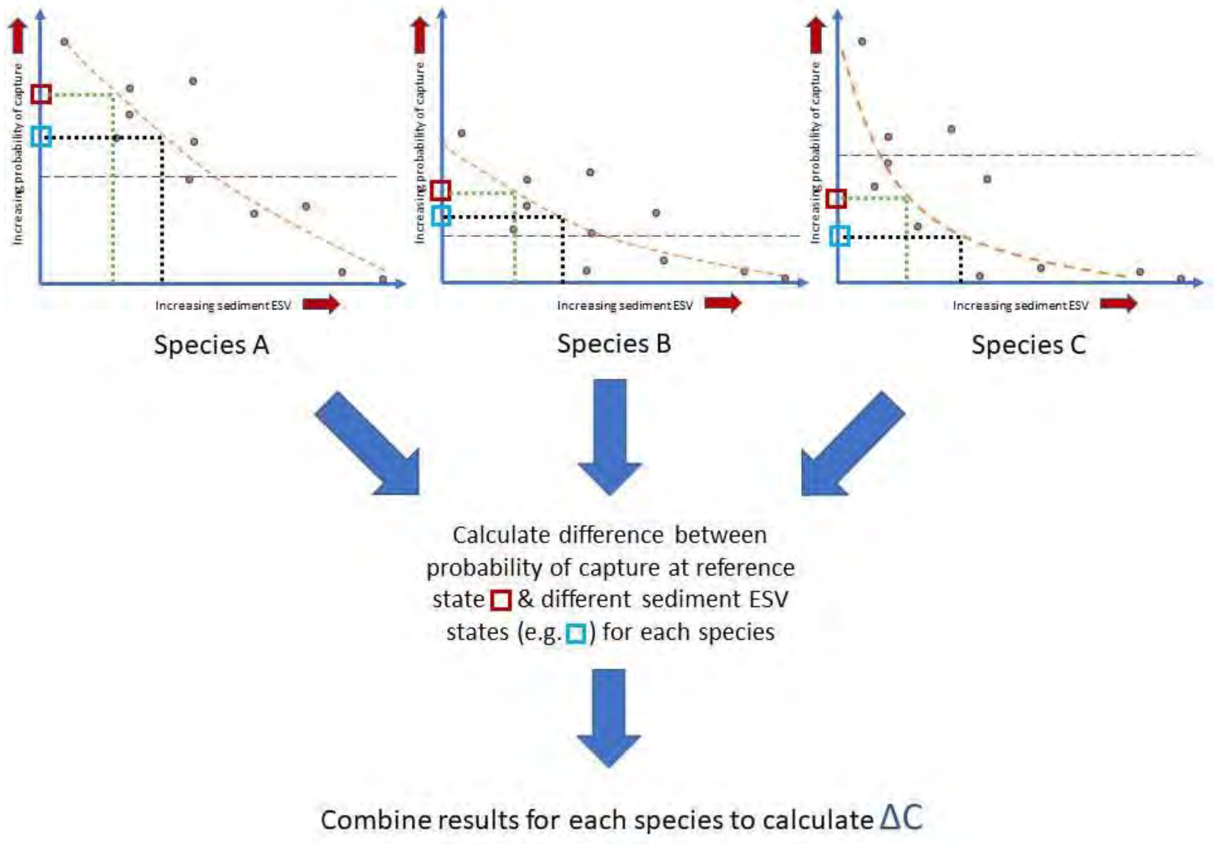


Figure 4-8: Summary of key steps involved in calculating ΔC. For a detailed explanation see Figure J-15 and associated text in Appendix J.

ΔC is always zero at the reference sediment ESV state, because it represents deviation from the community expected at reference conditions. Negative values in ΔC represent a net loss in the community composition relative to reference conditions. Positive values in ΔC represent net gains in community composition across species relative to reference conditions. ΔC, therefore, represents a deviation in community integrity relative to reference conditions.

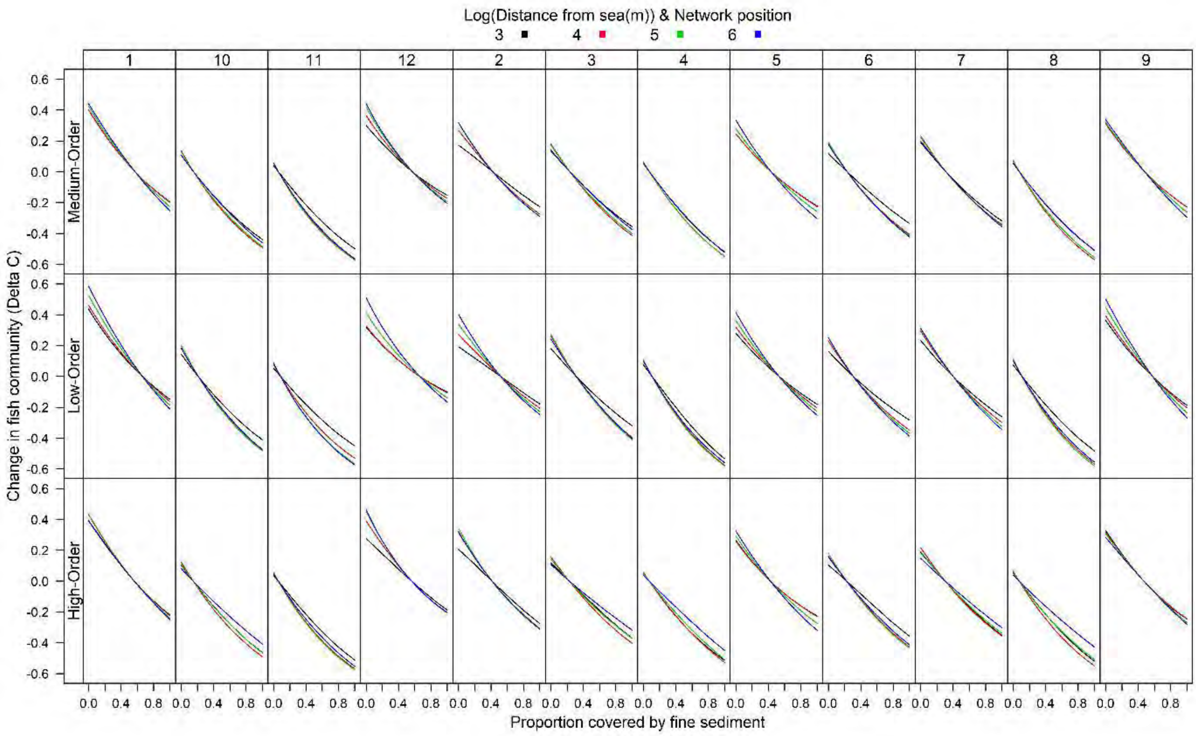


Figure 4-9: Response of the fish community deviation index, ΔC, to increasing proportion of the substrate covered by fine sediment. The individual lines in each colour represent the combinations of stream order and distance inland. Reference condition occurs when ΔC = 0. Negative values represent a decline in community integrity from reference condition.

4.3 Results: Deposited sediment

Here we report the results of the different analyses used to characterise ecosystem responses to increasing cover of deposited sediment. We found that macroinvertebrate communities were more sensitive to elevated deposited sediment than fish communities. Consequently, we recommend using macroinvertebrate responses (Section 4.3.2) as the basis for deriving bottom-lines and attribute bands for deposited sediment. The weight of evidence assessment (Section 5) resulted in the community deviation method being identified as the preferred method for basing thresholds on. Table 4-4 shows the results of the macroinvertebrate community deviation analysis that are used as the basis of defining the attribute tables presented in Section 6.2.

4.3.1 Fish

Calculated ΔC values from the community deviation method were used as the basis of deriving ESV bands that could potentially inform the development of the sediment NOF attribute. Because ΔC is a gradient response, as opposed to a threshold response, a risk-based approach was utilised to evaluate band thresholds. The greater the reduction in ΔC from reference, the greater the risk to fish community integrity. Consequently, increasing departure from reference state was considered to increase the risk of negative outcomes for fish communities. A 20% departure in fish community integrity from average reference state (ΔC = -0.20) was selected as the threshold for defining potential C/D bottom-line values. Similarly, intermediate deviations in ΔC of -0.066 and -0.133 (i.e., equally distributed between reference and the C/D threshold) were used to derive potential A/B and B/C band thresholds respectively. These were normative decisions and other ΔC values could be used to define thresholds.

Potential band thresholds for deposited fine sediment based on the fish community deviation (with a C/D threshold defined as ΔC = -20%) are presented for the 12 classes at the fourth level of aggregation in the SSC in Table 4-1.

Table 4-1: Potential band thresholds for deposited sediment based on the fish community deviation method. Thresholds are presented as proportions of the bed covered by fine sediment for the 12 classes at the fourth level of aggregation in the SSC. NA indicates that the thresholds exceed the maximum value of 1.0. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	Proportion of fine sediment cover		
		A/B threshold	B/C threshold	C/D threshold
L4.1	0.79	0.92	NA	NA
L4.2	0.04	0.09	0.15	0.21
L4.3	0.33	0.42	0.51	0.61
L4.4	0.07	0.12	0.17	0.23
L4.5	0.74	0.88	NA	NA
L4.6	0.22	0.32	0.42	0.54
L4.7	0.34	0.44	0.55	0.67
L4.8	0.13	0.22	0.33	0.45
L4.9	0.43	0.56	0.70	0.85
L4.10	0.09	0.15	0.22	0.29
L4.11	0.69	0.81	0.94	NA
L4.12	0.20	0.27	0.36	0.45

4.3.2 Macroinvertebrates

BRT analyses

A total of 996 samples across 558 sites were used to run the national BRT analysis using the % cover instream data. The ability to run independent BRT analyses within classes was limited by the number of macroinvertebrate-sediment observations available in each class at the different levels of SSC aggregation. At the 50% dissimilarity level (i.e., Level 1), there were sufficient data to proceed with BRT analyses in only one of the two classes (L1.2). At the 30% dissimilarity level (i.e., Level 2), there were sufficient data to proceed with flexible regression in two of the four classes (L2.2 and L2.3). At the 20% dissimilarity level (i.e., Level 3), there appeared to be sufficient data to proceed with flexible regression in four of the eight classes; however, exploratory analysis revealed poor model performance probably due to low sample numbers. Consequently, analyses were conducted on data in classes 2 and 3 at second level of aggregation only (i.e., L2.2 and L2.3).

Overall the BRT model fit ranged from 29% to 71% TDE (total deviance explained), and the CV correlation coefficient ranged from 0.52 to 0.83, indicating good predictive performance (Table 4-2). Deposited sediment was also a more important predictor of macroinvertebrate metrics in class L2.2 than L2.3.

Table 4-2: BRT model fit (TDE, total deviance explained) and mean CV correlation coefficient; CV=cross-validation for % deposited sediment instream.

Macroinvertebrate metric	Class	TDE (%)	CV correlation coefficient	Rank relative importance of deposited sediment	Rank relative importance of chlorophyll-*a*
MCI	Global	64	0.80	3	7
'EPT taxon richness'		60	0.77	1	5
'Sediment MCI'		69	0.83	2	11
'No. of decreasers'		71	0.83	2	3
MCI	L2.2	57	0.75	1	5
'EPT taxon richness'		57	0.74	1	5
'Sediment MCI'		69	0.82	1	10
'No. of decreasers'		69	0.82	2	3
MCI	L2.3	50	0.71	15	3
'EPT taxon richness'		29	0.52	10	1
'Sediment MCI'		50	0.71	3	8
'No. of decreasers'		55	0.73	13	1

The partial plot for the global BRT model showed similar response shapes for the four different metrics (Figure 4-10). Visual inspection of the plot indicates that marked changes in the metrics do not occur until about 30% sediment cover, after which metrics continued to decline up to 100% sediment cover. 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. The upper end of the relationship (i.e., higher sediment cover) where the greatest response is observed may, therefore, be strongly influenced by relatively few data points.

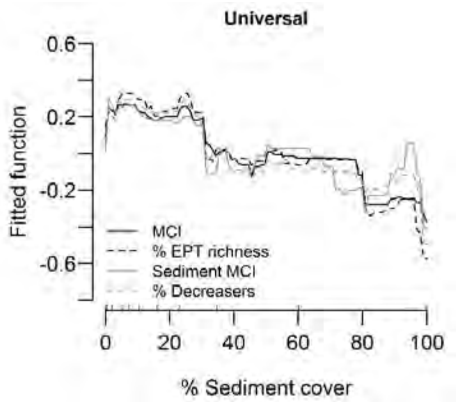


Figure 4-10: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of '% cover instream' at a national level. Note that the y-axis shows change from mean response values in units of standard deviation.

The partial dependence plots within classes further illustrate the relative effect of deposited sediment on macroinvertebrate metrics (Figure 4-11). Metrics show a strong negative response to deposited sediment in class L2.2 where deposited sediment was identified as an important predictor, but not in class L2.3 where it was not. Visual inspection of the partial dependence plots for class L2.2 shows a non-linear decrease in macroinvertebrate metric values from about 30% sediment cover through to 100% sediment cover similar to that observed for the global model. For class L2.3, there is a discernible decrease within the dominant distribution of data across the sediment gradient, but the magnitude of the signal is significantly smaller than in class L2.2. 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 and below 20% cover in class L2.3. An initial increase in metric values is discernible for both class L2.2 and class L2.3, up to approximately 10% and 5% sediment cover respectively, where a large proportion of the sample data are distributed.

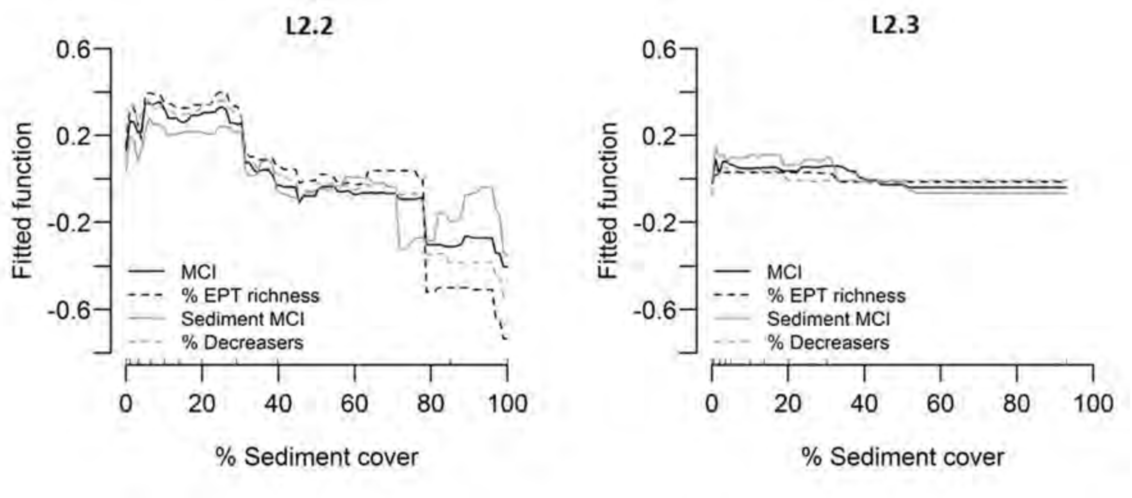


Figure 4-11: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of % sediment cover. Results are presented for class 2 and 3 at the second level of the SSC aggregation (i.e., L2.2 and L2.3).

GLM analyses

Patterns predicted by the GLM models generally conformed to expected relationships between macroinvertebrate metrics and ESV, with higher metric scores in "cleaner" conditions. The 'number of sediment decreasers' metric resulted in more environmentally conservative thresholds than the two MCI metrics (Table 4-3). This reflects the fact that MCI and Sediment MCI appear to be less sensitive to increasing deposited sediment. This is consistent with results from exploratory analyses using quantile regression reported in Depree et al. (2018). Consequently, the slope of the relationship between the metric and ESV was relatively flat and C/D thresholds often sat at the 'dirtiest' ESV state (e.g., 100% cover of fine sediment).

Table 4-3: Summary of potential ESV thresholds for total fines (% cover) for the 12 class SSC. Thresholds are defined at Level 4 of the SSC. NA values indicate insufficient data. Where the response curve is not intercepted by the deviation value, the maximum observed value is used (i.e., 100%).

Metric	Class	Predicted reference	A/B threshold	B/C threshold	C/D threshold
MCI	L4.1	79%	100%	100%	100%
MCI	L4.2	4%	17%	32%	49%
MCI	L4.3	33%	57%	72%	79%
MCI	L4.4	7%	21%	36%	53%
MCI	L4.5	74%	100%	100%	100%
MCI	L4.6	22%	79%	86%	93%
MCI	L4.7	34%	77%	88%	100%
MCI	L4.8	13%	NA	NA	NA
MCI	L4.9	43%	74%	100%	100%
MCI	L4.10	9%	50%	73%	80%
MCI	L4.11	69%	100%	100%	100%
MCI	L4.12	20%	62%	83%	88%
Sediment MCI	L4.1	79%	82%	86%	90%
Sediment MCI	L4.2	4%	13%	22%	31%
Sediment MCI	L4.3	33%	39%	45%	50%
Sediment MCI	L4.4	7%	16%	24%	32%
Sediment MCI	L4.5	74%	78%	81%	85%
Sediment MCI	L4.6	22%	27%	32%	37%
Sediment MCI	L4.7	34%	39%	43%	48%
Sediment MCI	L4.8	13%	NA	NA	NA
Sediment MCI	L4.9	43%	47%	50%	55%
Sediment MCI	L4.10	9%	18%	26%	34%
Sediment MCI	L4.11	69%	73%	77%	82%
Sediment MCI	L4.12	20%	26%	32%	39%
Sediment decreasers	L4.1	79%	100%	100%	100%
Sediment decreasers	L4.2	4%	15%	28%	43%
Sediment decreasers	L4.3	33%	75%	88%	90%
Sediment decreasers	L4.4	7%	19%	32%	47%
Sediment decreasers	L4.5	74%	96%	100%	100%
Sediment decreasers	L4.6	22%	30%	39%	49%
Sediment decreasers	L4.7	34%	43%	54%	65%
Sediment decreasers	L4.8	13%	NA	NA	NA
Sediment decreasers	L4.9	43%	63%	86%	100%
Sediment decreasers	L4.10	9%	57%	71%	78%
Sediment decreasers	L4.11	69%	NA	NA	NA
Sediment decreasers	L4.12	20%	63%	80%	82%

Community deviation analyses

A 20% departure in macroinvertebrate community integrity from average reference state (ΔC = -0.20) was again selected as the threshold for defining potential C/D bottom-line values. Similarly, intermediate deviations in ΔC of -0.066 and -0.133 (i.e., equally distributed between reference and the C/D threshold) were used to derive potential A/B and B/C band thresholds respectively. Other deviations could be used to define thresholds.

Potential band thresholds for deposited fine sediment based on the macroinvertebrate community deviation (with a C/D threshold defined as ΔC = -20%) are presented for the 12 classes at the fourth level of aggregation in the SSC in Table 4-4.

It should be noted that for macroinvertebrates there is a mismatch between the deposited sediment ESV measure used to define reference state (NZFFD % total fines) and the deposited sediment ESV measure used to describe the stressor-response relationship (SAM2 % cover instream). In the absence of paired observations of NZFFD % total fines and SAM2 % cover instream required to develop a conversion factor, for the purposes of this analysis we assume that the two measures are correlated and equivalent (see Appendix N for more discussion of this).

Table 4-4: Potential band thresholds for deposited sediment based on the macroinvertebrate community deviation method. Thresholds are presented as proportions of the bed covered by fine sediment for the 12 classes at the fourth level of aggregation in the SSC. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	Proportion of fine sediment cover		
		A/B threshold	B/C threshold	C/D threshold
L4.1	0.79	0.84	0.90	0.97
L4.2	0.04	0.09	0.15	0.21
L4.3	0.33	0.42	0.50	0.60
L4.4	0.07	0.12	0.17	0.23
L4.5	0.74	0.80	0.86	0.92
L4.6	0.22	0.30	0.38	0.46
L4.7	0.34	0.41	0.48	0.56
L4.8	0.13	0.22	0.33	0.45
L4.9	0.43	0.48	0.54	0.61
L4.10	0.09	0.15	0.22	0.29
L4.11	0.69	0.76	0.82	0.89
L4.12	0.20	0.27	0.36	0.45

4.3.3 Key findings

Deposited sediment data have not been routinely collected in New Zealand, which limited the scope and range of techniques that we could utilise for analyses of this ESV. Based on the analyses that were undertaken, we are confident that both macroinvertebrate and fish communities as a whole respond negatively to increases in the cover of deposited fine sediment. However, response trajectories vary between individual species and in the magnitude of the response across different landscape settings.

Slightly different measures of deposited sediment had to be used for the fish and macroinvertebrate analyses - NZFFD % fines and SAM2 % cover instream respectively. Ideally, analyses of different ecological endpoints would be carried out relative to the same driver variable, but paired observations (i.e., taken from the same place at the same time) for fish were not available with the SAM2 data and vice versa for macroinvertebrates and NZFFD records. No paired observations of NZFFD % fines and SAM2 % cover instream are available to determine the equivalence of the two measures, but there is a strong, almost 1:1 correlation, between SAM2 % cover instream and SAM1 % cover bankside, which is a very similar measure of deposited sediment cover to the NZFFD % fines method (see Appendix N). We therefore consider it reasonable to assume that the two different measures are sufficiently equivalent to allow direct comparison between results for both fish and macroinvertebrates.

We found that stressor-specific macroinvertebrate community metrics (e.g., sediment MCI) were more sensitive to increases in deposited fine sediment cover than generic macroinvertebrate community metrics (e.g., MCI). Furthermore, macroinvertebrates communities were more sensitive to elevated fine sediment cover than fish communities. The community deviation method also resulted in more protective threshold values than did the GLM analyses of macroinvertebrate community metrics (Figure 4-12).

Both the GLM and community deviation methods can be utilised to identify thresholds for attribute bands in the form of a deviation from reference state. However, this is dependent on a subjective decision on the 'acceptable' magnitude of deviation from the reference state and the intermediate cuts that equate to the A/B and B/C band thresholds. Results are presented here for a 20% reduction in the ecological response variable compared to reference state (for defining a C/D threshold). This was considered by the team to be a reasonable balance between not being overly permissive (i.e., allowing a larger change) or overly restrictive (i.e., limiting to a smaller change). Intermediate band thresholds were derived at equidistant cuts between the C/D threshold and reference state. This was a pragmatic solution that requires further consideration before final implementation. Different deviations could be used for determining all thresholds resulting in either more restrictive or permissive thresholds (see Appendix K for examples).

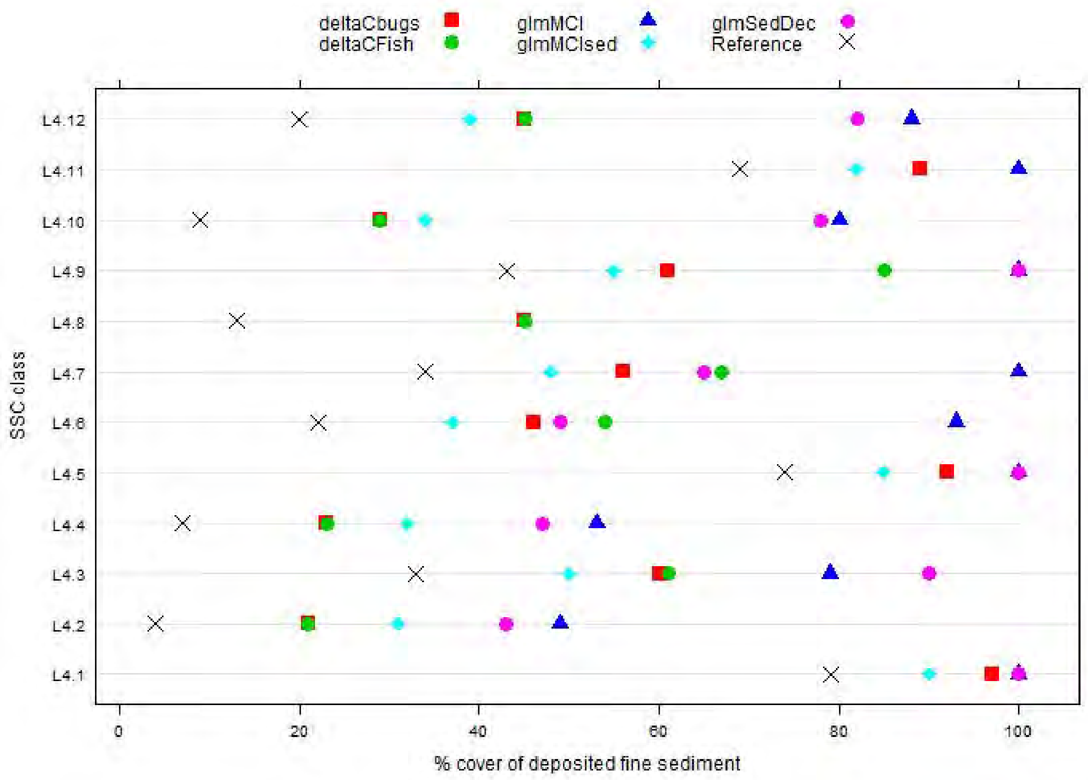


Figure 4-12: Comparison of derived C/D thresholds from each of the analytical methods for deposited sediment. The predicted reference state for each class is indicated by the black cross. For details of each method see the main text. deltaC = community deviation method, glm = generalised linear modelling method. All thresholds were derived using a 20% deviation from predicted reference state.

4.4 Results: Suspended sediment

Here we report the results of the different analyses used to characterise ecosystem responses to increasing cover of suspended sediment. We found that fish communities were more sensitive to elevated suspended sediment than macroinvertebrate communities. Consequently, we recommend using fish responses (Section 4.4.1) as the basis for deriving bottom-lines and attribute bands for the suspended sediment ESVs. The weight of evidence assessment (Section 5) resulted in the community deviation method being identified as the preferred method for basing thresholds on. Table 4-5 and Table 4-6 show the results of the fish community deviation analysis that are used as the basis of defining the attribute tables for turbidity and visual clarity, respectively, that are presented in Sections 6.3 (turbidity) and 6.4 (visual clarity).

4.4.1 Fish

A 20% departure in fish community integrity from average reference state (ΔC = -0.20) was again selected as the threshold for defining potential C/D bottom-line values. Similarly, intermediate deviations in ΔC of -0.066 and -0.133 were used to derive potential A/B and B/C band thresholds respectively.

Potential band thresholds for suspended fine sediment based on the fish community deviation are presented for the 12 classes at the fourth level of aggregation in the SSC in Table 4-5 for turbidity and Table 4-6 for visual clarity.

Table 4-5: Potential band thresholds for turbidity based on the fish community deviation method. Thresholds are presented turbidity (NTU) for the 12 classes at the fourth level of aggregation in the SSC. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	Turbidity (NTU)		
		A/B threshold	B/C threshold	C/D threshold
L4.1	1.62	1.99	2.50	3.21
L4.2	4.90	6.15	7.90	10.45
L4.3	1.08	1.32	1.62	2.02
L4.4	2.73	3.25	3.93	4.83
L4.5	5.93	7.53	9.79	13.11
L4.6	3.83	4.84	6.25	8.29
L4.7	1.99	2.33	2.76	3.32
L4.8	3.63	4.33	5.23	6.42
L4.9	1.00	1.16	1.35	1.60
L4.10	0.90	1.05	1.25	1.49
L4.11	0.88	1.06	1.28	1.56
L4.12	2.16	2.43	2.74	3.14

Table 4-6: Potential band thresholds for visual clarity based on the fish community deviation method. Thresholds are presented visual clarity (m) for the 12 classes at the fourth level of aggregation in the SSC. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	Visual clarity (m)		
		A/B threshold	B/C threshold	C/D threshold
L4.1	2.65	2.25	1.88	1.55
L4.2	2.86	2.43	2.02	1.65
L4.3	1.72	1.45	1.21	1.00
L4.4	1.66	1.43	1.22	1.02
L4.5	0.80	0.66	0.53	0.42
L4.6	1.27	1.06	0.87	0.70
L4.7	2.05	1.78	1.53	1.30
L4.8	0.74	0.63	0.53	0.44
L4.9	3.52	3.10	2.71	2.35
L4.10	3.86	3.38	2.93	2.51
L4.11	3.28	2.84	2.43	2.06
L4.12	3.09	2.79	2.51	2.23

4.4.2 Macroinvertebrates

BRT analyses

A total of 4005 samples across 665 sites were used to run the global BRT analysis using turbidity data. The ability to run independent BRT analyses within classes was again limited by the number of macroinvertebrate-sediment observations available in each class at the different levels of SSC aggregation. At SSC Levels 1 and 2 there were sufficient paired macroinvertebrate-turbidity data to proceed with BRT analyses in all classes. At SSC Level 3 there were enough data to proceed with flexible regression in five of the eight classes. For consistency with the deposited sediment analyses, subsequent BRT models were developed using data grouped at SSC Level 2.

Overall the BRT model fit ranged from 48% to 76% TDE (total deviance explained), and the CV correlation coefficient ranged from 0.68 to 0.87, indicating good to very good predictive performance (Table 4-7). Turbidity was more important than chlorophyll *a* as a predictor of macroinvertebrate metrics in all classes except L2.3, where Sediment MCI was still more strongly driven by turbidity than chlorophyll *a*, but not the other three metrics (Table 4-7).

Table 4-7: BRT model fit (TDE, total deviance explained) and mean CV correlation coefficient; CV=cross-validation for turbidity.

Macroinvertebrate metric	Class	TDE (%)	CV correlation coefficient	Rank relative importance of turbidity[11]	Rank relative importance of chlorophyll-*a*[11]
MCI	Global	73	0.85	8	6
'EPT taxon richness'		63	0.79	8	6
'Sediment MCI'		62	0.79	7	17
'No. of decreasers'		71	0.84	7	4
MCI	L2.1	76	0.87	3	14
'EPT taxon richness'		66	0.82	1	14
'Sediment MCI'		48	0.69	2	12
'No. of decreasers'		73	0.86	1	11
MCI	L2.2	65	0.81	9	13
'EPT taxon richness'		59	0.77	12	16
'Sediment MCI'		50	0.71	6	16
'No. of decreasers'		66	0.81	8	14
MCI	L2.3	74	0.86	12	5
'EPT taxon richness'		58	0.76	8	3
'Sediment MCI'		72	0.85	5	11
'No. of decreasers'		68	0.82	9	5
MCI	L2.4	57	0.74	4	9
'EPT taxon richness'		48	0.68	6	7
'Sediment MCI'		54	0.72	6	15
'No. of decreasers'		55	0.73	4	2

The partial plot for the global BRT model showed that the four macroinvertebrates metrics responded similarly to turbidity, although the sediment MCI metric varied somewhat from the others (Figure 4-13). Visual inspection of the plot indicates an immediate negative response of metrics to increasing turbidity that continues across the full turbidity gradient, with Sediment MCI exhibiting a lower slope. However, it is noted that approximately 90% of the data used to build the model occur in the range of 0-10 NTU. Therefore, any response after 10 NTU may be strongly influenced by relatively few data points.

[11] From a total of 16 variables

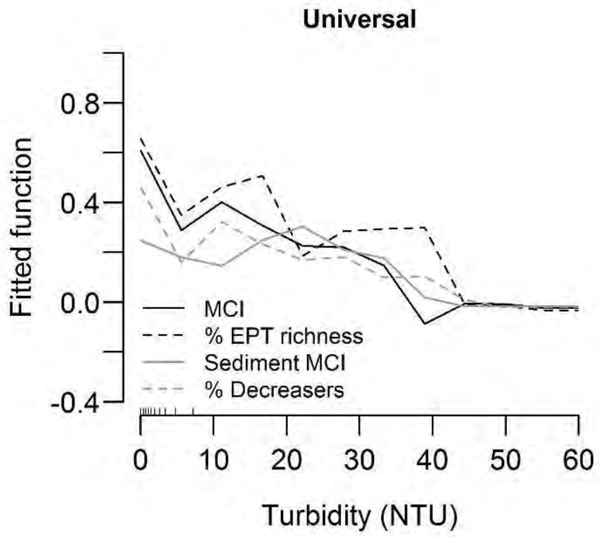


Figure 4-13: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of turbidity at a national scale. Note that the y-axis shows change from mean response values in units of standard deviation.

Despite strong model predictive performance and the relatively high importance of turbidity as a predictor variable, the partial dependence plots illustrate inconsistent responses of macroinvertebrate metrics to the turbidity gradient within classes (Figure 4-14). The majority of data are distributed at below 10 NTU in all classes, and a consistent negative response in this turbidity range is only observed for class L2.4. Visual inspection of the partial dependence plots does not provide impact initiation or cessation thresholds. A lack of response after approximately 20 NTU in all classes is likely due to a lack of data in this range.

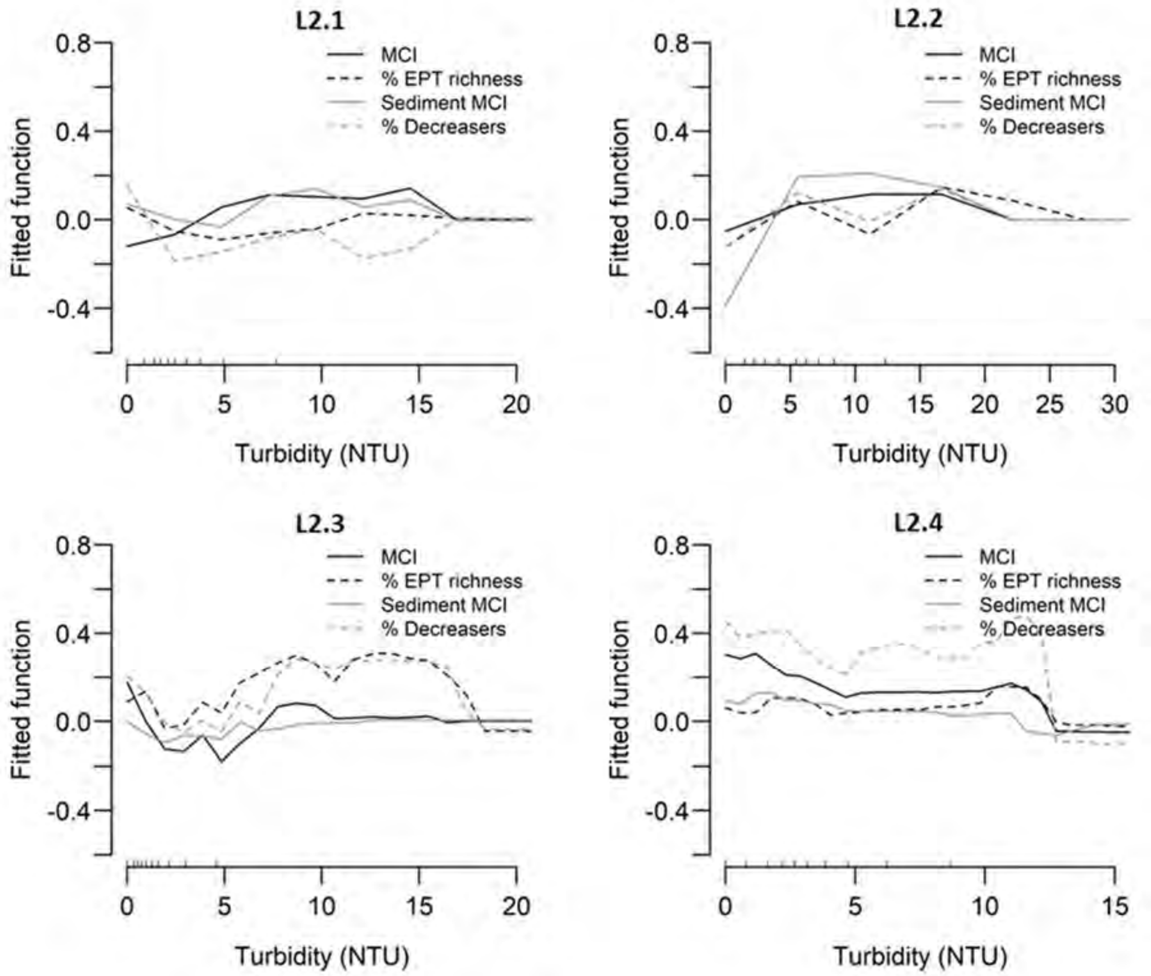


Figure 4-14: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of turbidity. Plots are shown for all four classes at the second level of aggregation in the suspended sediment SSC. The x-axis range is truncated to only show the gradient where the fitted function ≠ zero. Note the different ranges of the x-axis between plots.

Quantile regression analyses

MCI, QMCI and %EPT were modelled with log-linear quantile regressions. The other metrics were fitted with a subsidy/stress Ricker model. An example of the quantile regression relationships for MCI for a range of quantiles (99%, 95%, 90%, 80% and 50%) is shown in Figure 4-15 for both visual clarity and turbidity. Summary tables of the visual clarity and turbidity thresholds corresponding with a 30% reduction from the ESV state at the maxima of the biotic response are shown in Table 4-8 and Table 4-9 respectively.

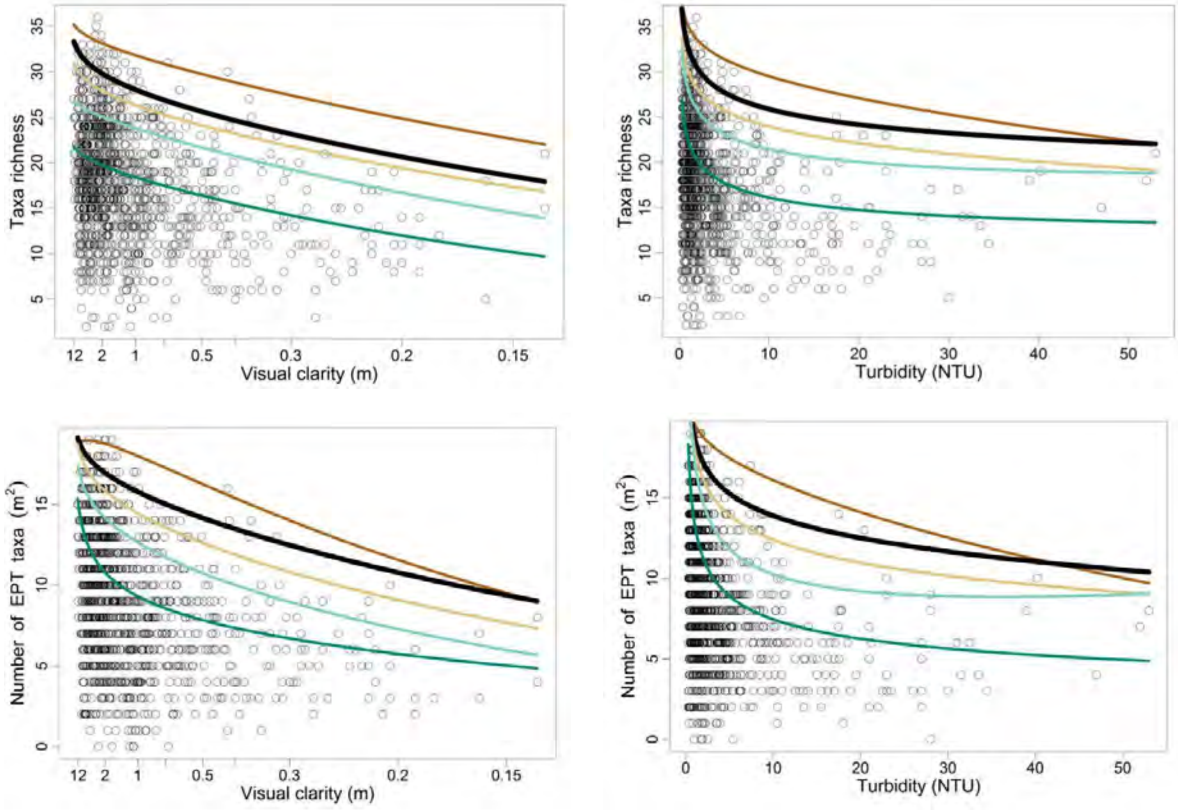


Figure 4-15: Example of fitted quantile regressions for the macroinvertebrate community index values (taxa richness, number of EPT taxa) versus visual clarity (left) and turbidity (right). Based on NRWQN data. Quantile lines are plotted for the 99% (brown), 95% (black), 90% (light brown), 80% (light turquoise) and 50% (turquoise) percentiles.

Table 4-8: Summary of 30% effect thresholds for visual clarity based on the 95th percentile quantile relationships. Threshold based on reduction from a nominal reference value of 12 m visual clarity. The blue highlighted variables are derived from log-linear regressions.

Biotic variable	Departure from reference or biotic maxima	Visual clarity threshold for 30% reduction (m)
Taxa richness	Reference	0.26
Density	Maxima	0.33
MCI	Reference	<0.15
QMCI	Reference	<0.15
EPT taxa	Reference	0.33
EPT individuals	Maxima	0.52
%EPT	Reference	<0.15

Table 4-9: Summary of 30% effect thresholds for turbidity based on the 95th percentile quantile relationships. Threshold based on reduction from a nominal reference value of 0.5 NTU turbidity. The blue highlighted variables are derived from log-linear regressions. NA indicates model fit not suitable for use in effects determination.

Biotic variable	Departure from reference or biotic maxima	Turbidity threshold for 30% reduction (NTU)
Taxa richness	Reference	17.0
Density	Maxima	19.0
MCI	Reference	>50
QMCI	Reference	>50
EPT taxa	Reference	8.2
EPT individuals	Maxima	12.2
%EPT	Reference	NA

GLM analyses

Patterns predicted by the GLM models generally conformed to expected relationships between macroinvertebrate metrics and ESV with higher metric scores in “cleaner” conditions. The ‘number of sediment decreasers’ metric again resulted in more environmentally conservative thresholds than the two MCI metrics (Table 4-10 and Table 4-11), reflecting the relatively flat slope of the relationship between the MCI metrics and ESVs. As a consequence, the C/D thresholds derived from the two MCI metrics were equivalent to the ‘dirtiest’ observed ESV state for several classes (e.g., L4.1). All three metrics were less sensitive to the suspended sediment ESVs than to the deposited sediment ESVs.

Table 4-10: Summary of potential ESV thresholds for turbidity (NTU) for the 12 class SSC. Thresholds are defined at Level 4 of the SSC. NA values indicate insufficient data. Where the response curve is not intercepted by the deviation value, the maximum observed value is used (i.e., 562.3 NTU).

Metric	Class	Predicted reference	Turbidity (NTU)		
			A/B threshold	B/C threshold	C/D threshold
MCI	L4.1	1.6	9.6	70.8	562.3
MCI	L4.2	4.9	11.4	28.9	81.6
MCI	L4.3	1.1	562.3	562.3	562.3
MCI	L4.4	2.7	16.6	125.3	562.3
MCI	L4.5	5.9	13.8	35.3	100.5
MCI	L4.6	3.8	NA	NA	NA
MCI	L4.7	2.0	41.1	111.3	343.5
MCI	L4.8	3.6	562.3	562.3	562.3
MCI	L4.9	1.0	4.6	25.2	173.1
MCI	L4.10	0.9	17.2	472.5	562.3
MCI	L4.11	0.9	5.1	36.4	337.3
MCI	L4.12	2.2	13.0	98.6	442.9
Sediment MCI	L4.1	1.6	3.7	8.8	21.7
Sediment MCI	L4.2	4.9	8.8	16.5	32.4
Sediment MCI	L4.3	1.1	2.5	5.8	14.1
Sediment MCI	L4.4	2.7	8.3	25.1	77.2
Sediment MCI	L4.5	5.9	12.0	25.0	54.6
Sediment MCI	L4.6	3.8	8.2	18.0	41.4
Sediment MCI	L4.7	2.0	6.5	21.1	67.7
Sediment MCI	L4.8	3.6	7.9	18.0	42.4
Sediment MCI	L4.9	1.0	3.9	14.6	53.5
Sediment MCI	L4.10	0.9	2.9	9.5	31.0
Sediment MCI	L4.11	0.9	2.9	9.3	30.2
Sediment MCI	L4.12	2.2	10.6	47.3	201.7
Sediment decreasers	L4.1	1.6	22.6	442.5	562.3
Sediment decreasers	L4.2	4.9	6.6	9.3	13.4
Sediment decreasers	L4.3	1.1	NA	NA	NA
Sediment decreasers	L4.4	2.7	39.3	562.3	562.3
Sediment decreasers	L4.5	5.9	8.1	11.3	16.4
Sediment decreasers	L4.6	3.8	NA	NA	NA
Sediment decreasers	L4.7	2.0	16.3	174.6	562.3
Sediment decreasers	L4.8	3.6	NA	NA	NA
Sediment decreasers	L4.9	1.0	9.2	111.5	562.3
Sediment decreasers	L4.10	0.9	NA	NA	NA
Sediment decreasers	L4.11	0.9	12.0	228.7	562.3
Sediment decreasers	L4.12	2.2	9.7	52.1	204.5

Table 4-11: Summary of potential ESV thresholds for visual clarity (m) for the 12 class SSC. Thresholds are defined at Level 4 of the SSC. NA values indicate insufficient data. Where the response curve is not intercepted by the deviation value, the minimum observed value is used (i.e., 0.02 m).

Metric	Class	Predicted reference	Visual clarity (m)		
			A/B threshold	B/C threshold	C/D threshold
MCI	L4.1	2.65	0.35	0.04	0.02
MCI	L4.2	NA	NA	NA	NA
MCI	L4.3	NA	NA	NA	NA
MCI	L4.4	1.66	0.69	0.26	0.09
MCI	L4.5	0.80	0.11	0.03	0.02
MCI	L4.6	1.27	0.13	0.02	0.02
MCI	L4.7	2.05	0.51	0.11	0.05
MCI	L4.8	0.74	0.08	0.02	0.02
MCI	L4.9	3.52	1.61	0.67	0.25
MCI	L4.10	3.86	NA	NA	NA
MCI	L4.11	3.28	1.41	0.54	0.19
MCI	L4.12	3.09	1.22	0.43	0.13
Sediment MCI	L4.1	2.65	2.06	1.58	1.21
Sediment MCI	L4.2	NA	NA	NA	NA
Sediment MCI	L4.3	NA	NA	NA	NA
Sediment MCI	L4.4	1.66	1.21	0.89	0.64
Sediment MCI	L4.5	0.80	0.65	0.53	0.42
Sediment MCI	L4.6	1.27	1.02	0.81	0.64
Sediment MCI	L4.7	2.05	1.45	1.02	0.72
Sediment MCI	L4.8	0.74	0.60	0.48	0.38
Sediment MCI	L4.9	3.52	2.38	1.63	1.12
Sediment MCI	L4.10	3.86	2.09	1.20	0.72
Sediment MCI	L4.11	3.28	2.30	1.62	1.14
Sediment MCI	L4.12	3.09	1.75	1.04	0.64
Sediment decreasers	L4.1	2.65	1.15	0.45	0.15
Sediment decreasers	L4.2	NA	NA	NA	NA
Sediment decreasers	L4.3	NA	NA	NA	NA
Sediment decreasers	L4.4	1.66	0.38	0.07	0.02
Sediment decreasers	L4.5	0.80	0.43	0.21	0.10
Sediment decreasers	L4.6	1.27	0.59	0.25	0.09
Sediment decreasers	L4.7	2.05	0.20	0.02	0.02
Sediment decreasers	L4.8	0.74	0.06	0.02	0.02
Sediment decreasers	L4.9	3.52	0.65	0.10	0.04
Sediment decreasers	L4.10	3.86	NA	NA	NA
Sediment decreasers	L4.11	3.28	0.78	0.15	0.02
Sediment decreasers	L4.12	3.09	0.45	0.05	0.02

Extirpation analyses

Sufficient data were available to derive extirpation-based thresholds for five of the eight classes at Level 3 of the suspended SSC for both turbidity (Table 4-12) and visual clarity (Table 4-13). Thresholds for those classes with insufficient data were determined using the most similar class in the hierarchical SSC. For further details of the results using this method see Appendix H.

Table 4-12: Threshold turbidity values (NTUs) at which 1%, 2.5% and 5% of macroinvertebrate taxa are extirpated from the community within suspended sediment classes. Classes are at Level 3 of the suspended sediment classification; sufficient data for four of eight classes; classes with insufficient data indicated by an asterisk, and thresholds assigned to those based on the method described in the text.

NOF band threshold		A/B	B/C	C/D
Sed. Class		1%	2.5%	5%
L3.1*	L4.1	3.9	4.7	5.5
L3.2	L4.5 L4.8	10.0	11.2	12.3
L3.3*	L4.6	10.0	11.2	12.3
L3.4	L4.7 L4.10 L4.11	2.5	3.3	4.1
L3.5	L4.9	3.0	3.7	4.3
L3.6	L4.12	3.9	4.7	5.5
L3.7*	L4.2	3.9	4.7	5.5
L3.8	L4.3 L4.4	6.1	7.0	7.8
	Global mean	5.4	6.3	7.2

Table 4-13: Threshold visual clarity values (m) at which 1%, 2.5% and 5% of macroinvertebrate taxa are extirpated from the community within suspended sediment classes. Classes at Level 3 of the suspended sediment classification; sufficient data for four of eight classes; classes with insufficient data indicated by an asterisk, and thresholds assigned to those based on use of the method described in the text.

NOF band threshold		A/B	B/C	C/D
Sed. Class		1%	2.5%	5%
L3.1*	L4.1	1.32	1.06	0.90
L3.2	L4.5 L4.8	0.58	0.50	0.45
L3.3*	L4.6	0.58	0.50	0.45
L3.4	L4.7 L4.10 L4.11	1.97	1.47	1.16
L3.5	L4.9	1.98	1.55	1.27
L3.6	L4.12	1.32	1.06	0.90
L3.7*	L7.2	1.32	1.06	0.90
L3.8	L4.3 L4.4	0.92	0.79	0.71
	Global mean	1.25	1.00	0.84

Community deviation analyses

A 20% departure in macroinvertebrate community integrity from average reference state (ΔC = -0.20) was again selected as the threshold for defining potential C/D bottom-line values. Similarly, intermediate deviations in ΔC of -0.066 and -0.133 (i.e., equally distributed between reference and the C/D threshold) were used to derive potential A/B and B/C band thresholds respectively. Other deviations could be used to define thresholds.

Potential band thresholds for the two suspended sediment ESVs based on the macroinvertebrate community deviation (with a C/D threshold defined as ΔC = -20%) are presented for the 12 classes at the fourth level of aggregation in the SSC in Table 4-5 and Table 4-6 for turbidity and visual clarity respectively.

Table 4-14: Potential band thresholds for turbidity based on the macroinvertebrate community deviation method. Thresholds are presented using measured turbidity (NTU) for the 12 classes at the fourth level of aggregation in the SSC. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC class	Predicted reference state	Turbidity (NTU)		
		A/B threshold	B/C threshold	C/D threshold
L4.1	1.62	2.70	4.71	8.62
L4.2	4.90	8.66	16.07	31.68
L4.3	1.08	2.06	4.12	8.71
L4.4	2.73	4.76	8.70	16.86
L4.5	5.93	10.09	17.95	33.65
L4.6	3.83	6.38	11.10	20.25
L4.7	1.99	3.53	6.50	12.60
L4.8	3.63	6.90	13.85	29.63
L4.9	1.00	1.74	3.11	5.83
L4.10	0.90	1.54	2.77	5.26
L4.11	0.88	1.52	2.74	5.19
L4.12	2.16	3.79	6.94	13.45

Table 4-15: Potential band thresholds for visual clarity based on the macroinvertebrate community deviation method. Thresholds are presented using visual clarity (m) for the 12 classes at the fourth level of aggregation in the SSC. NAs occur where insufficient data were available within the class to predict reference state. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC class	Predicted reference state	Visual clarity (m)		
		A/B threshold	B/C threshold	C/D threshold
L4.1	2.65	1.90	1.32	0.89
L4.2	2.86	2.26	1.76	1.34
L4.3	1.72	1.20	0.83	0.55
L4.4	1.66	1.17	0.80	0.52
L4.5	0.80	0.58	0.41	0.28
L4.6	1.27	0.90	0.62	0.41
L4.7	2.05	1.44	0.99	0.66
L4.8	0.74	0.52	0.36	0.25
L4.9	3.52	2.48	1.72	1.15
L4.10	3.86	2.82	2.00	1.38
L4.11	3.28	2.33	1.61	1.07
L4.12	3.09	2.16	1.48	0.97

4.4.3 Key findings

Analysis of fish community responses was limited to using modelled ESV data due to the absence of paired observations (in space and time) of fish and turbidity or visual clarity. In contrast, macroinvertebrate and suspended sediment samples are frequently collected from the same place during routine state of the environment monitoring. Analysis of macroinvertebrate community responses was, therefore, undertaken using paired observations.

We found that BRT, quantile regression and GLM analyses of macroinvertebrate community metrics consistently showed that macroinvertebrate communities are less sensitive to increasing suspended sediment relative to deposited sediment (i.e., the slope of the stressor-response relationship is flatter). We also found that for both suspended sediment ESVs, fish were more sensitive than any of the macroinvertebrate metrics (Figure 4-16 and Figure 4-17). The extirpation method produced slightly more permissive thresholds than the community deviation method, but this is consistent with a theoretically more severe ecological endpoint (i.e., species disappearing vs. decreases in probability of occurrence).

All methods demonstrated significant spatial variation in the relationship between ecological response metrics and suspended sediment supporting the derivation of different thresholds for different spatially differentiated classes.

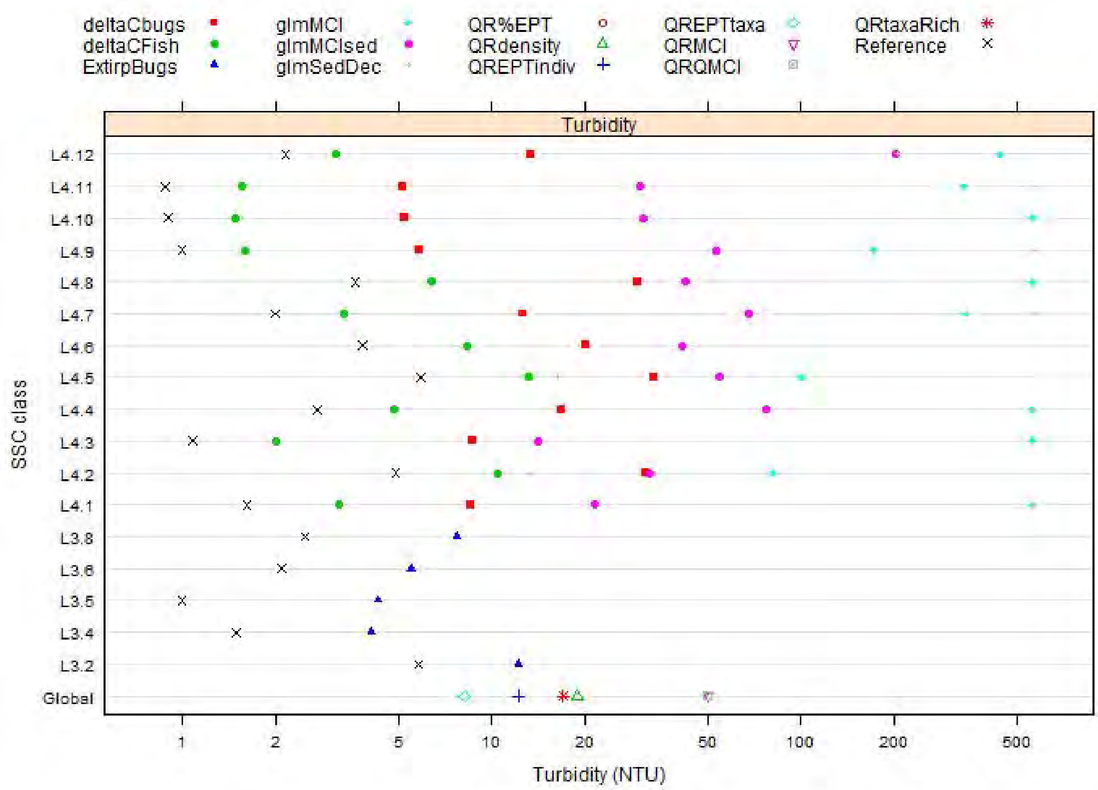


Figure 4-16: Comparison of derived C/D thresholds from each of the analytical methods for turbidity. The predicted reference state for each class is indicated by the black cross. For details of each method see the main text. deltaC = community deviation method, glm = generalised linear modelling method, QR = quantile regression method, Extirp = extirpation SSD method.

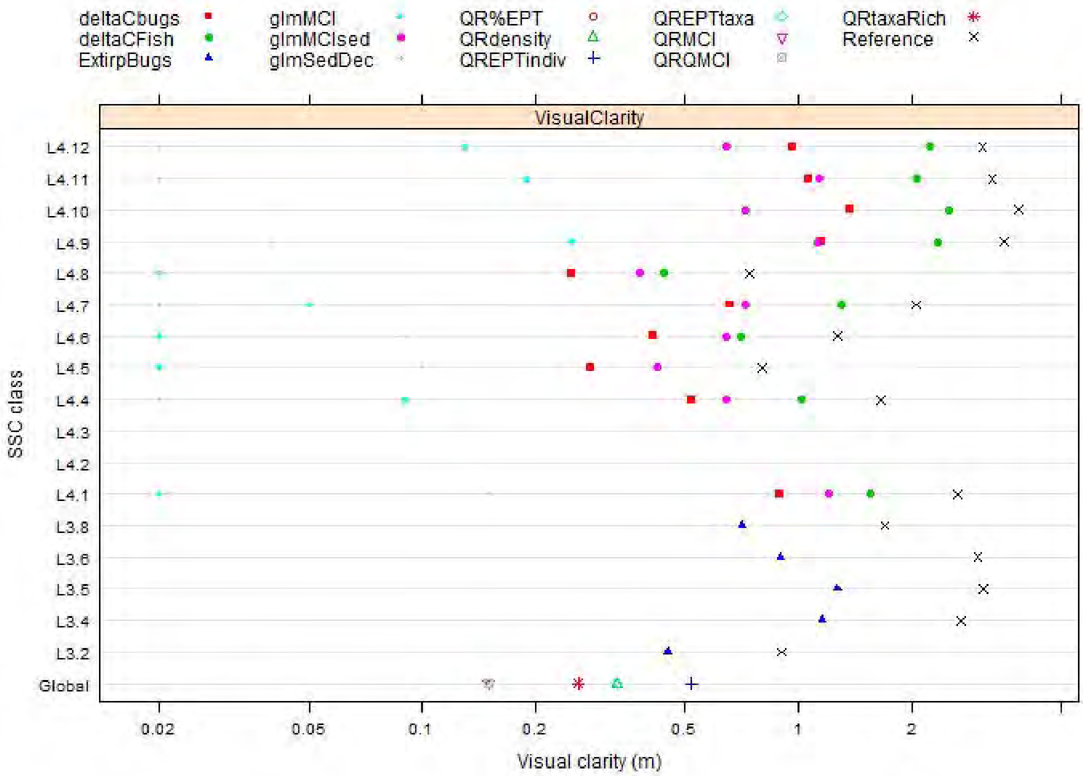


Figure 4-17: Comparison of derived C/D thresholds from each of the analytical methods for visual clarity. The predicted reference state for each class is indicated by the black cross. For details of each method see the main text. deltaC = community deviation method, glm = generalised linear modelling method, QR = quantile regression method, Extirp = extirpation SSD method.

5 Utilising multiple lines of evidence

- Multiple lines of evidence have been developed to support the development of robust and transparent fine sediment limits.
- We used a formal weight of evidence approach to evaluate the different lines of evidence (Section 5.1).
- The team assessed the community deviation method as having the greatest weight and recommended the use of this method as the basis of deriving fine sediment limits (Section 5.2 and Appendix L).

The work undertaken by Depree et al. (2018) and the workstreams continued in this project have generated multiple pieces of evidence that could be used to develop fine sediment NOF attributes. This indicates a requirement to consider a range of fine sediment ESVs, multiple impact pathways, and different indicators and measures of ecosystem health response. Generation of multiple pieces of evidence of different types is common in environmental assessments. Often none of the multiple estimates available represent the unknown true value, but each provides evidence regarding the true value (Suter et al. 2017). To determine the best-supported value of the threshold concentration or effect level it is necessary to weigh the different lines of evidence. Use of a formal framework for evaluating lines of evidence can provide greater transparency and reproducibility relative to ad hoc weighing of evidence (US EPA 2016b; Suter et al. 2017). We used the US EPA (2016b) framework for weight of evidence in ecological assessment as the basis for evaluating and combining the multiple lines of evidence for this project.

The weight of evidence approach inevitably involves subjective expert judgements. However, careful planning, use of a standardised framework, working in groups to achieve consensus, and trying to make objective and unbiased conclusions following consideration of the evidence all help to minimise the risk of achieving biased or arbitrary outcomes. A key benefit of a formal weight of evidence approach is that it allows incorporation of all relevant and reliable evidence, including expert knowledge. It can also increase the defensibility and transparency of the process and outcomes by demonstrating that all evidence has been considered and due process has been followed in evaluating and critiquing the evidence before making decisions.

5.1 Weight of evidence methodology

The weight of evidence approach involves three core steps (US EPA 2016b):

1. Assembling evidence.
2. Weighing evidence with respect to its properties.
3. Weighing the integrated body of evidence.

5.1.1 Assembling evidence

For this study it was necessary to assemble evidence that:

1. Confirms causality (i.e., criterion 1 for establishing attributes).

2. Allows identification of numeric criteria to achieve acceptable protection levels (i.e., criterion 2 for establishing attributes).

Two primary bodies of evidence were gathered to support these tasks - existing literature and new analyses of data relating fine sediment ESVs to ecosystem health responses.

The results of the literature review were reported in Depree et al. (2018) and are available in Appendix A to Appendix C of this report. The causality pathways identified from the literature review are broadly summarised in Section 2.2.2 and Figure 2-1 of this report. This body of evidence was used as the basis of evaluating modes of impact and confirming causality (i.e., fulfilling criterion 1).

Data on fine sediment ESV state and indicators of ecosystem health were subsequently collated and used to characterise exposure-response relationships to identify numeric criteria. A data driven approach was considered the most transparent and reproducible way of deriving numeric criteria. This body of evidence is described in Section 4 of this report and was subjected to the formal weight of evidence evaluation described below.

5.1.2 Weighing evidence

A formal weight of evidence approach weighs each piece of evidence to determine its degree of influence in the overall body of evidence. Each piece of evidence is given a weight based on an evaluation of defined properties and then allocated an explicit score by assessors. This provides a transparent and reproducible process by which the relative influence of different pieces of evidence are incorporated into the final assessment.

Properties to be weighted

The influence of a piece of evidence on the overall body of evidence is affected by its properties. For this study we evaluated the relevance (Table 5-1), reliability (Table 5-2), and suitability (Table 5-3). Relevance describes the correspondence between the piece of evidence and the assessment endpoint to which it is applied (US EPA 2016b). For example, an acute lethality test carried out in a laboratory on one life-stage of a single species may have low relevance for deriving thresholds for ecosystem health in the field. The assessment of reliability is based on inherent properties that make the evidence convincing (US EPA 2016b). A well designed and executed study is more convincing than one that is poorly executed. The suitability criteria were specifically derived for the purpose of this study and were used to evaluate the suitability of different pieces of evidence for deriving thresholds that are consistent with the requirements of a NOF attribute. We weighted a piece of evidence that allowed derivation of bottom-line and band thresholds that also accounted for spatial variations in natural state of the sediment ESV more strongly than a piece of evidence that could not.

Table 5-1: Description of relevance properties. Adapted from Box 5-1 in US EPA (2016b).

Relevance measure	Description
Biological	Correspondence among taxa, life stages, and processes measured or observed and the assessment endpoint.
Physical/chemical	Correspondence between the physical or chemical agent tested or measured at the study site and the physical or chemical stressor of concern.
Environmental	Correspondence between test conditions/conditions at assessed site/environmental conditions in studied system and the conditions in the region of concern.

Table 5-2: Description of reliability properties. Adapted from Box 5-3 in US EPA (2016b).

Reliability measure	Description
Design and execution	Evidence that is generated with good study design that is well performed is more reliable.
Abundance	Evidence from more numerous data is more reliable because it reflects greater replication or resolution.
Minimised confounding	Evidence is more reliable when the sampling design or analysis controls for unconnected correlates.
Specificity	Evidence specific to one cause is more reliable.
Potential for bias	Evidence from a study not funded by an interested party, is not produced for advocacy, and is not produced by an investigator with conflicts of interest is more reliable.
Standardisation	A standard method decreases the risk that evidence is biased, or analyses are inaccurate.
Corroboration	Using indicators or methods that have been verified by many studies and are accepted technical practice can increase reliability.
Transparency	Complete description of methods, logic and availability of data provides the means to check results and are presumed to increase reliability.
Peer review	An independent peer review increases reliability of a source of information.
Consistency	When evidence does not vary significantly in repeated instances within a study, it is an indicator of the reliability of a piece of evidence.
Consilience	Evidence that is shown to be consistent with accepted scientific knowledge and theory is more reliable.

Table 5-3: Description of suitability properties.

Suitability measure	Description
Bottom-line	Is the method suitable for identifying a bottom-line threshold?
Bands	Is the method suitable for identifying attribute band thresholds?
Global v classes	Can thresholds be derived for different sediment state classes?
Reproducible	Is the method of deriving thresholds easily reproducible? E.g., does it rely on expert interpretation or quantitative deviations?
Departure from reference	Can thresholds be derived as a departure from reference state using the method?

Scoring

We used a qualitative scoring system using symbols to represent the weights of each piece of evidence (Table 5-4). The use of symbols is intuitive and is preferable because it implies that they cannot be numerically combined (US EPA 2016b).

Table 5-4: Scoring system used for symbolising evidence weightings.

Symbol	Description
+++	Convincing evidence
++	Strong evidence
+	Moderate evidence
0	Weak evidence

5.1.3 Weighing the body of evidence

Once weights are allocated to each piece of evidence, there are two basic approaches to deriving quantitative thresholds from the overall body of evidence: combining the quantitative evidence or choosing the best quantitative evidence. Where the range of relevance and reliability scores are similar between pieces of evidence it may be possible to combine these numerically using meta-analysis (Suter et al. 2017). However, where the range of relevance or reliability of the different pieces of evidence is large, choosing the best value rather than combining them is preferable (US EPA 2016b; Suter et al. 2017).

5.2 Weight of evidence results

A two-day workshop was convened by the project team to evaluate all the lines of evidence that had been created through this project. The workshop was held on 26-27 November 2018 and was facilitated by an expert in the weight of evidence approach (Glenn Suter, US EPA). The workshop was attended by the project team and representatives of the Ministry for the Environment.

At the workshop we documented each of the lines of evidence created during this project to characterise the exposure-response relationship between fine sediment ESVs and indicators of ecosystem health and evaluated them for relevance, reliability, and suitability. The scoring process was documented and collated into a table of scores, presented in Appendix L. At the end of this process, the team reviewed the table of scores considering the equivalence and consistency of different pieces of evidence. There was consensus within the team regarding a lack of equivalency between pieces of evidence caused by differences in the available measures of fine sediment ESVs and the ecological indicators used. This restricted our ability to merge lines of evidence and so it was agreed that the line of evidence with the greatest weight would be used as the basis for defining thresholds. The other lines of evidence were then used as supporting information to ensure that the resulting thresholds were consistent with the overall body of evidence.

The weightiest piece of evidence was the community deviation method that was developed for this project (Table 5-5). This method had both strong relevance and reliability scores, and was also suitable for deriving numeric criteria across spatial classes and for interim attribute bands (as well as the bottom-line threshold).

Table 5-5: Summary of weight of evidence scores for the different lines of evidence. A / denotes where scores were between classes, for example */** means between a score of * and **. For details of scores see Appendix L.

	Method	Relevance	Reliability	Suitability
Deposited sediment	BRT	+/++	+/++	+
	GLM	+/++	+/++	++/+++
	Community deviation	++	++	++/+++
Suspended sediment	BRT	+	+/++	+
	QR	+	++	++
	Extirpation	++	++	++/+++
	GLM	+/++	+/++	++/+++
	Community deviation	++	++	+++

6 Defining potential fine sediment attributes

The primary purpose of this project is to provide MfE with the information required to define transparent and defensible fine sediment attribute thresholds with respect to protecting ecosystem health. The link between elevated fine sediment (both suspended and deposited) and the status of multiple aspects of ecosystem health are well documented in the literature and are also demonstrated in the results of the analyses undertaken as part of this project. Given this consensus, this project has focussed on how to determine fine sediment attributes and numeric thresholds for protecting ecosystem health. The lack of correlation between deposited and suspended sediment ESV measures and the different mechanisms hypothesised to drive ecosystem responses to each stressor, leads us to recommend that independent fine sediment attributes are defined for deposited and suspended fine sediment. This section describes the proposed fine sediment attributes and replaces Chapter 7 of Depree et al. (2018).

6.1 Guiding principles

We adopted the following guiding principles for attribute development set out by MfE.

- Te Mana o Te Wai must be recognised by putting the needs of the waterbody first; attributes contribute to how councils safeguard the life-supporting capacity of the waterbody and associated ecosystems with regard to national values.
- Prioritise recognition of the needs of indigenous species over introduced species.
- Describe bottom lines for ecosystem health in terms of ecological effects and/or departure from an estimated natural state free from alterations resulting from human activity.
- Base bottom lines on the least acceptable state of ecosystem health and/or the state prior to irreversible degradation occurring (the former is a normative and subjective judgment, the latter, given adequate information, is not).
- Note that information will never be perfect, and in the face of uncertainty and on the balance of probability, avoid potentially significant[1] adverse ecosystem effects.
- Be transparent about what the bottom line does, and does not, protect, as well as the multiple sources of evidence used in their development.

In defining potential fine sediment attribute thresholds for each fine sediment ESV we also adhered to the following criteria.

- **Define four bands, labelled A to D**. Numeric thresholds must define the boundaries between bands (i.e., A/B, B/C and C/D).
- **Account for spatial patterns in ecological distributions**. Freshwater communities vary naturally across New Zealand due to its wide variability in climatic, hydrological and physical conditions. The distributions of many native freshwater fish species in rivers are also influenced by landscape setting (including distance to the ocean, river slope and their ability to penetrate inland) because their life-histories include a marine

[1] Significance is based on irreversibility, severity, duration, frequency, and spatial extent of the effect of which the national bottom line applies.

phase. Expected spatial variations in ecological distributions must be accounted for within the thresholds.

- **Bands must be defined for each group of a previously defined Sediment State Classification (SSC)**. Sediment states also vary spatially across New Zealand, even under reference conditions. Appendix D describes how groups from the third level of the REC classification (climate-topography-geology) were combined into a smaller number of groups with similar characteristics in relation to suspended or deposited fine sediment. Thresholds for % cover of deposited fine sediment were required for each group of the 12-class deposited SSC. Thresholds for turbidity and clarity were required for each group of a 12-class suspended SSC. The 12-class level of each classification was applied as these were recommended to maximise between-group discrimination while minimising within-group differences of observed ESV distributions with respect to climate-topography-geology classes.
- **Thresholds will be defined as deviations away from reference conditions**. Reference conditions were previously calculated for each class of each SSC (Appendix D). Monotonic deviations away from these reference conditions were required to define thresholds for all four attribute bands. Thresholds for turbidity and total fines must be increases above reference conditions. Thresholds for clarity must be decreases below reference conditions. Use of a “deviations away from reference conditions” approach was necessary to ensure that the A-band was attainable in all locations and because of absence of data suitable for defining absolute effects thresholds.
- **Thresholds must be protective of the most vulnerable trophic level**. Species are known to exhibit dependencies across trophic levels. For example, trophic cascades can result when a species is removed from an otherwise stable food-web structure. Thresholds designed to maintain or improve ecosystem health must be designed to provide a quantified level of protection to the part of the ecological community that would potentially be most strongly impacted by the state of the ESV.
- **Thresholds must be developed using existing data**. One of the principles for NOF development provided by MfE describes the need to act even when data are uncertain. Thresholds must, therefore, be developed while recognising and within the constraints of existing datasets.

In each case, the results of the weight of evidence assessment were used to identify the most relevant, reliable and fit-for-purpose method. The ‘weightiest’ piece of evidence, the community deviation method, was then used as the basis for defining potential attribute thresholds in a way that was consistent with the guiding principles (e.g., prioritising native over exotic species, acting in the face of uncertainty and accounting for spatial variability). Consistency with the rest of the body of evidence was also checked to ensure the proposed thresholds were realistic.

6.2 Deposited fine sediment

The community deviation method applied to macroinvertebrate data was used as the basis for defining the deposited fine sediment attribute thresholds. This was because application of the community deviation method indicated that macroinvertebrate communities were more sensitive to increasing deposited fine sediment than fish communities and this finding was supported by the wider body of evidence. Consequently, macroinvertebrates represent the more vulnerable trophic level.

6.2.1 How is it measured?

Clapcott et al. (2011) set out protocols and guidelines for measurement of instream deposited sediment in wadeable streams and rivers in New Zealand. BRT analyses indicate that the SAM2 (instream visual) measure is a better predictor of macroinvertebrate metrics than the SAM1 (bankside visual) measure, when spatial variation is accounted for by including environmental predictors (Depree et al. 2018). Using available data we recommend, therefore, that the SAM2 instream visual method is used for measuring deposited sediment. It should, however, be noted that the SAM1 measure is strongly correlated with SAM2 (effectively a 1:1 relationship) and so is an acceptable alternative measure (see Appendix N).

The SAM2 method is a semi-quantitative measure of the relative cover of fine sediment in comparison to other substrate classes in run habitats (Clapcott et al. 2011). Application of the method is limited to wadeable streams and rivers, but could potentially be adapted for deeper streams where water clarity is sufficient to observe the stream bed from the water surface.

We accept that the median is not necessarily the most important metric in relation to ecological state. However, our analyses have focused on characterising long-term average responses (due to the type of data available) and so we propose that the site median is the most appropriate metric to represent long-term site characteristics. Analysis of temporal variation in the SAM1 and SAM2 methods (Appendix M) indicated that 24 monthly samples would enable estimation of the mean sediment cover using the SAM2 method sufficiently accurately, i.e., with a fine sediment cover margin of error of ±5%. This level of accuracy was considered appropriate given that individual observations are only considered precise to the nearest 5% due to variation between observers (Clapcott et al. 2011). We propose that the median of 24 consecutive monthly samples of % cover of fine sediment measured using the SAM2 method should be the basis for determining the attribute band for the deposited sediment attribute.

6.2.2 Where does it apply?

The proposed attribute thresholds are appropriate for all rivers and streams in New Zealand. However, the practical requirements for measurement of fine sediment cover dictate that it will primarily be applied to wadeable streams.

Some streams are naturally dominated by fine sediment and are often referred to as ‘soft-bottomed’ streams (e.g., Clapcott et al. 2011). We are not proposing inclusion of an exception for ‘soft-bottomed’ streams for two main reasons:

1. it is difficult to delineate naturally ‘soft-bottomed’ streams in the absence of reference data, and
2. the data analyses did not indicate an impact cessation threshold as deposited sediment cover increases to 100%. Consequently, continued increases in deposited fine sediment cover are expected to further impact ecosystem health, even in a ‘soft-bottomed’ stream.

6.2.3 What does it protect?

The bottom-line thresholds are anticipated to provide sufficient protection on average to avoid significant adverse effects on macroinvertebrate and fish communities. Specifically, they are defined to restrict overall decreases in the macroinvertebrate community deviation metric to <20% relative to the reference state. However, it should be recognised that the thresholds may not always be sufficient to protection specific life-stages or habitat requirements of individual species. For example, salmonid spawning habitats require <10% fine sediment cover (Kemp et al. 2011) and kākahi are impacted by substrate composition (Kusabs et al. 2015). These thresholds are also not intended to provide protection from shorter-term episodic events that may have acute impacts on specific species.

6.2.4 Proposed attribute table

Value	Ecosystem Health												
Freshwater Body Type	Rivers												
Attribute	Deposited fine sediment												
Attribute Unit	% fine sediment cover (percentage cover of the streambed in a run habitat determined by the instream visual method, SAM2)												
Attribute State	**SSC class[1]**												**Narrative Attribute State**
	1	2	3	4	5	6	7	8	9	10	11	12	
	Site median[2]												
A	<84	<9	<42	<12	<80	<30	<41	<22	<48	<15	<76	<27	Minimal impact of deposited fine sediment on instream biota. Ecological communities are similar to those observed in natural reference conditions.
B	<90	<15	<50	<17	<86	<38	<48	<33	<54	<22	<82	<36	Low to moderate impact of deposited fine sediment on instream biota. Abundance of sensitive macroinvertebrate species may be reduced.
C	≤97	≤21	≤60	≤23	≤92	≤46	≤56	≤45	≤61	≤29	≤89	≤45	Moderate to high impact of deposited fine sediment on instream biota. Sensitive macroinvertebrate and fish species may be lost.
National Bottom Line[3]	97	21	60	23	92	46	56	45	61	29	89	45	
D	>97	>21	>60	>23	>92	>46	>56	>45	>61	>29	>89	>45	High impact of deposited fine sediment on instream biota. Ecological communities are significantly altered and sensitive fish and macroinvertebrate species are lost or at high risk of being lost.

[1] Classes are streams and rivers defined according to the fourth level of aggregation (L4) of the deposited sediment State Classification (SSC).

[2] The minimum record length for grading a site based on an instream visual assessment of % fine sediment cover (SAM2) is 2 years based on a monthly monitoring regime.

[3] Bottom-line thresholds are anticipated to provide a sufficient level of protection at an overall macroinvertebrate community level (i.e., will cause <20% decrease in the *macroinvertebrate community deviation metric*). Bottom-line thresholds may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations for certain biota. For example, salmonid spawning habitats may require sediment cover of <10%. Fine sediments with high organic enrichment may also result in higher levels of impacts on macroinvertebrate communities or sensitive fish life-stages.

6.2.5 Limitations and assumptions

The proposed thresholds were derived using currently available deposited sediment data on structural measures of ecosystem health. There are very few cases where deposited fine sediment cover has been measured repeatedly over time in the same location, particularly in conjunction with potential ecological response variables. Consequently, our analyses were reliant on a space-for-time substitution and could not explicitly investigate temporal variations in fine sediment cover or ecological responses. By using structural measures of ecosystem health we have also not directly considered potential impacts on ecosystem function.

Different measures of fine sediment cover were used for the fish and macroinvertebrate analyses (NZFFD % fines and SAM2 % cover instream respectively) based on data availability. This reflects that macroinvertebrate samples are routinely sampled by most councils at their SOE sites (where water quality and other data are also collected), whereas sampling of fish is not routinely incorporated in SOE monitoring programmes. For the purposes of this project we made an assumption that the SAM2 % cover instream and NZFFD % fines ESV measures were sufficiently equivalent to allow a direct 1:1 comparison. We do not have data to prove that this assumption is valid because NZFFD % fines and SAM2 % cover instream have not been measured at the same time in the same place. However, the SAM1 % cover bankside method is fundamentally equivalent to the NZFFD % fines method with data showing a strong, close to 1:1 correlation between SAM1 % cover bankside and SAM2 % cover instream when they are measured in the same place at the same time (see Appendix N). We therefore consider this assumption justifiable.

Our analyses have focused on trying to characterise community level responses in macroinvertebrates and fish to increasing deposited fine sediment. Community level responses are complex because they integrate the diverse responses of multiple species. However, in our view basing thresholds for ecosystem health on a community level response was more consistent with the guiding principles than using the sensitivity of individual sentinel species.

Because there are species and life-stage specific differences in sensitivity to elevated cover of fine sediment we recognise that the proposed thresholds may not provide adequate protection for individual highly valued species or critical life-stages. A good example is the high sensitivity of salmonid spawning habitats to elevated fine sediment cover. It is our view that where such sensitive species or life-stages are identified as being of value, then either objectives can be set to achieve a higher attribute band, or value-specific attribute thresholds can be defined.

We have assumed that a 20% deviation in community state from the predicted reference state represents a reasonable boundary for the least acceptable state of ecosystem health. However, we accept that this is a normative decision. On average, this results in an absolute difference between reference state and the C/D threshold of 21% (% fine sediment cover) across the 12 classes, with a range from 16 (L4.4) to 32 (L4.8). It is our view that an average increase in fine sediment cover of 21% over the reference state is ecologically significant and highly likely to result in significant adverse effects on ecosystem health. This is towards the upper bound of existing guidelines for deposited sediment elsewhere in the world (Clapcott et al. 2011). The outcome of selecting a larger community deviation for the resulting thresholds are illustrated in Appendix K.

It is our view that thresholds for deposited sediment should be implemented at Level 4 of the SSC. This reflects significant differences in the identified references states and thresholds between classes, meaning that bias will be introduced by aggregating classes. By this we mean that when classes with different reference conditions and thresholds are combined at a higher level of aggregation in the SSC, the 'averaging' is likely to cause reference condition at a site to be further away from the estimated reference state for the class. This means that the magnitude of acceptable change in some classes becomes greater, while in others it becomes smaller. This results in more variable ecological outcomes. While the reference states and thresholds are similar in some classes where, we do not recommend combining classes purely for the benefit of simplifying the classification. The classes represent different sediment supply and retention characteristics - while they may have similar thresholds, the management actions required to achieve the limits may be different. It is noted that choosing a higher level of aggregation results in a more significant impact on the thresholds than changing the accepted degree of deviation (Appendix K).

6.3 Suspended fine sediment - Turbidity

The community deviation method applied to fish was used as the basis for defining the suspended fine sediment attribute thresholds for turbidity. This was because the body of evidence indicated that fish communities were generally more sensitive to increasing suspended fine sediment than macroinvertebrate species or communities and, therefore, fish represent the more environmentally conservative response measure.

6.3.1 How is it measured?

National standards for measuring turbidity are set out in NEMS (2017) and NEMS (2019) for continuous and discrete sampling respectively. We recommend that measurement of turbidity uses these standards as a basis for implementing the turbidity attribute. Historically, most turbidity data have been reported in nephelometric turbidity units (NTU), but measurement units vary between instruments according to the measurement principle. NEMS (2017) implements the ISO 7027 standard for which the measurement and reporting units are Formazin nephelometric units (FNU). While measurements taken using instruments with different standards may not be directly comparable, for the purposes of implementing this attribute we have assumed that FNU and NTU are equivalent. The required precision to meet the NEMS requirements (both for continuous and discrete field sampling) is 0.1 FNU, but it is noted that required accuracy to meet the standard for continuous turbidity measurement is ±3 FNU averaged across 10 samples. This is greater than the margin between attribute bands within some SSCs and so careful consideration must be given to achieving higher accuracy in these locations. In contrast, the accuracy standard for discrete water quality samples is ±0.3 FNU for field sensors and a detection limit of 0.05 FNU for laboratory measurements making data that comply with this standard compatible with evaluating compliance with the proposed limits.

Numeric thresholds for turbidity are defined as site medians in NTU/FNU. Estimates of medians become more accurate as the number of samples is increased, assuming that the sampling programme is bias-free. Figure 6-1 illustrates how McBride (2005) showed that uncertainty around estimates of the median declines with increasing sample size. At around 20 to 40 samples, a threshold of rapidly diminishing returns is reached, where improvements in the accuracy of the median estimate become increasingly small as sample numbers increase. It was, therefore, recommended that a sample size of 20-40 is suitable for defining medians (McBride 2005). Accordingly, for turbidity we recommend a minimum of 24 samples (i.e., two years of monthly measurements) for calculating site medians. However, we strongly recommend implementation of in-situ continuous measurement of turbidity calibrated to low sediment concentrations to help inform future refinements of suspended sediment attributes to better reflect the dynamics of suspended sediment concentrations and loads.

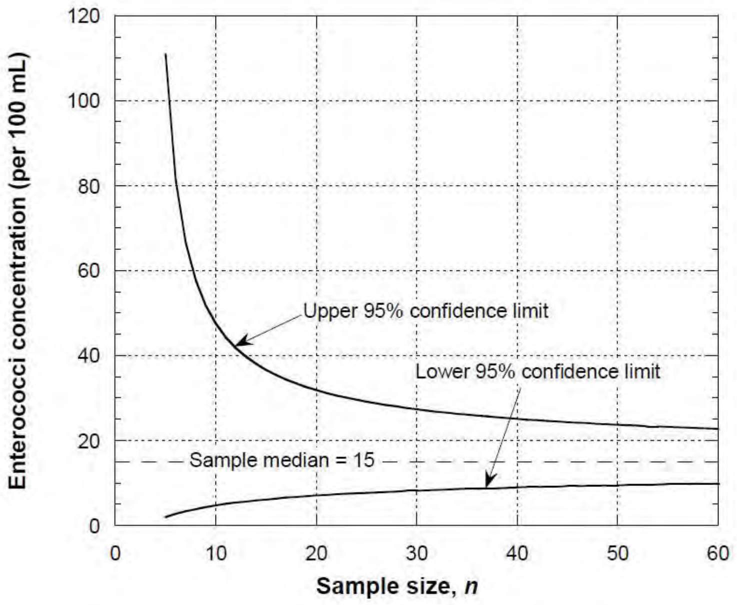


Figure 6-1: An example of the effect of sample size on confidence limits for the median. Source McBride (2005).

6.3.2 Where does it apply?

The proposed attribute thresholds are applicable to all rivers and streams in New Zealand with the exception of (i) naturally highly coloured brown-water streams; (ii) glacial flour affected streams and rivers; and (iii) selected lake-fed REC classes (particularly warm climate classes), where high turbidity may reflect autochthonous phytoplankton production (as opposed to organic/inorganic sediment derived from the catchment). Presently, these locations are not mapped, but the Glacial-Mountain and Lake REC classes are likely to provide a high level indication of where some of these situations may occur.

6.3.3 What does it protect?

The bottom-line thresholds are anticipated to provide sufficient protection on average to avoid significant adverse effects on the overall macroinvertebrate and fish communities. Specifically, they are defined to restrict overall decreases in the fish community deviation metric to <20% relative to the reference state. However, it should be recognised that the thresholds may not protect specific life-stages or the habitat requirements of individual species. For example, the effectiveness of salmonid feeding has been shown to be reduced as visibility in the water column decreases. These thresholds are also not designed to provide protection from shorter-term episodic events that may cause acute effects on specific species.

6.3.4 Proposed attribute table

Value	Ecosystem Health												
Freshwater Body Type	Rivers												
Attribute	Suspended fine sediment												
Attribute Unit	Turbidity (NTU/FNU)												
Attribute State	**SSC class[1]**												**Narrative Attribute State**
	1	2	3	4	5	6	7	8	9	10	11	12	
	Site median[2]												
A	<2.0	<6.2	<1.3	<3.3	<7.5	<4.8	<2.3	<4.3	<1.2	<1.1	<1.1	<2.4	Minimal likelihood of instream biota being impaired by median turbidity. Ecological communities equivalent to minimally disturbed sites in the absence of other confounding stressors.
B	<2.5	<7.9	<1.6	<3.9	<9.8	<6.3	<2.8	<5.2	<1.4	<1.3	<1.3	<2.7	Low to moderate likelihood of instream biota being impaired by median turbidity. Abundance of sensitive fish species reduced.
C	≤3.2	≤10.5	≤2.0	≤4.8	≤13.1	≤8.3	≤3.3	≤6.4	≤1.6	≤1.5	≤1.6	≤3.1	Moderate to high likelihood of instream biota being impaired by median turbidity. Risk of sensitive fish and macroinvertebrate species being lost and change in community composition.
National Bottom Line[3]	3.2	10.5	2.0	4.8	13.1	8.3	3.3	6.4	1.6	1.5	1.6	3.1	
D	>3.2	>10.5	>2.0	>4.8	>13.1	>8.3	>3.3	>6.4	>1.6	>1.5	>1.6	>3.1	High likelihood of instream biota being impaired due to median turbidity. High probability of loss of sensitive fish and macroinvertebrate species and change in community composition.

[1] Classes are streams and rivers defined according to the fourth level of aggregation (L4) of the suspended sediment State Classification (SSC).

[2] The minimum record length for grading a site is 2 years of monthly samples. Continuous turbidity[12] data may be used to calculate 2-year median turbidity.

[3] Bottom-line thresholds are anticipated to provide a sufficient level of protection at an overall fish community level (i.e., will cause <20% decrease in the *fish community deviation metric*). Bottom-line thresholds may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations for certain biota.

[12] Turbidity sensors should be calibrated to the range of interest defined by the attribute band thresholds for the relevant SSC class

6.3.5 Limitations and assumptions

The proposed thresholds were derived using available turbidity data and structural measures of ecosystem health. We have not directly considered potential impacts on ecosystem function. Analyses of macroinvertebrate responses were undertaken using annual at-a-site median turbidity values paired with yearly macroinvertebrate samples. Fish responses were analysed using one-off observations of fish paired with modelled at-a-site median turbidity values due to the absence of paired observations (at the same site at the same time) of both variables. All fish analyses were carried out using presence/absence data due to absence of consistently collected abundance data. Macroinvertebrate data collected at the NRWQN could be used to analyse changes in the abundance of species. However, due to inconsistencies in collection and processing methodologies between regional councils, SOE data had to be reduced to proportional relative abundance within samples or presence/absence for inclusion in analyses.

We would prefer to be able to undertake all analyses using quantitative abundance data because it is a more sensitive measure of ecological response than presence/absence (the number of individuals present will decline before they are locally extirpated). It would also be better to incorporate more information about temporal variation in both turbidity and ecological response variables (rather than using annual measures) to better understand how differences in the duration and frequency of exposure to elevated turbidity impact ecosystem health. However, such data are not available, particularly not across a wide enough gradient of landscape settings to make the results transferable.

As for deposited sediment, our analyses have focused on trying to characterise community level responses in macroinvertebrates and fish to increasing turbidity (see Section 6.2.5 for our explanation). However, because there are species and life-stage specific differences in sensitivity to elevated turbidity we recognise that the proposed thresholds may not provide adequate protection for individual valued species or critical life-stages.

We have assumed that a 20% deviation in community state from the predicted reference state represents a reasonable boundary for the least acceptable state of ecosystem health. However, we accept that this is a normative decision. On average, this results in an absolute difference between reference state and the C/D threshold of 2.4 NTU across the 12 classes, with a range from 0.59 NTU (L4.10) to 7.2 NTU (L4.5). It is our view that an average increase in turbidity of 2.4 NTU over the reference state is ecologically significant and likely to result in significant adverse effects on ecosystem health. The consequences for thresholds of selecting a larger community deviation are illustrated in Appendix K.

It is our view that thresholds for turbidity should also be implemented at Level 4 of the suspended sediment SSC for the same reasons set out in Section 6.2.5 for deposited sediment.

Implementation of a turbidity-based suspended sediment attribute is potentially problematic as it is a relative unit and not derived from or reported in standard SI units. Furthermore, there may be numerical differences in results derived from different instruments, even when they adhere to the same technical standards Davies-Colley et al. (In press). In contrast, spot measurements of visual clarity, which are reported in SI units, have been shown to be subject to less error and have greater numerical similarity between observers than turbidity measurements (West and Scott 2016; Davies-Colley et al. In press). However, turbidity can currently be monitored continuously more cost-effectively than visual clarity. It is also more practical to measure in non-wadeable streams. Continuous turbidity measurements are also often used as the basis of developing sediment rating curves and for total load estimation.

6.4 Suspended fine sediment – Visual clarity

The community deviation method applied to fish was also used as the basis for defining the suspended fine sediment attribute thresholds for visual clarity. The body of evidence indicated that fish communities were generally more sensitive to increasing suspended fine sediment than macroinvertebrate communities and, therefore, fish represent the more environmentally conservative response measure. The visual clarity thresholds were derived using a data-driven approach independently from the turbidity attribute.

6.4.1 How is it measured?

National standards for the measurement of visual clarity are set out in NEMS (2019). We recommend that measurement of visual clarity follow this standards for the purpose of implementing the visual clarity attribute.

As was the case for turbidity, we recommend a minimum of 24 samples (i.e., two years of monthly spot measurements) for calculating site median visual clarity for evaluating against the proposed attribute thresholds.

6.4.2 Where does it apply?

The proposed attribute thresholds are applicable to all rivers and streams in New Zealand with the exception of (i) naturally highly coloured brown-water streams; (ii) glacial flour affected streams and rivers; and (iii) selected lake-fed REC classes (particularly warm climate classes) where low visual clarity may reflect autochthonous phytoplankton production (as opposed to organic/inorganic sediment from the catchment).

6.4.3 What does it protect?

The bottom-line thresholds are anticipated to provide sufficient protection on average to avoid significant adverse effects on the overall macroinvertebrate and fish communities. Specifically, they are defined to restrict overall decreases in the fish community deviation metric to <20% relative to the reference state. However, it should be recognised that the thresholds may not always be sufficient for the protection specific life-stages or habitat requirements of individual species. For example, the effectiveness of salmonid feeding has been shown to reduce as visibility in the water column is decreased. These thresholds are also not designed to provide protection from shorter-term episodic events that may cause acute effects on specific species.

6.4.4 Proposed attribute table

Value	Ecosystem Health
Freshwater Body Type	Rivers
Attribute	Suspended fine sediment
Attribute Unit	Visual clarity (m)

Attribute State	**SSC class[1]**												**Narrative Attribute State**
	1	2	3	4	5	6	7	8	9	10	11	12	
	Site median[2]												
A	>2.25	>2.43	>1.45	>1.43	>0.66	>1.06	>1.78	>0.63	>3.10	>3.38	>2.84	>2.79	Minimal impact of suspended sediment on instream biota. Ecological communities are similar to those observed in natural reference conditions.
B	>1.88	>2.02	>1.21	>1.22	>0.53	>0.87	>1.53	>0.53	>2.71	>2.93	>2.43	>2.51	Low to moderate impact of suspended sediment on instream biota. Abundance of sensitive fish species may be reduced.
C	≥1.55	≥1.65	≥1.00	≥1.02	≥0.42	≥0.70	≥1.30	≥0.44	≥2.35	≥2.51	≥2.06	≥2.23	Moderate to high impact of suspended sediment on instream biota. Sensitive fish and macroinvertebrate species may be lost.
National Bottom Line[3]	1.55	1.65	1.00	1.02	0.42	0.70	1.30	0.44	2.35	2.51	2.06	2.23	
D	<1.55	<1.65	<1.00	<1.02	<0.42	<0.70	<1.30	<0.44	<2.35	<2.51	<2.06	<2.23	High impact of suspended sediment on instream biota. Ecological communities are significantly altered and sensitive fish and macroinvertebrate species are lost or at high risk of being lost.

[1] Classes are streams and rivers defined according to the fourth level of aggregation (L4) of the suspended sediment Sediment State Classification (SSC).

[2] The minimum record length for grading a site is 2 years of monthly samples. Continuous visual clarity data may be used to calculate a 2-year median visual clarity.

[3] Bottom-line thresholds are anticipated to provide a sufficient level of protection at an overall fish community level (i.e., will cause <20% decrease in the *fish community deviation metric*). Bottom-line thresholds may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations for certain biota.

6.4.5 Limitations and assumptions

The limitations and assumptions are similar to those identified for turbidity in Section 6.3.5. However, there are fewer visual clarity data available (it is not currently routinely monitored by all regional councils) relative to turbidity.

We have again assumed that a 20% deviation in community state from the predicted reference state represents a reasonable boundary for the least acceptable state of ecosystem health. However, we acknowledge that this is a normative decision. On average, this results in an absolute difference between reference state and the C/D threshold of 0.8 m across the 12 classes, with a range from 0.3 m (L4.8) to 1.35 m (L4.10). It is our view that an average decrease in visual clarity of 0.8 m over the reference state is ecologically significant and likely to result in significant adverse effects on ecosystem health. The consequences for thresholds of selecting a larger community deviation are illustrated in Appendix K.

Consideration should be given to implementing the suspended sediment attribute as visual clarity (rather than turbidity) because it is measured in standard SI units that are meaningful to the public, i.e., people can understand what visual clarity of 1 m might look like, but may have no idea what 5 NTU looks like. There is also evidence to demonstrate that spot measurements of visual clarity are subject to less error and greater reproducibility than turbidity spot measurements (West and Scott 2016; Davies-Colley et al. In press). One potential disadvantage is that implementation of continuous monitoring of visual clarity is more challenging and costly than continuous monitoring of turbidity. However, evidence suggests that visual clarity is strongly correlated to turbidity and it has been recommended that turbidity not be treated as an absolute quantity. Instead it is recommended that turbidity records be converted to an alternative sediment variable (e.g., visual clarity or suspended sediment concentration) based on empirical (local) correlations. We think there is a strong argument for continuous monitoring of suspended sediment (e.g., continuous turbidity measurements or beam transmissivity for visual clarity) to better understand sediment dynamics, potential ecological impacts, and for evaluating the effectiveness of future management interventions. Developing local empirical relationships between continuous turbidity measurements and visual clarity may offer an opportunity to achieve this cost-effectively.

7 Outstanding issues

We have taken a data-driven approach to deriving potential numeric thresholds for deposited and suspended sediment ESVs. The analyses have focused on establishing robust and defensible bottom-line thresholds for ecosystem health, while also demonstrating suitability of these methods to characterising interim band thresholds. While we believe that using a data-driven approach provides a transparent and reproducible methodology, the types of analysis that can be applied and the way in which exposure-response relationships can be characterised are necessarily dictated by the data available.

Sediment delivery and transport processes are highly dynamic and, therefore, ecosystem stressor exposure is a function of not only long-term averages, but also the impact of short-term events. However, the currently available data dictate that only the impacts of long-term average conditions can be considered. Consequently, all thresholds derived through this process can only be considered to be protective of changes to the long-term average condition. To establish the impacts of shorter-term sediment dynamics on ecosystem health, it will be necessary to collect data on both sediment ESVs and ecological response variables at a greater temporal resolution. The spatial coverage of existing data also limited our ability to effectively account for spatial variations in environmental gradients in some cases. Given that implementation of sediment attributes should result in more monitoring effort for the sediment ESVs, we believe that there would be value in revisiting these analyses as these data become available. This would enable validation of the models developed as part of this project and if monitoring is designed and conducted in a coordinated and standardised way (e.g., collecting fully quantitative macroinvertebrate data and collecting fish and sediment data concurrently), our ability to characterise exposure-response relationships would be strengthened.

We have developed potential attributes for two different suspended sediment ESVs, namely turbidity and visual clarity. There are mechanistic and pragmatic arguments that favour the adoption of visual clarity as the preferred suspended sediment attribute. Turbidity is considered a good proxy variable for several sediment-related variables (including visual clarity), but evidence showing that turbidity measurements are instrument dependent has led to recommendations that the use of nephelometric turbidity units as an absolute quantity should be abandoned (Hughes et al. 2019). Turbidity is also measured in relative units (NTU/FNU) and not standardised SI units whereas visual clarity (m) is, making evaluation of compliance more robust for visual clarity. At present, however, turbidity is monitored routinely by all councils, whereas visual clarity is not. Consequently, more turbidity than visual clarity data are available, meaning that we were able to better characterise ecological responses to turbidity (ignoring potentially significant issues associated with the comparability of data collected using different instruments) than visual clarity. It is also currently more cost effective to monitor turbidity continuously and across a larger range, than it is to measure visual clarity continuously (using a beam transmissometer). A number of councils are currently conducting continuous measurement of turbidity at a limited number of sites for sediment load calculation or sediment rating curve derivation. These data could potentially be used for evaluating compliance with the proposed limits, but sensors are typically calibrated to higher sediment concentrations resulting in greater measurement uncertainty at concentrations in the range of the proposed limits. Furthermore, current national standards for continuous turbidity data collection set a required accuracy of ±3 FNU (NEMS 2017), which is not high enough for measuring compliance with the proposed suspended sediment limits.

Turbidity and visual clarity are typically strongly correlated (Appendix O; Davies-Colley et al. (2015)). However, for the purposes of this project the turbidity and visual clarity attribute thresholds were derived independently using currently available data. Comparison of the results show that for some classes the thresholds for turbidity and visual clarity are not numerically similar when using published equations for converting between the two ESVs. This is likely to be the result of fewer data being available for visual clarity to derive exposure-response relationships. If there is a preference to implement a suspended sediment attribute using visual clarity as the ESV, some consideration should be given as to whether the visual clarity data-derived thresholds should be used, or whether the turbidity thresholds (which were derived using more data and therefore, should be more robust) should be used to derive a visual clarity attribute after applying a conversion factor. An alternative would be to implement both suspended sediment attributes, but further guidance would be required to indicate whether they are to be managed separately or whether one ESV takes precedence (i.e., management using the most sensitive attribute).

We propose that attribute thresholds be assessed against site medians derived from at least 24 samples - this is based on independent work that has demonstrated that a site median may be reliably estimated using two years of monthly samples. In our view this provides a reasonable balance between representing contemporary conditions, while moderating the effects of natural inter-annual variability at a site. However, we recognise that at some sites the magnitude of inter-annual variability in the annual median state of suspended sediment may be greater than the magnitude of differences between proposed band thresholds. Management of compliance in this context must be considered further, including determination of the extent to which inter-annual variability is a consequence of anthropogenic or natural causes.

Some normative decisions were integral to the determination of the proposed thresholds, particularly the magnitude of acceptable deviation from reference in the different community metrics. Ideally these results would be validated with independent data, but data limitations prevented this. We have endeavoured to be transparent about where normative decisions have been made, and where possible have demonstrated how thresholds would vary if those normative decisions were different (e.g., Appendix K).

We have not explicitly calculated uncertainty in our analyses. As far as practicable we have, however, indicated where uncertainties exist. The most critical steps are around the estimation of reference state, from which the deviation from reference is calculated, and characterisation of the ecological responses. It should be noted that the estimates of reference state represent the median for a class. This means that mathematically 50% of sites will have a reference state ‘cleaner’ than this estimate and 50% of sites will have a reference state that is ‘dirtier’ than this. This means that within a class, sites that have a ‘cleaner’ reference state can degrade (compared to reference) more than sites that have a ‘dirtier’ reference state. This is one of our justifications for developing thresholds at Level 4 of the SSCs, because the potential range of variability in reference state that may exist within a class increases as classes are aggregated. Consequently, bias increases with aggregation of classes, resulting in more variable and potentially unachievable outcomes. This variation around the median reference state is also justification for setting the A/B band threshold at a state that is slightly ‘dirtier’ than that predicted for reference state.

Throughout these analyses the interactions between ESVs have not been considered. It is possible that if you manage for one ESV, that will also influence the state of the other ESVs. However, it was outside the scope of this project to determine how the different ESV measures correlate with different management actions and how their responses may be interrelated. We have also not considered any interactions with other stressors (e.g., temperature, flows or nutrients).

8 Conclusions

This report presents numeric thresholds that could potentially be used as the basis of defining NOF attributes for fine sediment ESVs. Our ability to characterise stressor-response relationships between fine sediment ESVs and indicators of ecosystem health were strongly influenced by the types of data that exist. We developed a sediment state classification system to account for spatial patterns in sediment ESV state and we used this as the basis for deriving numeric thresholds using a departure from reference approach. The additional technical analyses undertaken in this phase of the project have helped address equivalency between the multiple lines of evidence, for example by applying the community deviation method to both macroinvertebrate and fish communities. Our data-driven approach, in combination with carrying out a formal weight of evidence evaluation, has provided a transparent and reproducible process for determining the relative strength of different lines of evidence and integrating them to define potential attribute thresholds. We have highlighted some remaining uncertainties in this report, but we believe that the proposed attribute tables for % cover of deposited fine sediment, turbidity and visual clarity provide a strong foundation for establishment of fine sediment attributes.

9 Acknowledgements

We gratefully acknowledge data provided by:

- Northland Regional Council, Auckland Council, Waikato Regional Council, Bay of Plenty Regional Council, Gisborne District Council, Hawke's Bay Regional Council, Greater Wellington Regional Council, Horizons Regional Council, Tasman District Council, Nelson City Council, West Coast Regional Council, Otago Regional Council, and Southland Regional Council. Without help from these organisations, the project simply would not have gotten off the ground.
- Various research datasets collected/collated by NIWA and Cawthron Institute. In particular, Jon Harding, Christoph Matthaei, and John Quinn provided access to, and permission to use, research datasets for macroinvertebrate metric development and ecological thresholds analysis.

We also wish to extend our thanks to Glenn Suter who generously shared his time, expertise and experience to assist our team in implementing the weight of evidence evaluation. This stage was a critical component of the project and the benefit from Glenn's experience was invaluable.

Finally, we thank Neale Hudson for his significant contribution in reviewing this report.

10 Glossary of abbreviations and terms

BEA	Biological Extirpation Analysis.
Bottom line	The minimum acceptable state for an attribute in the NPS-FM. Also referred to as Band D.
Compulsory values	The national values defined in the National Objectives Framework of the NPS-FM that must always be used. Currently these are defined as ecosystem health and human health for recreation (MfE 2017).
CTG	Combined Climate-Topography-Geology (CTG) classes from the River Environment Classification
Deposited sediment	Fine sediment (<2 mm) deposited on the stream bed.
Ecosystem health	A broad term generally used to describe the condition of an ecosystem.
EPT	Ephemeroptera, Plecoptera and Trichoptera – three core orders of freshwater insects that are typically intolerant of pollution and are used as an indicator of water quality.
ESV	Environmental state variable: a variable that captures an aspect of the state of the physical, chemical, or ecological environment.
Extirpation	Is operationally defined as the point above which only 5% of the observations of a genus occurs (Cormier and Suter 2013). Effectively this represents local disappearance of a species.
Fine sediment	<2 mm particle size.
FNU	Formazin Nephelometric Units.
MCI	Macroinvertebrate community index. A biotic index based on the sensitivity of individual macroinvertebrate taxa to organic pollution.
NOF	National Objectives Framework.
NPS-FM	National Policy Statement for Freshwater Management.
NRWQN	National River Water Quality Network. A monitoring network of 77 river sites run by NIWA since 1989, with an aggregate catchment of about 50% of NZ's land area.
NTU	Nephelometric Turbidity Units.
NZReach	Individual river segment within RECv1, with associated environmental information available.
NZSegment	Individual river segment within RECv2, with associated environmental information available.
POM	Particulate organic matter
Quantile regression	Quantile regression models the relationship between a specified conditional quantile (or percentile) of a dependent (response) variable and one or more independent (explanatory) variables (Cade and Noon 2003).

REC	River Environment Classification.
SAM1	Sediment Assessment Method 1: Bankside visual estimate of % sediment cover. Rapid qualitative assessment of the surface area of the streambed covered by sediment in pool, riffle and run habitats.
SAM2	Sediment Assessment Method 2: Instream visual estimate of % sediment cover. Semi-quantitative assessment of the surface area of the streambed covered by sediment. At least 20 readings are made within a single habitat (runs).
Sediment MCI	A sediment specific macroinvertebrate community index. That is a biotic index based on the sensitivity of individual macroinvertebrate taxa to deposited sediment.
SSC	Sediment state classification derived for deposited sediment and suspended sediment. Used for estimating reference ESV state and as the spatial unit for defining limits.
SSD	Species Sensitivity Distribution – mathematical model fitted to the distribution of species extirpation values.
Strahler stream order	Numerical measure of the branching complexity of a stream and its upstream tributaries. For example, a second order stream reach is formed below the confluence of two first order reaches and a third order stream reach is formed below the confluence of two second order reaches.
Suspended sediment	Fine sediment (<2 mm) suspended in the water column.
TSS	Total suspended sediment (concentration) – measured by filtration of a subsample of a water sample, in contrast to SSC which is measured by filtration of the whole sample. Ideally TSS would equal SSC, but if the subsampling is not representative, typically owing to rapid settling sand, TSS may differ (and be biased).
Turbidity	A relative measure of suspended sediment. Units NTU/FNU.
US EPA	United States Environmental Protection Agency.
Visual clarity	A measure of how clear the water is measured using black disc visibility (in the horizontal direction). Unit metres.
Weight of Evidence (WoE)	WoE is defined as an inferential process that assembles, evaluates, and integrates evidence to perform a technical inference in an assessment (US EPA 2016a)
XC95	95th percentile extirpation concentration derived from the species sensitivity distribution (SSD) composed of all of the extirpation values for species in a suspended sediment classification class.

11 References

Akbaripasand, A., Nichol, E.C., Lokman, P.M., Closs, G.P. (2011) Microhabitat use of a native New Zealand galaxiid fish, *Galaxias fasciatus*. *New Zealand Journal of Marine and Freshwater Research*, 45(1): 135-144. 10.1080/00288330.2010.544044

Alabaster, J.S., Lloyd, R. (1982) *Water quality criteria for freshwater fish*. Butterworth Scientific, Boston.

Allouche, S. (2002) Nature and functions of cover for riverine fish. *Bull. Fr. Pêche Piscic.,* (365-366): 297-324.

ANZECC (2000) *Australian and New Zealand guidelines for fresh and marine water quality*. National Water Quality Management Strategy Paper No. 4, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, Australia.

Baker, C.F. (2003) Effect of adult pheromones on the avoidance of suspended sediment by migratory banded kokopu juveniles. *Journal of Fish Biology*, 62(2): 386-394. 10.1046/j.1095-8649.2003.00029.x

Baker, C.F., Montgomery, J.C. (2001) Species-specific attraction of migratory banded kokopu juveniles to adult pheromones. *Journal of Fish Biology*, 58(5): 1221-1229.

Barrett, J.C., Grossman, G.D., Rosenfeld, J. (1992) Turbidity-Induced Changes in Reactive Distance of Rainbow Trout. *Transactions of the American Fisheries Society*, 121(4): 437-443. 10.1577/1548-8659(1992)121<0437:TICIRD>2.3.CO;2

Bilotta, G.S., Brazier, R.E. (2008) Understanding the influence of suspended solids on water quality and aquatic biota. *Water Research*, 42(12): 2849-2861. 10.1016/j.watres.2008.03.018

Booker, D. (2013) Spatial and temporal patterns in the frequency of events exceeding three times the median flow (FRE3) across New Zealand. *Journal of Hydrology (NZ)*, 52: 1-25.

Booker, D.J., Dunbar, M., Ibbotson, A.T. (2004) Predicting juvenile salmonid drift-feeding habitat quality using a three-dimensional hydraulic-bioenergetic model. *Ecological Modelling*, 177: 157-177.

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

Borsuk, M.E., Reichert, P., Peter, A., Schager, E., Burkhardt-Holm, P. (2006) Assessing the decline of brown trout (*Salmo trutta*) in Swiss rivers using a Bayesian probability network. *Ecological Modelling*, 192(1): 224-244. https://doi.org/10.1016/j.ecolmodel.2005.07.006

Boubee, J.A.T., Dean, T.L., West, D.W., Barrier, R.F.G. (1997) Avoidance of suspended sediment by the juvenile migratory stage of six New Zealand native fish species. *New Zealand Journal of Marine and Freshwater Research*, 31: 61-69.

Burdon, F.J., McIntosh, A.R., Harding, J.S. (2013) Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. *Ecological Applications*, 23(5): 1036-1047. 10.1890/12-1190.1

Burnham, K.P., Anderson, D.R. (2002) *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York.

Cade, B.S., Guo, Q.F. (2000) Estimating effects of constraints on plant performance with regression quantiles. *Oikos*, 91(2): 245-254. DOI 10.1034/j.1600-0706.2000.910205.x

Cade, B.S., Noon, B.R. (2003) A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, 1(8): 412–420.

Cantilli, R., Stevens, R., Sweitlik, W., Berry, W., Kaufman, P., Paul, J., Spehar, R., Cormier, S., Norton, D. (2006) Framework for developing suspended and bedded-sediments (SABS) water quality criteria.

Cavanagh, J.E., Hogsden, K.L., Harding, J. (2014) Effects of suspended sediment on freshwater fish. *Landcare Research Contract Report*, LC1986: 29.

Chapman, J.M., Proulx, C.L., Veilleux, M.A.N., Levert, C., Bliss, S., André, M.-È., Lapointe, N.W.R., Cooke, S.J. (2014) Clear as mud: A meta-analysis on the effects of sedimentation on freshwater fish and the effectiveness of sediment-control measures. *Water Research*, 56: 190-202. http://dx.doi.org/10.1016/j.watres.2014.02.047

Clapcott, J.E., Goodwin, E.O. (2017) Technical report on developing a deposited sediment classification for New Zealand streams. Prepared for Ministry for the Environment, Cawthron Report No. 2994: 36.

Clapcott, J.E., Young, R.G., Harding, J.S., Matthaei, C.D., Quinn, J.M., Death, R.G. (2011) *Sediment Assessment Methods: Protocols and guidelines for assessing the effects of deposited fine sediment on in-stream values*. Cawthron Institute, Nelson, New Zealand.

Cohen, J. (1960) A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1): 37-46. doi:10.1177/001316446002000104

Collins, A.L., Naden, P.S., Sear, D.A., Jones, J.I., Foster, I.D.L., Morrow, K. (2011) Sediment targets for informing river catchment management: international experience and prospects. *Hydrological Processes*, 25(13): 2112-2129. 10.1002/hyp.7965

Cormier, S.M., Paul, J.F., Spehar, R.L., Shaw-Allen, P., Berry, W.J., Suter, G.W. (2008) Using field data and weight of evidence to develop water quality criteria. *Integrated Environmental Assessment and Management*, 4(4): 490–504.

Cormier, S.M., Zheng, L., Leppo, E.W., Hamilton, A. (2018) Step-by-step calculation and spreadsheet tools for predicting stressor levels that extirpate genera and species. *Integrated Environmental Assessment and Management*, 14(2): 174-180. doi:10.1002/ieam.1993

Crisp, D.T., Carling, P.A. (1989) Observations on siting, dimensions and structure of salmonid redds. *Journal of Fish Biology*, 34(1): 119-134. 10.1111/j.1095-8649.1989.tb02962.x

Crow, S.K., Booker, D., Sykes, J., Unwin, M., Shankar, U. (2014) Predicting distributions on New Zealand freshwater fishes. *NIWA Client Report*, CHC2014-145: 100.

Crow, S.K., Snelder, T., Jellyman, P.G., Greenwood, M.J., Dunn, A. (2016) Temporal trends in the relative abundance of New Zealand freshwater fishes: Analysis of New Zealand Freshwater Fish Database records. *NIWA Client Report*, CHC2016-049.

Crowe, A., Hay, J. (2004) Effects of fine sediment on river biota. *Cawthron Report*, 951: 40.

Cumming, H., Herbert, N.A. (2016) Gill structural change in response to turbidity has no effect on the oxygen uptake of a juvenile sparid fish. *Conservation Physiology*, 4(1): cow033-cow033. 10.1093/conphys/cow033

Davies-Colley, R., Milne, J., Heath, M.W. (In press) Reproducibility of river water quality measurements: Inter-agency comparisons for quality assurance. *New Zealand Journal of Marine and Freshwater Research*.

Davies-Colley, R.J., Hickey, C.W., Quinn, J.M., Ryan, P.A. (1992) Effects of clay discharges on streams, 1. Optical properties and epilithon. *Hydrobiologia*, 248: 215-234.

Davies-Colley, R.J., Hicks, M., Hughes, A., Clapcott, J., Kelly, D., Wagenhoff, A. (2015) Fine sediment effects on freshwaters, and the relationship of environmental state to sediment load: a literature review. *NIWA Client Report*, HAM2015-104: 105.

Davies-Colley, R.J., Smith, D.G. (2001) Turbidity, suspended sediment, and water clarity: A review. *Journal of the American Water Resources Association*, 37(5): 1085-1101.

De'ath, G. (2007) Boosted trees for ecological modeling and prediction. *Ecology*, 88(1): 243-251. 10.1890/0012-9658(2007)88[243:btfema]2.0.co;2

Depree, C. (2017) Sediment Attribute Stage 1b. Proposed classification system: 75.

Depree, C., Clapcott, J., Booker, D., Franklin, P., Hickey, C., Wagenhoff, A., Matheson, F., Shelley, J., Unwin, M., Wadhwa, S., Goodwin, E., Mackman, J., Rabel, H. (2018) Development of ecosystem health bottom-line thresholds for suspended and deposited sediment in New Zealand rivers and streams: 190.

Depree, C., Unwin, M., Clapcott, J. (2016) Sediment stage 1a report: summary of data. *NIWA Client Report*, 2016124HN: 27.

Dodds, W.K., Oakes, R.M. (2004) A technique for establishing reference nutrient concentrations across watersheds affected by humans. *Limnology and Oceanography: Methods*, 2(10): 333-341. doi:10.4319/lom.2004.2.333

Doeg, T., Milledge, G. (1991) Effect of experimentally increasing concentration of suspended sediment on macroinvertebrate drift. *Marine and Freshwater Research*, 42(5): 519-526. https://doi.org/10.1071/MF9910519

Elith, J., Leathwick, J. (2017) Boosted Regression Trees for ecological modeling. Vignette for R dismo package version 1.1-4. http://cran.r-project.org/web/packages/dismo/.

Elith, J., Leathwick, J.R., Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, 77: 802-813.

Goldes, S.A., Ferguson, H.W., moccia, R.D., Daoust, P.-Y. (1988) Histological effects of the inert suspended clay kaolin on the gills of juvenile rainbow trout, *Salmo gairdneri* Richardson. *Journal of Fish Diseases*, 11(1): 23-33. doi:10.1111/j.1365-2761.1988.tb00520.x

Grace, J.B., Adler, P.B., Harpole, W.S., Borer, E.T., Seabloom, E.W. (2014) Causal networks clarify productivity-richness interrelations, bivariate plots do not. *Functional Ecology*, 28(4): 787-798. 10.1111/1365-2435.12269

Gray, L.J., Ward, J.V. (1982) Effects of sediment releases from a reservoir on stream macroinvertebrates. *Hydrobiologia*, 96(2): 177-184. 10.1007/bf02185433

Greer, M.J.C. (2014) The effects of macrophyte control on freshwater fish communities and water quality in New Zealand streams. University of Otago, Dunedin: 158.

Greer, M.J.C., Crow, S.K., Hicks, A.S., Closs, G.P. (2015) The effects of suspended sediment on brown trout (*Salmo trutta*) feeding and respiration after macrophyte control. *New Zealand Journal of Marine and Freshwater Research*, 49(2): 278-285. 10.1080/00288330.2015.1013140

Greet, J., Webb, J.A., Cousens, R.D. (2011) The importance of seasonal flow timing for riparian vegetation dynamics: a systematic review using causal criteria analysis. *Freshwater Biology*, 56: 1231-1247.

Gregory, R.S. (1993) Effect of Turbidity on the Predator Avoidance Behaviour of Juvenile Chinook Salmon (*Oncorhynchus tshawytscha*). *Canadian Journal of Fisheries and Aquatic Sciences*, 50(2): 241-246. 10.1139/f93-027

Gregory, R.S., Levings, C.D. (1998) Turbidity Reduces Predation on Migrating Juvenile Pacific Salmon. *Transactions of the American Fisheries Society*, 127(2): 275-285. 10.1577/1548-8659(1998)127<0275:TRPOMJ>2.0.CO;2

Grove, M.K., Bilotta, G.S., Woockman, R.R., Schwartz, J.S. (2015) Suspended sediment regimes in contrasting reference-condition freshwater ecosystems: Implications for water quality guidelines and management. *Science of The Total Environment*, 502: 481-492. https://doi.org/10.1016/j.scitotenv.2014.09.054

Harrison, E.T. (2010) Fine Sediment in Rivers: Scale of Ecological Outcomes. University of Canberra, Canberra, Australia.

Harvey, B.C., White, J.L. (2008) Use of Benthic Prey by Salmonids under Turbid Conditions in a Laboratory Stream. *Transactions of the American Fisheries Society*, 137(6): 1756-1763. 10.1577/T08-039.1

Hay, J. (2004) Survival of trout eggs in relation to sediment composition of redds in Waikakahi Stream. *Cawthron Report*, 965.

Hayes, J., Stark, J.D., Shearer, K.A. (2000) Development and Test of a Whole-Lifetime Foraging and Bioenergetics Growth Model for Drift-Feeding Brown Trout. *Transactions of the American Fisheries Society*, 129(2): 315-332. 10.1577/1548-8659(2000)129<0315:DATOAW>2.0.CO;2

Hayes, J.W., Goodwin, E., Shearer, K.A., Hay, J., Kelly, L. (2016) Can Weighted Useable Area Predict Flow Requirements of Drift-Feeding Salmonids? Comparison with a Net Rate of Energy Intake Model Incorporating Drift–Flow Processes. *Transactions of the American Fisheries Society*, 145(3): 589-609. 10.1080/00028487.2015.1121923

Henley, W.F., Patterson, M.A., Neves, R.J., Lemly, A.D. (2000) Effects of sedimentation and turbidity on lotic food webs: a concise review for natural resource managers. *Reviews in Fisheries Science*, 8(2): 125-139.

Herbert, D.W., Merkins, J.C. (1961) The effect of suspended sediment mineral solids on the survival of trout. *International Journal of Air and Water Pollution*, 5: 46-55.

Hickford, M., Schiel, D. (2011) Population sinks resulting from degraded habitats of an obligate life-history pathway. *Oecologia*, 166(1): 131-140. 10.1007/s00442-010-1834-7

Hicks, D.M., Clapcott, J., Davies-Colley, R., Dymond, J., Greenwood, M., Hughes, A., Shankar, U., Walter, K. (2016) Sediment Attributes Stage 1, *NIWA Client Report* No: CHC2016-058: 189.

Hicks, D.M., Haddadchi, A., Whitehead, A.L., Shankar, U. (2019) Sediment load reductions to meet suspended and deposited sediment thresholds. *NIWA Client Report* No: 2019100CH.

Hughes, A., Davies-Colley, R., Heubeck, S. (2019) Comparability of ISO 7027 compliant turbidity sensors. *NIWA Internal Report*, 2019125HN: 27.

Jones, J.I., Murphy, J.F., Collins, A.L., Sear, D.A., Naden, P.S., Armitage, P.D. (2012) The impact of fine sediment on macro-invertebrates. *River Research and Applications*, 28(8): 1055-1071.

Jowett, I.G., Boustead, N. (2001) Effects of substrate and sedimentation on the abundance of upland bullies (*Gobiomorphus breviceps*). *New Zealand Journal of Marine and Freshwater Research*, 35(3): 605-613. 10.1080/00288330.2001.9517026

Jowett, I.G., Richardson, J., McDowall, R.M. (1996) Relative effects of in-stream habitat and land use on fish distribution and abundance in tributaries of the Grey River, New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 30(4): 463-475. 10.1080/00288330.1996.9516735

Kefford, B.J., Zalizniak, L., Dunlop, J.E., Nugegoda, D., Choy, S.C. (2010) How are macroinvertebrates of slow flowing lotic systems directly affected by suspended and deposited sediments? *Environmental Pollution*, 158(2): 543-550. 10.1016/j.envpol.2009.08.008

Kemp, P., Sear, D., Collins, A., Naden, P., Jones, I. (2011) The impacts of fine sediment on riverine fish. *Hydrological Processes*, 25(11): 1800-1821.

Khan, K.S., Kunz, R., Kleijnen, J., Antes, G. (2003) Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 96(3): 118-121.

Kjelland, M.E., Woodley, C.M., Swannack, T.M., Smith, D.L. (2015) A review of the potential effects of suspended sediment on fishes: potential dredging-related physiological, behavioral, and transgenerational implications. *Environment Systems and Decisions*, 35(3): 334-350. 10.1007/s10669-015-9557-2

Koenker, R. (2013) quantreg: Quantile Regression. *R package* version 5.05.

Kusabs, I.A., Quinn, J.M., Hamilton, D.P. (2015) Effects of benthic substrate, nutrient enrichment and predatory fish on freshwater crayfish (kōura, *Paranephrops planifrons*) population characteristics in seven Te Arawa (Rotorua) lakes, North Island, New Zealand. *Marine and Freshwater Research*, 66(7): 631-643. https://doi.org/10.1071/MF14148

Lake, R.G., Hinch, S.G. (1999) Acute effects of suspended sediment angularity on juvenile coho salmon (*Oncorhynchus kisutch*). *Canadian Journal of Fisheries and Aquatic Sciences*, 56(5): 862-867. 10.1139/f99-024

Lange, K., Townsend, C.R., Gabrielsson, R., Chanut, P.C.M., Matthaei, C.D. (2014) Responses of stream fish populations to farming intensity and water abstraction in an agricultural catchment. *Freshwater Biology*, 59(2): 286-299. 10.1111/fwb.12264

Larned, S.T., Schallenberg, M. (2018) Stressor-response relationships and the prospective management of aquatic ecosystems. *New Zealand Journal of Marine and Freshwater Research*: in press. 10.1080/00288330.2018.1524388

Leathwick, J.R., Snelder, T., Chadderton, W.L., Elith, J., Julian, K., Ferrier, S. (2011) Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications. *Freshwater Biology*, 56(1): 21-38. doi:10.1111/j.1365-2427.2010.02414.x

Lehtiniemi, M., Engström-Öst, J., Viitasalo, M. (2005) Turbidity decreases anti-predator behaviour in pike larvae, *Esox lucius*. *Environmental Biology of Fishes*, 73(1): 1-8. 10.1007/s10641-004-5568-4

Lewis, W.M., Melack, J.M., McDowell, W.H., McClain, M., Richey, J.E. (1999) Nitrogen yields from undisturbed watersheds in the Americas. *Biogeochemistry*, 46(1): 149-162. 10.1007/bf01007577

Louhi, P., Ovaska, M., Mäki-Petäys, A., Erkinaro, J., Muotka, T. (2011) Does fine sediment constrain salmonid alevin development and survival? *Canadian Journal of Fisheries and Aquatic Sciences*, 68(10): 1819-1826. 10.1139/f2011-106

Manel, S., Williams, H.C., Ormerod, S.J. (2001) Evaluating presence–absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38(5): 921-931. 10.1046/j.1365-2664.2001.00647.x

McBride, G. (2005) *Using statistical methods for water quality management: issues, problems and solutions*. John Wiley & Sons Inc., Hoboken, NJ.

McDowall, R.M. (1990) *New Zealand freshwater fishes: A natural history and guide*. Heinemann Reed, Auckland: 553.

McDowell, R.W., Snelder, T.H., Cox, N., Booker, D.J., Wilcock, R.J. (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): 387-400. http://dx.doi.org/10.1071/MF12153

McEwan, A.J. (2009) *Fine scale spatial behaviour of indigenous riverine fish in a small New Zealand stream.* Massey University, Palmerston North: 95.

McEwan, A.J., Joy, M.K. (2014a) Habitat use of redfin bullies (*Gobiomorphus huttoni*) in a small upland stream in Manawatu, New Zealand. *Environmental Biology of Fishes*, 97(2): 121-132. 10.1007/s10641-013-0131-9

McEwan, A.J., Joy, M.K. (2014b) Diel habitat use of two sympatric galaxiid fishes (*Galaxias brevipinnis* and *G. postvectis*) at two spatial scales in a small upland stream in Manawatu, New Zealand. *Environmental Biology of Fishes*, 97(8): 897-907. 10.1007/s10641-013-0191-x

MfE (2017) National Policy Statement for Freshwater Management 2014 (amended 2017): 47.

Mitchell, S.B., West, J.R., Guymer, I. (1999) Dissolved-Oxygen/Suspended-Solids Concentration Relationships in the Upper Humber Estuary. *Water and Environment Journal*, 13(5): 327-337. 10.1111/j.1747-6593.1999.tb01057.x

NEMS (2017) National Environmental Monitoring Standards: Turbidity - Measurement, processing and archiving of turbidity data, Version 1.2: 65.

NEMS (2019) National Environmental Monitoring Standards: Water Quality - Part 2 of 4: Sampling, measuring, processing and archiving of discrete river water quality data, Version 0.1: 111.

Newcombe, C.P. (2003) Impact assessment model for clear water fishes exposed to excessively cloudy water. *Journal of the American Water Resources Association*, 39(3): 529-544. 10.1111/j.1752-1688.2003.tb03674.x

Newcombe, C.P., Macdonald, D.D. (1991) Effects of Suspended Sediments on Aquatic Ecosystems. *North American Journal of Fisheries Management*, 11(1): 72-82. 10.1577/1548-8675(1991)011<0072:EOSSOA>2.3.CO;2

Nichols, S., Webb, A., Norris, R., Stewardson, M. (2011) *Eco Evidence analysis methods manual: a systematic approach to evaluate causality in environmental science*.

Norris, R.H., Webb, J.A., Nichols, S.J., Stewardson, M.J., Harrison, E.T. (2012) Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. *Freshwater Science*, 31: 5-21.

Nuttall, P.M., Bielby, G.H. (1973) The effect of China-clay wastes on stream invertebrates. *Environmental Pollution (1970)*, 5(2): 77-86. http://dx.doi.org/10.1016/0013-9327(73)90013-X

Oksanen, J. (2015) Multivariate Analysis of Ecological Communities in R: vegan tutorial, http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., E., S., Wagner, H. (2018) vegan: Community Ecology Package. *R package* version 2.5-3, https://CRAN.R-project.org/package=vegan.

Olsson, T., Persson, B.-G. (1988) Effects of deposited sand on ova survival and alevin emergence in brown trout (*Salmo trutta* L.). *Archiv fuer Hydrobiologie*, 113(4): 621-627.

Owens, P.N., Batalla, R.J., Collins, A.J., Gomez, B., Hicks, D.M., Horowitz, A.J., Kondolf, G.M., Marden, M., Page, M.J., Peacock, D.H., Petticrew, E.L., Salomons, W., Trustrum, N.A. (2005) Fine-grained sediment in river systems: environmental significance and management issues. *River Research and Applications*, 21(7): 693-717. 10.1002/rra.878

Posthuma, L., Suter, G.W., Traas, T.P. (2002) General introduction to species sensitivity distributions. In: L. Posthuma, G.W. Suter & T.P. Traas (Eds). *Species sensitivity distributions in ecotoxicology*. CRC Press, Boca Raton, FL: 3-10.

Qian, S.S. (2014) Ecological threshold and environmental management: A note on statistical methods for detecting thresholds. *Ecological Indicators*, 38 192-197.

Qian, S.S., Cuffney, T.F. (2018) The multiple-comparison trap and the Raven's paradox—perils of using null hypothesis testing in environmental assessment. *Environmental Monitoring and Assessment*, 190(7): 409. 10.1007/s10661-018-6793-1

Quinn, J.M. (2000) Effects of pastoral development. In: K.J. Collier & M.J. Winterbourn (Eds). *New Zealand stream invertebrates : ecology and implications for management*. New Zealand Limnological Society, Hamilton, NZ: 208-229.

Quinn, J.M., Davies-Colley, R.J., Hickey, C.W., Vickers, M.L., Ryan, P.A. (1992) Effects of clay discharges on streams 2. Benthic invertebrates. *Hydrobiologia*, 248(3): 235-247. Doi 10.1007/Bf00006150

Quinn, J.M., Hickey, C.W. (1993) Effects of sewage stabilization lagoon effluent on stream invertebrates. *Journal of Aquatic Ecosystem Health*, 2: 205-219.

Ramezani, J., Rennebeck, L., Closs, G.P., Matthaei, C.D. (2014) Effects of fine sediment addition and removal on stream invertebrates and fish: a reach-scale experiment. *Freshwater Biology*, 59(12): 2584-2604. 10.1111/fwb.12456

Reid, D., Quinn, J. (2011) *Preliminary information for developing sediment guidelines for streams of the West Coast, New Zealand*: 21.

Richardson, J., Rowe, D.K., Smith, J.P. (2001) Effects of turbidity on the migration of juvenile banded kokopu (*Galaxias fasciatus*) in a natural stream. *New Zealand Journal of Marine and Freshwater Research*, 35(1): 191-196. 10.1080/00288330.2001.9516989

Robertson, M. (1957) The effects of suspended materials on the reproductive rate of *Daphnia magna*. *Publ. Inst. Mar. Sci. Univ. Tex.*, 4: 265–277.

Rosenberg, D.M., Wiens, A.P. (1978) Effects of sedimentation on macrobenthic invertebrates on a northern Canadian river. *Water Research*, 12: 753–763.

Rowe, D.K., Dean, T.L. (1998) Effects of turbidity on the feeding ability of the juvenile migrant stage of six New Zealand freshwater fish species. *New Zealand Journal of Marine and Freshwater Research*, 32(1): 21-29. 10.1080/00288330.1998.9516803

Rowe, D.K., Hicks, M., Richardson, J. (2000) Reduced abundance of banded kokopu (*Galaxias fasciatus*) and other native fish in turbid rivers of the North Island of New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 34(3): 547-558. 10.1080/00288330.2000.9516956

Rowe, D.K., Hicks, M., Smith, J., Williams, E. (2009) Lethal concentrations of suspended solids for common native fish species that are rare in New Zealand rivers with high suspended solids loads. *New Zealand Journal of Marine and Freshwater Research*, 43(5): 1029-1038. 10.1080/00288330.2009.9626526

Rowe, D.K., Smith, J., Williams, E. (2002a) Effects of turbidity on the feeding ability of adult, riverine smelt (*Retropinna retropinna*) and inanga (*Galaxias maculatus*). *New Zealand Journal of Marine and Freshwater Research*, 36(1): 143-150. 10.1080/00288330.2002.9517077

Rowe, D.K., Suren, A.M., Martin, M., Smith, J., Smith, B., Williams, E. (2002b) Lethal turbidity levels for common freshwater fish and invertebrates in Auckland streams. *Auckland Regional Council Technical Publication*, 337: 37.

Ryan, P.A. (1991) Environmental effects of sediment on New Zealand streams: a review. *New Zealand Journal of Marine and Freshwater Research*, 25: 207-221.

Shaw, E.A., Richardson, J.S. (2001) Direct and indirect effects of sediment pulse duration on stream invertebrate assemblages and rainbow trout (*Oncorhynchus mykiss*) growth and survival. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(11): 2213-2221. 10.1139/f01-160

Slaney, P.A., Halsey, T.G., Tautz, A.F. (1977) Effects of forest harvesting practices on spawning habitat of stream salmonids in the Centennial Creek watershed, British Columbia: 35.

Snelder, T.H., Biggs, B.J.F. (2002) Multiscale River Environment Classification for water resources management. *Journal of the American Water Resources Association*, 38(5): 1225-1239. 10.1111/j.1752-1688.2002.tb04344.x

Snelder, T.H., Biggs, B.J.F., Woods, R.A. (2005) Improved eco-hydrological classification of rivers. *River Research and Applications*, 21(6): 609-628. 10.1002/rra.826

State of Oregon Department of Environmental Quality (2010) *Turbidity Technical Review: Summary of Sources, Effects, and Issues Related to Revising the Statewide Water Quality Standard for Turbidity:* 85.

Stillman, R.A., Wood, K.A., Goss-Custard, J.D. (2016) Deriving simple predictions from complex models to support environmental decision-making. *Ecological Modelling*, 326: 134-141. https://doi.org/10.1016/j.ecolmodel.2015.04.014

Stoddard, J.L., Larsen, D.P., Hawkins, C.P., Johnson, R.K., Norris, R.H. (2006) Setting expectations for the ecological condition of streams: the concept of reference condition. *Ecological Applications*, 16(4): 1267-1276. doi:10.1890/1051-0761(2006)016[1267:SEFTEC]2.0.CO;2

Suren, A.M., Martin, M.L., Smith, B.J. (2005) Short-Term Effects of High Suspended Sediments on Six Common New Zealand Stream Invertebrates. *Hydrobiologia*, 548(1): 67-74. 10.1007/s10750-005-4167-5

Suter, G., Cormier, S., Barron, M. (2017) A weight of evidence framework for environmental assessments: Inferring quantities. *Integrated Environmental Assessment and Management*, 13(6): 1045-1051. doi:10.1002/ieam.1953

Sutherland, A.B., Meyer, J.L. (2007) Effects of increased suspended sediment on growth rate and gill condition of two southern Appalachian minnows. *Environmental Biology of Fishes*, 80(4): 389-403. 10.1007/s10641-006-9139-8

Suttle, K.B., Power, M.E., Levine, J.M., McNeely, C. (2004) How fine sediment in riverbeds impairs growth and survival of juvenile salmonids. *Ecological Applications*, 14(4): 969-974. 10.1890/03-5190

Sweka, J.A., Hartman, K.J. (2001) Influence of turbidity on brook trout reactive distance and foraging success. *Transactions of the American Fisheries Society*, 130(1): 138-146. 10.1577/1548-8659(2001)130<0138:IOTOBT>2.0.CO;2

Uncles, R.J., Joint, I., Stephens, J.A. (1998) Transport and retention of suspended particulate matter and bacteria in the Humber-Ouse Estuary, United Kingdom, and their relationship to hypoxia and anoxia. *Estuaries*, 21(4): 597-612. 10.2307/1353298

Unwin, M., Larned, S.T. (2013) Statistical models, indicators and trend analyses for reporting national-scale river water quality. *NIWA Client Report* CHC2013-033: 71.

US EPA (2016a) *Weight of evidence in ecological assessment,* EPA/100/R-16/001.

US EPA (2016b) Generic ecological assessment endpoints (GEAEs) for ecological risk assessment and Technical Background Paper. *Risk Assessment Forum*, EPA/100/F15/005.

US EPA (2017) CADDIS - stressor response analysis: Quantile regression. US Environmental Protection Agency, Causal Analysis/Diagnosis Decision Information System (CADDIS), Washington D.C.: Data analysis reference material.

Usio, N., Townsend, C.R. (2001) The significance of the crayfish *Paranephrops zealandicus* as shredders in a new zealand headwater stream. *Journal of Crustacean Biology*, 21(2): 354-359. 10.1651/0278-0372(2001)021[0354:TSOTCP]2.0.CO;2

Vant, B. (2011) Water quality of the Hauraki Rivers and Southern Firth of Thames. *Waikato Regional Council Technical Report*, 2011/06: 40.

Vant, B. (2013) Water quality of the lower Piako River 2011-13. *Waikato Regional Council Internal Series*, 2013/15: 20.

Velleman, P.F. (1989) *The Data Desk handbook. Version 6.3*. Data Description Inc., Ithica, New York.

Wagener, S.M., LaPerriere, J.D. (1985) Effects of Placer Mining on the Invertebrate Communities of Interior Alaska Streams. *Freshwater Invertebrate Biology*, 4(4): 208-214. 10.2307/1467163

Wagenhoff, A., Liess, A., Pastor, A., Clapcott, J.E., Goodwin, E.O., Young, R.G. (2017) Thresholds in ecosystem structural and functional responses to agricultural stressors can inform limit setting in streams. *Freshwater Science*, 36(1): 178-194

Waters, T.F. (1995) *Sediment in streams: sources, biological effects, and control*. American Fisheries Society, Bethesda, MA: 251.

Webb, A.J., Miller, K.A., King, E.L., de Little, S.C., Stewardson, M.J., Zimmerman, J.K.H., LeRoy Poff, N. (2013) Squeezing the most out of existing literature: a systematic re-analysis of published evidence on ecological responses to altered flows. *Freshwater Biology*, 58(12): 2439-2451. 10.1111/fwb.12234

Webb, J.A., Miller, K.A., Stewardson, M.J., de Little, S.C., Nichols, S.J., Wealands, S.R. (2015) An online database and desktop assessment software to simplify systematic reviews in environmental science. *Environmental Modelling & Software*, 64: 72-79. http://dx.doi.org/10.1016/j.envsoft.2014.11.011

West, A.O., Scott, J.T. (2016) Black disk visibility, turbidity, and total suspended solids in rivers: A comparative evaluation. *Limnology and Oceanography: Methods*, 14(10): 658-667. doi:10.1002/lom3.10120

Whitehead, A.L. (2019) Spatial modelling of river water-quality state: incorporating monitoring data from 2013 to 2017. *NIWA Client Report* 2018360CH: 42.

Wilkes, M.A., Gittins, J.R., Mathers, K.L., Mason, R., Casas-Mulet, R., Vanzo, D., Mckenzie, M., Murray-Bligh, J., England, J., Gurnell, A., Jones, J.I. (2019) Physical and biological controls on fine sediment transport and storage in rivers. *Wiley Interdisciplinary Reviews: Water*, 6(2): e1331. doi:10.1002/wat2.1331

Wilson, K.A., Westphal, M.I., Possingham, H.P., Elith, J. (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biological Conservation*, 122(1): 99-112. http://dx.doi.org/10.1016/j.biocon.2004.07.004

Wood, P.J., Armitage, P.D. (1997) Biological effects of fine sediment in the lotic environment. *Environmental Management*, 21(2): 203-217.

Appendix A Literature review: The effects of deposited fine sediment on macroinvertebrates

Introduction

The effects of deposited fine sediment on stream biota have been studied extensively. Reviews identify strong predictive relationships, with increasing deposited fine sediment decreasing ecological health (Waters 1995; Wood and Armitage 1997; Clapcott et al. 2011; Collins et al. 2011; Kemp et al. 2011; Jones et al. 2012). In New Zealand and elsewhere, deposited sediment is measured in a variety of ways leading to estimates of various deposited sediment ESVs. There also is a range of ecological indicators that have been used to quantify stressor-response relationships with sediment ESVs (Davies-Colley et al. 2015). This lack of uniformity makes it difficult 1) to extract from the literature what are the overall most consistent and reliable ecological indicators that can be used to assess and compare the effects of deposited sediment on the ecological state within as well as among streams, and 2) to merge datasets and identify robust sediment management thresholds that prevent or reduce detrimental effects. As narrative reviews do not have any rules regarding evidence interpretation (Norris et al. 2012), there is no guarantee they can reach consistent, or even correct, conclusions. We revisited international literature, specifically focusing on benthic macroinvertebrate responses, and conducted a formal causal criteria analysis to identify what ecological evidence exists to inform sediment-specific metric development and support development of management thresholds. The aims were to:

- Determine the responses of macroinvertebrates to fine sediment addition, ranging from individual species/taxa to community-level metrics.
- Determine if different measures (i.e., ESVs) of suspended and deposited sediment used in studies show the same results.
- Test whether overall results from the causal criteria analysis are consistent with individual studies and the current scientific consensus identified by more descriptive, narrative reviews.

We considered the Eco Evidence software suitable to test these hypotheses as it has already been used in a number of systematic reviews on river flows (Greet et al. 2011; Webb et al. 2013) and sediment (Harrison 2010) that have provided more definitive conclusions than those provided by earlier narrative reviews.

Eco Evidence

Eco Evidence is a form of systematic review that is based upon causal criteria analysis (Webb et al. 2015). Systematic reviews are in contrast to narrative reviews as they treat relevant literature as data (Khan et al. 2003), and employ statistical analysis to succinctly analyse and summarise a large body of literature, testing the level of support for hypotheses across numerous studies (Webb et al. 2015). Though currently uncommon in environmental science, a systematic synthesis improves the defence and transparency of decision making, Eco Evidence may help increase scientific input into the setting of resource limits and freshwater targets/objectives (Webb et al. 2013). This would not only fulfil legal requirements to create 'evidence based' environmental management, but could in turn improve environmental outcomes.

Two key features allow the Eco Evidence software to do to this. The first is an open-access online database that stores causal evidence from systematic reviews, thus simplifying data extraction from the literature by allowing evidence to be reused. The second is an analysis tool with a standardised 8-step form of causal criteria analysis that produces a transparent report of the level of support for specific cause-effect hypotheses reviewed. Eco Evidence could therefore, be used in several ways to advance the understanding of the effect of stressors on macroinvertebrates, including: to identify knowledge gaps, establish the scientific consensus prior to research, evaluate how effective management decisions have been, improve environmental review standards, and as in this study, test cause-effect hypotheses found in a body of literature (Webb et al. 2015).

Method

The Eco Evidence framework adopted in this study consisted of eight steps (Norris et al. 2012) that were used to assess evidence on the effect of sediment on macroinvertebrates in the causal criteria analysis:

1. *Problem definition*. Many anthropological activities degrade terrestrial and riparian environments in such a way that they increase the amount of fine sediment found in streams and rivers. Freshwater macroinvertebrates are sensitive to levels of both fine sediment suspended in the water column and deposited on the benthos, with the direct and indirect addition of anthropogenic sediment affecting habitat and food availability, as well as their direct biological functioning.

2. *Research question*. 'What are the effects of anthropogenic sedimentation on macroinvertebrates in freshwater systems?'

3. *Conceptual model*. Figure A-1.

4. *Cause-effect hypotheses.* Entries consisted of a term (an entity) and an attribute (a property of the entity), which were structured 'term (attribute)' e.g., *Deleatidium* (abundance). Classifications (drop down lists) were then used to assign hypothesised trajectories of both the cause and effect terms. From the conceptual model, the identified causes were an increase in deposited and suspended sediment and the measures used to quantify them (e.g., percentage cover of fine sediment), whilst the identified effects were a change in both hypothesised sensitive and non-sensitive individual taxa, as well as changes in more general community structure indicators.

5. *Review literature and extract evidence*. A search for all combinations of cause and effect terms was primarily conducted on Web of Science and Google Scholar. Reference lists of relevant studies and those of previous narrative reviews, along with lists of studies that had cited papers with evidence items relevant to any of the hypotheses were also reviewed. Studies were only included if they generated primary data (to eliminate the risk of double counting a data set), and to avoid misinterpretation by citing authors. Furthermore, only studies that proved statistical significance (or insignificance) of evidence items were retained (as guided by Norris et al. (2012)).

6. *Revise*. Both the cause-effect hypotheses and conceptual model were revised throughout the analysis as more causes and effects were discovered in the literature, with these being added to the analysis.

7. *Catalogue and weight the evidence*. A total of 65 studies (see References section in this Appendix) with varying numbers of evidence items were found that were relevant to the ecological effect of fine sediment addition on macroinvertebrates, and were entered into the software for analysis. The weight of evidence assigned to the item was determined from the experimental design and the level of sample replication. These components were summed to give an overall study weight (Figure A-2), with greater weighting assigned to research having study design that controlled confounding influences and had greater replication of both controls and treatments.

8. *Assess the level of support for the research question*. In the weighting of evidence items, three causal criteria were used to test for a potential cause-effect relationship. These were: **Response** (the presence of a response), **Dose Response** (if a response is present whether there is a dose relationship between the cause and effect), and **Consistency of Association** (the same results amongst numerous studies) (Nichols et al. 2011). High levels of evidence for the Response and Dose Response criteria display an association between the cause and effect, with this occurring when the summed weight for an evidence item is ≥ 20. A summed weight <20 shows a low level of evidence for the Response and Dose Response criteria. This means as few as three studies with a high quality, robust design may provide enough evidence to support a cause-effect hypothesis, whereas seven poorly-designed studies may not (Norris et al. 2012). This association was only developed into support for a causal link if high Consistency of Association for the cause-effect hypothesis existed as well. For this the weighting of all the studies that *did not* support the hypothesised cause-effect linkage were summed, and if the summed value was ≥ 20, this was considered to indicate lack of consistency and hence low support for causality. A value <20 therefore, indicated high consistency of association and a high level of support for causality (Nichols et al. 2011). The three causal criteria were then collated for each cause-effect relationship to see the level of support for the hypotheses under investigation.

After an evidence item had been weighted, its trajectory was then compared to that of the cause-effect linkage to assess if it contributed to supporting or refuting the hypothesis. When this had been done for all linkages in relevant citations, the weighting values for all evidence items that supported the hypotheses were summed, as were those refuting it. These two totals were then compared to a threshold value (again with a default of 20 summed points) to see the overall strength and direction of evidence, thus reaching one of four conclusions for the hypothesis (Table A-1) (Webb et al. 2013).

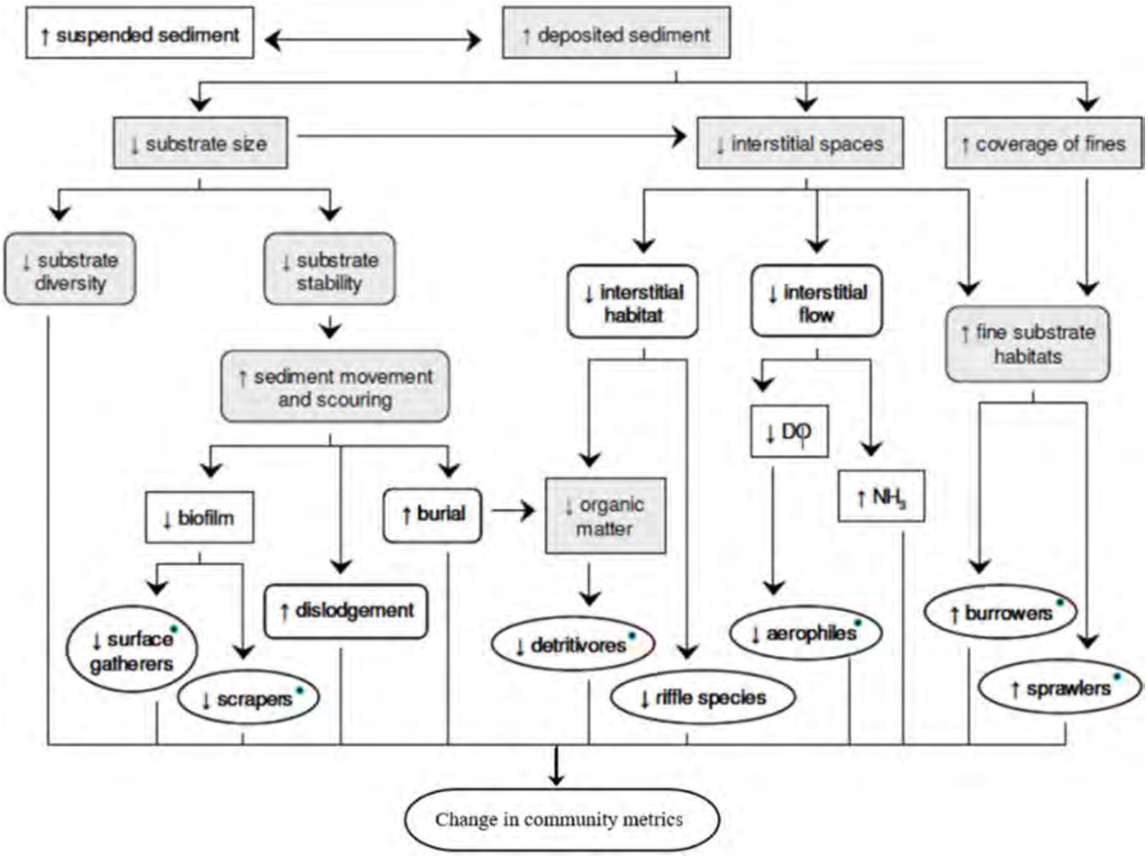


Figure A-1: Conceptual model for the effect of fine sediment on macroinvertebrates. Rectangular boxes are used for stressors, rounded rectangular boxes show an additional step in the causal pathway, and ovals are used for responses. Responses with blue-black dots indicate individual species are included within these responses. Image adapted from Cantilli et al. (2006).

Study design component	Weight
Study design type	
After impact only	1
Reference/control vs impact with no before data	2
Before vs after with no reference/control location(s)	2
Gradient response model	3
BACI, BARI, MBACI, or beyond MBACI	4
Replication of factorial designs	
Number of reference/control sampling units	
0	0
1	2
>1	3
Number of impact/treatment sampling units	
1	0
2	2
>2	3
Replication of gradient-response models	
<4	0
4	2
5	4
>5	6

Figure A-2: The weightings of different components of an evidence item. Each evidence item consists of a study design weighting and a weighting for the number of controls and treatments used, except for gradient response studies; these are weighted using the replication of gradient-response models. Image from Nichols et al. (2011).

Table A-1: The four possible outcomes of the Eco Evidence Causal Criteria Analysis.

Conclusion	Weighting Supporting Hypothesis	Weighting Refuting Hypothesis	Implications
Support for Hypothesis	≥20	<20	The evidence verifies the hypothesis.
Support for Alternate Hypothesis	<20	≥20	The evidence fails to verify the hypothesis.
Inconsistent Evidence	≥20	≥20	The evidence fails to verify the hypothesis, though a subset of the hypothesis may be supported.
Insufficient Evidence	<20	<20	There is too little data to test the hypothesis and may also indicate a literature gap.

Results

Overall, 655 cause-effect hypotheses were tested, with these containing 1858 individual linkages/items of evidence that were unevenly distributed between the hypotheses. Most hypotheses had insufficient evidence to test the cause-effect relationship, with only 111 of the 655 hypotheses returning sufficient support for a conclusion other than insufficient evidence (Table A-2).

In response to a general increase in deposited fine sediment, 14 cause-effect hypotheses were supported by the analysis including a decrease in 8 taxa, 3 species traits and 3 community metrics (i.e., EPT density, %EPT abundance, MCI). Eleven alternate hypotheses were supported by the analysis including an increase in 2 taxa, 1 trait and 1 metric, and a decrease in a further 4 taxa and 3 traits (Table A-2).

There was little consistency among responses when comparing patch-scale and reach-scale measures of deposited fine sediment, other than for decreases in EPT richness and abundance (Table A-2). There was no overlap between deposited sediment and suspended sediment in supported hypotheses. An increase in suspended sediment causing a decrease in macroinvertebrate abundance was the only causal relationship for suspended sediment supported by the literature.

Discussion

The Eco Evidence systematic review confirmed 25 conceptual hypotheses (original or alternate) of the effect of sediment on benthic invertebrates. In particular, EPT metrics were a good indicator of deposited fine sediment effects. There was also significant ecological evidence of the effect of deposited fine sediment on the MCI metric. These results showed 544 hypotheses had insufficient evidence and 86 hypotheses had inconsistent findings.

The Eco Evidence approach may limit findings in part due to the way causal criteria are assigned. As also observed by Harrison (2010), several hypotheses showed very strong support for a response, but the outcome was considered inconsistent due to a small number of studies showing support for an alternate hypothesis. For example, the hypothesis that an increase in deposited sediment caused a decrease in EPT richness had a response of 166. The consistency of association score of 60 was sufficient to make the outcome inconsistent, even though the level of support for the hypothesis was over 2.5 times greater than support for an alternate outcome. This suggests the total number of studies that do not support the causal hypothesis disproportionately influence the outcome.

Table A-2: Cause-effect hypotheses that contained sufficient evidence from the literature to reach an outcome other than insufficient evidence. * = taxa not present in New Zealand, ↑ and ↓ represent increasing and decreasing responses to increasing fine sediment, respectively.

Metric	Support	Alternate	Inconsistent
Increasing (↑) Deposited fine sediment	↓%EPT abundance ↓clinger ↓*Deleatidium* ↓*Ecdyonurus** ↓Elmidae ↓Ephemeroptera ↓EPT density ↓*Leuctra** ↓low body flexibility ↓MCI ↓Orthocladiinae ↓*Paraleptophlebia** ↓Plecoptera ↓surface egg laying	↑Baetidae* ↑macroinvertebrate biomass ↑*Potamopyrgus antipodarum* ↑respires using gills ↓%crawlers ↓Cladocera ↓Copepoda ↓*Oxyethira* ↓scraper ↓shredder ↓Tanypodinae	↑burrower ↑*Hexatoma** ↑macroinvertebrate density ↑Nematoda ↓%EPT ↓Chironomidae ↓EPT abundance ↓EPT richness ↓filter-feeder ↓*Glossosoma** ↓*Hesperoperla pacifica* ↓macroinvertebrate abundance ↓macroinvertebrate diversity ↓macroinvertebrate species richness ↓Oligochaeta
↑% cover	↑burrower ↓%EPT abundance ↓clinger ↓*Deleatidium* ↓Ephemeroptera ↓EPT density ↓low body flexibility ↓MCI ↓*Paraleptophlebia** ↓Plecoptera ↓surface egg laying	↑Baetidae* ↑macroinvertebrate biomass ↑*Potamopyrgus antipodarum* ↓%crawlers ↓Cladocera ↓Copepoda ↓Oligochaeta ↓scrapers ↓shredders ↓Tanypodinae	↑*Hexatoma** ↑macroinvertebrate density ↑Nematoda ↓%EPT ↓Chironomidae ↓EPT abundance ↓EPT richness ↓*Glossosoma** ↓macroinvertebrate abundance ↓macroinvertebrate diversity ↓macroinvertebrate species richness ↓*Neophylax**
↑% cover (patch)	↑burrower ↑nematoda ↓%EPT ↓*Deleatidium* ↓Ephemeroptera ↓EPT abundance ↓EPT density ↓*Paraleptophlebia** ↓Plecoptera	↑Baetidae* ↑*Potamopyrgus antipodarum* ↓Cladocera ↓Copepoda ↓Tanypodinae	↑macroinvertebrate density ↑Oligochaeta ↓Chironomidae ↓EPT richness ↓macroinvertebrate abundance ↓macroinvertebrate diversity ↓macroinvertebrate species richness ↓*Neophylax** ↓scrapers ↓shredders
↑% cover (reach)	↓%EPT abundance ↓EPT density ↓EPT richness ↑macroinvertebrate density	↓chironomidae ↓macroinvertebrate diversity ↓Oligochaeta ↓shredder	↓macroinvertebrate abundance ↓macroinvertebrate biomass ↓macroinvertebrate species richness
↑Suspended sediment	↓macroinvertebrate abundance		↓macroinvertebrate species richness ↓EPT richness

Within the Eco Evidence framework, there is an option to redefine the consistency of association threshold, making it possible to raise this threshold when there is strong evidence for a hypothesised response. However, as no guidelines currently exist for redefining this threshold, any manipulation would be both subjective and arbitrary, and the validity of conclusions reached questionable. The best option to manage this sensitivity may therefore, be to incorporate a ratio aspect into the consistency of association criteria, as well as the current threshold. This could work in the same way the current framework does, except when the consistency of association threshold is exceeded, the proportion of evidence for and against the hypothesised response is compared, and if there is sufficient evidence (e.g., twice as much) for the response versus refuting it, the low consistency of association is overruled. This would therefore, allow widely-used responses such as EPT richness to be analysed, whilst also indicating why the inconsistencies in a cause-effect hypothesis are occurring.

Another limitation of the Eco Evidence framework is that it lacks any gauge of the strength of association between a cause and effect, and hence the magnitude of an impact. This creates uncertainty as to whether the effect is significant but potentially manageable, or catastrophic. For example, Eco Evidence support for an increase in deposited fine sediment causing a decrease in *Deleatidium* could indicate a small but significant drop in abundance, or it could indicate complete elimination of the population, but there is no indication as to which end of the spectrum the impact will be. This limits the utility of Eco Evidence because prediction of the magnitude of response is key for management. This suggests that some gauge of magnitude needs to be incorporated for the software to have more widespread use.

In summary, we consider this approach to be potentially very useful, but have identified that improvements are required before widespread use of an Eco Evidence systematic review to inform management objectives/targets may be recommended. Currently, the results support use of EPT metrics for investigating the effects of deposited sediment on benthic macroinvertebrates, and provide further support for the development of a sediment-specific metric based on taxa sensitivity.

References included in Eco Evidence review

Angradi, T.R. (1999) Fine sediment and macroinvertebrate assemblages in Appalachian streams: a field experiment with biomonitoring applications. *Journal of the North American Benthological Society*, 18(1): 49–66.

Benoy, G.A., Sutherland, A.B., Culp, J.M., Brua, R.B. (2012) Physical and ecological thresholds for deposited sediments in streams in agricultural landscapes. *Journal of Environmental Quality*, 41(1): 31–40.

Blettler, M., Amsler, M.L., Ezcurra de Drago, I., Espinola, L.A., Eberle, E., Paira, A., Best, J.L., Parsons, D.R., Drago, E.E. (2015) The impact of significant input of fine sediment on benthic fauna at tributary junctions: a case study of the Bermejo–Paraguay River confluence, Argentina. *Ecohydrology*, 8(2): 340–352.

Bo, T., Fenoglio, S., Malacarne, G., Pessino, M., Sgariboldi, F. (2007) Effects of clogging on stream macroinvertebrates: an experimental approach. *Limnologica-Ecology and Management of Inland Waters*, 37(2): 186–192.

Braccia, A., Voshell, Jr J.R. (2006) Environmental factors accounting for benthic macroinvertebrate assemblage structure at the sample scale in streams subjected to a gradient of cattle grazing. *Hydrobiologia*, 573(1): 55–73.

Buendia, C., Gibbins, C.N., Vericat, D., Lopez-Tarazon, J.A., Batalla, R.J. (2011) Influence of Naturally High Fine Sediment Loads on Aquatic Insect Larvae in a Montane River. *Scottish Geographical Journal*, 127(4): 315–334.

Buendia, C., Gibbins, C.N., Vericat, D., Batalla, R.J., Douglas, A. (2013) Detecting the structural and functional impacts of fine sediment on stream invertebrates. *Ecological Indicators*, 25: 184–196.

Burdon, F.J., McIntosh, A.R., Harding, J.S. (2013) Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. *Ecological Applications,* 23(5): 1036–1047.

Chou, L., Yu, J., Loh, T. (2004) Impacts of sedimentation on soft-bottom benthic communities in the southern islands of Singapore. *Hydrobiologia,* 515(1-3): 91–106.

Ciesielka, I.K., Bailey, R.C. (2001) Scale-specific effects of sediment burial on benthic macroinvertebrate communities. *Journal of Freshwater Ecology*, 16(1): 73–81.

Conroy, E., Turner, J.N., Rymszewicz, A., Bruen, M., Kelly-Quinn, M. (2016) An evaluation of visual and measurement-based methods for estimating deposited fine sediment. *International Journal of Sediment Research*, 31(4): 368–375.

Conroy, E., Turner, J.N., Rymszewicz, A., Bruen, M., O'Sullivan, J.J., Lawler, D.M., Lally, H., Kelly-Quinn, M. (2016) Evaluating the relationship between biotic and sediment metrics using mesocosms and field studies. *Science of the Total Environment*, 568: 1092–1101.

Couceiro, S.R.M., Hamada, N., Forsberg, B.R., Padovesi-Fonseca, C. (2010) Effects of anthropogenic silt on aquatic macroinvertebrates and abiotic variables in streams in the Brazilian Amazon. *Journal of Soils and Sediments*, 10(1): 89–103.

Cover, M.R., May, C.L., Dietrich, W.E., Resh, V.H. (2008) Quantitative linkages among sediment supply, streambed fine sediment, and benthic macroinvertebrates in northern California streams. *Journal of the North American Benthological Society*, 27(1): 135–149.

Culp, J.M., Wrona, F.J., Davies, R.W. (1986) Response of stream benthos and drift to fine sediment deposition versus transport. *Canadian Journal of Zoology-Revue Canadienne De Zoologie*, 64(6): 1345–1351.

De Castro Vasconcelos, M., Melo, A.S. (2008) An experimental test of the effects of inorganic sediment addition on benthic macroinvertebrates of a subtropical stream. *Hydrobiologia*, 610(1): 321–329.

Descloux, S., Datry, T., Marmonier, P. (2013) Benthic and hyporheic invertebrate assemblages along a gradient of increasing streambed colmation by fine sediment. *Aquatic Sciences*, 75(4): 493–507.

Descloux, S., Datry, T., Usseglio-Polatera, P. (2014) Trait-based structure of invertebrates along a gradient of sediment colmation: Benthos versus hyporheos responses. *Science of the total environment*, 466: 265–276.

Doeg, T., Koehn, J. (1994) Effects of draining and desilting a small weir on downstream fish and macroinvertebrates. *Regulated Rivers: Research & Management*, 9(4): 263–277.

Downes, B.J., Lake, P., Glaister, A., Bond, N.R. (2006) Effects of sand sedimentation on the macroinvertebrate fauna of lowland streams: are the effects consistent*? Freshwater Biology*, 51(1): 144–160.

Elbrecht, V., Beermann, A.J., Goessler, G., Neumann, J., Tollrian, R., Wagner, R., Wlecklik, A., Piggott, J.J., Matthaei, C.D., Leese, F. (2016) Multiple-stressor effects on stream invertebrates: a mesocosm experiment manipulating nutrients, fine sediment and flow velocity. *Freshwater Biology,* 61(4): 362–375.

Evans-White, M.A., Dodds, W.K., Huggins, D.G., Baker, D.S. (2009) Thresholds in macroinvertebrate biodiversity and stoichiometry across water-quality gradients in Central Plains (USA) streams. *Journal of the North American Benthological Society*, 28(4): 855–868.

Gray, L.J., Ward, J.V. (1982) Effects of sediment releases from a reservoir on stream macroinvertebrates. *Hydrobiologia*, 96(2): 177–184.

Gurtz, M.E., Wallace, J.B. (1984) Substrate-Mediated Response of Stream Invertebrates to Disturbance. *Ecology,* 65(5): 1556–1569.

Harrison, E. (2010) Fine sediment in rivers: scale of ecological outcomes. *Unpublished thesis*, University of Canberra.

Harrison, E., Norris, R., Wilkinson, S. (2008) Can an indicator of river health be related to assessments from a catchment-scale sediment model? *Hydrobiologia,* 600(1): 49–64.

Jones, I., Growns, I., Arnold, A., McCall, S., Bowes, M. (2015) The effects of increased flow and fine sediment on hyporheic invertebrates and nutrients in stream mesocosms. *Freshwater Biology*, 60(4): 813–826.

Kaller, M.D., Hartman, K.J. (2004) Evidence of a threshold level of fine sediment accumulation for altering benthic macroinvertebrate communities. *Hydrobiologia,* 518(1-3): 95–104.

Kreutzweiser, D.P., Capell, S.S., Good, K.P. (2005) Effects of fine sediment inputs from a logging road on stream insect communities: a large-scale experimental approach in a Canadian headwater stream. *Aquatic Ecology*, 39(1): 55–66.

Larsen, S., Ormerod, S.J. (2010) Low-level effects of inert sediments on temperate stream invertebrates. *Freshwater Biology,* 55(2): 476–486.

Larsen, S., Vaughan, I.P., Ormerod, S.J. (2009) Scale-dependent effects of fine sediments on temperate headwater invertebrates. *Freshwater Biology*, 54(1): 203–219.

Larsen, S., Pace, G., Ormerod, S.J. (2011) Experimental effects of sediment deposition on the structure and function of macroinvertebrate assemblages in temperate streams. *River Research and Applications*, 27(2): 11.

Lenat, D.R., Penrose, D.L., Eagleson, K.W. (1981) Variable effects of sediment addition on stream benthos. *Hydrobiologia*, 79(2): 187–194.

Logan, O.D. (2007) Effects of fine sediment deposition on benthic invertebrate communities. *Unpublished thesis,* University of New Brunswick.

Longing, S.D., Voshell, J.R., Dolloff, C.A., Roghair, C.N. (2010) Relationships of sedimentation and benthic macroinvertebrate assemblages in headwater streams using systematic longitudinal sampling at the reach scale. *Environmental Monitoring And Assessment,* 161(1-4): 517–530.

Magbanua, F.S. (2012) Agricultural intensification and stream health: combined impacts of pesticide and sediment. *Unpublished thesis*, University of Otago.

Magbanua, F.S., Townsend, C.R., Hageman, K.J., Matthaei, C.D. (2013) Individual and combined effects of fine sediment and the herbicide glyphosate on benthic macroinvertebrates and stream ecosystem function. *Freshwater Biology*, 58(8): 1729–1744.

Mathers, K.L., Millett, J., Robertson, A.L., Stubbington, R., Wood, P.J. (2014) Faunal response to benthic and hyporheic sedimentation varies with direction of vertical hydrological exchange. *Freshwater Biology*, 59(11): 2278–2289.

Matthaei, C.D., Piggott, J.J., Townsend, C.R. (2010) Multiple stressors in agricultural streams: interactions among sediment addition, nutrient enrichment and water abstraction. *Journal of Applied Ecology*, 47(3): 639–649.

Matthaei, C.D., Weller, F., Kelly, D.W., Townsend, C.R. (2006) Impacts of fine sediment addition to tussock, pasture, dairy and deer farming streams in New Zealand. *Freshwater Biology*, 51(11): 2154–2172.

McIntyre, P.B., Michel, E., France, K., Rivers, A., Hakizimana, P., Cohen, A.S. (2005) Individual- and assemblage-level effects of anthropogenic sedimentation on snails in Lake Tanganyika. *Conservation Biology,* 19(1): 171–181.

Molinos, J.G., Donohue, I. (2010) Interactions among temporal patterns determine the effects of multiple stressors. *Ecological Applications,* 20(7): 1794–1800.

Niyogi, D.K., Koren, M., Arbuckle, C.J., Townsend, C.R. (2007) Longitudinal changes in biota along four New Zealand streams: declines and improvements in stream health related to land use. *New Zealand Journal of Marine and Freshwater Research*, 41: 63–75.

Niyogi, D.K., Koren, M., Arbuckle, C.J., Townsend, C.R. (2007) Stream communities along a catchment land-use gradient: subsidy-stress responses to pastoral development. *Environmental Management*, 39(2): 213–225.

Nuttall, P., Bielby, G. (1973) The effect of china-clay wastes on stream invertebrates. *Environmental Pollution*, (1970) 5(2): 77–86.

Peckarsky, B.L. (1985) Do predaceous stoneflies and siltation affect the structure of stream insect communities colonizing enclosures? *Canadian Journal of Zoology*, 63(7): 1519–1530.

Pedersen, M.L., Friberg, N., Larsen, S.E. (2004) Physical habitat structure in Danish lowland streams. *River Research and Applications*, 20(6): 653–669.

Piggott, J.J., Townsend, C.R., Matthaei, C.D. (2015) Climate warming and agricultural stressors interact to determine stream macroinvertebrate community dynamics. *Global Change Biology*, 21(5): 1887–1906.

Piggott, J.J., Lange, K., Townsend, C.R., Matthaei, C.D. (2012) Multiple stressors in agricultural streams: a mesocosm study of interactions among raised water temperature, sediment addition and nutrient enrichment. *Plos One*, 7(11).

Poff, N.L., Zimmerman, J.K. (2010) Ecological responses to altered flow regimes: a literature review to inform the science and management of environmental flows. *Freshwater Biology,* 55(1): 194–205.

Pollard, A.I., Yuan, L.L. (2010) Assessing the consistency of response metrics of the invertebrate benthos: a comparison of trait- and identity-based measures. *Freshwater Biology*, 55(7): 1420–1429.

Quinn, J.M., Davies-Colley, R.J., Hickey, C.W., Vickers, M.L., Ryan, P.A. (1992) Effects of clay discharges on streams. 2. Benthic invertebrates. *Hydrobiologia*, 248(3): 235–247.

Rabeni, C.F., Doisy, K.E., Zweig, L.D. (2005) Stream invertebrate community functional responses to deposited sediment. *Aquatic Sciences*, 67(4): 395–402.

Ramezani, J., Rennebeck, L., Closs, G.P., Matthaei, C.D. (2014) Effects of fine sediment addition and removal on stream invertebrates and fish: a reach-scale experiment. *Freshwater Biology*, 59(12): 2584–2604.

Reed, J. (1977) Stream community response to road construction sediments. *Virginia Water Resources Research Centre Bulletin*, No. 97. Virginia Polytechnic Institute and State University.

Relyea, C.D. (2007) Fine inorganic sediment effects on stream macroinvertebrates. *Unpublished thesis*, Idaho State University.

Relyea, C.D., Minshall, G.W., Danehy, R.J. (2000) Stream insects as bioindicators of fine sediment. *Proceedings of the Water Environment Federation*, 2000(6): 663–686.

Rosenberg, D.M., Wiens, A.P. (1978) Effects of sediment addition on macrobenthic invertebrates in a northern Canadian river. *Water Research*, 12(10): 753–763.

Schofield, K.A., Pringle, C.M., Meyer, J.L. (2004) Effects of increased bedload on algal- and detrital-based stream food webs: Experimental manipulation of sediment and macroconsumers. *Limnology and Oceanography*, 49(4): 900–909.

Shaw, E.A., Richardson, J.S. (2001) Direct and indirect effects of sediment pulse duration on stream invertebrate assemblages and rainbow trout (Oncorhynchus mykiss) growth and survival. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(11): 2213–2221.

Strand, R.M., Merritt, R.W. (1997) Effects of episodic sedimentation on the net-spinning caddisflies *Hydropsyche betteni* and *Ceratopsyche sparna* (Trichoptera: Hydropsychidae). *Environmental Pollution,* 98(1): 129–134.

Suren, A.M., Jowett, I.G. (2001) Effects of deposited sediment on invertebrate drift: an experimental study. *New Zealand Journal of Marine and Freshwater Research,* 35(4): 725–737.

Suttle, K.B., Power, M.E., Levine, J.M., McNeely, C. (2004) How fine sediment in riverbeds impairs growth and survival of juvenile salmonids. *Ecological Applications*, 14(4): 969–974.

Townsend, C.R., Uhlmann, S.S., Matthaei, C.D. (2008) Individual and combined responses of stream ecosystems to multiple stressors. *Journal of Applied Ecology*, 45(6): 1810–1819.

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

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

Zweig, L.D., Rabeni, C.F. (2001) Biomonitoring for deposited sediment using benthic invertebrates: A test on 4 Missouri streams. *Journal of the North American Benthological Society*, 20(4): 643–657.

Appendix B Literature review: The effects of suspended fine sediment on macroinvertebrates

Introduction

Suspended sediments are often regarded as the single most important pollutant of freshwaters, in terms of the quantities discharged and the damage that they cause to aquatic ecosystems (Henley et al. 2000; Owens et al. 2005). Some suspended particulate matter arises from point sources such as sewage outfalls, mining, industrial wastes and stormwater drains, but in New Zealand most is contributed from diffuse land runoff due to soil erosion (ANZECC 2000). Sediment may be deposited on stream beds or remain in suspension. Most of the suspended sediment is <2 mm (Owens et al. 2005), with suspended particle size distribution dependent on flow velocities and source characteristics.

The functioning and productivity of streams can be altered by suspended sediment, which can reduce photosynthesis of in-stream autotrophs, clog the filter feeding structures of certain invertebrates, and increase invertebrate drift (Ryan 1991; Wood and Armitage 1997; Henley et al. 2000). The impact on aquatic biota depends on the species and life-stages present in communities, and the concentration and duration of exposure (Newcombe and Macdonald 1991). Continuous high level inputs of sediment are likely to have most deleterious effects on aquatic communities, as some sediment input is natural and necessary for ecosystems, and animals are presumably adapted to cope with smaller pulsed inputs like those that occur naturally (Ryan 1991; Grove et al. 2015).

The results of published studies on stream macroinvertebrate response to suspended sediment in New Zealand and elsewhere are outlined below.

New Zealand studies on suspended sediment effects on macroinvertebrates

Organic suspended solids

Field gradients of POM have identified a subsidy-stress response in macroinvertebrate gradients downstream of wastewater lagoon effluent discharges (Quinn and Hickey 1993). These discharges contain both particulate organic SS and nutrients which result in stimulation of benthic periphyton growths (Figure B-1). The field gradient showed a pronounced subsidy/stress response for macroinvertebrate metrics – with the threshold for marked declines in the EPT abundance in the range 5-10 mg/L increased in SS concentration. This study confirms that a subsidy/stress response should be expected for riverine macroinvertebrate communities from particulate organic SS inputs, and provides an indicative range of concentrations where the adverse effects might be expected.

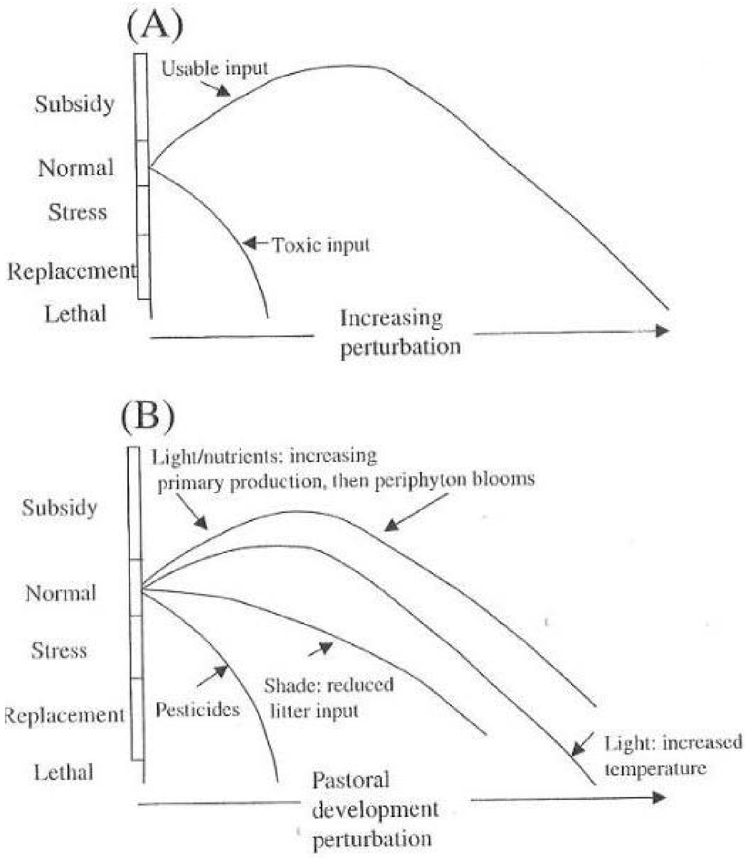


Figure B-1: Conceptual models for subsidy/stress effects of environmental perturbations (A), and as adapted to summarise the key effects of pastoral agriculture on stream macroinvertebrates (B). From Quinn (2000).

Inorganic suspended solids

Relatively few New Zealand studies have examined impacts of inorganic suspended sediment on stream invertebrates, and it is often difficult to separate the effects of increased suspended sediments from those of other pollutants resulting from intensified landuse (Ryan 1991). Relevant studies included ones conducted in the early 1990s on West Coast streams impacted by placer mining (fine, clay inputs), where the main impact is almost exclusively due to elevated suspended sediment (Davies-Colley et al. 1992; Quinn et al. 1992).

Densities of invertebrates downstream of the mining activities were negatively correlated with the logarithm of the turbidity loading (r = -0.82), with densities at downstream sites ranging from 9% to 45% (median 26%) of those at matched upstream sites (Quinn et al. 1992). These reductions in invertebrate densities were associated with as little as 7 NTU increase in turbidity above background. Taxon richness was significantly lower at four sites that had mean turbidity increases between 23 and 154 NTU. Reduced invertebrate densities below mining activities may have been due to a combination of lower periphyton biomass and productivity, degraded food quality, reduced bed permeability and interstitial dissolved oxygen, and increased downstream drift (Quinn et al. 1992). Total invertebrate density provided a better indicator of sediment pollution than either changes in taxon richness or densities of particular species, except for *Deleatidium* (Quinn et al. 1992). Quinn et al. (1992) recommended that average increases be limited to <5 mg/ L suspended sediments or turbidity to <5 NTU to prevent substantial impacts on invertebrate communities of West Coast streams. If the aim is to protect taxa richness, but not abundance, then evidence in Quinn et al.

(1992) suggests <20 NTU increase above reference would be an appropriate limit (Reid and Quinn 2011). A laboratory study investigated acute effects of suspended sediment on stream invertebrates (Suren et al. 2005), which investigated responses of five common native stream insects and a native crayfish that are supposedly sensitive to fine sediment. They showed that even very high clay concentrations (up to ~20,000 NTU), were not toxic over relatively short durations (24 hours). Furthermore, there were no detectable toxic effects on the mayfly *Deleatidium* compared to controls with exposure to 1000 NTU of clay in 4-hr 'pulses' for up to 14 days. They interpreted these null findings as suggesting that absence of these animals from eroding catchments does not express direct toxicity, but must result either from behavioural avoidance (increased drift), or deposition of fine sediment degrading their benthic habitat, or perhaps other indirect effects such as reduced food quality.

International studies on suspended sediment effects on macroinvertebrates

In their review of the influence of suspended sediment on water quality and aquatic biota, Bilotta,Brazier (2008) provided a summary table of studies worldwide that have documented the effects of suspended sediment on stream invertebrates. The local studies of Quinn et al. (1992) and Suren et al. (2005) are included. Of note, most of these studies document acute, rather than chronic, exposure effects. We have updated this table with data from three further studies, two of which document chronic exposure effects (Table B-1). The report by the State of Oregon Department of Environmental Quality (2010) documents the findings of two internal investigation and conclude that benthic macroinvertebrate impairment occurs at (chronic) turbidity levels in the range of 7 NTU to 10 NTU.

Table B-1: Summary of study results on effects of suspended sediments on invertebrates. Adapted from Bilotta and Brazier (2008).

Organism	SS concentration (mg/L or NTU)	exposure (h)	Effect on organism	Country of study	Reference
Benthic invertebrates	8 mg/L	2.5	Increased rate of drift	Canada	Rosenberg,Wiens (1978)
Invertebrates	8–177 mg/L	1344	Reduced invertebrate density by 26%	NZ	Quinn et al. (1992)
Benthic invertebrates	62 mg/L	2400	77% reduction in population size	USA	Wagener,LaPerriere (1985)
Stream invertebrates	130 mg/L	8760	40% reduction in species diversity	England	Nuttall,Bielby (1973)
Macro-invertebrates	133 mg/L	1.5	Seven-fold increase in drifting invertebrates	Australia	Doeg,Milledge (1991)
Cladocera	82–392 mg/L	72	Survival and reproduction harmed	USA	Robertson (1957)
Cladocera and Copepoda	300–500 mg/L	72	Gills and gut clogged	Germany	Alabaster,Lloyd (1982)
Chironomids	300 mg/L	2016	90% decrease in population size	USA	Gray,Ward (1982)
Benthic invertebrates	743 mg/L	2400	85% reduction in population size	USA	Wagener,LaPerriere (1985)
Mayfly (leptophlebiid)	1000 NTU	336	No increased mortality	NZ	Suren et al. (2005)
Invertebrates	20,000 NTU	24	No increased mortality	NZ	Suren et al. (2005)
Invertebrates	25,000 mg/L	8760	Reduction or elimination of populations	England	Nuttall,Bielby (1973)
Macro-invertebrates	1000-1500 NTU	552-576	Reduced visual feeding of ⅓ of test species. Survival of test species increased, growth and feeding unaffected	Australia	Kefford et al. (2010)
Macro-invertebrates	8 (±2) NTU 9 (±2.2) NTU	Winter dataset Not stated	Moderate impairment of riffle macroinvertebrate scores 20% decrease in PREDATOR[a] score	USA	State of Oregon Department of Environmental Quality (2010)

[a] The Predictive Assessment Tool for Oregon (PREDATOR), compares observed macroinvertebrate taxa versus expected taxa.

Appendix C Literature review: The effects of fine sediment on fish

Introduction

Sediment plays a pivotal role in determining the biological integrity of fish communities (Ryan 1991; Bilotta and Brazier 2008; Kemp et al. 2011). Suspended and deposited sediments impact on fish directly through physical effects and indirectly through impacts on habitat, food supply, migratory cues and behaviour. The effects are most often chronic and sub-lethal, leading to a decline in fish growth and condition, curtailed migration, redistribution of populations and changes in population demographics. However, acute, lethal impacts may also occur in extreme circumstances. Regardless of the impact pathway, a reduction in survivorship and consequently, the population size of the affected species is the inferred conclusion. The effects of sediment on fish communities are dependent on several characteristics; the sediment concentration, the duration and frequency that aquatic environments are exposed to the elevated sediment levels and the particle-size distribution of the sediments (Bilotta and Brazier 2008; Collins et al. 2011).

Several comprehensive reviews have been published within New Zealand detailing the effects of sediment in aquatic systems (Ryan 1991; Crowe and Hay 2004; Reid and Quinn 2011; Cavanagh et al. 2014; Davies-Colley et al. 2015). Here, we provide a concise overview of the effects of sediments on fish in lotic (stream) environments, with a primary focus on studies undertaken on New Zealand native fish species. Literature on the introduced brown trout (*Salmo trutta*) and rainbow trout (*Oncorhynchus mykiss*) is also included, as there is a considerable amount of data available on fine sediment impacts on these species, which may help to guide the setting of thresholds for native species. Freshwater crayfish, 130nang (*Paranephrops sp.*), are also included in this review as MfE specified that they should be evaluated as an indicator species. In addition to the literature summarised in these previous reviews, additional novel research that has been produced more recently is also included.

A variety of measures are used to quantify changes in both suspended and deposited sediments in aquatic environments (Bilotta and Brazier 2008; Clapcott et al. 2011; Cavanagh et al. 2014) complicating the interpretation and comparison of results from different studies reported in the literature. As far as possible, these differences are identified, distinguished and accounted for in interpreting the literature reviewed, with all different measures (suspended sediment concentration/suspended solids, turbidity, visual clarity, % cover of deposited sediment on the stream bed, embeddedness etc.,) reported.

Impacts of deposited sediment on fish

Deposited fine sediment impacts riverine fish mainly through reducing overall habitat quality and quantity, particularly for spawning, and through impacts on food supply (Ryan 1991; Kemp et al. 2011). The impact on fish may be direct, particularly through mortality at early life stages, or indirect through declines in reproductive success, growth rates and fish condition. Increases in deposited fine sediment may also cause fish to relocate temporarily, causing short-term, localised declines in population sizes, or permanently causing long-term changes in community composition (e.g., Jowett,Boustead (2001)). These impacts have been well documented overseas, but have received limited attention in New Zealand (Newcombe and Macdonald 1991; Wood and Armitage 1997; Bilotta and Brazier 2008; Kemp et al. 2011). A summary of the key findings relevant to New Zealand is provided below. For more details of the individual studies, see Table C-1.

Fine sediment filtering into the interstitial spaces (gaps) between rocks in the river bed is one of the primary mechanisms through which elevated deposited sediment can impact on fish. The interstitial spaces act as important refuge habitat for small species, as well as juveniles of larger species. The degree to which fine sediments surround coarse substrates on the surface of a streambed is known as embeddedness. New Zealand native fish and crayfish species are mostly associated with the benthos, i.e., are bottom dwelling (McDowall 1990), and broadly favour habitats with larger substrate sizes and, thus, larger interstitial spaces. For instance, upland bully (*Gobiomorphus breviceps*) (Jowett and Boustead 2001), redfin bully (*Gobiomorphus huttoni*) (McEwan and Joy 2014b) bluegill bully (*Gobiomorphus hubbsi*) (Jowett et al. 1996), torrentfish (*Cheimarrichthys fosteri*) (Jowett et al. 1996), adult banded kōkopu (*Galaxias fasciatus*) (Akbaripasand et al. 2011), kōaro (*Galaxias breviceps*) (McEwan and Joy 2014a), shortjaw kōkopu (*Galaxias postvectis*) (McEwan and Joy 2014a), dwarf galaxias (*Galaxias divergens*) (Jowett et al. 1996) and kōura (Usio and Townsend 2001; Kusabs et al. 2015) have all been shown to have an association with these habitats and may, therefore, be negatively impacted by the infilling of interstitial spaces. In contrast, there are some native fish species that are relatively tolerant of deposited sediment, for example, shortfin eels (*Anguilla australis*), and the ammocoete larval life stage of the native lamprey (*Geotria australis*) utilise deposited fine sediments as a key habitat within streams (McDowall 1990).

Experiments where fine sediment was added/removed to natural streams found that the abundance and/or density of bullies (*Gobiomorphus sp.*), eels (*Anguilla* sp.) and brown trout was lowest in the sediment addition reach, and highest in the sediment removal reach after a 27–34 day period (Ramezani et al. 2014). However, no measure of deposited sediment was reported in this study meaning it is of little value for the purposes of informing possible thresholds. In a separate experiment, Jowett,Boustead (2001) evaluated the effects of sediment additions on upland bully densities and found that increasing fine sediment loading resulted in significant declines in fish density, with the primary mechanism thought to be loss of cover habitat due to infilling of interstitial spaces (i.e., increased embeddedness). Sediment loading was reported in this study in terms of mass per unit area, with the treatments being 2.49 kg m^{-2}, 7.46 kg m^{-2} and 14.93 kg m^{-2} with the highest sediment loading essentially representing a condition of 100% embeddedness.

Growth and condition of brown trout (Ramezani et al. 2014) and rainbow trout (Suttle et al. 2004) has also been show to decline in stream reaches affected by high deposited sediment loads. Suttle et al. (2004) experimentally evaluated the consequences of increasing substrate embeddedness (0–100% in 20% increments) and found growth of juvenile rainbow trout declined significantly in response to the direct manipulation of substrate embeddedness. Furthermore, they observed increasing levels of intraspecific aggression as prey availability and visual separation between fish decreased with higher deposited fine sediment levels.

Many fish lay eggs in interstitial spaces within the substrate. Deposition of fine sediment can clog that microhabitat (i.e., increased embeddedness) or smother the eggs themselves. When the eggs are smothered, this disrupts the supply of oxygen to the egg and embryo leading to physiological impacts such as reduced length and weight, or mortality due to hypoxia (Olsson and Persson 1988; Wood and Armitage 1997; Kemp et al. 2011; Louhi et al. 2011) and may lead to premature hatching (Olsson and Persson 1988). Another mechanism via which deposited sediment can impact on breeding success is through emerging fry being trapped in the substrate when they hatch, leading to mortality (Collins et al. 2011).

Some quantitative studies are available that describe the relationship between spawning rate and the degree of fine sediment cover. In these studies, fine sediment measurements are either presented as the fine sediment observed on the surface layer of the streambed (% sediment cover) or the fraction of surface and subsurface sediment that has filtered into the interstitial spaces (% sediment volume). In a study of brown trout alevins (newly spawned trout still carrying the yolk) in English streams, Olsson,Persson (1988) found that 0–10% volume of deposited fine sediment was associated with greater than 88% embryo survival and no premature hatching, 10-20% sediment volume with 28% survival and 55% premature hatching, and greater than 20% sediment volume with 4% survival and 100% premature hatching. Similarly, in a study of Canadian stream channels impacted by logging, Slaney et al. (1977) also found that 19% volume of deposited fine sediment lead to a 30% reduction in the survival of rainbow trout eggs. Crisp,Carling (1989) also found that optimal spawning habitat for brown trout, rainbow trout and Atlantic salmon (*Salmo salar*) was characterised by having less than 10% fine sediment cover.

In New Zealand, low egg survivorship was observed for brown trout in the Waikakahi Stream where fine sediment volume was low (<10%), although these results were likely also influenced by other factors such as low dissolved oxygen levels, high ammonia, and nutrient levels (Hay 2004).

To build a statistical model that describes the influence of habitat change on brown trout populations in Switzerland, Borsuk et al. (2006) determined categories ('low', 'medium' and 'high' impact) that describe the relationship between spawning rate and fine sediment clogging based on the advice of three independent inland fisheries experts. The experts based their rating on research experience and the literature and the categories relate to both an informal test of fine sediment cover and methods used to determine fine sediment volume. The consensus was that 0–10% sediment cover/volume has a low impact, 10–20% has a moderate impact, and >20% has a high impact. Similarly, Clapcott et al. (2011) proposed a limit of <20% sediment cover to support fishery values in New Zealand based on a review of international literature, which provides some guidance on setting criteria for native freshwater values.

Complicating the issue, smaller silt-clay particles may be responsible for suffocating eggs, with Louhi et al. (2011) showing decreased embryo survival and condition in rainbow trout being related to a change in fine sediment (<0.074 mm) from less than 0.5% to 1.5% of total sediment volume. Any deposited fine sediment limits based on % sediment cover would have to assume that these fine particles, in addition to sub-surface sediment that has filtered into the interstitial spaces, are accounted for.

Salmonids are particularly susceptible to deposited sediment impacts, and have been the focus of studies in the international literature (Clapcott et al. 2011). However, many New Zealand native fishes also lay their eggs in the cobbled beds of streams and at the base of aquatic plants (McDowall 1990) and, thus, may be similarly impacted. A stream-based study by Hickford,Schiel (2011) illustrated that fine sediment significantly reduced the availability of spawning habitat for 132nanga (*Galaxias maculatus*), likely by clogging the interstitial spaces in riparian grasses where they lay their eggs. However, there are few examples explicitly addressing the impacts on spawning habitats and spawning success for native fish species.

Elevated sediment deposition is widely recognised to negatively impact macroinvertebrate communities, reducing the availability of food for fish (Ryan 1991; Kemp et al. 2011). This can take the form of an overall decrease in macroinvertebrate abundance, or a change in community composition towards less preferred and more difficult to detect prey, i.e., a reduction in drifting species and an increase in burrowing species (Bilotta and Brazier 2008). Suttle et al. (2004), for example, showed a significant reduction in 'vulnerable prey' (i.e., epibenthic grazers and predators) and replacement by unavailable burrowing macroinvertebrate species as substrate embeddedness increased, particularly above 60%. A reduction in food quality and supply, combined with reduced feeding efficiency from elevated suspended sediments, can reduce fish growth rates and overall condition (Hayes et al. 2000; Collins et al. 2011).

Table C-1: Summary of the documented relationships between deposited fine sediment and fish species found in New Zealand. Where possible, sediment ESV metrics are identified and details of responses are summarised. In many cases, insufficient information is included in the published literature to clarify exact responses in a form consistent with existing sediment ESV measures.

Species	Life Stage	Cause/Effect	Hypothesised Mechanism	Frequency/ duration	Location of study	Sediment ESV metric	Reference
Density/abundance							
Brown trout	Juvenile & Adult	Decreased density with sediment addition, increased with sediment removal	Reduction in habitat and prey abundance	27-34 days	NZ, modified stream channel	SIS 800-1,200 g m^{-2} (exact values not reported)	Ramezani et al. (2014)
Bullies sp.	Juvenile & Adult	Decreased density with sediment addition, increased with sediment removal	Reduction in habitat and prey abundance	27-34 days	NZ, modified stream channel	SIS 800-1,200 g m^{-2} (exact values not reported)	Ramezani et al. (2014)
Eels sp.	Juvenile & Adult	Decreased density with sediment addition, increased with sediment removal	Reduction in habitat and prey abundance	27-34 days	NZ, modified stream channel	SIS 800-1,200 g m^{-2} (exact values not reported)	Ramezani et al. (2014)
Upland bully	Adult	>50% decline in abundance relative to reference condition	Reduction in habitat and prey abundance	6 days	NZ, modified stream channel	Sediment load 2.48 – 14.9 kg m^{-2}	Jowett,Boustead (2001)
Condition/growth							
Brown trout	Juvenile & Adult	Condition (K) lower at sites with sediment added than sites without sediment	Reduced prey abundance and reduced detectability	27-34 days	NZ, modified stream channel	SIS 800-1,200 g m^{-2} (exact values not reported)	Ramezani et al. (2014)
Rainbow trout	Juvenile	Linear reduction in growth with increasing embeddedness	Reduction of available surface prey	46 days	USA, modified stream channel	0, 20, 40, 60, 80, 100% embeddedness	Suttle et al. (2004)
Survival							
Brown trout	Eggs	Decrease in survival	Reduced dissolved oxygen transfer to smothered eggs	8 mon	Canada, lab	1.5% volume of fine sediment (<0.074 mm) in stream gravel	Louhi et al. (2011)

Species	Life Stage	Cause/Effect	Hypothesised Mechanism	Frequency/ duration	Location of study	Sediment ESV metric	Reference
Brown trout	Eggs	65% reduction in survival	Reduced dissolved oxygen transfer to smothered eggs	128 days	UK, experimental stream channel	60% volume of fine sediment (peat material) in stream gravel	Olsson,Persson (1988)
Brown trout	Eggs	~3% reduction in survival @ 10% fine sediment, ~63% reduction in survival @ 20% fine sediment, ~87% reduction in survival @ 40% fine sediment	Reduced dissolved oxygen transfer to smothered eggs, alevins trapped below sediment	126 days	UK, experimental stream channel	0, 5, 10, 20, 40% volume of fine sediment (sand) in stream gravel	Olsson,Persson (1988)
Rainbow trout	Eggs	30% reduction in survival	Reduced dissolved oxygen transfer to smothered eggs	48 days	Canada, modified stream channel	18.7% volume of fine sediment (<0.297 mm) in stream gravel	Slaney et al. (1977)
Habitat association							
Redfin bully	Juvenile & Adult	Presence associated with gravel and larger substrates in day but spread out at night	Likely relates to predation pressure day vs. night	N/A	NZ, survey of a natural stream	0.5 mm as part of substrate index	McEwan,Joy (2014b)
Banded kōkopu	Juvenile & Adult	Size-based microhabitat selection; juveniles associated with fine (<2mm) substrates, adults associated with coarse (>2 mm substrates)	Natural habitat	N/A	NZ, survey of a natural stream	2 mm and as part of a substrate index	Akbaripasand et al. (2011)
Redfin bully	Juvenile & Adult	Presence associated with larger substrates day and night	Natural habitat	N/A	NZ, survey of a natural stream	0.5 mm as part of substrate index	McEwan (2009)
Kōaro	Juvenile & Adult	Presence associated with larger substrates day and night	Natural habitat	N/A	NZ, survey of a natural stream	0.5 mm as part of substrate index	McEwan (2009)
Kōaro & shortjaw kōkopu	Juvenile & Adult	Presence associated with larger substrates day and night	Natural habitat	N/A	NZ, survey of a natural stream	Substrate index	McEwan,Joy (2014a)

Species	Life Stage	Cause/Effect	Hypothesised Mechanism	Frequency/ duration	Location of study	Sediment ESV metric	Reference
Brown trout	Juvenile & Adult	Presence and density negatively correlated with fine sediment depth	Natural habitat	N/A	NZ, survey of natural streams	SIS (exact values not reported)	Lange et al. (2014)
Upland bully	Juvenile & Adult	Presence and density unaffected by fine sediment depth	Natural habitat	N/A	NZ, survey of natural streams	SIS (exact values not reported)	Lange et al. (2014)
Bluegill bully	Juvenile & Adult	Presence associated with gravel and larger substrates	Natural habitat	N/A	NZ, survey of natural lakes	Substrate index	Jowett et al. (1996)
Torrentfish	Juvenile & Adult	Presence associated with gravel and larger substrates	Natural habitat	N/A	NZ, survey of natural lakes	Substrate index	Jowett et al. (1996)
Kōura	Juvenile & Adult	Presence associated with gravel and larger substrates	Natural habitat	N/A	NZ, survey of natural lakes	Substrate index	Kusabs et al. (2015)
Kōura	Juvenile & Adult	Presence associated with gravel and larger substrates	Natural habitat	N/A	NZ, survey of natural streams	Substrate index	Usio,Townsend (2001)

Impacts of suspended sediment on fish

Documented responses of fish to suspended sediment relevant to New Zealand's fish communities are summarised below. Further details of the individual studies are provided in Table C-2.

Direct impacts

Most direct physical effects of elevated suspended sediments are attributed to the clogging, thickening and damaging of the fishes' gills. This reduces respiration leading to declines in growth, greater susceptibility to disease (Waters 1995), and even mortality due to suffocation or stress (Ryan 1991; Wood and Armitage 1997). The type of sediment can further exacerbate the issue, with small, angular, sediment particles found to be more damaging to the gills of juvenile coho salmon (*Oncorhynchus kisutch*) than larger or rounded ones (Lake and Hinch 1999). In a meta-analysis of the data available from the literature, reporting on sediment impacts on aquatic organisms (fish, insects, plants), Newcombe,Macdonald (1991) found that suspended sediment concentration alone was a poor predictor of impacts (r^2 = 0.14, not statistically significant), but concentration and duration combined was a good predictor (r^2 = 0.64, P <0.01).

Research has been conducted on six New Zealand native fish species to determine lethal concentrations of suspended fine sediment (Rowe et al. 2002b; Rowe et al. 2009). These experiments primarily measured the level of turbidity required to cause 50% mortality in a population (referred to as the LC_{50}) over a 24-hour period. Survival rates of banded kōkopu and redfin bully were generally close to 100% irrespective of turbidity levels up to the maximum tested (40,000 NTU), suggesting that fish have adapted resilience to short-term elevated suspended solids that occur during floods. Kōura (*Paranephrops planifrons*) were also tolerant of concentrations >20,000 NTU. In contrast, survival rates for common smelt (*Retropinna retropinna*) and ināanga were around 100% up to a turbidity of 1000 NTU, but declined with increasing turbidity above this level. LC_{50} thresholds were around 3050 NTU for smelt and 20,000 NTU for ināanga, with 100% mortality at around 15,000 NTU and 30,000 NTU for smelt and ināanga respectively (Rowe et al. 2002b). In further analyses of the length of time to 50% mortality under different levels of turbidity (LT50), smelt were also shown to be highly sensitive to relatively short duration (<5 h) high turbidity (>5,000 NTU) events (Rowe et al. 2002b). Similarly, Rowe et al. (2009) reported that survival of banded kōkopu and redfin bully was not reduced by suspended sediment concentrations up to 43,000 g m^{-3} (24-h exposure), but that survival of smelt was reduced by suspended sediment concentrations of over 1000 g m^{-3}. These values are, however, extremely high relative to typical ranges of turbidity and suspended sediment in New Zealand streams.

Longer exposure times to lower levels of suspended sediment have also been shown to cause moderate gill damage (Sutherland and Meyer 2007; Cumming and Herbert 2016) and physiological stress (Herbert and Merkins 1961) leading to lower growth rates, and greater susceptibility to infection, parasitism and disease (e.g., fin rot). Sutherland,Meyer (2007) found moderate and severe gill damage in minnows (a North American species) at suspended sediment concentrations of 100 and 500 g m^{-3}, respectively; and reduced growth rates at suspended sediment concentrations of 25–50 g m^{-3} (21-day exposures). This indicates that some small fish species can be more susceptible to the impacts of elevated suspended sediment concentrations than salmonids. In New Zealand, Rowe et al. (2009) exposed common smelt to sub-lethal suspended sediment levels (c.1000 g m^{-3}) for 4 hours every 2 to 3 days over 2-3 week periods to test prolonged exposure to sub-lethal suspended sediment. The authors recorded no mortality and no outward signs of physiological stress; however, no measurements of growth rate or gill state were taken and so sub-lethal impacts cannot be ruled out at this exposure level. Except for this study by Rowe et al. (2009), tests with other native fishes looking at prolonged exposure to lower levels of suspended sediments are lacking.

Indirect impacts

Indirectly, suspended sediments affect fish through decreases in the visual clarity of water (i.e., increased cloudiness/turbidity), which can alter feeding success and consequently habitat quantity and quality. Movement or migration patterns can also be impacted either due to the changing distribution of suitable habitat or through suspended sediments altering behaviour or blocking migratory cues. When a given threshold for a species is reached, these effects lead to decreased growth rates and changes in community structure and population sizes (Kemp et al. 2011).

Increased turbidity (i.e., reduced visual clarity) has been shown to alter feeding activity, the ability to detect prey, feeding efficiency and the amount and quality of food available to both benthic and drift-feeding fish (e.g. Barrett et al. 1992; Rowe and Dean 1998; Harvey and White 2008). Significant changes to fish feeding rates have been observed at a relatively wide range of turbidity values (15-640 NTU), depending on the species.

Fish reactive distance has been defined in several ways, but essentially describes the distance over which a fish can detect and subsequently intercept prey in flowing waters. Reactive distance is influenced by water velocity, temperature, prey size and fish size (Hayes et al. 2000; Booker et al. 2004) and is highly sensitive to these parameters. Barrett et al. (1992) observed that the reactive distance of rainbow trout (87–185 mm length) was reduced by 20% at 15 NTU over a 1 hour period (and up to 55% at 30 NTU) in a laboratory study, when compared to ambient turbidity of 4-6 NTU. In contrast, using a bioenergetics model Hayes et al. (2016) predicted that the reactive distance of 520 mm brown trout would be reduced by 49% at 10 NTU over a 24-hour period. It is hypothesised that a decrease in reactive distance will reduce feeding efficiency with consequences for fish growth, with the greatest impact on visual feeders. Newcombe (2003) proposed thresholds for visual clarity to protect fish based, in part, on fish reactive distance. This attempted to combine measures of duration of exposure and reduced visual clarity, with a severity of effects score to recommend protection levels. However, the underlying model for the severity of effects score has been criticised for its subjectivity, low statistical power and lack of validation (Kjelland et al. 2015). Furthermore, the model of reactive distance fails to account for differences in water velocity, temperature, fish size and prey availability. However, the general conceptual framework of combining duration of exposure and concentration is valid (Bilotta and Brazier 2008; Chapman et al. 2014).

Amongst the native New Zealand fish fauna, laboratory tank experiments have indicated that fish feeding efficiency is reduced by increasing turbidity for five of six species evaluated (Rowe and Dean 1998). In these tests fish were acclimated at the test turbidity (0, 10, 20, 40, 80, 160, 320, 640 NTU) for two hours prior to the feeding trial commencing. Feeding efficiency was then evaluated over a 30 minute period. Juvenile banded kōkopu, smelt, inānga, common bully and redfin bully all displayed reduced feeding rates at elevated turbidity. Banded kōkopu were concluded to be the most sensitive species with a significant ($p<0.05$) decrease in mean feeding rates at 10 NTU compared to the control (0 NTU). Common bully (160 NTU) and inānga (640 NTU) were the only other species where mean feeding rates were significantly different to the control trial. However, mean feeding rates for both these species began to decline at around 40 NTU. While no statistically significant difference in mean feeding rate was detected for smelt, this in part was due to high individual variation in feeding rates within treatments and overall, this species demonstrated the greatest average reduction in mean feeding rate across the full range of treatments (59%) and showed initial declines from 10 NTU. Redfin bully showed a subsidy-response relationship, with feeding rates initially increasing as turbidity increased from 0 to 40 NTU, but subsequently declining as turbidity was increased above 40 NTU. In contrast, kōaro showed no trend in response over the gradient of turbidity treatments. A later study by Rowe et al. (2002a) with adult inānga and smelt over a turbidity range of 0 to 160 NTU showed a similar negative relationship for inānga, but no significant trend for smelt. However, it was noted that most of the smelt used in this trial had mature gonads and were ready to spawn, a stage when many fish significantly reduce or cease feeding limiting the value of this study for informing thresholds. Greer et al. (2015) also evaluated the impacts of elevated suspended sediment on brown trout in New Zealand. They observed statistically significant decreases in feeding rates at 450 g m^{-3} and 600 g m^{-3} of suspended sediment.

There is a significant gap in the literature on NZ native species (and elsewhere) addressing the longer-term impacts of lower levels of suspended sediment on fish condition. The only study of suspended sediment impacts on fish growth for native New Zealand fishes is reported by Cavanagh et al. (2014). Experimental trials in tanks were used to evaluate the impact of elevated turbidity (0, 5, 15, 50 and 200 NTU) over 21 days on inānga, kōaro, eels and brown trout. Inānga showed a significant decrease in growth rates with increasing turbidity, particularly as turbidity increased from 5 to 15 NTU. The growth rates of kōaro were more resilient, with negative impacts on growth rate only observed when turbidity increased from 15 to 50 NTU. No difference in length or weight of eels was observed over the 21-day trial period and the results for trout were inconclusive (Cavanagh et al. 2014). In the international literature, significant declines in growth rates have been recorded from 10 NTU in brook trout (*Salvelinus fontinalis*) (Sweka and Hartman 2001). Shaw,Richardson (2001) also evaluated the impact of suspended sediment pulses (average concentration 704 g m^{-3}) of varying duration (0-6 hours) on growth of rainbow trout fry. They found that trout length and mass was negatively correlated with pulse length over the 19-day trial period, again highlighting the important influence of duration and frequency of exposure.

Fish are highly mobile and can avoid high sediment concentrations by moving into unimpacted stream reaches (Wood and Armitage 1997; Kemp et al. 2011). Avoidance responses are observed in different species at varying levels of suspended sediment concentration or turbidity, with this considered to be indicative of the overall sensitivity of the species to suspended sediment impacts (Rowe et al. 2000). Boubee et al. (1997) evaluated avoidance of suspended sediment by the juvenile migratory stage of six New Zealand native fish species in laboratory experiments. Banded kōkopu were the most sensitive species, demonstrating a 50% avoidance response at a turbidity of around 25 NTU (20 min exposure time). The thresholds for a 50% avoidance response in kōaro and inānga were 70 and 420 NTU respectively. In contrast, redfin bully and shortfin and longfin eel elvers showed no avoidance behaviour even at the highest turbidity levels evaluated (1100 NTU). In conclusion, Boubee et al. (1997) recommended a limit of 15 NTU to ensure that the upstream migration of key native species was not impacted. Similarly, Rowe et al. (2000) found that banded kōkopu abundance was lower in rivers that were turbid (defined as suspended sediment concentrations >120 g m^{-3} for >20% of the time) during the migration season when compared to clear streams (suspended sediment concentrations >120 g m^{-3} for <10% of the time). Furthermore, Richardson et al. (2001) undertook a field test of banded kōkopu avoidance behaviour and showed that significantly fewer fish migrated upstream within a given period when turbidity exceeded 25 NTU, resulting in recruitment limitation. However, in a series of choice experiments, Baker (2003) found that the threshold for avoidance response to suspended sediment in juvenile banded kōkopu was moderated by the presence of adult banded kōkopu pheromones. Baker,Montgomery (2001) had previously shown that banded kōkopu whitebait exhibited a species-specific attraction to adult pheromones during their migratory phase. Baker (2003) found that despite juvenile banded kōkopu displaying avoidance of 25 NTU water in isolation, when combined with an adult odour, a preference for water with turbidity of up to 35 NTU was shown compared to a control. However, when turbidity was increased to 50 NTU and paired with the adult odour, avoidance behaviour was once again observed. These results indicate that diadromous fishes may be more susceptible to suspended sediment impacts than non-diadromous species. This behaviour could either be the result of avoidance of poor habitat (highly turbid) conditions or the blocking of olfactory senses.

Effects of sediment on other water quality parameters may also have an impact on fish communities (Ryan 1991). Where sediment has a high organic content, dissolved oxygen can be reduced because of decomposition of the organic matter in the water column. For example, Greer (2014) observed significant reductions in dissolved oxygen in response to sediment mobilisation during mechanical macrophyte removal in New Zealand streams, resulting in increased exposure to moderate and severe hypoxia. This has also been observed in the tidal reaches of lowland rivers in New Zealand (Vant 2011; Vant 2013) and overseas(Uncles et al. 1998; Mitchell et al. 1999), where turbidity maxima are associated with zones of hypoxia.

There is also some evidence to indicate that elevated turbidity may impact predation of fish. Gregory,Levings (1998), for example, found evidence for reduced predation of migrating juvenile Pacific salmon in a turbid river (27-108 NTU) compared to a clear tributary (≤1 NTU). Predator avoidance behaviour has also been observed to reduce under conditions of elevated turbidity in juvenile chinook salmon (*Oncorhynchus tshawytscha*) (Gregory 1993) and northern pike (*Esox lucius*) (Lehtiniemi et al. 2005) This has not been documented for any New Zealand species, although there is some anecdotal evidence of increased capture rates of some species in West Coast streams with elevated turbidity compared to nearby clear water reaches (John Quinn, NIWA, pers. com.). Elevated turbidity has been hypothesised to act as cover (Allouche 2002), which is consistent with reduced predator avoidance and the observations of increased capture rates.

Table C-2: Summary of the direct and indirect effects of suspended sediment (SS) on freshwater fish species found in New Zealand. The SS measure (concentration or NTU) reflects the level at which significant effects were observed, unless followed by an * in which case the results showed a trend, although it was not statistically significant. Studies are ordered by increasing SS measure within effect type (e.g., gill damage).

Species	Life Stage	Cause/Effect	Hypothesised Mechanism	Frequency/ duration	Location of study	Suspended Solids Threshold	Reference
Gill Damage							
Brown trout	Juvenile	Gill thickening	Response to physical abrasion	21 day	England, lab tank	810 g m^{-3}	Herbert,Merkins (1961)
Rainbow trout	Juvenile	Slight gill thickening	Response to physical abrasion	64 day	Canada, lab tank	4,887 g m^{-3}	Goldes et al. (1988)
Feeding/foraging success							
Rainbow trout	Juvenile	Reduced reactive distance (20% @ 15 NTU, 55% @ 30 NTU)	Reduced visual clarity	1 hrs	USA, artificial channel	15–30 NTU	Barrett et al. (1992)
Banded kōkopu	Juvenile	Reduction in feeding rate (45%)	Reduced ability to detect prey	2 hrs	NZ, lab tank	20 NTU	Rowe,Dean (1998)
Redfin bully	Juvenile	Reduction in feeding rate (50%)	Reduced ability to detect prey	2 hrs	NZ, lab tank	40 NTU	Rowe,Dean (1998)
Rainbow trout	Adult	No significant effect on feeding rate		30 min	NZ, lab tank	160 NTU	Rowe et al. (2003)
Common bully	Juvenile	Reduced feeding rate (% not stated)	Reduced ability to detect prey	2 hrs	NZ, lab tank	160 NTU	Rowe,Dean (1998)
Inānga	Adult	No significant effect on feeding rate		1 hrs	NZ, lab tank	160 NTU	Rowe et al. (2002a)
Smelt	Adult	No significant effect on feeding rate		1 hrs	NZ, lab tank	160 NTU	Rowe et al. (2002a)
Brown trout	Juvenile	Reduction in feed rate (22%)	Reduced ability to detect prey	90 min	NZ, lab tank	450 g m^{-3}	Greer et al. (2015)

Species	Life Stage	Cause/Effect	Hypothesised Mechanism	Frequency/ duration	Location of study	Suspended Solids Threshold	Reference
Inānga	Juvenile	Reduced feeding rate (% not stated)	Reduced ability to detect prey	2 hrs	NZ, lab tank	640 NTU	Rowe,Dean (1998)
Smelt	Juvenile	Reduction in feeding rate (59%)	Reduced ability to detect prey	2 hrs	NZ, lab tank	640 NTU*	Rowe,Dean (1998)
Kōaro	Juvenile	No significant effect on feeding rate		2 hrs	NZ, lab tank	640 NTU	Rowe,Dean (1998)
Growth							
Inānga	Juvenile (assumed)	No effect on growth, no effect on weight	Reduced feeding efficiency	21 days	NZ, lab tank	15 NTU	Cavanagh et al. (2014)
Kōaro	Juvenile (assumed)	Growth slowed, no effect on weight	Reduced feeding efficiency	21 days	NZ, lab tank	50 NTU	Cavanagh et al. (2014)
Eel sp.	Juvenile (assumed)	No effect on growth, no effect on weight	Reduced feeding efficiency	21 days	NZ, lab tank	200 NTU	Cavanagh et al. (2014)
Rainbow trout	Juvenile	Reduced growth	Reduced feeding efficiency	4-5 Pulses, every second day, for 19 days	Canada, in-stream	700 g m^{-3}	Shaw,Richardson (2001)
Survival							
Inānga	Juvenile (assumed)	No mortality		21 days	NZ, lab tank	15 NTU	Cavanagh et al. (2014)
Kōaro	Juvenile (assumed)	No mortality		21 days	NZ, lab tank	50 NTU	Cavanagh et al. (2014)
Eel sp.	Juvenile (assumed)	No mortality		21 days	NZ, lab tank	200 NTU	Cavanagh et al. (2014)

Species	Life Stage	Cause/Effect	Hypothesised Mechanism	Frequency/ duration	Location of study	Suspended Solids Threshold	Reference
Smelt	Juvenile	No mortality		4 hrs every 2-3 days over 2-3 weeks	NZ, lab tank	1,000 NTU	Rowe et al. (2002b)
Smelt	Adult	LC50	Gill damage	24 hrs	NZ, lab tank	3,000 g m^{-3}	Rowe et al. (2009)
Smelt	Juvenile	LC50	Gill damage	24 hrs	NZ, lab tank	3,050 NTU	Rowe et al. (2002b)
Kōura	Adult	No mortality		24 hrs	NZ, lab tank	20,000 NTU	Rowe et al. (2002b)
Inānga	Juvenile	LC50	Gill damage	24 hrs	NZ, lab tank	20,235 NTU	Rowe et al. (2002b)
Redfin bully	Adult	No mortality		24 hrs	NZ, lab tank	40,000 NTU	Rowe et al. (2002b)
Banded kōkopu	Juvenile	No mortality		24 hrs	NZ, lab tank	40,000 NTU	Rowe et al. (2002b)
Redfin bully	YOY	Mortality (15%)	Gill damage	24 hrs	NZ, lab tank	43,000 g m^{-3} *	Rowe et al. (2009)
Banded kōkopu	Juvenile	Mortality (10%)	Gill damage	24 hrs	NZ, lab tank	43,000 g m^{-3} *	Rowe et al. (2009)
Behaviour							
Banded kōkopu	Juvenile	Avoidance response (50%)		20 min	NZ, lab tank	17–25 NTU	Boubee et al. (1997)
Banded kōkopu	Juvenile	Reduced upstream migration (100%)		100 sec	NZ, in-stream	25 NTU	Richardson et al. (2001)
Banded kōkopu	Juvenile	37% fewer fish attracted to adult odour (migratory cue)		10 min per treatment	NZ, lab tank	50 NTU	Baker (2003)
Kōaro	Juvenile	Avoidance response (50%)		20 min	NZ, lab tank	70 NTU	Boubee et al. (1997)

Species	Life Stage	Cause/Effect	Hypothesised Mechanism	Frequency/ duration	Location of study	Suspended Solids Threshold	Reference
Banded kōkopu	Juvenile & Adult	Reduced upstream migration (89.5%)		5 mon	NZ, in-stream	120 g m^{-3}, >20% of the time	Rowe et al. (2000)
Ināanga	Juvenile	Avoidance response (50%)		20 min	NZ, lab tank	420 NTU	Boubee et al. (1997)
Redfin bully	Juvenile	No avoidance		20 min	NZ, lab tank	1,110 NTU	Boubee et al. (1997)
Shortfin eel	Juvenile	No avoidance		20 min	NZ, lab tank	1,110 NTU	Boubee et al. (1997)
Brown trout	Juvenile	Reduction in abundance (85%)		361 days	England, in-stream	5,838 g m^{-3}	Herbert,Merkins (1961)

Summary of expected fish-sediment ESV response mechanisms

A range of mechanisms have been identified through which elevated sediment can impact on fish communities (Bilotta and Brazier 2008; Collins et al. 2011; Kemp et al. 2011; Chapman et al. 2014; Kjelland et al. 2015). To date, the main New Zealand studies have focused on lethal thresholds, impacts on feeding efficiency and avoidance behaviour. However, relatively few species have been evaluated and there remains a lack of understanding on the long-term effects of elevated sediment exposure. This is consistent in the international literature (Kjelland et al. 2015). The main mechanisms by which fish are thought to be impacted by elevated sediments are summarised in Figure C-1. Sediment impacts occur primarily at a sub-lethal level for most life stages through changes in behaviour, food availability and habitat quality and quantity.

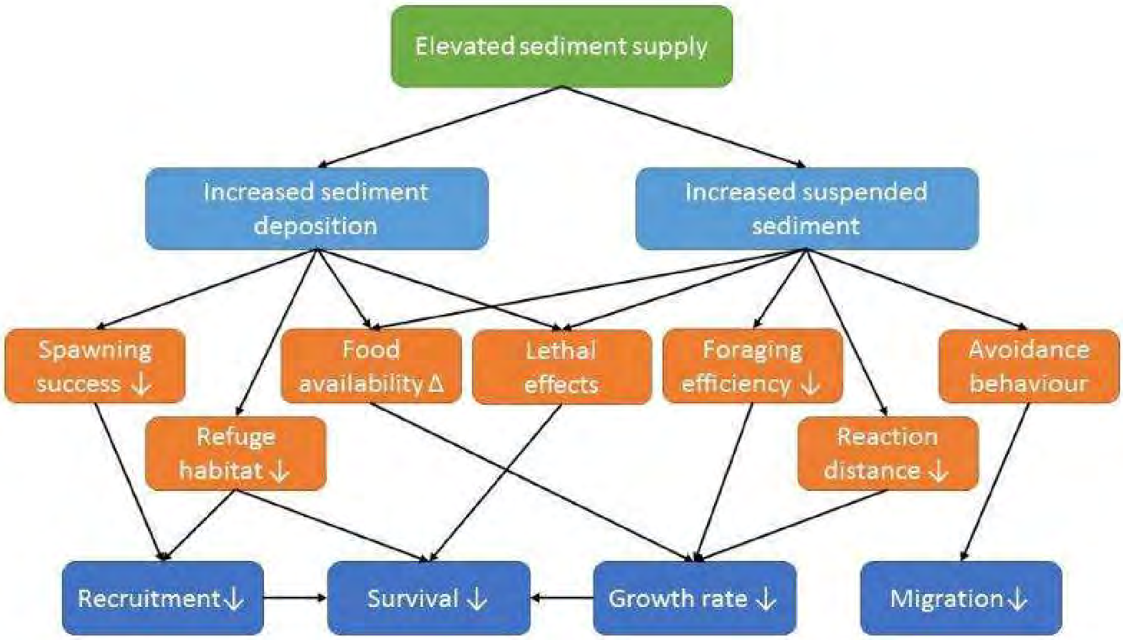


Figure C-1: Summary of key mechanisms governing impacts of elevated sediments on freshwater fish.

Based on the literature review and our knowledge of the ecology of New Zealand's fish species, we have evaluated the expected sensitivity of a range of the more common fish species to chronic elevated fine sediment inputs leading to both elevated suspended and deposited fine sediment levels (Table C-3). This was used to inform our statistical analyses of fish-sediment ESV relationships.

Deposited sediment thresholds in the region of 10-30% cover have commonly been cited as having quantifiable negative effects on specific fish life stages. However, it is rare that studies specifically evaluate consequences at <10% cover and so responses in the 0-10% range are uncertain, particularly over longer durations. For suspended sediment, most of the studies available for NZ species are based on responses to turbidity. Significant effects have been detected for short durations of elevated levels in the range from 5-25 NTU for the more sensitive species. However, studies have rarely evaluated the consequences of elevated turbidity in the range of 0-15 NTU leaving significant uncertainty in responses over this lower range, particularly at longer exposure durations.

Table C-3: Expected sensitivity, based on expert knowledge, of New Zealand's main fish species to elevated fine sediment inputs. *The non-migratory galaxiids grouping is intended to be representative of the expected response of this important group of generally range-restricted endemic taxa. +Exotic species.

Species	Sensitivity to elevated sediment	Hypothesised mechanism(s)
Banded kōkopu	High	Avoidance, reduced feeding.
Kōaro	Medium	Reduced habitat suitability, avoidance, reduced growth.
Inānga	Medium	Reduced feeding and growth.
Shortfin eel	Low	
Longfin eel	Medium	Reduced habitat suitability.
Torrentfish	High	Reduced habitat suitability.
Common bully	Low	
Redfin bully	High	Reduced habitat suitability.
Upland bully	High	Reduced habitat suitability.
Bluegill bully	Medium	Reduced habitat suitability.
Smelt	Medium	Reduced feeding and growth.
Non-migratory galaxiids*	High	Reduced habitat suitability, reduced feeding and growth.
Rainbow trout+	High	Reduced habitat suitability, reduced feeding and growth, reduced spawning success.
Brown trout+	High	Reduced habitat suitability, reduced feeding and growth, reduced spawning success.
Kōura	Medium	

Appendix D Defining sediment reference states of New Zealand catchments

Rationale

The general aim of this analysis was to determine reference states for instream deposited fine sediment and suspended sediment across segments of the New Zealand national river network. For the purposes of this investigation, the reference state of a segment was broadly defined as the levels of deposited and suspended sediment within that segment, on the average through time, assuming minimal urban, agricultural and forestry development within the catchment upstream. The levels of deposited and suspended sediment a segment would experience in its reference state depends on its climatic, topographic and geological context. These factors interact to influence both supply and retention of sediment.

It follows, therefore, that any sediment management objectives—in our case, values delineating the A, B, C and D management bands of the NOF—must take into consideration landscape-scale variability. We require a classification of New Zealand streams such that segments within each class can be assigned a sediment reference state. We require reference states throughout New Zealand for three environmental state variables (ESVs): deposited fine sediment (sediment <2 mm diameter; proportion of streambed covered); turbidity (NTUs); and visual clarity (m). The specific objectives of this analysis were:

- Develop a sediment state classification (SSC) for New Zealand rivers. The SSC will sort New Zealand river segments into groups or 'sediment classes' that have different sediment supply and retention characteristics. As such, the SSC will subdivide the catchments of New Zealand into regions with different sediment supply and retention characteristics.

- Within each sediment class, estimate the reference state for each ESV.

Our approach to meeting these two objectives was guided by the following five principles:

1. **The reference state classification should achieve the right balance between generality, hence ease of use, and sensitivity to any change in the sediment status of steams**. If we have too few classes, then streams that naturally have different sediment characteristics are combined in the one class, leading to the situation where reference conditions are biased. Biased reference conditions, in turn, result in management bands that may either (a) not provide the protective and/or restorative direction required, or (b) result in management objectives that are not achievable. By contrast, if we have too many classes then we may yield a classification system that is complicated and impractical to use, with managers having to frequently refer to new sediment management bands as they move among streams within regions. Moreover, the classification system developed herein will be based on data, and so the number of classes will be constrained by the amount of data available to define each class.

2. **The classification should build on existing river classification systems used in New Zealand, particularly those that have been used to inform catchment policy and management.** There is value in building on a classification system that already exists within scientific literature, and that managers and policy makers are already familiar with. By using a familiar classification system we aim to streamline both adoption and use.

3. **The classification should be (a) based on the key geomorphological and climatological variables that drive sediment supply and retention; and (b) also be based on observed deposited and suspended sediment data, hence capture real differences in the sediment characteristics of rivers.** If key climatological and geomorphological variables that drive sediment supply and retention are used as a basis for our classification, then the classification system will be intuitive to the user. That is, streams in different sediment classes will also have different climatological and/or geomorphological settings. We aim to avoid a classification whereby streams within the one sediment class have obviously contrasting geomorphological or climatological settings—such properties in a classification system may erode confidence in the classification. Equally, we do not wish to have a classification that subdivides streams in an intuitive way yet results in different classes that have indistinguishable sediment characteristics. Therefore, the classification must also be based on real sediment data.

4. **The classification should group stream segments at a spatial resolution reflecting likely changes in the geomorphological and climatological variables driving sediment supply and retention.** If we select too fine a resolution for analysis we generate a risk of yielding a classification whereby sediment management bands switch back and forth frequently as one moves up- or down-stream. A classification with too coarse a resolution would result in whole regions/catchments being grouped together, hence streams with different natural sediment states being treated as equal, in turn resulting in biased reference states.

5. **Estimates of reference state within all regions of New Zealand should result in NOF management bands — hence management targets — that are achievable.** Reference states of ESVs within each sediment class should not be so stringent that management bands are not achievable. As such, reference state estimates within each SS class need to be representative of the streams within that class as a whole.

Development of the Sediment State Classification

Two SSCs were developed; one for deposited fine sediment (SSC_Dep) and one for suspended sediment (SSC_Sus). A separate SSC for deposited and suspended sediment was deemed necessary since, while turbidity and visual clarity are strongly correlated within New Zealand river segments, turbidity and deposited fine sediment are not (Figure D-1). Given turbidity and visual clarity are strongly correlated, we used the turbidity data to develop an SSC for suspended sediment. Turbidity was chosen for development of the SSC_Sus as sites at which turbidity were monitored were more numerous and had greater spatial coverage than those for visual clarity.

To satisfy Principles 2, 3a and 4 presented in the Rationale, we used the New Zealand River Environment Classification (REC; Snelder and Biggs 2002) as a basis for our SSCs. Specifically, the first step of developing our SSCs was to group streams by their REC climate, topography and geology values (REC variables: CLIMATE; SRC_OF_FLW; GEOLOGY). These REC variables were selected as three variables likely to drive supply and retention of both deposited and suspended sediment in New Zealand streams (Table D-1). Combined, these Climate-Topography-Geology (CTG) classes form our 'least aggregated' classification—our starting point—grouping streams that should experience contrasting sediment supply and retention processes.

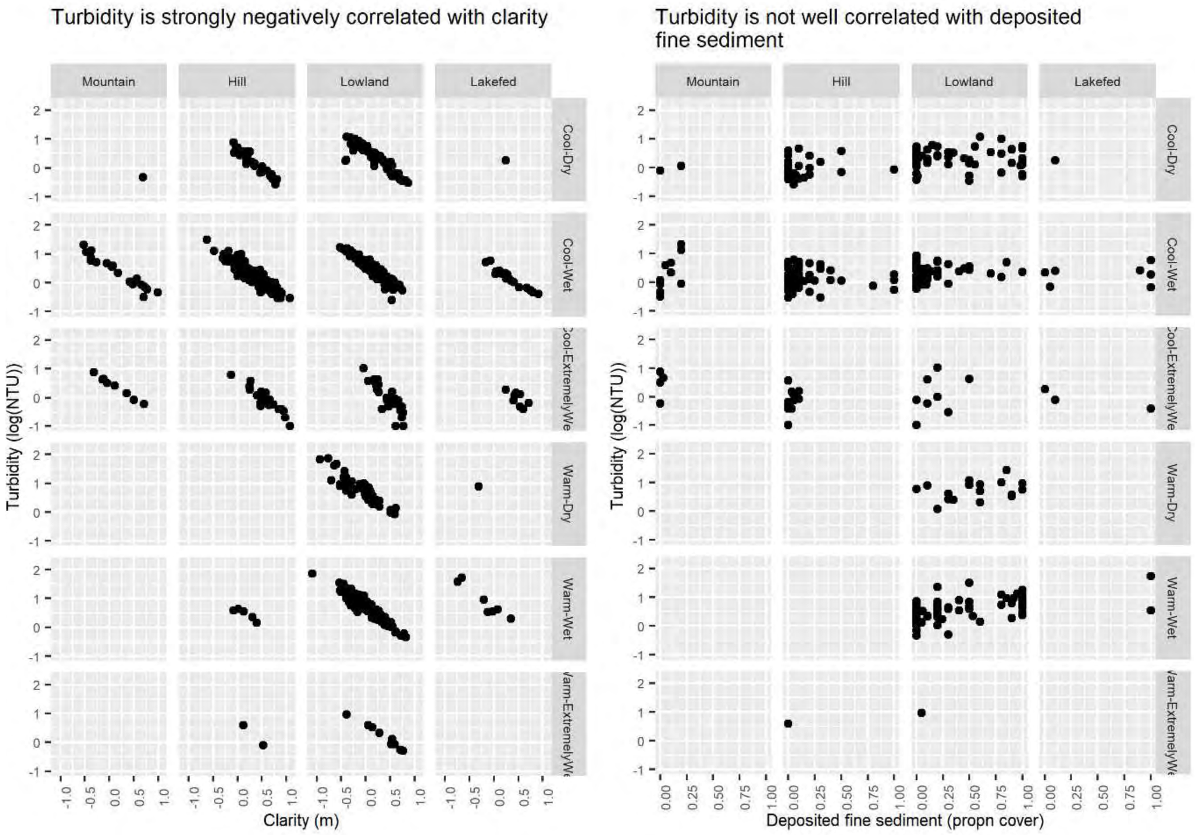


Figure D-1: Relationships between median turbidity and median visual clarity (left plot) and median turbidity and median proportional cover of deposited fine sediment (right plot), within each CTG (Climate-Topography-Geology) class of the New Zealand River Environment Classification (REC). Each point corresponds to a median value from an individual monitoring site (see Methods). Turbidity and visual clarity data sourced from the National River Water Quality Network, while deposited fine sediment data sourced from the New Zealand Freshwater Fish Database.

Table D-1: Explanation of how REC Climate, Topography and Geology classes were aggregated prior to running the SSC algorithm. This further aggregates the resulting CTG classes based on the similarity of their average sediment states. Abbreviations used to define CTG names also presented.

REC variable	Values	Aggregation to form new CTG classes, prior to running SSC algorithm
Climate	Warm-Wet Warm-Extremely Wet Warm-Dry Cold-Wet Cold-Extremely Wet Cold-Dry	Wet and Extremely Wet were combined given these two climatic classes are both characterised by generally high runoff. Hence six values were aggregated to four: Warm-Wet (WW) Warm-Dry (WD) Cold-Wet (CW) Cold-Dry (CD)
Topography (SRC_OF_FLW)	Lowland Lakefed Hill Mountain Glacial Mountain	Mountain and Glacial Mountain classes combined on the basis of them both being associated with rivers of high gradient, hence low sediment retention. Yielding four topography classes: Lowland (Low) Lakefed (Lake) Hill (Hill) Mountain (Mount)
Geology	Soft Sedimentary Hard Sedimentary Alluvium Plutonic Volcanic Miscellaneous Volcanic Basic Volcanic Acidic	Plutonic Volcanic and Miscellaneous were aggregated with Soft Sedimentary, based on exploration of the frequency histograms of sediment values within CTG classes, and consultation with expert geologists. Volcanic Basic and Volcanic Acidic combined to form Volcanic – geology resistant to erosion. This aggregation yielded four geological classes: Soft Sedimentary (SS) Hard Sedimentary (HS) Alluvium (Al) Volcanic (VA)

Using the CTG classification as a basis, we then implemented the following steps towards meeting Objectives 1 and 2:

1. Characterise each CTG class as a 'vector' (an ordered list of numbers) defining the frequency distribution of observed ESV values within that CTG class; if there are insufficient data in initial CTG classes, aggregate in a logical fashion until the frequency distribution of the ESV can be defined in each CTG class.

2. Based on the frequency distribution of ESV values within each CTG class, use multivariate analysis to determine the dissimilarity of ESV frequency distributions among classes, such that we may aggregate CTG classes into sediment classes based on the similarity of their ESV distributions.

3. Within each sediment class, estimate the ESV reference states, determine which level of aggregation provides the most parsimonious description of reference states for each ESV, and map the spatial distribution of sediment classes, hence reference states, to all river reaches of the New Zealand river network.

Step 1: Characterising CTG classes by their ESV characteristics

In accordance with Principles 3 and 5, we wished to develop a SSC that was based on observed sediment data, so we needed to define what a ‘sample’ was. Herein, an individual sample was the median of all ESV values recorded at a monitoring site. What constituted a monitoring site varied between ESVs. For deposited fine sediment a ‘site’ was an individual reach (NZReach) within the New Zealand Freshwater Fish Database (NZFFD). For the two suspended sediment ESVs, a monitoring site was an individual monitoring station within the regional council monitoring network, which comprises the State of the Environment (SOE) monitoring. Specifically, the suspended sediment data used herein was the data collated by the 2018 MfE State and Trends Projects (Whitehead 2019).

If we were to do no aggregation of the CTG classes in Table D-1 we have up to 6 (climate classes) x 5 (topography classes) x 7 (geological classes) = 210 possible CTG classes. If we undertook no further aggregation then our sediment classes would be our 210 CTG classes (assuming all CTG classes are represented within New Zealand).

We defined the ESV characteristics of each CTG class as the frequency distribution of ESV values within that CTG class. To estimate a frequency distribution we obviously require a ‘reasonable’ number of samples. We selected N = 20 samples as the minimum sample size for histogram estimation. This value is somewhat arbitrary, but its selection was based on exploration of the data and seeking a balance between a minimum N that was too stringent (too high, resulting in too many CTGs being excluded from the SSC) and too lenient (too low, resulting in an imprecise characterisation of the sediment state of a CTG class).

The routine for defining the histogram bins was common to each ESV: 11 bins were established for each ESV, with 10 breakpoints defined as a sequence from the minimum value, to the maximum value, with a step size of range/10. The minimum, maximum and range were estimated using the global dataset for each ESV.

Prior to moving onto Step 2, we aggregated certain CTG classes if (a) one of a pair of CTG classes had N <20; (b) the two CTG classes were likely to experience similar sediment supply and retention characteristics. The CTG classes resulting from this first step of aggregation are presented in Table D-1. Most CTG classes containing Lakefed topographies were associated with very few (<20; often <10) monitoring sites. We could not, however, simply exclude them from the SSC based on low sample size, as Lakefed topographies contained some of the most socio-economically important rivers (e.g., Clutha; Waitaki; Waiau in Southland). Accordingly, the CTG classes containing Lakefed topographies required special treatment: We assumed no variation in the effect of geology on either deposited or suspended sediment, within any CTG class containing the Lakefed topography. As an example, consider the four Cold-Wet, Lakefed CTG classes, differentiated by different geologies CW_Lake_SS, CW_Lake_HS, CW_Lake_Al and CW_Lake_VA; these four CTG classes are pooled into a single CTG class (assuming geology has no impact): CW_Lake_Any. Our assumption was based on the reasoning that lakes are excellent sediment traps, and so the geology underpinning rivers flowing from lakes should have minor effects on sediment supply and retention, relative to the lakes themselves.

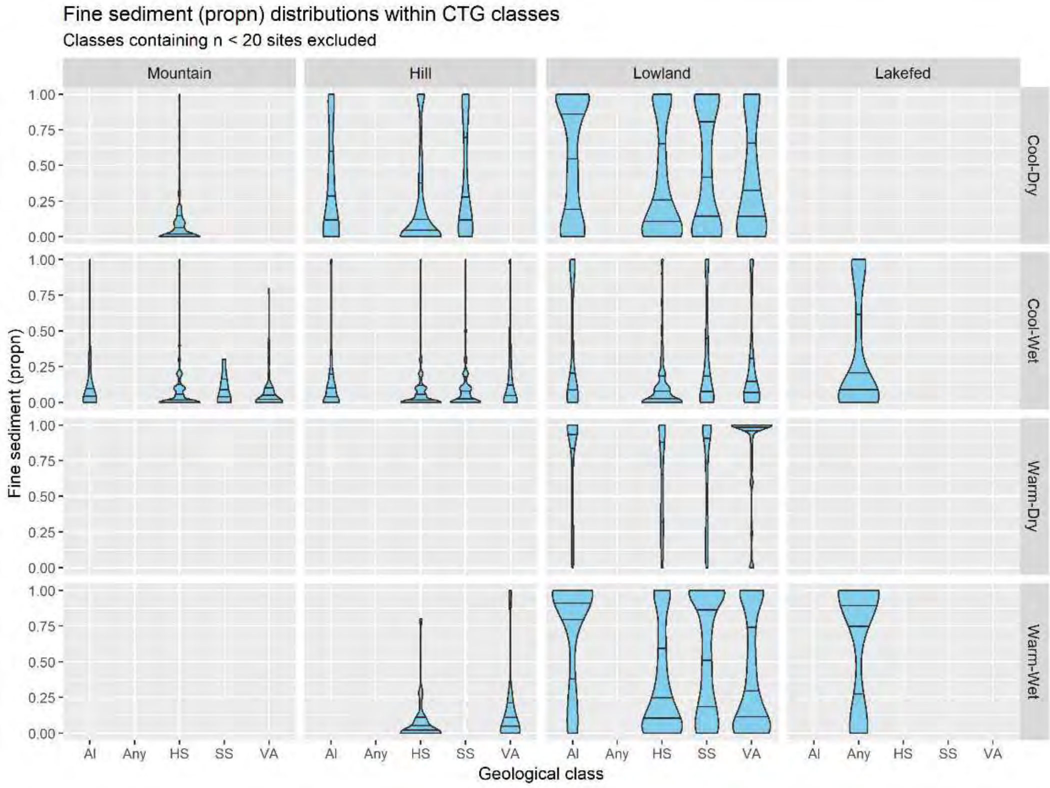


Figure D-2: Violin plots describing the frequency distributions of deposited fine sediment within all REC CTG (Climate-Topography-Geology) classes considered.

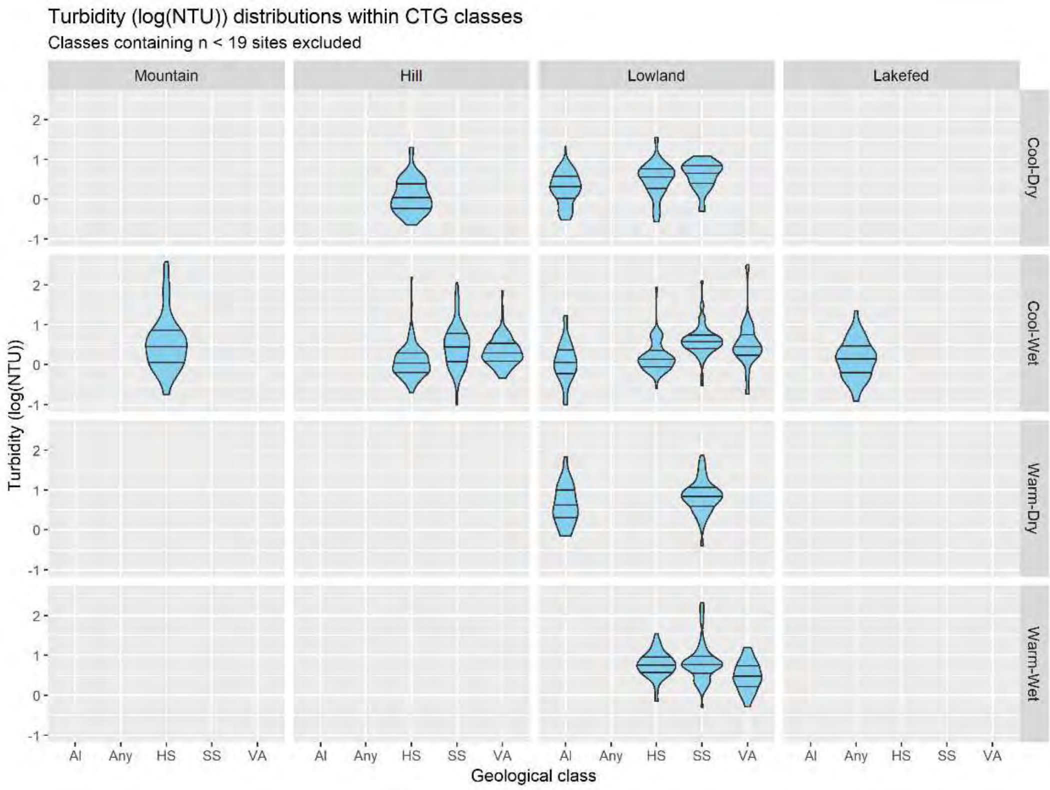


Figure D-3: Violin plots describing the frequency distributions of suspended fine sediment (turbidity) within all REC CTG (Climate-Topography-Geology) classes considered.

Table D-2: Climate-Topography-Geology (CTG) classes that were either mapped or unmapped to the SSC for both deposited and suspended sediment. CTG classes unmapped contained less than 20 sites/samples to use for defining the ESV histogram.

	Deposited SSC		Suspended SSC	
	Mapped	Unmapped	Mapped	Unmapped
1	CD_Hill_Al	CD_Hill_VA	CD_Hill_HS	CD_Hill_Al
2	CD_Hill_HS	CD_Lake_Any	CD_Low_Al	CD_Hill_SS
3	CD_Hill_SS	CD_Mount_Al	CD_Low_HS	CD_Hill_VA
4	CD_Low_Al	CD_Mount_SS	CD_Low_SS	CD_Lake_Any
5	CD_Low_HS	CD_Mount_VA	CW_Hill_HS	CD_Low_VA
6	CD_Low_SS	WD_Hill_VA	CW_Hill_SS	CD_Mount_Al
7	CD_Low_VA	WD_Lake_Any	CW_Hill_VA	CD_Mount_HS
8	CD_Mount_HS	WW_Hill_SS	CW_Lake_Any	CD_Mount_SS
9	CW_Hill_Al		CW_Low_Al	CD_Mount_VA
10	CW_Hill_HS		CW_Low_HS	CW_Hill_Al
11	CW_Hill_SS		CW_Low_SS	CW_Mount_Al
12	CW_Hill_VA		CW_Low_VA	CW_Mount_SS
13	CW_Lake_Any		CW_Mount_HS	CW_Mount_VA
14	CW_Low_Al		WD_Low_Al	WD_Hill_VA
15	CW_Low_HS		WD_Low_SS	WD_Lake_Any
16	CW_Low_SS		WW_Low_HS	WD_Low_HS
17	CW_Low_VA		WW_Low_SS	WD_Low_VA
18	CW_Mount_Al		WW_Low_VA	WW_Hill_HS
19	CW_Mount_HS			WW_Hill_SS
20	CW_Mount_SS			WW_Hill_VA
21	CW_Mount_VA			WW_Lake_Any
22	WD_Low_Al			WW_Low_Al
23	WD_Low_HS			
24	WD_Low_SS			
25	WD_Low_VA			
26	WW_Hill_HS			
27	WW_Hill_VA			
28	WW_Lake_Any			
29	WW_Low_Al			
30	WW_Low_HS			
31	WW_Low_SS			
32	WW_Low_VA			

Within the REC there are a total of 52 CTG classes represented (combinations of certain classes in column 3 of Table D-1). Of these, we had sufficient data to include 34 CTG classes for deposited sediment, and 18 CTG classes for turbidity. As we will see below (Step 3), although we had insufficient data to include a large proportion of the total CTG classes, the CTG classes included comprise a majority of the New Zealand stream network. The CTG classes for which we had sufficient deposited and suspended sediment data are presented in Figure D-2and Figure D-3 respectively. Although variation in ESV composition among CTG classes is evident for both deposited and suspended sediment in these figures, many CTG classes exhibit similar ESV composition, justifying further aggregation of the SSCs (Figure D-2 and Figure D-3). For those CTG classes that could not be classified, we developed a spatial mapping procedure to allocate them to a class (see Appendix E for details).

Within Table D-2 the CTG classes for which we had sufficient data (n ≥ 20) and insufficient data are listed, for both the deposited and suspended SSC.

Step 2: Aggregation of CTG classes using cluster analysis

Following Step 1 the set of CTG classes for both deposited and fine sediment was characterised as a 'CTG class' x 'histogram-bin' matrix, thus permitting the estimation of multivariate similarity in the frequency distribution of sediment values among CTG classes. Bray-Curtis similarity between CTG classes was estimated prior to classification analysis using hierarchic clustering. Clustering was performed using average linkage, which tends to preserve the structure of dissimilarity among samples better than complete and single linkage algorithms (Oksanen 2015) R package *vegan* was used for all multivariate analysis (Oksanen et al. 2018).

In hierarchical clustering there are fewer clusters at a higher level of dissimilarity, while at a lower level of dissimilarity more clusters are produced. Thus the classification method used herein yields SSCs containing different numbers of sediment classes, depending on the level of dissimilarity selected to aggregate CTG classes into sediment classes. We generated four SSCs for both the deposited and suspended ESVs; one each for sediment classes grouped at (1) 50%; (2) 30%; (3) 20%; and (4) 15% dissimilarity. For both deposited and suspended sediment, these dissimilarities yielded 2, 4, 8 and 12 sediment classes. For ease of communication we hereafter refer to these different critical dissimilarities as 'levels of aggregation', with Aggregation Levels 1, 2, 3 and 4 corresponding to sediment classes aggregated at 50%, 30%, 20% and 15% dissimilarity respectively. Individual sediment classes within each level are referred to in a manner such that the level of aggregation is explicit; for example, sediment classes L1.1 and L4.3 are, respectively, sediment classes 1 at Aggregation Level 1, and 3 at Aggregation Level 4.

For both deposited and suspended fine sediment the cluster analysis yielded sediment classes that clearly had different climatic, topographical and geological characteristics (Figure D-4, Figure D-5 and Table D-3). Examination of the frequency distributions of values within sediment classes showed very strong differences in distributions at Aggregation Level 1 for both deposited (Figure D-6) and suspended sediment (Figure D-7). Differences in the distributions of sediment values among classes became more nuanced through Levels 2 – 4 for both deposited and suspended sediment (Figure D-6 and Figure D-7). Results of the cluster analysis are summarised in Table D-3.

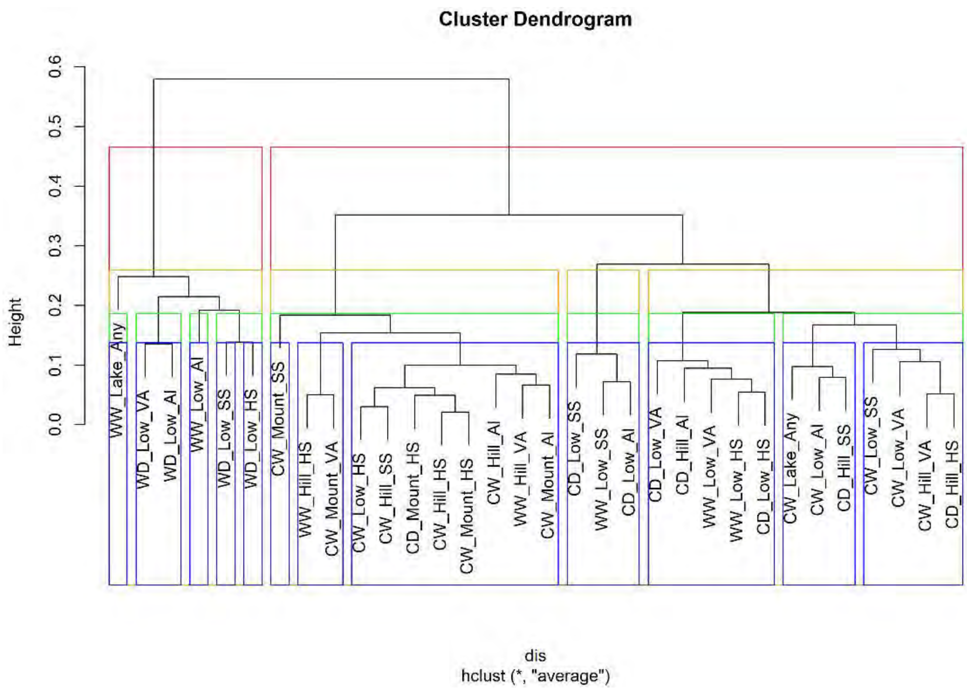


Figure D-4: Dendrogram showing four levels of aggregation of CTG classes based on the (Bray-Curtis) similarity of their frequency distributions of deposited fine sediment. Boxes outline sediment classes at Aggregation Levels 1 (red; two groups); 2 (orange; four groups); 3 (green; eight groups); and 4 (blue; 12 groups).

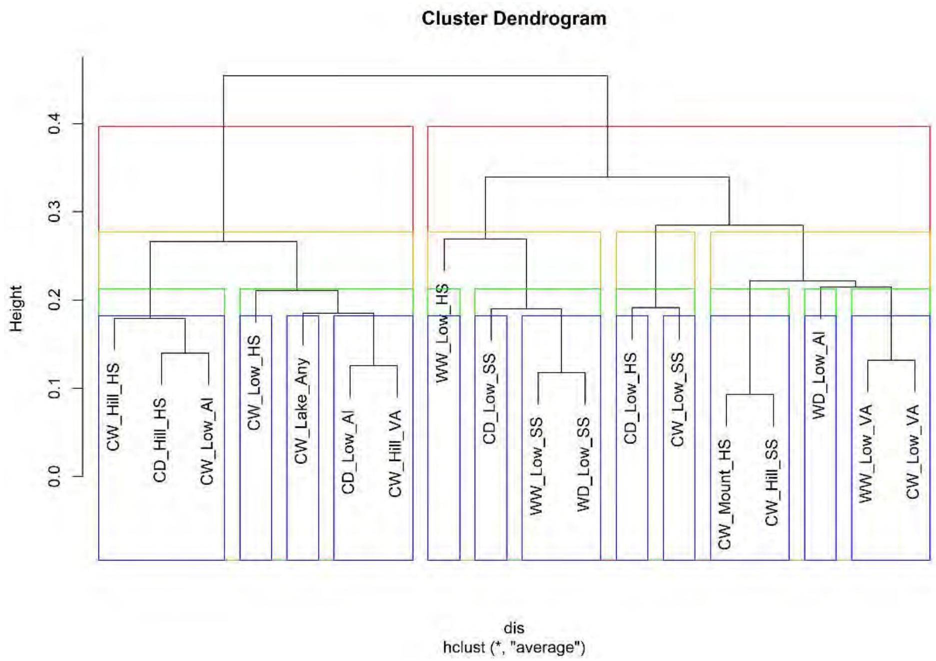


Figure D-5: Dendrogram showing four levels of aggregation of CTG classes based on the (Bray-Curtis) similarity of their frequency distributions of turbidity values (NTUs). Boxes outline sediment classes at Aggregation Levels 1 (red; two groups); 2 (orange; four groups); 3 (green; eight groups); and 4 (blue; 12 groups).

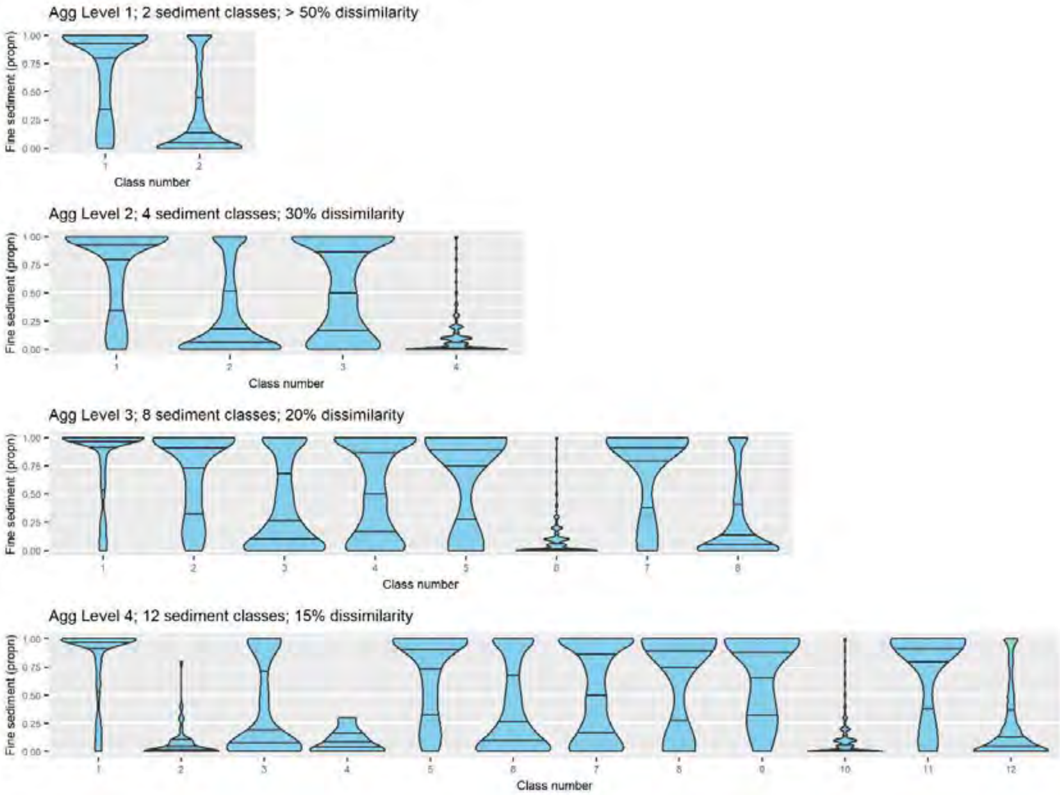


Figure D-6: Violin plots describing the frequency distributions of deposited fine sediment within sediment classes at different levels of aggregation.

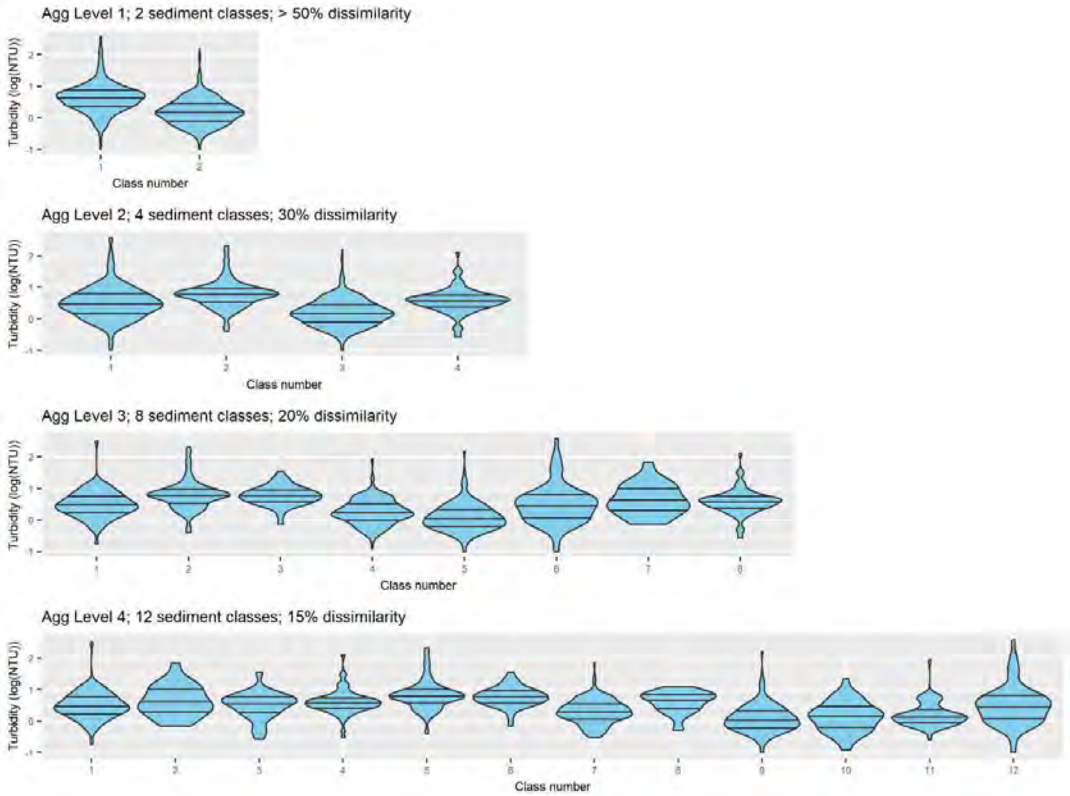


Figure D-7: Violin plots describing the frequency distributions of suspended sediment (turbidity) within sediment classes at different levels of aggregation.

Table D-3: Class membership hierarchy for both the deposited and suspended sediment classes at different levels of aggregation (Aggregation Levels 1-4). CTG = Climate-Topography-Geology classes. Rates of sediment supply and retention have been included: very high; high; medium; low; very low.

<table>
<tr><th colspan="6">Deposited fine sediment class hierarchy</th><th colspan="6">Suspended sediment class hierarchy</th></tr>
<tr><th>Agg L1</th><th>Agg L2</th><th>Agg L3</th><th>Agg L4</th><th>CTG Classes</th><th>Supply, retention</th><th>Agg L1</th><th>Agg L2</th><th>Agg L3</th><th>Agg L4</th><th>CTG Classes</th><th>Supply</th></tr>
<tr><td rowspan="5">1</td><td rowspan="5">1</td><td>1</td><td>1</td><td>WD_Low_VA; WD_Low_Al</td><td>Very high</td><td rowspan="8">1</td><td rowspan="3">1</td><td>1</td><td>1</td><td>WW_Low_VA; CW_Low_VA</td><td>Med</td></tr>
<tr><td rowspan="2">2</td><td>5</td><td>WD_Low_SS</td><td>Very high</td><td>6</td><td>12</td><td>CW_Mount_HS; CW_Hill_SS</td><td>High</td></tr>
<tr><td>9</td><td>WD_Low_HS</td><td>High</td><td>7</td><td>2</td><td>WD_Low_Al</td><td>Very high</td></tr>
<tr><td>5</td><td>8</td><td>WW_Lake_Any</td><td>High</td><td rowspan="3">2</td><td rowspan="2">2</td><td>5</td><td>WW_Low_SS; WD_Low_SS</td><td>Very high</td></tr>
<tr><td>7</td><td>11</td><td>WW_Low_Al</td><td>Very high</td><td>8</td><td>CD_Low_SS</td><td>High</td></tr>
<tr><td rowspan="7">2</td><td rowspan="3">2</td><td>3</td><td>6</td><td>WW_Low_VA; WW_Low_HS; CD_Low_VA; CD_Hill_Al; CD_Low_HS</td><td>Low</td><td>3</td><td>6</td><td>WW_Low_HS</td><td>High</td></tr>
<tr><td rowspan="2">8</td><td>12</td><td>CW_Hill_VA; CW_Low_VA; CW_Low_SS; CD_Hill_HS</td><td>Very low</td><td rowspan="2">4</td><td rowspan="2">8</td><td>3</td><td>CD_Low_HS</td><td>High</td></tr>
<tr><td>3</td><td>CW_Lake_Any; CW_Low_Al; CD_Hill_SS</td><td>Low</td><td>4</td><td>CW_Low_SS</td><td>High</td></tr>
<tr><td>3</td><td>4</td><td>7</td><td>WW_Low_SS; CD_Low_SS; CD_Low_Al</td><td>Med</td><td rowspan="4">2</td><td rowspan="4">3</td><td rowspan="3">4</td><td>7</td><td>CD_Low_Al; CW_Hill_VA</td><td>Med</td></tr>
<tr><td rowspan="3">4</td><td rowspan="3">6</td><td>10</td><td>WW_Hill_VA; CW_Hill_HS; CW_Low_HS; CW_Mount_HS; CW_Hill_SS; CW_Hill_Al; CD_Mount_HS; CW_Mount_Al</td><td>Very low</td><td>10</td><td>CW_Lake_Any</td><td>Very low</td></tr>
<tr><td>2</td><td>WW_Hill_HS; CW_Mount_VA</td><td>Very low</td><td>11</td><td>CW_Low_HS</td><td>Low</td></tr>
<tr><td>4</td><td>CW_Mount_SS</td><td>Low</td><td>5</td><td>9</td><td>CW_Hill_HS; CD_Hill_HS; CW_Low_Al</td><td>Very low</td></tr>
</table>

Step 3: Estimating ESV reference states

Two broad approaches to estimating reference state were considered: The first approach involves estimating the state of an ESV within river segments that have no history of significant anthropogenic disturbance upstream—the 'reference site' approach. Under this approach, the reference state is often referred to as the 'minimally disturbed condition' (Lewis et al. 1999; Stoddard et al. 2006). An advantage of the reference site approach is its simplicity; the definition of reference state is intuitive and its calculation requires little to no statistical sophistication and so is easy to explain.

However, minimally-disturbed river segments are usually rare, resulting in very few replicate reference sites per sediment class, which may in turn lead to biased estimates of reference state (McDowell et al. 2013). That is, the lower the number of replicate reference sites the greater the risk of having reference states that are not representative of the broader region we wish to manage.

The second approach we considered for estimating reference states of ESVs was that of Dodds,Oakes (2004). This approach involves (a) selecting a model that describes ESV state as a function of covariates that describe the magnitude of anthropogenic disturbance within a region; and (b) using that model to estimate predicted ESV state at zero anthropogenic disturbance. We refer to this approach as the 'model-based' approach. The model-based approach involves using all the data available within a region, and so it follows that (a) if the sites from which data are obtained are randomly distributed throughout the region we wish to manage; and (b) if our model is a good fit to the data, then we obtain a least biased estimate of reference state.

A disadvantage of the model-based approach is that it is more complex than the reference site approach, and so may be more difficult for various stakeholders to understand.

In the present study we used the model-based approach, due to the small number and restricted distribution of reference sites for deposited and suspended sediment. We sought parsimonious models of reference states within sediment classes. Towards that end the following set of candidate models was fitted to each ESV, at each aggregation level:

$$ESV = \beta_0 + \beta_1 P + \beta_2 C + \beta_3 PC + \varepsilon \qquad \text{Model 1}$$

$$ESV = \beta_0 + \beta_1 P + \beta_2 C + \beta_3 PC + \beta_4 E + \beta_5 EC + \varepsilon \qquad \text{Model 2}$$

$$ESV = \beta_0 + \beta_1 P + \beta_2 C + \beta_3 PC + \beta_4 U + \beta_5 UC + \varepsilon \qquad \text{Model 3}$$

$$ESV = \beta_0 + \beta_1 P + \beta_2 C + \beta_3 PC + \beta_4 E + \beta_5 EC + \beta_6 U + \beta_7 UC + \varepsilon \qquad \text{Model 4}$$

In the above equations the β values are parameters and ε is error. The covariates P, E and U are continuous covariates with domain [0,1] describing the proportions of the catchment upstream comprised of heavy pasture, exotic vegetation (mostly pine forests) and urban development, respectively. C is a categorical, fixed covariate referring to the sediment class. The number of values of C is dependent on the aggregation level: at Level 1, C has two values (one for each of two sediment classes); at Level 2, C has 4 values; at Level 3, C has 8 values; at Level 4, C has 12 values. When the ESV was deposited fine sediment (proportion) we used binomial linear models, but when the ESV was either turbidity or visual clarity, Gaussian linear models were fitted.

For each ESV, the Akaike Information Criterion (AIC; Burnham and Anderson 2002) was used to select the most parsimonious candidate from Models 1-4, within each aggregation level. Consequently, for each ESV we generated four possible models of reference state; one at each level of aggregation. To obtain the reference state within each sediment class, within each aggregation level, we obtained the predicted value within each level of C, with other covariates set to zero. Occasionally the slope of the fitted model within a certain sediment class was approximately zero and in the direction opposite to that expected (e.g., turbidity actually decreasing as anthropogenic pressure increases). When this occurred the reference state for this class was estimated as the median ESV value.

The final step of selecting an appropriate model of reference state was to choose the level of aggregation of sediment classes.

Within this project, sediment reference states are passed to models of biological response to each ESV, which in turn are used to estimate NOF management bands. Accordingly, the decisive factor determining which aggregation level to use may be the availability of either ESV or biological data in sediment classes. In any case, to assist decisions concerning the level of aggregation to use, we provided three further outputs:

First, for each ESV the optimal models (from Models 1-4) across each of the four levels of aggregation differed considerably in their complexity. Suppose, for example, that Model 4 is the most likely model of reference state for Aggregation Levels 1 (average dissimilarity between classes = 50%) and 4 (average dissimilarity = 15%). Then at Level 1 the most likely model of reference states has 8 parameters while at Level 4 the most likely model has 48 parameters. The Level 4 model is likely to yield less biased estimates of reference states, because the reference state estimation is allowed to vary across a fine-resolution decomposition of sediment classes throughout New Zealand. But any reduction in bias comes at the cost of many more parameters. Thus we have a standard model selection problem of the need to find an appropriate balance between model complexity and simplicity. We employed information-theoretic statistics to help find that balance. Specifically, for each ESV, we estimated the following statistics for the most likely models at each of the four levels of aggregation: (a) AIC; (b) the AIC model rank: $\Delta_i = AIC_i - \min(AIC)$; and (c) w_i, the Akaike weight of model i, interpreted as the approximate probability that Model i is the best model in the candidate set, given the data (Burnham and Anderson 2002).

Second, for each ESV we generated plots to compare and contrast estimates of reference state with:

- The median and interquartile range of the ESV, within the subset of the data where heavy pasture values were less than the lowest decile of all heavy pasture values (ESV_HPd1). This statistic provides a ‘check’ on the alignment between our modelled reference estimate and the distribution of observed ESV values under minimal anthropogenic disturbance, within each sediment class.
- The median and interquartile range of the ESV as measured at reference sites within each sediment class (Reference). In this study a reference site was a site with the following catchment characteristics upstream, as estimated within the NZ REC (Snelder and Biggs 2002): <90% cover of native vegetation; 0% coverage of urban development; <5% exotic vegetation (hence <5% commercial forestry).

- The median and interquartile range of all ESV data within each sediment class (ESV_allData), such that we may see how our modelled reference states contrast with the contemporary, observed state of that ESV throughout regions defined by each sediment class. One could suggest that, if our modelled reference states are useful, then within sediment classes associated with agricultural development we would ideally see (i) reference states below the median of ESV_allData; but (ii) reference estimates that are not so far below the IQR of ESV_allData that entire regions of NZ are set unachievable management objectives.

Third, for each ESV we generated plots showing how biased our estimates of reference state might be when we use a higher level of aggregation, when we group together more streams that may have different sediment states. These plots were designed to demonstrate the direction and magnitude of change in estimated reference state—hence the magnitude and direction of bias—as we move from a lower level of aggregation (e.g., Level 2; 30% dissimilarity between sediment classes) to a higher level of aggregation (e.g., Level 1; 50% average dissimilarity between sediment classes). Our SSCs are hierarchical, so multiple reference states within a lower level of aggregation may correspond to a single reference state at the next highest level of aggregation. In these plots we will see just how much several estimates of reference states at lower levels of aggregation are pulled towards the 'average' reference states at the higher levels of aggregation.

Deposited fine sediment

For deposited fine sediment, Model 4 provided the most parsimonious description of the data at all levels of aggregation. Hence, given the data and candidate models, we found variation in deposited fine sediment throughout New Zealand is best explained by the additive effects of heavy pasture, urbanisation and forestry, and how those three drivers interact with sediment classes of the New Zealand landscape. Using Nagelkerke's R^2 for generalised linear models, the R^2 values for the fit of Model 4 to the deposited sediment data were: 0.25 (Level 1); 0.31 (Level 2); 0.33 (Level 3); 0.34 (Level 4).

The fitted optimal models for deposited fine sediment are presented in Figure D-8. Table D-4 presents the reference states (intercepts) for, and the proportion of the NZ REC covered by, each sediment class, at each level of aggregation. In one instance (Class L3.6, which is also Class L4.9; Table D-4) a counterintuitive slope was returned (Figure D-8, Agg. Level 3 and 4), resulting in the reference state for that class being estimated as the median deposited sediment value in that class (Table D-4). For most classes we had a good range of heavy pasture values for regression, irrespective of level of aggregation (Figure D-8).

It is clear from Figure D-8 and Table D-4 that the variation in reference state across sediment classes increases as we move from Aggregation Level 1 through to Level 4. This can also be seen in Figure D-9, which presents the comparisons of our reference state estimates for deposited fine sediment with ESV_HPd1, the value of the ESV at reference sites and ESV_allData. The following inferences may be gleaned from Figure D-9:

- Using the method of classification derived here, less than 2% of the New Zealand river network was unclassified.

- Irrespective of the level of aggregation, model-based estimates of reference state were always higher (equating to higher levels of deposited sediment) than the reference site estimates. This could mean either (a) reference site estimates of deposited fine sediment reference state are too restrictive; or (b) model-based estimates are too liberal. We will know which of these explanations is closest to the truth when these reference states are used to derive management bands in subsequent chapters; for example, if, using these model-based reference values, no rivers fall into the C-D Management Band—such that no management action is required anywhere—then it is likely our model-based estimates of reference state are too liberal.
- At lower levels of aggregation, the number of reference sites is often very low (<10; Figure D-9 Level 3).

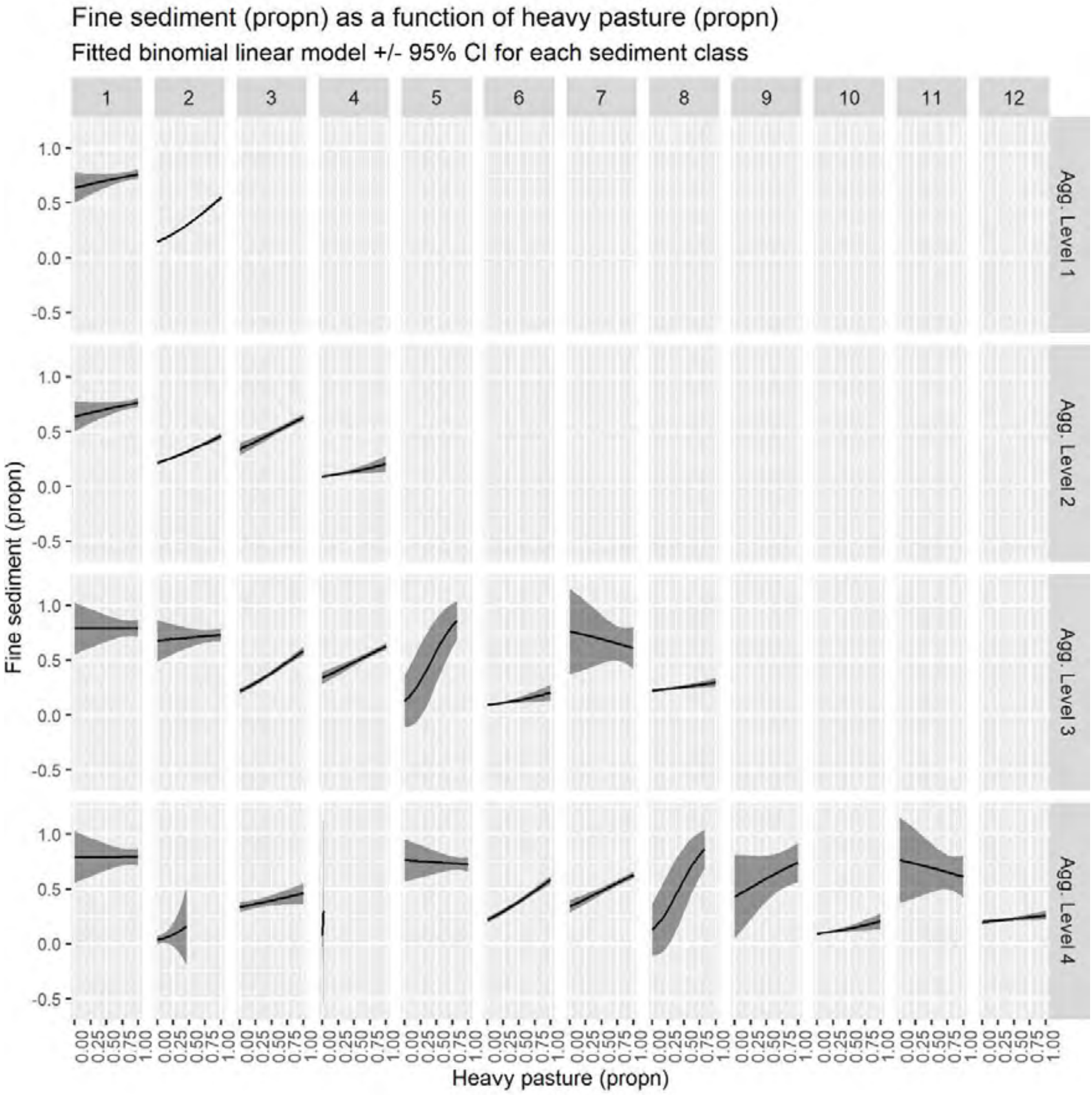


Figure D-8: Binomial linear regression lines of Model 4, describing proportion of fine sediment as a function of proportion of heavy pasture, within each sediment class at four different levels of aggregation (dissimilarity between) of the REC CTG classes. These fitted model traces were obtained by setting covariates *U* and *E* to zero, thus focusing on the impact of heavy pasture in a hypothetical catchment with no forestry and urban development.

Table D-4: Reference values (Ref) for proportion cover of deposited fine sediment for each sediment class, at each level of aggregation. Also presented are the percentages of the New Zealand river network allocated to each class (% River Net.), at each level of aggregation.

<table>
<tr><th>Agg. L1</th><th>Ref</th><th>% River Net.</th><th>Agg. L2</th><th>Ref</th><th>% River Net.</th><th>Agg. L3</th><th>Ref</th><th>% River Net.</th><th>Agg. L4</th><th>Ref</th><th>% River Net.</th><th>CTG Classes</th></tr>
<tr><td rowspan="5">1</td><td rowspan="5">0.64</td><td rowspan="5">5.88</td><td rowspan="5">1</td><td rowspan="5">0.64</td><td rowspan="5">5.88</td><td>1</td><td>0.79</td><td>1.88</td><td>1</td><td>0.79</td><td>1.88</td><td>WD_Low_VA;
WD_Low_Al</td></tr>
<tr><td rowspan="2">2</td><td rowspan="2">0.68</td><td rowspan="2">3.42</td><td>5</td><td>0.74</td><td>3.05</td><td>WD_Low_SS</td></tr>
<tr><td>9</td><td>0.43</td><td>0.36</td><td>WD_Low_HS</td></tr>
<tr><td>5</td><td>0.13</td><td>0.14</td><td>8</td><td>0.13</td><td>0.14</td><td>WW_Lake_Any</td></tr>
<tr><td>7</td><td>0.69</td><td>0.45</td><td>11</td><td>0.69</td><td>0.45</td><td>WW_Low_Al</td></tr>
<tr><td rowspan="7">2</td><td rowspan="7">0.15</td><td rowspan="7">93.05</td><td rowspan="3">2</td><td rowspan="3">0.21</td><td rowspan="3">37.73</td><td>3</td><td>0.22</td><td>13.32</td><td>6</td><td>0.22</td><td>13.32</td><td>WW_Low_VA;
WW_Low_HS;
CD_Low_VA;
CD_Hill_Al;
CD_Low_HS</td></tr>
<tr><td rowspan="2">8</td><td rowspan="2">0.22</td><td rowspan="2">24.41</td><td>12</td><td>0.20</td><td>19.73</td><td>CW_Hill_VA;
CW_Low_VA;
CW_Low_SS;
CD_Hill_HS</td></tr>
<tr><td>3</td><td>0.33</td><td>4.68</td><td>CW_Lake_Any;
CW_Low_Al;
CD_Hill_SS</td></tr>
<tr><td>3</td><td>0.34</td><td>15.51</td><td>4</td><td>0.34</td><td>15.51</td><td>7</td><td>0.34</td><td>15.51</td><td>WW_Low_SS;
CD_Low_SS;
CD_Low_Al</td></tr>
<tr><td rowspan="3">4</td><td rowspan="3">0.09</td><td rowspan="3">39.82</td><td rowspan="3">6</td><td rowspan="3">0.09</td><td rowspan="3">39.82</td><td>10</td><td>0.09</td><td>36.41</td><td>WW_Hill_VA;
CW_Hill_HS;
CW_Low_HS;
CW_Mount_HS;
CW_Hill_SS;
CW_Hill_Al;
CD_Mount_HS;
CW_Mount_Al</td></tr>
<tr><td>2</td><td>0.04</td><td>1.46</td><td>WW_Hill_HS;
CW_Mount_VA</td></tr>
<tr><td>4</td><td>0.07</td><td>1.95</td><td>CW_Mount_SS</td></tr>
</table>

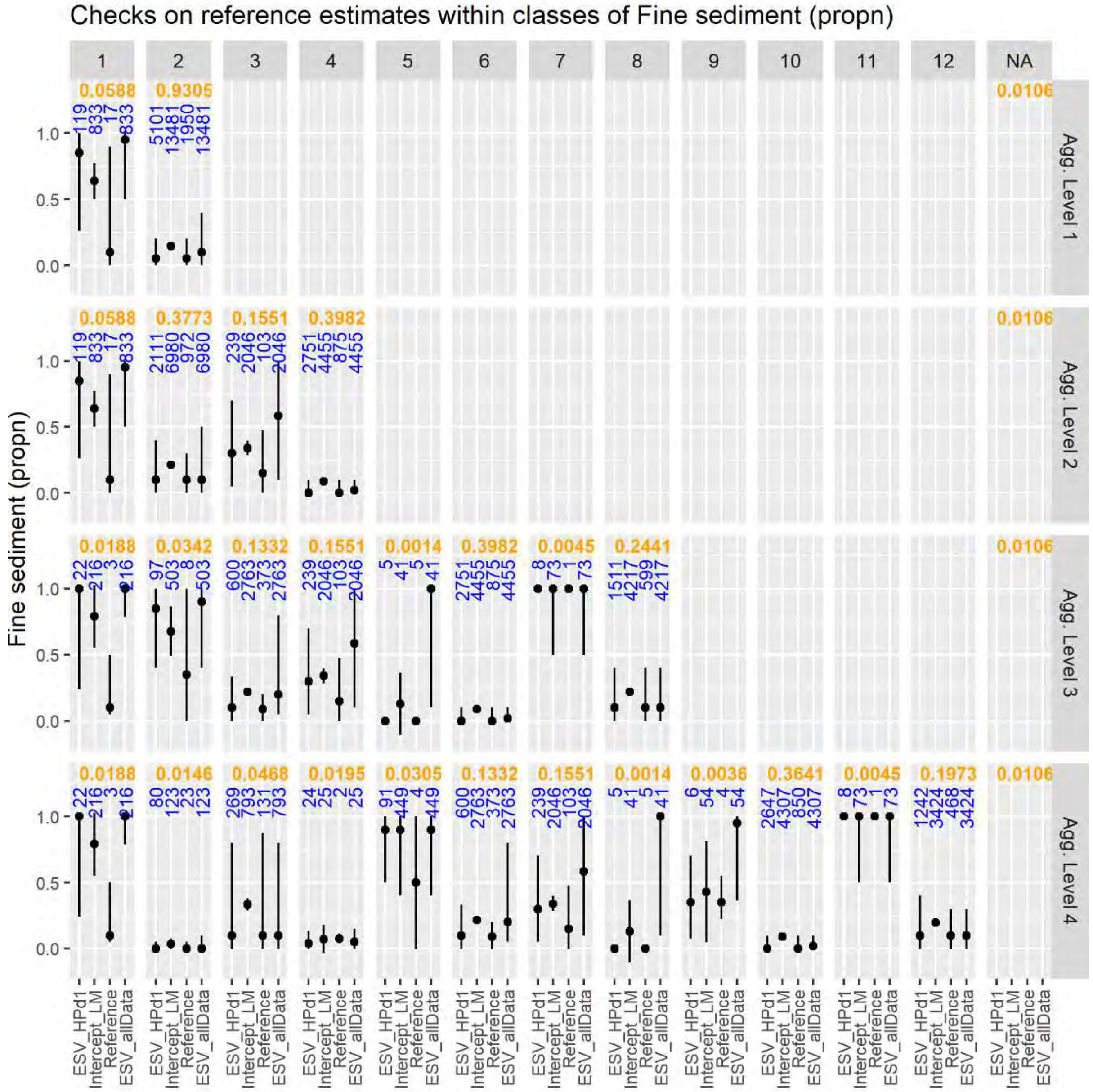


Figure D-9: Comparison of reference state estimates for deposited fine sediment within each class at four levels of aggregation. ESV_HPd1: The median sediment value within the lowest decile of heavy pasture (error = IQR); Intercept_LM: the estimated y-intercept of Model 4 (error = 95% CI); Reference: The median fine sediment value obtained only from reference sites, within the sediment class (error = IQR). ESV_allData: The median deposited fine sediment value for all data within that class (error = IQR). Blue numbers above each point indicate the number of sites contributing data to each statistic. Orange numbers indicate the proportion of the entire New Zealand REC comprised of each sediment class. NA indicates the 'undefined' class; CTG classes containing insufficient ESV data to enter the classification algorithm.

Figure D-10 presents the direction and magnitude of change in reference state estimates as we further aggregate sediment classes from one level in our classification hierarchy to the next highest level. It is clear that the higher the level of aggregation we use the more biased our reference estimates. For example, at Aggregation Level 4, Classes L4.5 and L4.9 have, respectively, reference states of 0.90 and 0.43 (proportionate coverage). At Level 3 these two classes are aggregated yielding a reference state of 0.68 (Class L3.2). Thus, aggregating from the lowest level to the next highest level in the classification hierarchy can have very significant consequences for management.

Specifically, if we opted for the Level 3 classification, any stream that would have been in Class L4.9 at the Level 4 classification (estimated reference at Level 4 of 0.43) may be allowed a significant level of further degradation (as the reference for those streams at Level 3 is now 0.68; appreciably higher than 0.43). Conversely, any stream in Class L4.5, with an estimated reference of 0.9 at the Level 4 classification, is assigned a reference state of 0.68 at the Level 3 classification, possibly resulting in many streams being incorrectly deemed as degraded and in need of management action.

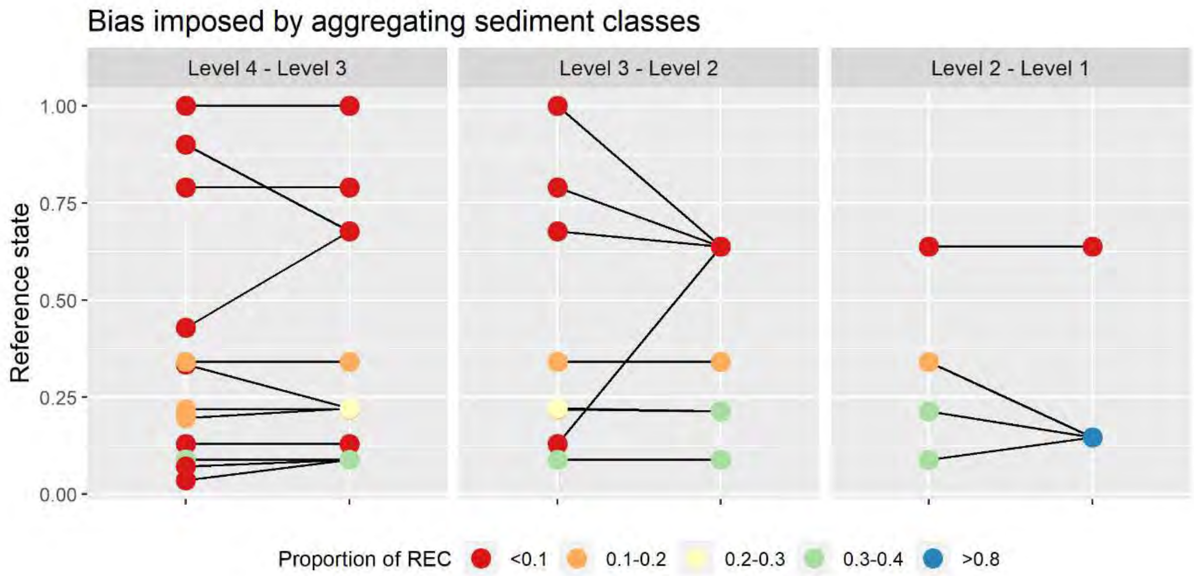


Figure D-10: Change in reference state of deposited fine sediment as classes at one aggregation level are further aggregated into classes at the next highest level in our classification hierarchy.

The AIC statistics for Model 4 fitted to the deposited fine sediment data at Levels 1 through to 4 are presented in Table D-5. Despite the large number of parameters, the most parsimonious model of deposited fine sediment as a function of anthropogenic development is the one that includes interactions between covariates and sediment classes at the lowest level of aggregation (12 classes). Indeed, relative to the other three levels of aggregation in the hierarchy, there is a probability of 1 that Level 4 is the most likely model in the candidate set. Thus the data very strongly indicate that the lowest level of aggregation provides the most parsimonious description of deposited fine sediment reference states in New Zealand.

Table D-5: AIC statistics for Model 4 fitted to deposited fine sediment data at all four levels of aggregation in the classification hierarchy. K is the number of parameters in the regression model; AICc is the corrected AIC statistic; Δi = AICi – min(AIC) is known as the model rank; *wi*, is the Akaike weight of model *i*, interpreted as the approximate probability that Model *i* is the best model in the candidate set, given the data; LL is log-likelihood of each model; Cum.Wt is the cumulative model weight of the ranked models.

Agg. Level	K	AICc	Δ_i	w_i	LL	Cum.Wt
4	48	12822.23	0	1	-6362.95	1
3	32	12886.25	64.01785	1.26E-14	-6411.05	1
2	16	13049.62	227.3884	4.20E-50	-6508.79	1
1	8	13731.36	909.1235	3.86E-198	-6857.67	1

The spatial distribution of deposited fine sediment classes at each of the four levels of aggregation is presented in Figure D-11.

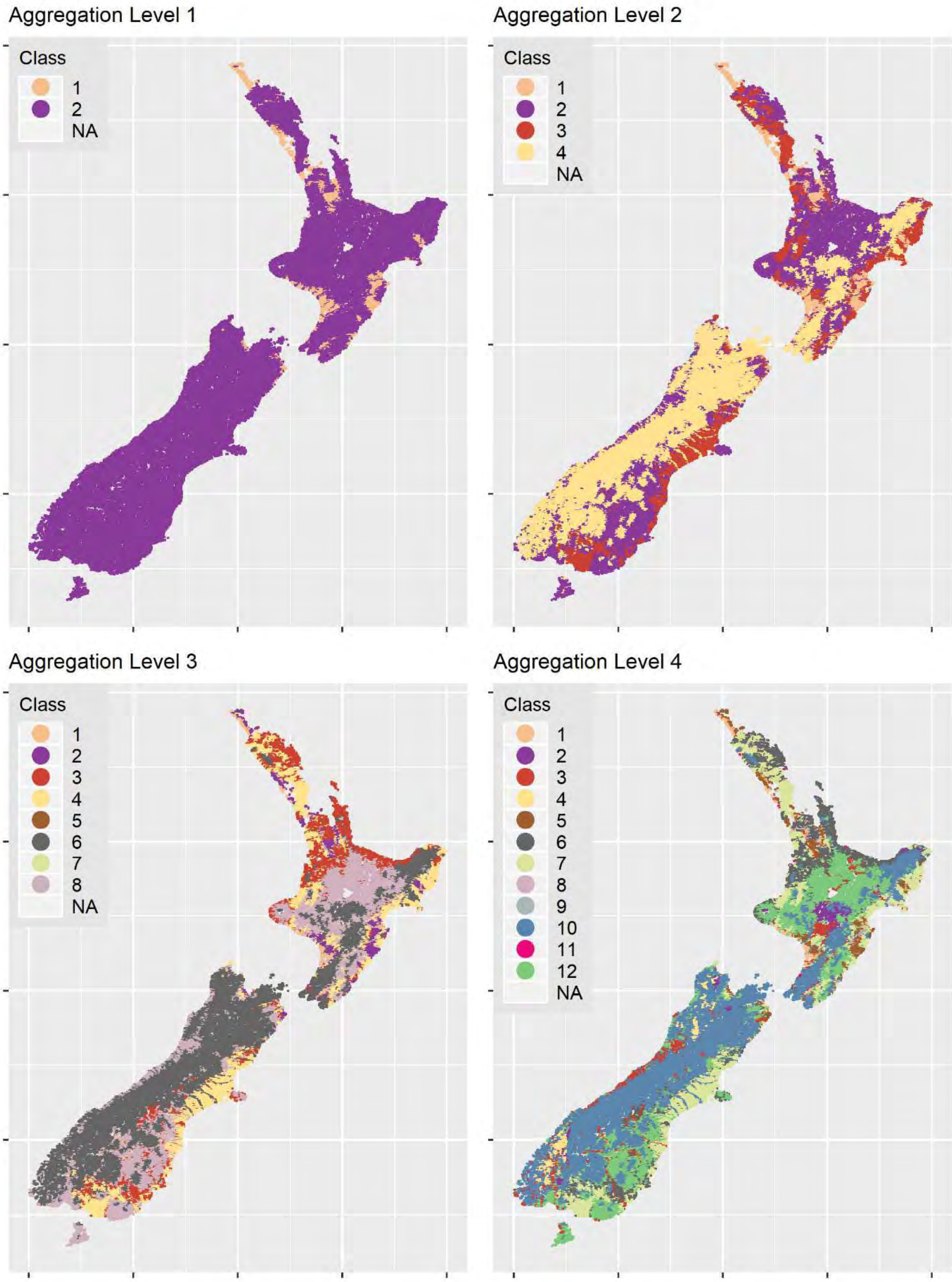


Figure D-11: Spatial distribution of the deposited fine sediment classes under four different levels of aggregation of the REC CTG classes. See Table D-3 for description of sediment classes.

Turbidity

The model that best explained variation in turbidity as a function of our stressor covariates was dependent on the level of aggregation. At Aggregation Level 1 (50% dissimilarity between 2 sediment classes), Model 4 provided the most parsimonious description of the data. By contrast, at Aggregations Levels 3-1 the most parsimonious model of turbidity as a function of our stressor covariates was the simplest model in the set, Model 1. Hence, given the data and candidate models, at the highest level of aggregation we found variation in turbidity throughout New Zealand is best explained by the additive effects of heavy pasture, urbanisation and forestry, and how those three drivers interact with the two sediment classes of the New Zealand landscape. At finer levels of aggregation the model containing interactions between sediment classes and heavy pasture alone was optimal.

Using Nagelkerke's R^2 for generalised linear models, the R^2 values for the fit of the most parsimonious models to the turbidity data were: 0.27 (Level 1; Model 4); 0.28 (Level 2; Model 1); 0.30 (Level 3; Model 1); 0.34 (Level 4; Model 1).

The fitted optimal models for turbidity are presented in Figure D-12. Table D-6 presents the reference states (intercepts) for, and the proportion of the NZ REC covered by, each sediment class, at each level of aggregation. At Aggregation Level 3, Class L3.7 (which was also Class L4.2) returned a counterintuitive slope, so the reference state for that sediment class was estimated as the median turbidity value for all data in that class (Table D-6). At Aggregation Level 4, in addition to L4.2, the reference state of Class L4.4 was estimated as the median turbidity value within that class (Table D-6). For most classes we had a good range of heavy pasture values for regression, irrespective of level of aggregation (Figure D-12).

Based on Figure D-12and Table D-6, the variation in reference state across sediment classes increases as we move from Aggregation Level 1 through to Level 4. This can also be seen in Figure D-13, which presents the comparisons of our reference state estimates for turbidity with ESV_HPd1, the value of the ESV at reference sites and ESV_allData. The following inferences may be gleaned from Figure D-13:

- Using the method of classification derived here, less than 12% of the New Zealand river network was unclassified for the turbidity ESV.
- Irrespective of the level of aggregation there was generally good agreement between model-based reference estimates and those based on reference sites alone. When there was discordance between the model-based and reference site estimates, model-based estimates were not necessarily always higher than those based on reference sites. For example, at Aggregation Level 4, the model-based estimate was lower than that based on reference sites for Classes L4.6 and L4.10, while the reverse may be true for Class L4.12 (Figure D-13).
- At lower levels of aggregation, the number of reference sites is often very low (<5; Figure D-13 Level 4) and often classes are without reference sites at Aggregation Levels 3 and 4 (Figure D-13).

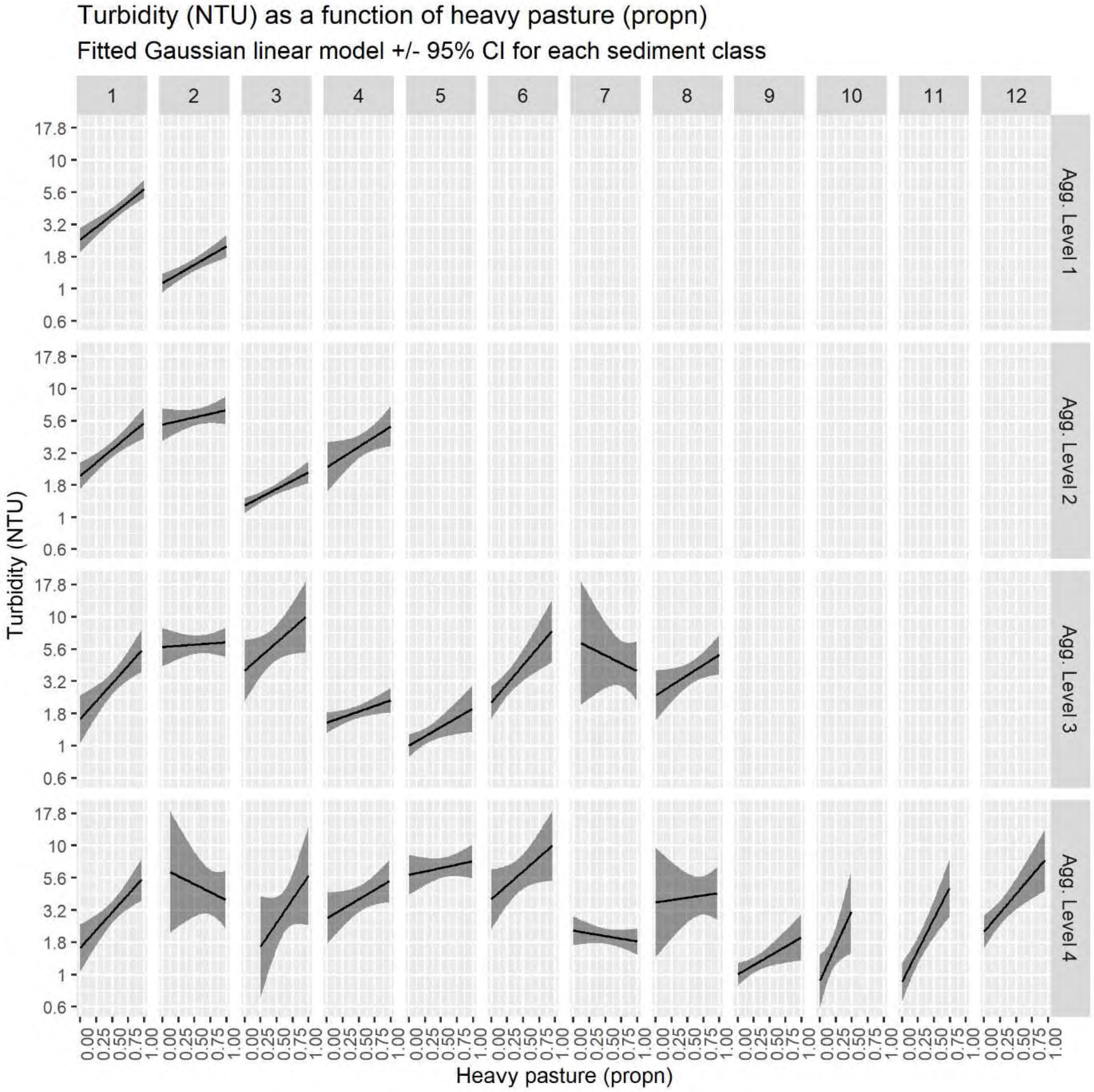


Figure D-12: Gaussian linear regression lines of either Model 4 (Agg. Level 1) or Model 1 (Agg. Levels 2-4), describing turbidity as a function of proportion of heavy pasture, within each sediment class, at four different levels of aggregation (dissimilarity between) of the REC CTG classes. In the case of Agg. Level 1, these fitted model traces were obtained by setting covariates *U* and *E* to zero, thus focusing on the impact of heavy pasture in a hypothetical catchment with no forestry and urban development.

Table D-6: Reference values (Ref) for turbidity (NTUs) for each sediment class, at each level of aggregation. Also presented are the percentages of the New Zealand river network (% River Net.) allocated to each class, at each level of aggregation.

<table>
<tr><th>Agg. L1</th><th>Ref</th><th>% River Net.</th><th>Agg. L2</th><th>Ref</th><th>% River Net.</th><th>Agg. L3</th><th>Ref</th><th>% River Net.</th><th>Agg. L4</th><th>Ref</th><th>% River Net.</th><th>CTG Classes</th></tr>
<tr><td rowspan="8">1</td><td rowspan="8">2.4</td><td rowspan="8">56.82</td><td rowspan="3">1</td><td rowspan="3">2.1</td><td rowspan="3">30.83</td><td>1</td><td>1.6</td><td>7.05</td><td>1</td><td>1.6</td><td>7.05</td><td>WW_Low_VA;
CW_Low_VA</td></tr>
<tr><td>6</td><td>2.1</td><td>22.37</td><td>12</td><td>2.2</td><td>22.37</td><td>CW_Mount_HS;
CW_Hill_SS</td></tr>
<tr><td>7</td><td>4.9</td><td>1.42</td><td>2</td><td>4.9</td><td>1.42</td><td>WD_Low_Al</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">5.2</td><td rowspan="3">17.26</td><td rowspan="2">2</td><td rowspan="2">5.8</td><td rowspan="2">14.42</td><td>5</td><td>5.9</td><td>10.81</td><td>WW_Low_SS;
WD_Low_SS</td></tr>
<tr><td>8</td><td>3.6</td><td>3.61</td><td>CD_Low_SS</td></tr>
<tr><td>3</td><td>3.8</td><td>2.84</td><td>6</td><td>3.8</td><td>2.84</td><td>WW_Low_HS</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2">2.5</td><td rowspan="2">8.72</td><td rowspan="2">8</td><td rowspan="2">2.5</td><td rowspan="2">8.72</td><td>3</td><td>1.1</td><td>2.72</td><td>CD_Low_HS</td></tr>
<tr><td>4</td><td>2.7</td><td>6.01</td><td>CW_Low_SS</td></tr>
<tr><td rowspan="4">2</td><td rowspan="4">1.1</td><td rowspan="4">31.70</td><td rowspan="4">3</td><td rowspan="4">1.2</td><td rowspan="4">31.70</td><td rowspan="3">4</td><td rowspan="3">1.5</td><td rowspan="3">14.58</td><td>7</td><td>2</td><td>10.92</td><td>CD_Low_Al;
CW_Hill_VA</td></tr>
<tr><td>10</td><td>0.9</td><td>1.63</td><td>CW_Lake_Any</td></tr>
<tr><td>11</td><td>0.9</td><td>2.03</td><td>CW_Low_HS</td></tr>
<tr><td>5</td><td>1.0</td><td>17.12</td><td>9</td><td>1.0</td><td>17.12</td><td>CW_Hill_HS;
CD_Hill_HS;
CW_Low_Al</td></tr>
</table>

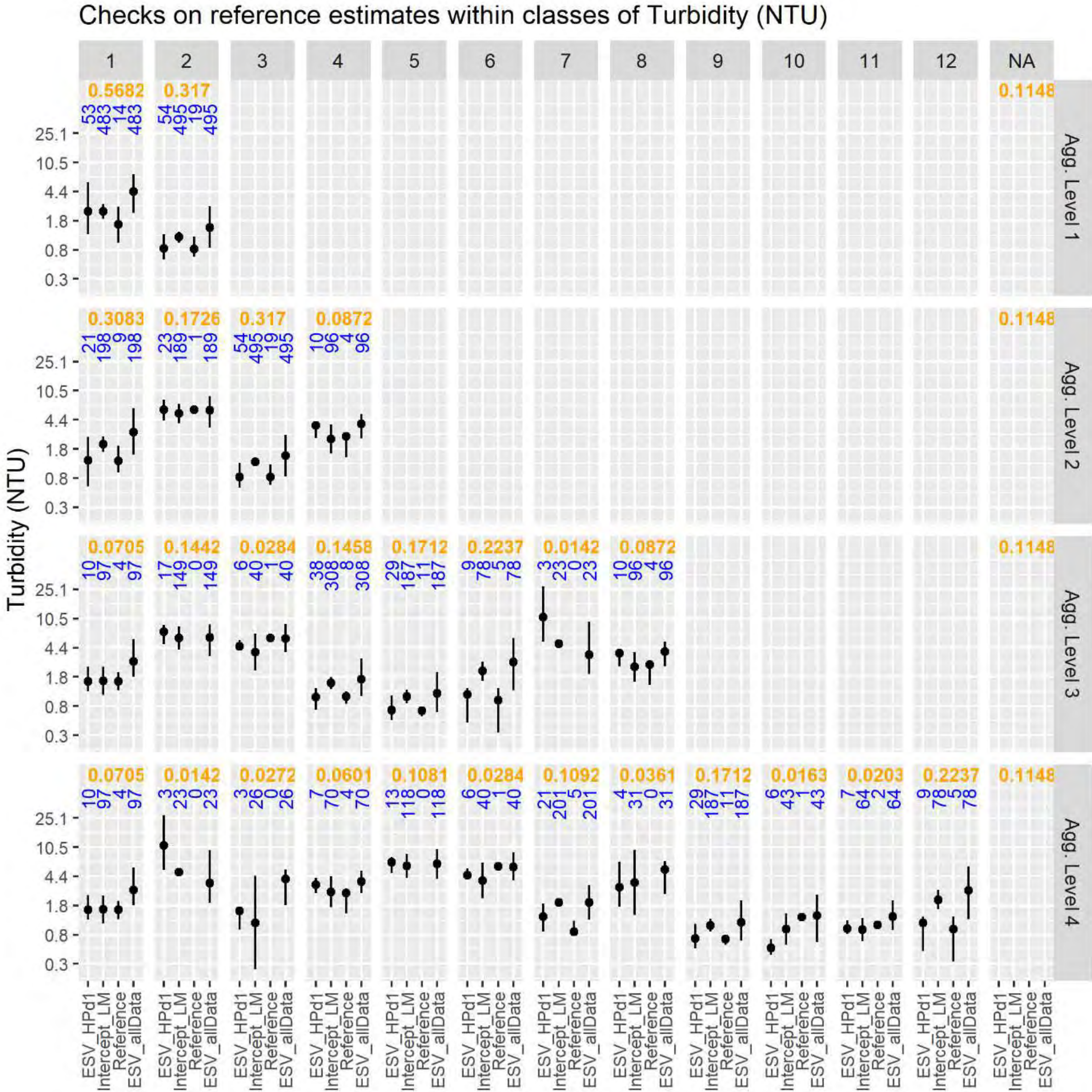


Figure D-13: Comparison of reference state estimates for turbidity within each class at four levels of aggregation. ESV_HPd1: The median turbidity value within the lowest decile of heavy pasture (error = IQR); Intercept_LM: the estimated y-intercept of the most parsimonious models of turbidity as a function of stressor covariates (error = 95% CI); Reference: The median turbidity value obtained only from reference sites, within the sediment class (error = IQR). ESV_allData: The median turbidity value for all data within that class (error = IQR). Blue numbers above each point indicate the number of sites contributing data to each statistic. Orange numbers indicate the proportion of the entire New Zealand REC comprised of each sediment class. NA indicates the 'undefined' class; CTG classes containing insufficient ESV data to enter the classification algorithm.

Figure D-14 presents the direction and magnitude of change in turbidity reference state estimates as we further aggregate sediment classes from one level in our classification hierarchy to the next highest level. As was the case for deposited fine sediment, the higher the level of aggregation we use the more biased our reference estimates. At Aggregation Level 1, consider Class L1.1, which is decomposed into Classes L2.1, L2.2 and L2.4 at Aggregation Level 2 (Table D-6; Figure D-14). The reference state for rivers in Class L1.1 is 2.4 NTUs. Using the next finest level of aggregation, those same rivers fall into three different classes with reference values ranging from 5.2 NTUs to 2.1 NTUs.

As was the case with deposited fine sediment, the coarser the level of aggregation of our SSC, the more sediment management decisions may be misguided by biased reference states. Of note is the fact that the bias estimates presented here are in log-scale, so the real bias in units of NTUs would be greater.

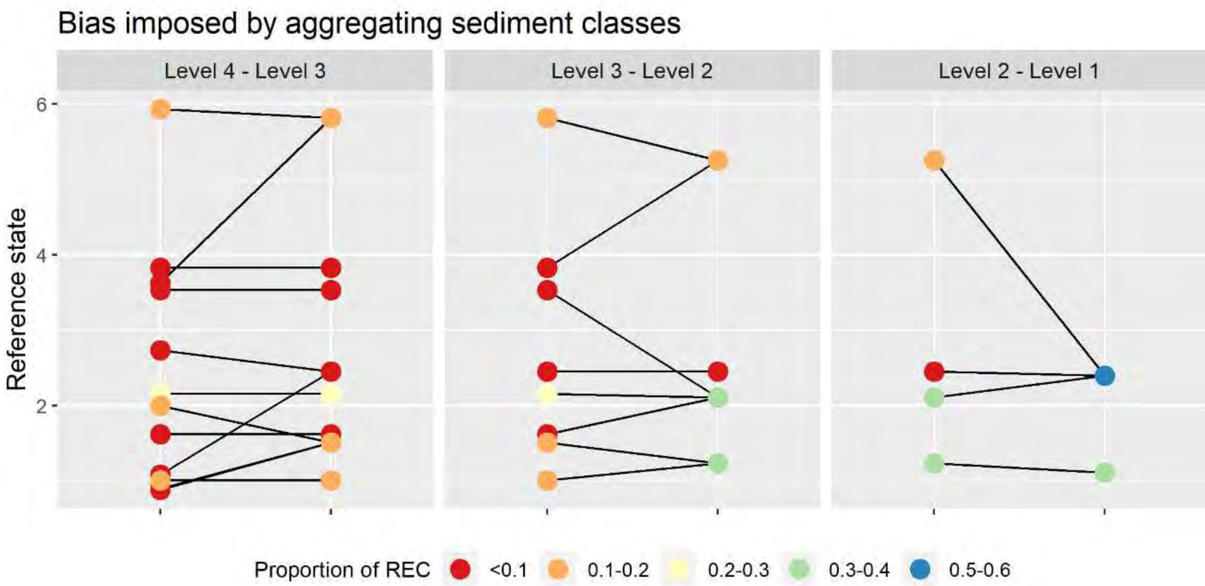


Figure D-14: Change in reference state of turbidity (NTUs) as classes at one aggregation level are further aggregated into classes at the next highest level in our classification hierarchy.

The AIC statistics for the optimal models of turbidity as a function of stressor covariates Levels 1 through to 4 are presented in Table D-7. As was the case for deposited fine sediment, and despite the large number of parameters, the most parsimonious model of turbidity as a function of anthropogenic development is the one that includes interactions between covariates and sediment classes at the lowest level of aggregation (Level 4; 12 classes). Indeed, relative to the other three levels of aggregation in the hierarchy, there is a probability of 1 that Level 4 is the most likely model in the candidate set. Thus the data very strongly indicate that the lowest level of aggregation provides the most parsimonious description of turbidity reference states in New Zealand.

Table D-7: AIC statistics for optimal models fitted to turbidity data at all four levels of aggregation in the classification hierarchy. K is the number of parameters in the regression model; AICc is the corrected AIC statistic; Δi = AICi – min(AIC) is known as the model rank; *wi*, is the Akaike weight of model *i*, interpreted as the approximate probability that Model *i* is the best model in the candidate set, given the data; LL is log-likelihood of each model; Cum.Wt is the cumulative model weight of the ranked models.

Agg. Level	K	AICc	Δ_i	w_i	LL	Cum.Wt
4	25	1114.24	0.00	1.00	-531.44	1.00
3	17	1136.49	22.25	0.00	-550.92	1.00
2	9	1142.81	28.57	0.00	-562.31	1.00
1	9	1162.02	47.79	0.00	-571.92	1.00

The spatial distribution of the turbidity sediment classes at all level of aggregation is presented in Figure D-15.

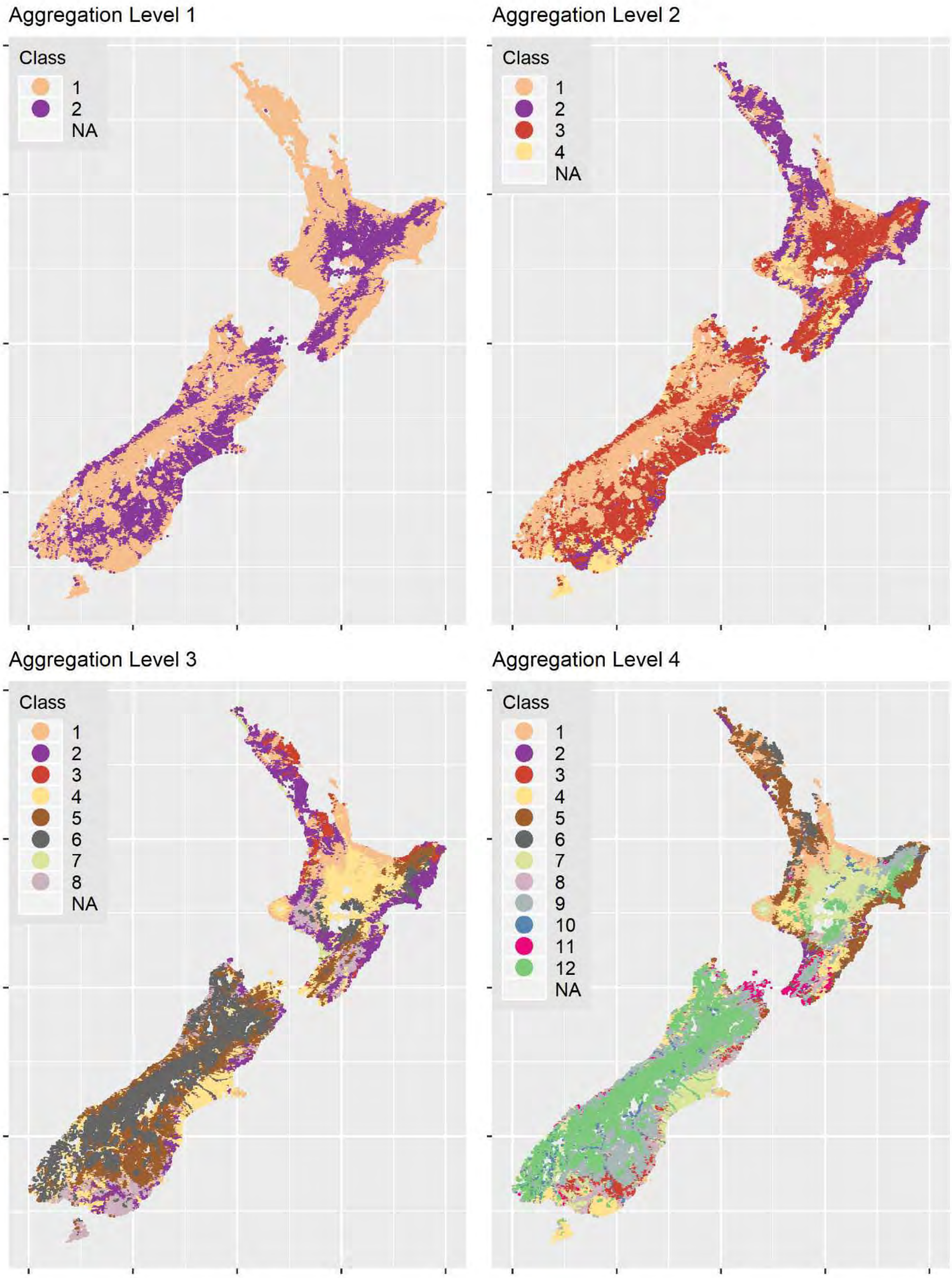


Figure D-15: Spatial distribution of the suspended sediment (turbidity) classes under four different levels of aggregation of the REC CTG classes. See Table D-3 for description of sediment classes.

Visual clarity

As stated at the beginning of this chapter, the sediment state classification for suspended sediment was developed using turbidity data. A consequence of this is that there may be suspended sediment classes for which we have insufficient visual clarity data for robust stressor-visual clarity relationships (using Models 1-4 presented earlier); there was less visual clarity data than turbidity data. Indeed, in Figure D-16 we can see a total of three suspended sediment classes for which we had insufficient visual clarity to define reference state (classes L3.7, L4.2 and L4.3; Figure D-16). Classes that contained insufficient data for reference estimation using our model-based approach were assigned the reference state of their 'parent class.' All classes within Levels 2-4 have a parent class such that Class LX.i (i = 1,…n_X, where n_X is the number of classes in Level X) is nested within Parent Class LX-1.i (i = 1,…,n_{X-1}). Thus, this is another advantage of using a hierarchical classification scheme.

The model that best explained variation in visual clarity as a function of our stressor covariates was dependent on the level of aggregation. At Aggregation Level 1 (50% dissimilarity between 2 sediment classes), Model 4 provided the most parsimonious description of the data. At Aggregation Levels 2 (30% dissimilarity between 4 classes) and 3 (20% dissimilarity between 8 classes), Models 3 and 2, respectively were the optimal models in the candidate set. By contrast, at Aggregation Level 4 Model 1 was the most parsimonious model of visual clarity as a function of anthropogenic stressor covariates. Thus all three stressor covariates (heavy pasture, forestry and urbanisation) were included in the best model at the coarsest level of aggregation. At intermediate levels of aggregation, when we incorporate a greater number of sediment classes, models including only two of the three stressor covariates are most parsimonious (heavy pasture + urbanisation; or heavy pasture + forestry). At the finest level of aggregation the 12 sediment classes subsume some of the variation in visual clarity due to anthropogenic stressors, and the most parsimonious model contained the single stressor of heavy pasture.

Using Nagelkerke's R^2 for generalised linear models, the R^2 values for the fit of the most parsimonious models to the visual clarity data were: 0.33 (Level 1; Model 4); 0.37 (Level 2; Model 3); 0.40 (Level 3; Model 2); 0.41 (Level 4; Model 1).

The fitted optimal models for visual clarity are presented in Figure D-16, where the strong negative effect of heavy pasture on visual clarity can be seen. Table D-8 presents the reference states (intercepts) for, and the proportion of the NZ REC covered by, each sediment class, at each level of aggregation. At Aggregation Level 3, Class L3.2 (which was split by the cluster analysis into Classes L4.5 and L4.8) returned a counterintuitive slope, as did Class L4.8, so the reference states for those suspended sediment classes were estimated as the median visual clarity value for all data in that class (Table D-8). For most classes we had a good range of heavy pasture values for regression, irrespective of level of aggregation (Figure D-16).

Based on Figure D-16 and Table D-8, the variation in reference state across sediment classes increases as we move from Aggregation Level 1 through to Level 4. This can also be seen in Figure D-17, which presents the comparisons of our reference state estimates for turbidity with ESV_HPd1, the value of the ESV at reference sites and ESV_allData. Similar to those gleaned for turbidity, the following inferences may be gleaned from Figure D-17:

- Using the method of classification derived here, less than 12% of the New Zealand river network was unclassified for the visual clarity ESV.

- Irrespective of the level of aggregation there was generally good agreement between model-based reference estimates and those based on reference sites alone.
- At lower levels of aggregation, the number of reference sites is often very low (<5; Figure D-17 Level 4) and often classes are without reference sites at Aggregation Levels 3 and 4 (Figure D-17).

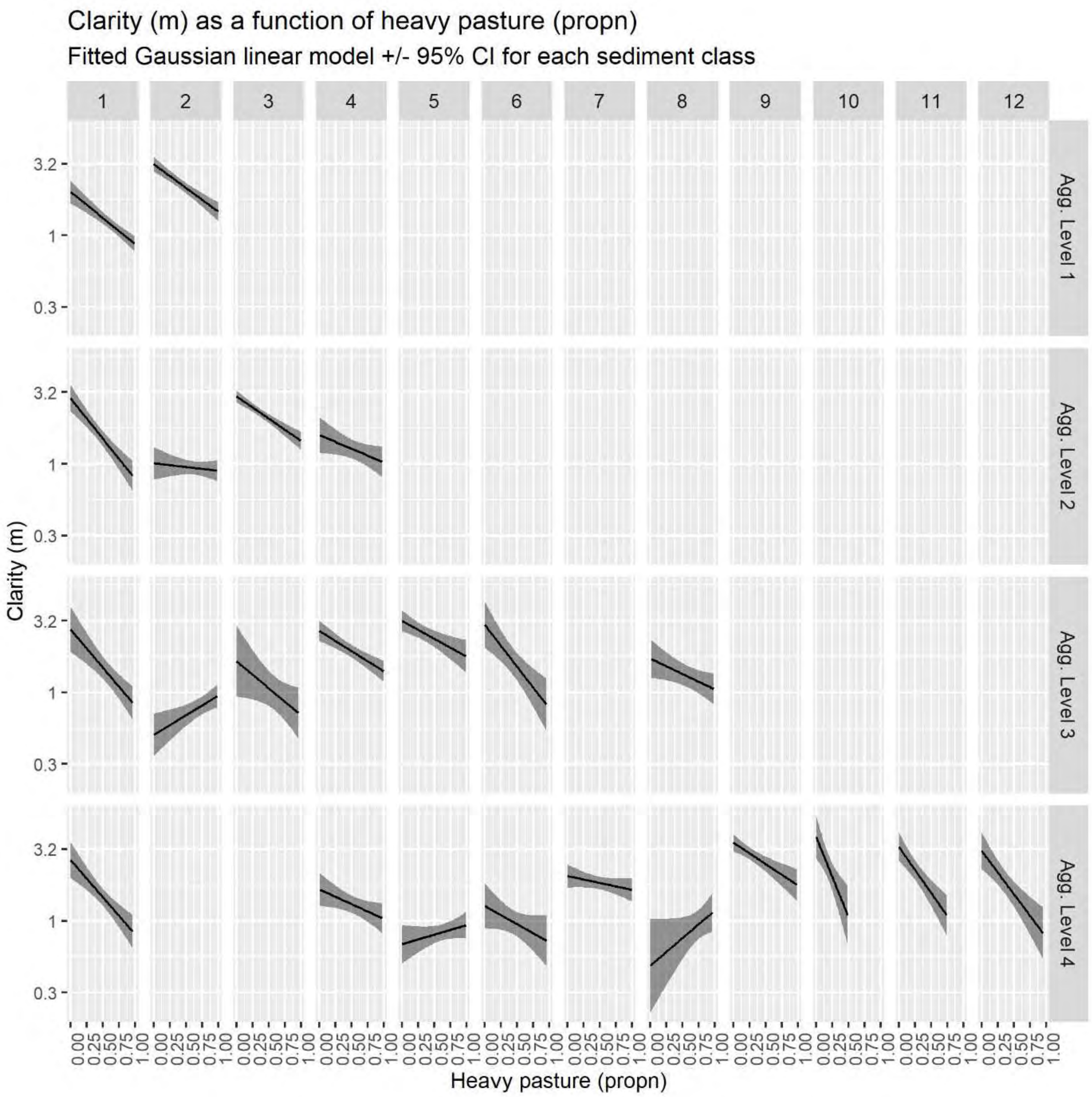


Figure D-16: Gaussian linear regression lines of either Model 4 (Agg. Level 1; top row), Model 3 (Agg. Level 2; row 2); Model 2 (Agg. Level 3; row 3), or Model 1 (Agg. Level 4; bottom row) describing log-transformed visual clarity as a function of proportion of heavy pasture, within each sediment class, at four different levels of aggregation (dissimilarity between) of the REC CTG classes. In the case of Agg. Levels 1-3, these fitted model traces were obtained by setting covariates *U* and *E* to zero, thus focusing on the impact of heavy pasture in a hypothetical catchment with no forestry and urban development.

Table D-8: Reference values (Ref) for visual clarity (m) for each sediment class, at each level of aggregation. Also presented are the percentages of the New Zealand river network (% River Net.) allocated to each class, at each level of aggregation. Classes whose reference state estimates were denoted by an Asterix (*) were assigned the reference state of their parent class, due to insufficient data within that class, at that level, for implementation of the model-based estimation.

<table>
<tr><th>Agg. L1</th><th>Ref</th><th>% River Net.</th><th>Agg. L2</th><th>Ref</th><th>% River Net.</th><th>Agg. L3</th><th>Ref</th><th>% River Net.</th><th>Agg. L4</th><th>Ref</th><th>% River Net.</th><th>CTG Classes</th></tr>
<tr><td rowspan="8">1</td><td rowspan="8">2.0</td><td rowspan="8">56.82</td><td rowspan="3">1</td><td rowspan="3">2.9</td><td rowspan="3">30.83</td><td>1</td><td>2.7</td><td>7.05</td><td>1</td><td>2.7</td><td>7.05</td><td>WW_Low_VA; CW_Low_VA</td></tr>
<tr><td>6</td><td>3.0</td><td>22.37</td><td>12</td><td>3.1</td><td>22.37</td><td>CW_Mount_HS; CW_Hill_SS</td></tr>
<tr><td>7</td><td>2.9*</td><td>1.42</td><td>2</td><td>2.9*</td><td>1.42</td><td>WD_Low_Al</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">1.0</td><td rowspan="3">17.26</td><td rowspan="2">2</td><td rowspan="2">0.9</td><td rowspan="2">14.42</td><td>5</td><td>0.8</td><td>10.81</td><td>WW_Low_SS; WD_Low_SS</td></tr>
<tr><td>8</td><td>0.7</td><td>3.61</td><td>CD_Low_SS</td></tr>
<tr><td>3</td><td>1.6</td><td>2.84</td><td>6</td><td>1.3</td><td>2.84</td><td>WW_Low_HS</td></tr>
<tr><td rowspan="2">4</td><td rowspan="2">1.6</td><td rowspan="2">8.72</td><td rowspan="2">8</td><td rowspan="2">1.7</td><td rowspan="2">8.72</td><td>3</td><td>1.7*</td><td>2.72</td><td>CD_Low_HS</td></tr>
<tr><td>4</td><td>1.7</td><td>6.01</td><td>CW_Low_SS</td></tr>
<tr><td rowspan="4">2</td><td rowspan="4">3.1</td><td rowspan="4">31.70</td><td rowspan="4">3</td><td rowspan="4">3.0</td><td rowspan="4">31.70</td><td rowspan="3">4</td><td rowspan="3">2.7</td><td rowspan="3">14.58</td><td>7</td><td>2.1</td><td>10.92</td><td>CD_Low_Al; CW_Hill_VA</td></tr>
<tr><td>10</td><td>3.9</td><td>1.63</td><td>CW_Lake_Any</td></tr>
<tr><td>11</td><td>3.3</td><td>2.03</td><td>CW_Low_HS</td></tr>
<tr><td>5</td><td>3.1</td><td>17.12</td><td>9</td><td>3.5</td><td>17.12</td><td>CW_Hill_HS; CD_Hill_HS; CW_Low_Al</td></tr>
</table>

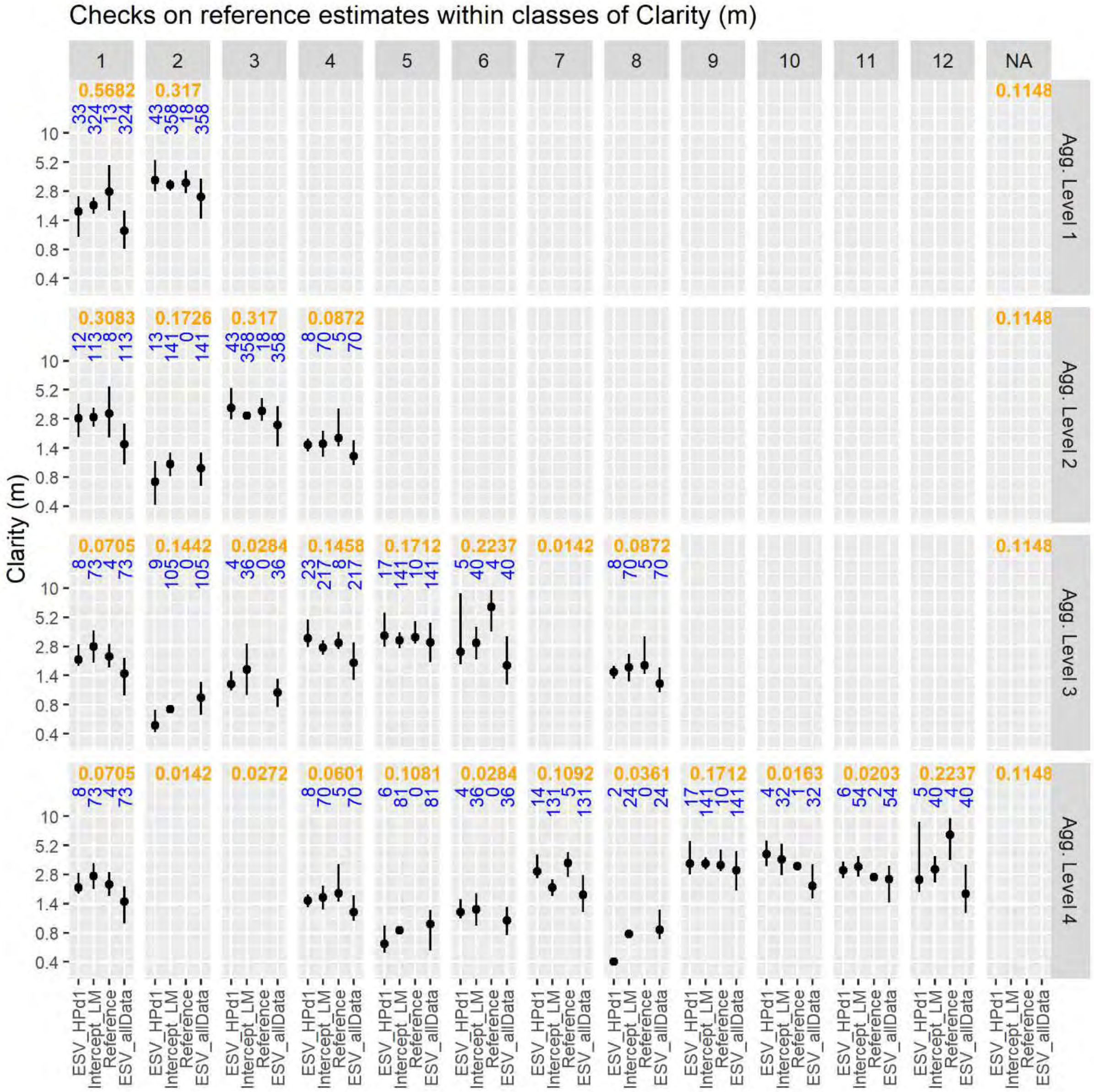


Figure D-17: Comparison of reference state estimates for visual clarity within each class at four levels of aggregation. ESV_HPd1: The median visual clarity value within the lowest decile of heavy pasture (error = IQR); Intercept_LM: the estimated y-intercept of the most parsimonious models of visual clarity as a function of stressor covariates (error = 95% CI); Reference: The median visual clarity value obtained only from reference sites, within the sediment class (error = IQR). ESV_allData: The median visual clarity value for all data within that class (error = IQR). Blue numbers above each point indicate the number of sites contributing data to each statistic. Orange numbers indicate the proportion of the entire New Zealand REC comprised of each sediment class. NA indicates the 'undefined' class; CTG classes containing insufficient ESV data to enter the classification algorithm. We did not have sufficient data to estimate reference state for visual clarity in classes L3.7, L4.2 and L4.3.

Figure D-18 presents the direction and magnitude of change in turbidity reference state estimates as we further aggregate sediment classes from one level in our classification hierarchy to the next highest level. As was the case for deposited fine sediment and turbidity, the higher the level of aggregation we use the more biased our reference estimates (see discussion for deposited fine sediment and turbidity for further exposition).

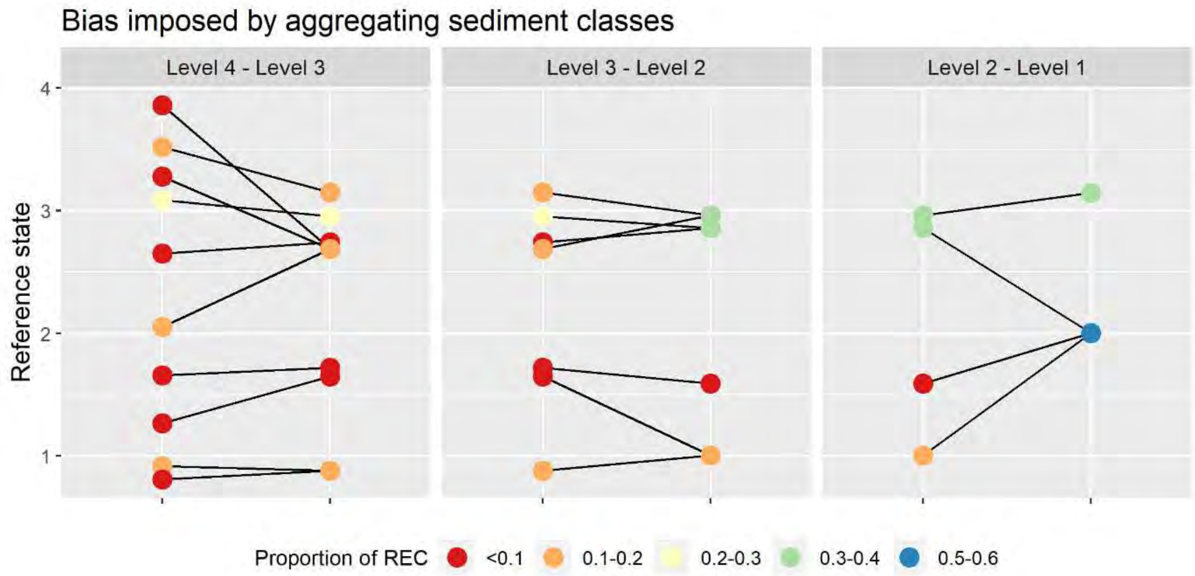


Figure D-18: Change in reference state of visual clarity (m) as classes at one aggregation level are further aggregated into classes at the next highest level in our classification hierarchy.

The AIC statistics for the optimal models of turbidity as a function of stressor covariates Levels 1 through to 4 are presented in Table D-9. As was the case for the other ESVs the most parsimonious model of visual clarity as a function of anthropogenic development is the one that includes interactions between covariates and sediment classes at the lowest level of aggregation (Level 4; 10 classes). Relative to the other three levels of aggregation in the hierarchy, there is a probability of 1 that Level 4 is the most likely model in the candidate set. Thus the data very strongly indicate that the lowest level of aggregation provides the most parsimonious description of turbidity reference states in New Zealand.

Table D-9: AIC statistics for optimal models fitted to visual clarity data at all four levels of aggregation in the classification hierarchy. K is the number of parameters in the regression model; AICc is the corrected AIC statistic; Δi = AICi – min(AIC) is known as the model rank; *wi*, is the Akaike weight of model *i*, interpreted as the approximate probability that Model *i* is the best model in the candidate set, given the data; LL is log-likelihood of each model; Cum.Wt is the cumulative model weight of the ranked models.

Agg. Level	K	AICc	Δ_i	w_i	LL	Cum.Wt
4	21	97.26	0.00	0.99	-26.93	0.99
3	22	107.47	10.22	0.01	-30.97	1.00
2	13	124.56	27.30	0.00	-49.01	1.00
1	9	160.80	63.54	0.00	-71.26	1.00

Appendix E Mapping unclassified reaches

When developing the Sediment State Classification we found that some Climate-Topography-Geology (CTG) classes could not be classified due to insufficient sediment ESV data being available within the class. We developed a set of spatial mapping rules to allocate classes to the segments of the river network that we were not able to classify during development of the SSC. Following application of these procedures, all river segments are now mapped to a class in the SSCs.

The procedure we used to classify unclassified segments was as follows:

4. Isolate each catchment containing any segments with unclassified segments.
5. For each unclassified segment, substitute the unclassified class with the class found in the next downstream classified segment. Label this method of substitution "Downstream".
6. For each remaining unclassified segment, substitute the unclassified class with the class found in the highest order upstream classified segment. Label this method of substitution "Upstream".
7. For each remaining unclassified segment, substitute the class found in the next downstream classified segment regardless of whether classified segments have been substituted in a previous step. Label this method of substitution "Downstream Round 2".
8. For catchments where all segments in that catchment are unclassified, for each segment, substitute the unclassified class with the class found in the nearest classified segment in Euclidian space. Label this method of substitution "Nearest Neighbour".

The above procedure was applied separately to the classifications of suspended sediment (SSC_Sus) and deposited fine sediment (SSC_Dep). For SSC_Sus, 11.0% of 593,548 segments were unclassified. The downstream substitution method was applied to 10.4% of segments. The nearest neighbour substitution method was applied 0.4% of segments. The downstream round 2 and upstream methods were each applied to 0.1% of segments (Figure E-1).

For SSC_Dep, 0.6% of segments were unclassified. The downstream substitution method was applied to substitute for nearly all unclassified segments. The nearest neighbour, downstream round 2 and upstream methods were each applied to less than 0.01% of segments (Figure E-2).

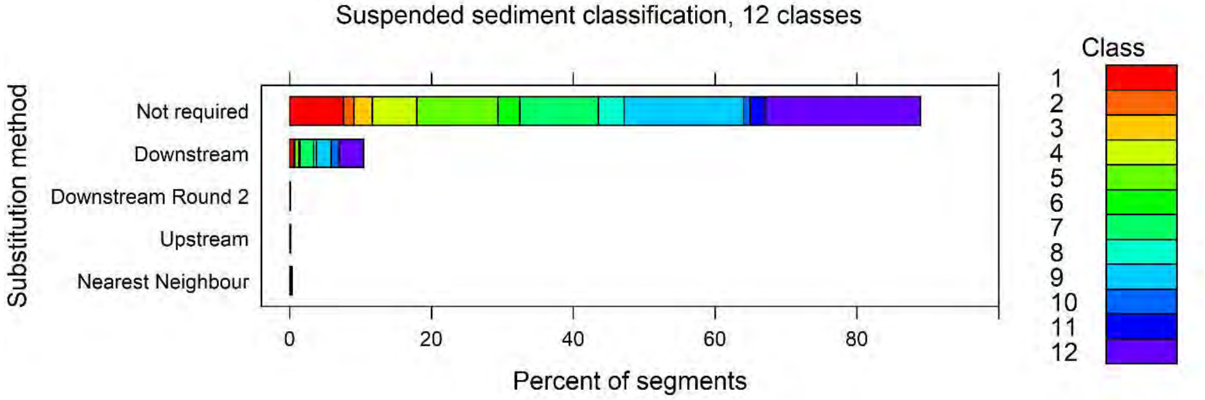


Figure E-1: Distribution of substitution method for the suspended sediment classification.

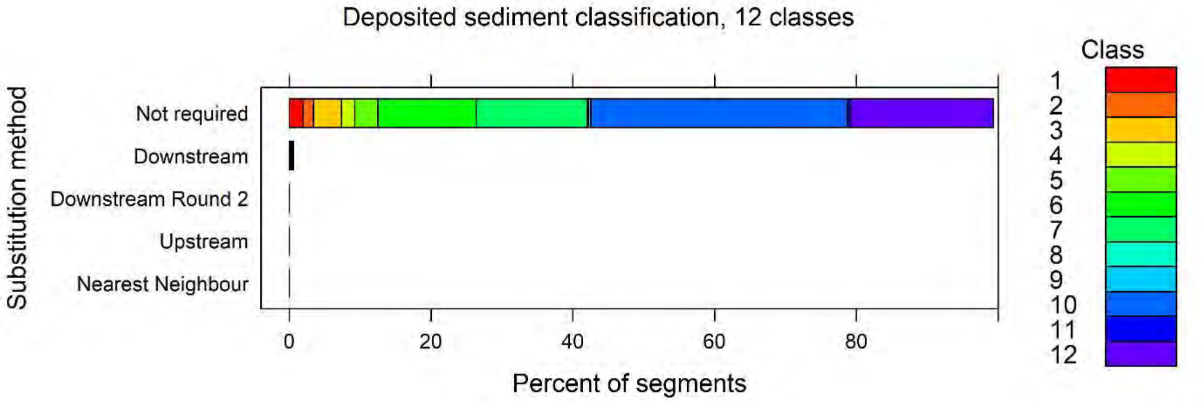


Figure E-2: Distribution of substitution method for the deposited sediment classification.

Appendix F Boosted regression tree analyses

Introduction

There are two different analytical approaches to defining effects-based thresholds. One is to decide on a set ecological target, for example, a 10% or a 20% deviation of a macroinvertebrate metric from the reference condition, and then to calculate with use of a statistical model the sediment threshold which likely allows reaching that target (Cormier et al. 2008). In this case, the ecological target is chosen independent of the shape of the stressor-response relationship, but typically a simple linear regression model or a quantile regression model are used to derive these thresholds.

The second effects-based approach, by contrast, assumes that the stressor-response relationship is of a non-linear type and potentially characterised by some abrupt ecological threshold at which a macroinvertebrate metric changes dramatically over a short increase in sediment. If so, definition of sediment thresholds for management should stay below such ecological thresholds (Larned and Schallenberg 2018). Statistical models such as step-function model or the piecewise linear model have been suggested but these have shown to often inaccurately model the stressor-response relationships of macroinvertebrate metrics (Qian 2014; Wagenhoff et al. 2017). Instead, the use of a flexible modelling approach, such as boosted regression tree (BRT) analysis, allows modelling of complex response shapes. BRT analysis is a flexible modelling approach that allows incorporation of multiple predictors. In contrast to multiple linear regression analysis, correlations between predictors are handled well automatically. BRT analysis is well described in the statistical literature for ecology (De'ath 2007; Elith et al. 2008). Wagenhoff et al. (2017) found that common macroinvertebrate metrics often show a sigmoidal shaped response with relatively gradual responses within certain points across the stressor gradient which they called ‘impact initiation’ and ‘impact cessation’ thresholds. BRT analysis can be used to characterise such response shapes and the impact initiation/cessation conceptual framework can be useful for definition of management thresholds.

Data

Deposited sediment

As described in Depree et al. (2018), a macroinvertebrate-stressor dataset was compiled from national SOE data, the National River Water Quality Network, and specific research studies. There were 1,039 samples of deposited sediment data measured using the SAM2 instream visual assessment protocol (‘% cover instream’).

A total of 602 samples across 354 sites were used to run the global BRT analysis using the % cover instream data. The ability to run independent BRT analyses within sediment state classes (SSCs) was limited by the number of macroinvertebrate-sediment observations available in each class at the different levels of SSC aggregation (Table F-1). At the 50% dissimilarity level (i.e., Level 1), there were sufficient data to proceed with BRT analyses in only one of the two classes (L1.2). At the 30% dissimilarity level (i.e., Level 2), there were sufficient data to proceed with flexible regression in two of the four classes (L2.2 and L2.3). At the 20% dissimilarity level (i.e., Level 3), there appeared to be sufficient data to proceed with flexible regression in four of the eight classes; however, exploratory analysis revealed poor model performance probably due to low sample numbers. Consequently, analyses were conducted on data in classes 2 and 3 at second level of aggregation (i.e., L2.2 and L2.3).

Table F-1: Number of macroinvertebrate-deposited sediment observations within the deposited sediment SSC classes at different levels of aggregation and % of the digital river network represented by each class. Number of independent NZReaches (RECv1) in parenthesis.

Class level	L1		L2		L3		L4	
	n	%	n	%	n	%	n	%
Class 1	40 (21)	5.77	40 (21)	5.74	14 (10)	1.88	14 (10)	1.88
Class 2	942 (535)	92.94	646 (385)	52.54	22 (10)	3.72	2 (2)	3.42
Class 3			296 (150)	40.39	124 (83)	0.45	52 (21)	0.45
Class 4			0	0.03	188 (137)	0.03	0 (0)	0.03
Class 5					296 (150)	16.37	22 (10)	13.32
Class 6					4 (1)	15.94	72 (62)	3.5
Class 7					334 (165)	20.23	186 (135)	15.51
Class 8					0	40.39	276 (143)	0.43
Class 9							4 (1)	20.23
Class 10							334 (165)	36.41
Class 11							20 (7)	2.04
Class 12							0 (0)	1.95

Sediment data were paired with macroinvertebrate metric data sampled from the same site as described in Depree et al. (2018). Macroinvertebrate data included 4 metrics previously selected from 16 candidate metrics based on their response to deposited sediment:

- Macroinvertebrate Community Index (MCI).
- The number of taxa from the orders of Ephemeroptera, Plecoptera and Tricoptera (EPT taxon richness).
- Sediment sensitive Macroinvertebrate Community Index (sediment MCI).
- The number of taxa that decline with increasing deposited sediment (No. of decreasers).

Suspended sediment

There were 4005 turbidity samples (belonging to 665 NZReaches) in the combined macroinvertebrate-stressor dataset that could be assigned to the suspended sediment SSC classes. 333 samples (from 55 NZReaches) in the combined macroinvertebrate-stressor dataset were excluded because the NZReaches were not assigned to a sediment class (see Appendix D).

The ability to run independent BRT analyses within classes was restricted by the number of macroinvertebrate-sediment observations available in each class at the different levels of SSC aggregation (Table F-2). At SSC Levels 1 and 2 there were sufficient paired macroinvertebrate-turbidity data to proceed with BRT analyses in all classes. At SSC Level 3 there were enough data to proceed with flexible regression in five of the eight classes. For consistency with the deposited sediment analyses, subsequent BRT models were developed using data grouped at SSC Level 2.

Table F-2: Number of paired macroinvertebrate-suspended sediment observations within the suspended sediment SSC classes at different levels of aggregation and % of the digital river network represented by each class . Number of independent NZReaches (RECv1) in parenthesis.

Class level	L1		L2		L3		L4	
	n	%	n	%	n	%	n	%
Class 1	1604 (297)	56.82	622 (101)	30.83	552 (98)	7.05	352 (49)	7.05
Class 2	2071 (313)	30.5	473 (117)	17.26	337 (99)	22.37	70 (12)	22.37
Class 3			2071 (313)	8.72	136 (28)	1.42	70 (16)	1.42
Class 4			509 (70)	30.5	715 (110)	14.42	439 (54)	10.81
Class 5					977 (143)	2.84	215 (67)	3.61
Class 6					379 (60)	8.72	136 (28)	2.84
Class 7					70 (12)	13.38	715 (110)	2.72
Class 8					509 (70)	17.12	122 (22)	6.01
Class 9							502 (81)	11.35
Class 10							379 (60)	2.03
Class 11							475 (62)	9.25
Class 12							200 (49)	7.87

Methods

For each sediment measure both globally (i.e., for all samples) and within each class where sufficient date were available, boosted regression tree models were built for each of the four macroinvertebrate metrics. Sixteen response predictors including, a single deposited sediment measure (% cover instream), chlorophyll *a* to account for the effect of nutrients via periphyton biomass, and a range of environmental descriptors from spatial datasets (Table F-3) were used in the BRT models. Environmental descriptors were chosen based on their high relative importance during exploratory analyses and kept consistent across deposited and suspended sediment models. The response variables were standardised by dividing by the standard deviation in order to make the effects of deposited sediment comparable among the macroinvertebrate metrics. Sample observations were equally weighted within sites (NZReach) to account for pseudo-replication at the spatial scale.

BRT model building allows missing values for the predictors. Potentially, missing values can lead to bias of predictor importance. We used a subset of the data for each sediment measure requiring non-missing values for chlorophyll *a*, which was considered a potentially important stressor variable. Missing values were allowed for all other predictors. Model parameterisation was done following the suggestions by Elith et al. (2008) using the Gaussian family; interaction depth was set to 3. We used *gbm* R package and modified functions based on procedures published by Elith,Leathwick (2017).

Table F-3: Set of 16 predictor variables used in BRT models along with their data source and description. Two flow statistics (Booker 2013; Booker and Woods 2014) were downloaded from the MfE website on 23 August 2016 (https://data.mfe.govt.nz/table/2536-natural-river-flow-statistics-predicted-for-all-river-reaches/), REC = River Environment Classification database (Snelder and Biggs 2002), FENZ = Freshwater Ecosystems New Zealand database (Leathwick et al. 2011).

Predictor	Source	Description
InstreamVis OR turbidity	measured	% cover instream OR turbidity (NTU).
CHLA	measured	Benthic chlorophyll *a* from rock scrapings.
ELEVATION	REC	Altitude of the stream segment.
SegSlope	FENZ	Segment slope.
SegSumT	FENZ	Summer air temperature for a segment.
SegSubstrate	FENZ	Proportional cover of bed substrate size for a segment.
SegTSeas	FENZ	Seasonal air temperature range for a segment.
SegShade	FENZ	Riparian shade for a segment.
USCalcium	FENZ	Average calcium concentration of underlying rocks.
USPhosphorus	FENZ	Average phosphorous concentration of underlying rocks.
USHardness	FENZ	Average hardness of underlying rocks.
USRainDays	FENZ	Number of rain days >25 mm in the catchment.
USSlope	FENZ	Average slope in the catchment..
SegFlowStability	FENZ	Ratio of mean annual low flow/ mean annual mean flow
MALFtoMeanF	MfE website	Specific mean annual low flow / Specific mean flow.
FRE3	MfE website	Annual frequency of flood events <3x median annual flow.

BRT model output included the percentage total deviance explained (%TDE) and a mean cross-validation (CV) coefficient. The %TDE is a measure of the goodness of fit of the model whereas the CV coefficient is a measure of the predictive performance of the model. BRT output also provides the relative contribution of the predictors as well as the predictors' partial dependence plots. In the partial dependence plots, the fitted functions depict the response shape across each of the predictors when all other predictors are held constant, typically at the mean value. These fitted functions were used for visual threshold definition. Inclusion of stressors in the model other than sediment ESV and the environmental predictors improves confidence that the fitted function is depicting the response to sediment rather than the response to another predictor that is correlated with increasing sediment.

Results

Deposited sediment

Overall the BRT model fit ranged from 29% to 69% TDE (total deviance explained), and the CV correlation coefficient ranged from 0.52 to 0.82, indicating good predictive performance (Table F-4). Deposited sediment was also a more important predictor of macroinvertebrate metrics in class L2.2 than L2.3.

Table F-4: BRT model fit (TDE, total deviance explained) and mean CV correlation coefficient. CV = cross-validation for % cover instream.

Macroinvertebrate metric	Class	TDE (%)	CV correlation coefficient	Rank relative importance of deposited sediment	Rank relative importance of chlorophyll-a
MCI	Global	64	0.80	3	7
'EPT taxon richness'		60	0.77	1	5
'Sediment MCI'		69	0.83	2	11
'No. of decreasers'		71	0.83	2	3
MCI	L2.2	57	0.75	1	5
'EPT taxon richness'		57	0.74	1	5
'Sediment MCI'		69	0.82	1	10
'No. of decreasers'		69	0.82	2	3
MCI	L2.3	50	0.71	15	3
'EPT taxon richness'		29	0.52	10	1
'Sediment MCI'		50	0.71	3	8
'No. of decreasers'		55	0.73	13	1

The partial plot for the global BRT model showed similar response shapes for the four different metrics (Figure F-1). Visual inspection of the plot indicates that no marked changes in the metrics occur until about 30% sediment cover, after which metrics continued to decline up to 100% sediment cover. However, it is noted that approximately 70% of the data used to build the model occur in the range of 0-25% deposited sediment cover. The upper end of the relationship where the greatest response is observed may, therefore, be strongly influenced by relatively few data points.

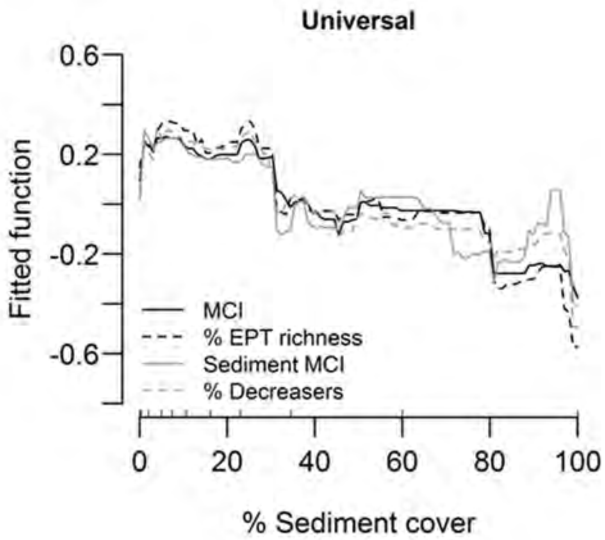


Figure F-1: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of '% cover instream'. Note that the y-axis shows change from mean response values in units of standard deviation. Also note the x-axis has been log-scaled to help with visual identification of thresholds.

The partial dependence plots within classes further illustrate the relative effect of deposited sediment on macroinvertebrate metrics (Figure F-2). Metrics show a strong negative response to deposited sediment in class L2.2 where deposited sediment was identified as an important predictor, but not in class L2.3 where it was not. Visual inspection of the partial dependence plots for class L2.2 shows a non-linear decrease in macroinvertebrate metric values from about 30% sediment cover through to 100% sediment cover similar to that observed for the global model. For class L2.3, there is a discernible decrease within the dominant distribution of data across the sediment gradient, but the magnitude of the signal is significantly smaller than in class L2.2. 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 and below 20% cover in class L2.3. An initial increase in metric values is discernible for both class L2.2 and class L2.3, up to approximately 10% and 5% sediment cover respectively, where a large proportion of the sample data are distributed.

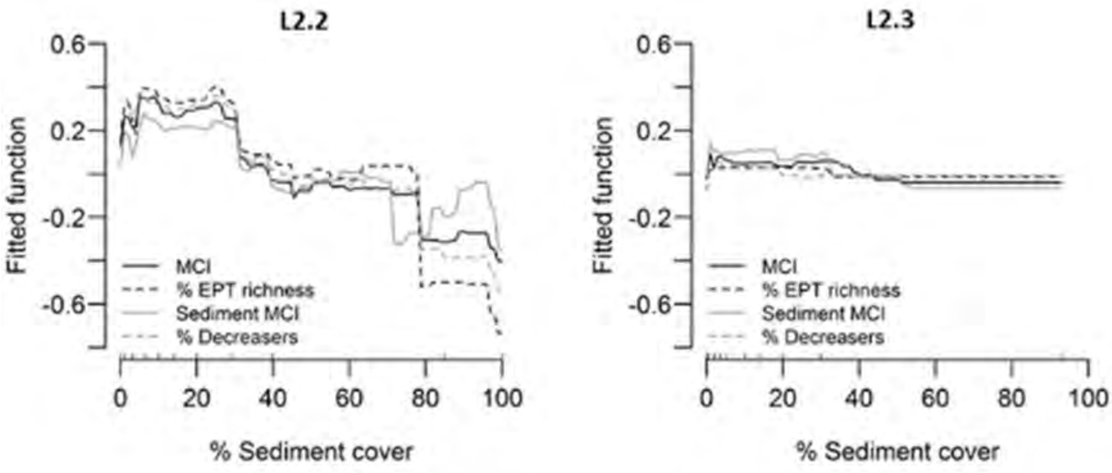


Figure F-2: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of % sediment cover. Results are presented for class 2 and 3 at the second level of the SSC aggregation (i.e., L2.2 and L2.3).

Suspended sediment

Overall the BRT model fit ranged from 48% to 76% TDE (total deviance explained), and the CV correlation coefficient ranged from 0.68 to 0.87, indicating good to very good predictive performance (Table F-5). Turbidity was more important than chlorophyll *a* as a predictor of macroinvertebrate metrics in all classes except L2.3, where Sediment MCI was still more strongly driven by turbidity than chlorophyll *a*, but not the other three metrics (Table F-5).

Table F-5: BRT model fit (TDE, total deviance explained) and mean CV correlation coefficient; CV=cross-validation for turbidity.

Macroinvertebrate metric	Class	TDE (%)	CV correlation coefficient	Rank relative importance of turbidity	Rank relative importance of chlorophyll-*a*
MCI	Global	73	0.85	8	6
'EPT taxon richness'		63	0.79	8	6
'Sediment MCI'		62	0.79	7	17
'No. of decreasers'		71	0.84	7	4
MCI	L2.1	76	0.87	3	14
'EPT taxon richness'		66	0.82	1	14
'Sediment MCI'		48	0.69	2	12
'No. of decreasers'		73	0.86	1	11
MCI	L2.2	65	0.81	9	13
'EPT taxon richness'		59	0.77	12	16
'Sediment MCI'		50	0.71	6	16
'No. of decreasers'		66	0.81	8	14
MCI	L2.3	74	0.86	12	5
'EPT taxon richness'		58	0.76	8	3
'Sediment MCI'		72	0.85	5	11
'No. of decreasers'		68	0.82	9	5
MCI	L2.4	57	0.74	4	9
'EPT taxon richness'		48	0.68	6	7
'Sediment MCI'		54	0.72	6	15
'No. of decreasers'		55	0.73	4	2

The partial plot for the global BRT model showed that the four macroinvertebrates metrics responded similarly to turbidity, although the sediment MCI metric varied somewhat from the others (Figure 4-13). Visual inspection of the plot indicates an immediate negative response of metrics to increasing turbidity that continues across the full turbidity gradient, with Sediment MCI exhibiting a lower slope. However, it is noted that approximately 90% of the data used to build the model occur in the range of 0-10 NTU. Therefore, any response after 10 NTU may be strongly influenced by relatively few data points.

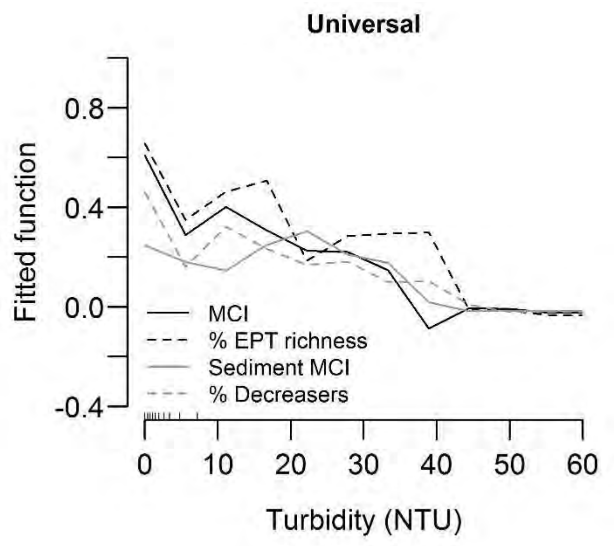


Figure F-3: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of turbidity. Note that the y-axis shows change from mean response values in units of standard deviation. Also note the x-axis has been log-scaled to help with visual identification of thresholds.

Despite strong model predictive performance and the relatively high importance of turbidity as a predictor variable, the partial dependence plots illustrate inconsistent responses of macroinvertebrate metrics to the turbidity gradient (Figure F-4). The majority of data are distributed at below 10 NTU in all classes, and a consistent negative response in this turbidity range is only observed for class L2.4. Visual inspection of the partial dependence plots does not provide impact initiation or cessation thresholds. A lack of response after approximately 20 NTU in all classes is likely due to a lack of data in this range.

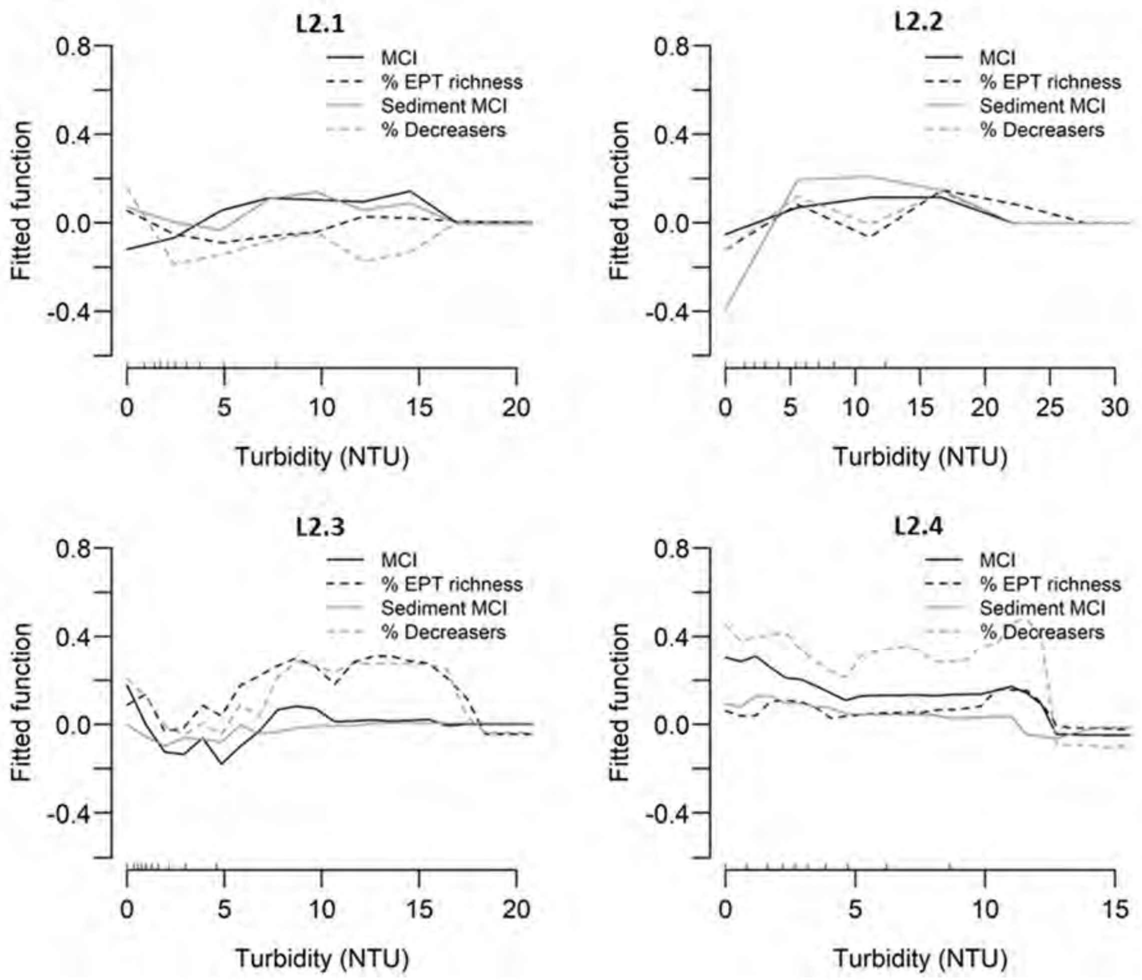


Figure F-4: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of turbidity. Plots are shown for all four classes at the second level of aggregation in the suspended sediment SSC. The x-axis range is truncated to only show the gradient where the fitted function ≠ zero. Note the different ranges of the x-axis between plots.

Outcome

Deposited sediment

We are confident that macroinvertebrate communities respond negatively to deposited sediment based on BRT model output including, model performance, the relative importance of deposited sediment compared to other predictors, the distribution of data, and the shape of the response curves. However, this relationship was not universal and was only observed for samples from class L2.2 (52% of the river network) but not from class L2.3 (40% of the river network). The relationship observed for class L2.2 is similar to that previously observed in the universal regression described in Depree et al. (2018). The metrics did not start to negatively respond until approximately 30% sediment cover but beyond this impact initiation threshold there was a negative response to increasing percent sediment cover. The response shape did not suggest that there is an impact cessation threshold, a point beyond which further increase in sediment cover does not further change community structure based on these metrics. Instead, the response shape suggests that negative effects continue up to 100% sediment cover. In both classes, an initial increase in macroinvertebrate metric values may be possible.

In support of threshold definition for deposited sediment, these results provide evidence that:

- There are sufficient data to quantify the response of macroinvertebrate metrics across a full deposited sediment gradient.
- Whilst illustrating simplified relationships, the partial dependence plots demonstrate a predominantly negative response in the presence of other predictors, at least for class L2.2.
- A significant change in community composition is likely at an impact initiation of approximately 30% deposited sediment cover, at least for class L2.2. Below 30% sediment cover, there is the likelihood of some community-level resilience to changes in deposited sediment, at least for class L2.2.
- There is significant spatial variation in the relationship between macroinvertebrate community metrics and deposited sediment.

Suspended sediment

The BRT partial dependence plots illustrate a complex relationship between turbidity and macroinvertebrate metrics and do not appear useful for determining suspended sediment thresholds. However, the BRT model output provides evidence to support threshold definition for suspended sediment as follows:

- There are sufficient data to quantify the response of macroinvertebrate metrics between 0 and 10 NTU, and possibly up to 20 NTU.
- The partial dependence plots demonstrate a predominantly negative response in the presence of other predictors only for class L2.4 (30% of the river network).
- There is significant spatial variation in the relationship between macroinvertebrate community metrics and suspended sediment.

Conclusion

Flexible regression techniques are exploratory methods that are useful for forming hypotheses and selecting appropriate parametric approaches to be able to make simple predictions about complex ecological relationships (Stillman et al. 2016; Qian and Cuffney 2018). The evidence provided from this BRT analysis of sediment-macroinvertebrate metrics should be used to help inform the application, and interpretation of results, of generalised linear models as follows:

- Support for investigation of metric-sediment relationships within classes.
- Likelihood that within some classes there may not be a strong or negative relationship, hence support for the inclusion of model terms that allow for flexible response shapes (i.e., non-linear).
- Caution should be taken when inferring negative linear responses at low levels of deposited sediment (i.e., <10% cover) or across the full turbidity gradient for the majority of the river network.

Appendix G Quantile regression analyses

Introduction

Quantile regression is a form of regression analysis that can be used to determine different measures of central tendency and statistical dispersion and, thus, obtain a more comprehensive analysis of the relationships between variables. Cade,Noon (2003) suggested that quantile regression is particularly suited to characterising ecological responses where typically all the factors that affect ecological processes are not measured and so cannot be included in predictive models. This is often observed as a data 'wedge' in a scatter plot of biological metrics (e.g., Figure G-1) and is interpreted as being the result of other stressors co-occurring with the modelled stressor, causing additional declines in the biological response over the stressor gradient. The upper boundary of the wedge is assumed to represent the control on the biological response imposed by the stressor gradient of interest. Quantile regression provides a means of estimating the location of the upper boundary of a scatter plot and so can be used to help characterise stressor-response relationships.

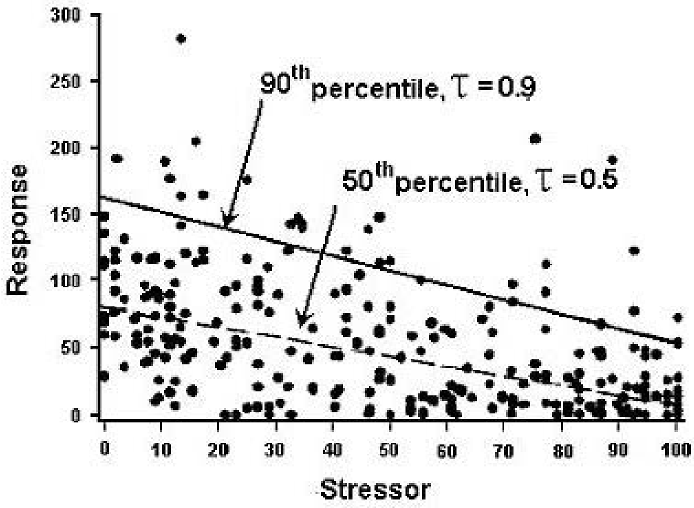


Figure G-1: Example of a 'wedge' shaped response to a stressor gradient. Quantile regression can be used to characterise the upper percentiles of the relationship that are assumed to represent the limiting boundary resulting from the stressor gradient of interest. Source: US EPA (2017) with 50th and 90th percentiles presented.

Depree et al. (2018) reported results of quantile regression analyses for both deposited sediment and suspended sediment ESVs. Poor quantile fits unsuitable for defining effects-based thresholds were found between macroinvertebrate metrics and deposited sediment. However, more robust quantile fits were derived between macroinvertebrate metrics and suspended sediment ESVs. Depree et al. (2018) also reported on quantile regression analyses of individual species' responses to suspended sediment. A non-linear Ricker model (Cade and Guo 2000; Grace et al. 2014) was used to fit the quantiles reflecting a 'subsidy-stress' type response observed in scatter plots of species abundance versus the suspended sediment ESVs. Robust relationships were described for a small sub-set of species (7) in the National River Water Quality Network (NRWQN) data, but further exploration of this approach indicated that it was difficult to apply effectively for species that are not widely distributed. This is because the derivation of quantiles is sensitive to the high number of zero counts in the data. The nature of the data in the larger 'SOE' dataset, with absolute species abundance not being available, also meant that the quantile regression approach could not be extended to individual species in that dataset. Consequently, results are only reported here for the analysis of macroinvertebrate metrics and suspended sediment ESVs.

Data

Suspended sediment

The primary stressor/response analysis was undertaken on the NRWQN dataset for the period 1990 to 2013 collated by Depree et al. (2018). These data were for 67 sites with monthly water quality monitoring and annual macroinvertebrate sampling giving a total of 1275 measurements. Additional summarising annual measures of turbidity and black disc clarity were included as required.

Pragmatically, the annual median of the monthly water quality monitoring data for each site provides a robust measure of the annual data. Conceptually, selecting a summary statistic away from the central tendency (median) of the data (such as the 80th percentile) may be valid for evaluating relationships with macroinvertebrates. This is based on the hypothesis that the higher suspended sediment levels experienced for a significant period will result in adverse effects – these low frequency, high exposure events are not 'captured' by the central tendency of the data. However, the classification analysis (Appendix D) used site medians and it was not known whether the assumption for medians would be valid if using 80th percentile values of suspended sediment. Thus, for pragmatic reasons, the annual median was also used for the effects analyses.

We identified a range of potential stressors at sites in the NRWQN dataset that may confound establishing a defining causative relationship with water column suspended sediment (or its surrogates). These include: the percentage of fines (sand fraction) in the substrate; periphyton cover; water temperature at time of macroinvertebrate sampling; pH; salt (i.e., salinity/electrical conductivity); dissolved colour; water velocity and flood frequency.

Investigation of the potential for these other stressors confounding the quantile regression analysis was undertaken using data visualisation software (DataDesk, Velleman (1989)). Relationships with each of the potential stressors were examined to determine:

- if there was an apparent stress effect with increasing concentration (or content), or
- whether the stressor leveraged the quantile regression region of the data cloud.

This analysis indicated that many of these stressors did result in apparent reductions in species or community metrics. However, the results of this analysis did not indicate that these other stressors were markedly influencing the upper quantiles of biological measures conditional on turbidity or clarity. Consequently, the quantile regression analysis for suspended sediment was conducted on the full NRWQN dataset consisting of 1275 samples collected at 67 sites.

Methods

Quantile regression analyses

All quantile regression analyses were performed using the 'quantreg' package (Koenker 2013) in R. Quantiles were fitted to the whole dataset using either a linear model or non-linear Ricker model (Cade and Guo 2000; Grace et al. 2014). The Ricker equation derives a linear regression on $\log_e$+1 transformed biological data and Log_{10} turbidity or inverse visual clarity with the curve being consistent with the subsidy/stress pattern of the stressor-response relationship observed for some density metrics and individual species abundance data. The curve increases in a convex fashion to its peak, then the curve declines in a concave fashion to some minimum – again consistent with the empirical form of the data. A range of macroinvertebrate metrics were used as the response variable.

Confidence intervals for the quantiles were calculated by inverting a rank test as described in (Koenker 2013). This provides standard error values for each of the equation coefficients and enables the 95% confidence interval for the equation to be calculated. The confidence intervals for point estimates of X can be calculated using each of the equation coefficients adjusted for their respective standard error values. Based on visual exploration of different quantiles, we selected the 95th percentile quantile for our quantile regression analysis. The 95th quantile provided a good bounding of the data cloud without high leveraging as might occur if the 99th percentile quantile was used for the dataset.

Defining effects-based thresholds

The turbidity and visual clarity values corresponding to a 30% reduction from either the reference ESV state (defined as either 0.5 NTU for turbidity and 6 m for visual clarity[13]) or the ESV state at the maxima of the biotic response were calculated from the 95th percentile quantile regression relationships. We considered this to be the best approximation of a C/D band threshold for macroinvertebrate metrics.

Results

MCI, QMCI and %EPT were modelled with log-linear quantile regressions. The other metrics were fitted with a subsidy/stress Ricker model on transformed variables. The quantile regression relationships for a range of quantiles (99%, 95%, 90%, 80% and 50%) in relation to visual clarity and turbidity are shown in Figure G-2 and Figure G-3. Summary tables of the visual clarity and turbidity thresholds corresponding with a 30% reduction from the ESV state at the maxima of the biotic response are shown in Table G-1 and Table G-2 respectively.

Table G-1: Summary of 30% effect thresholds for visual clarity based on the 95th percentile quantile relationships. All variables show variable maximum and corresponding visual clarity with model-derived 30% reduction. The blue highlighted variables are derived from log-linear regressions and a 30% reduction from a high-quality biotic condition.

Biotic variable	Maximum of biotic variable	Visual clarity at Maximum (m)	Maximum less 30%	Visual clarity threshold for 30% reduction (m)
Taxa richness	32.0	6	22.4	0.26
Density	13,910	0.81	9,737	0.33
MCI	136	6	95.1	<0.15
QMCI	7.7	6	5.4	<0.15
EPT taxa	18.3	6	12.8	0.33
EPT individuals	5,687	1.76	3981	0.52
%EPT	94.3	6	66.0	<0.15

[13] This analysis was transferred directly from Depree et al. (2018) unchanged. The default reference states used here were arbitrary and are not consistent with those derived in this stage of the project, but it was outside the scope to redo these analyses.

Table G-2: Summary of 30% effect thresholds for turbidity based on the 95th percentile quantile relationships. All variables show variable maximum and corresponding turbidity with model-derived 30% reduction. The blue highlighted variables are derived from log-linear regressions and a 30% reduction from a high-quality biotic condition corresponding to a low turbidity condition. NA indicates model fit not suitable for use in effects determination.

Biotic variable	Maximum of biotic variable	Turbidity at Maximum (NTU)	Maximum less 30%	Turbidity threshold for 30% reduction (NTU)
Taxa richness	35.1	0.5	24.5	17.0
Density	13,063	5.7	9144	19.0
MCI	136	0.5	95.5	>50
QMCI	7.7	0.5	5.4	>50
EPT taxa	20.4	0.5	14.3	8.2
EPT individuals	5,435	2.4	3805	12.2
%EPT	93.9	0.5	65.8	NA

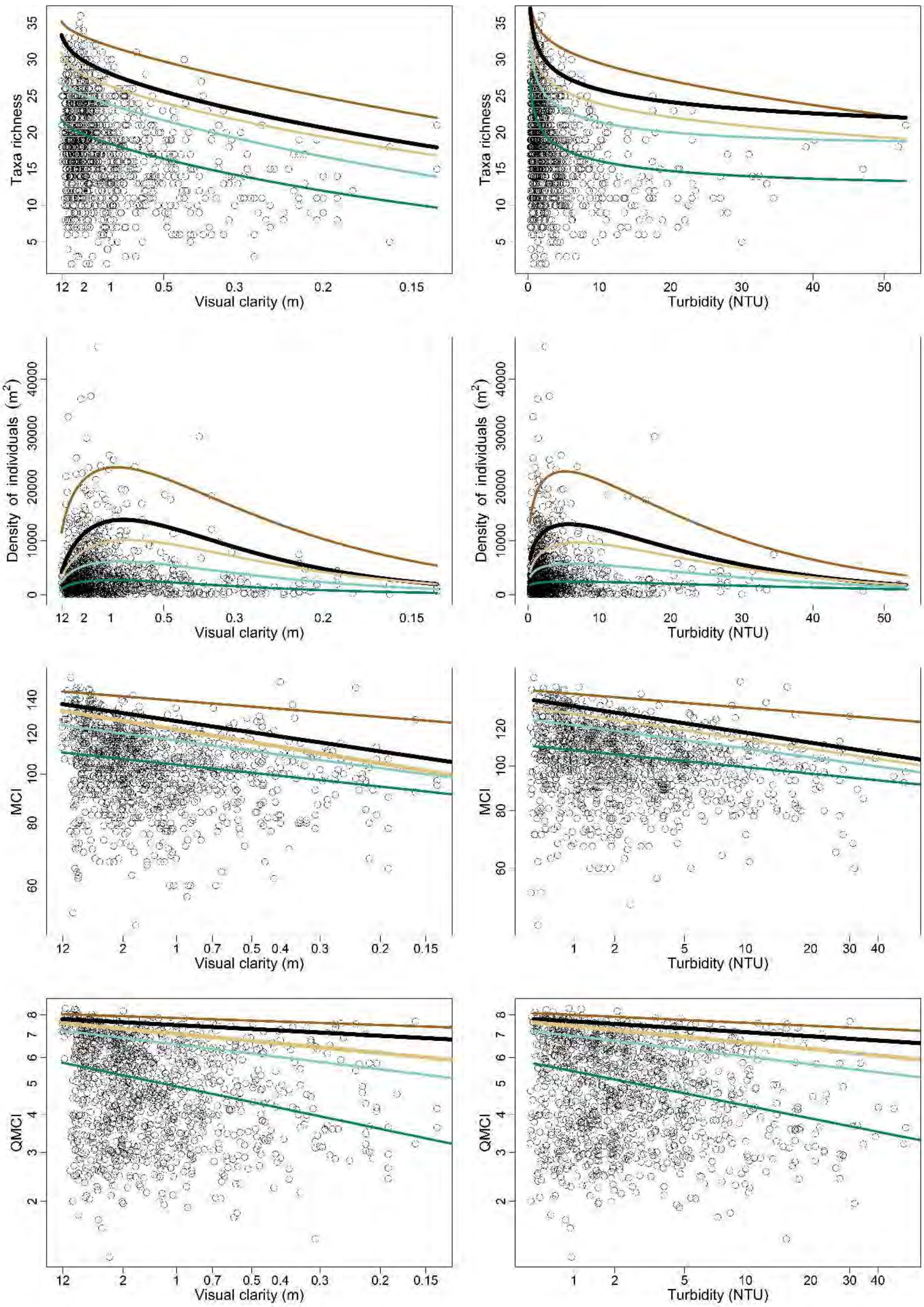


Figure G-2: Quantile regression fits for a range of macroinvertebrate metrics. Quantile lines are plotted for the 99% (brown), 95% (black), 90% (light brown), 80% (light turquoise) and 50% (turquoise) percentiles.

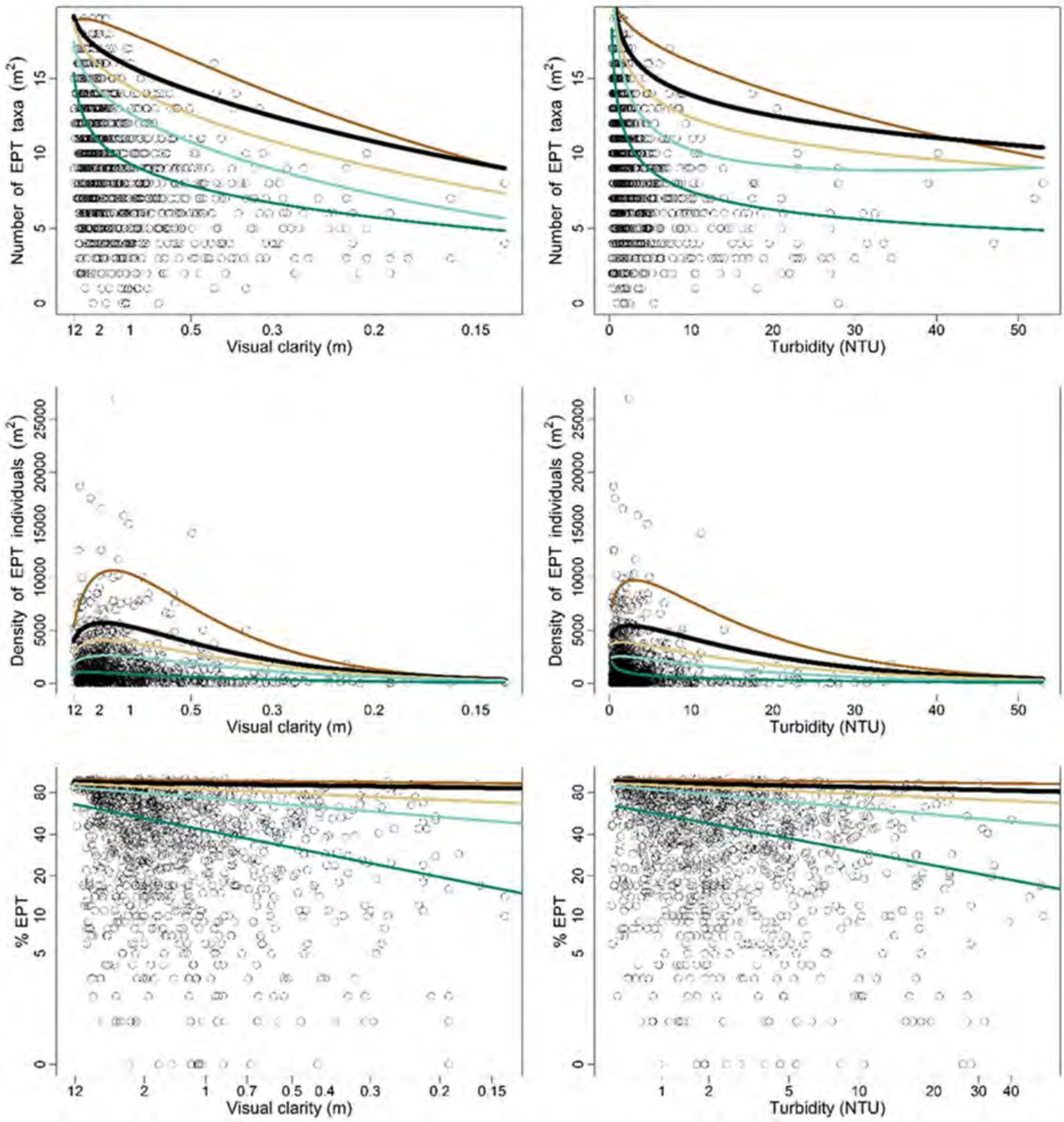


Figure G-3: Quantile regression fits for a range of macroinvertebrate metrics. Quantile lines are plotted for the 99% (brown), 95% (black), 90% (light brown), 80% (light turquoise) and 50% (turquoise) percentiles.

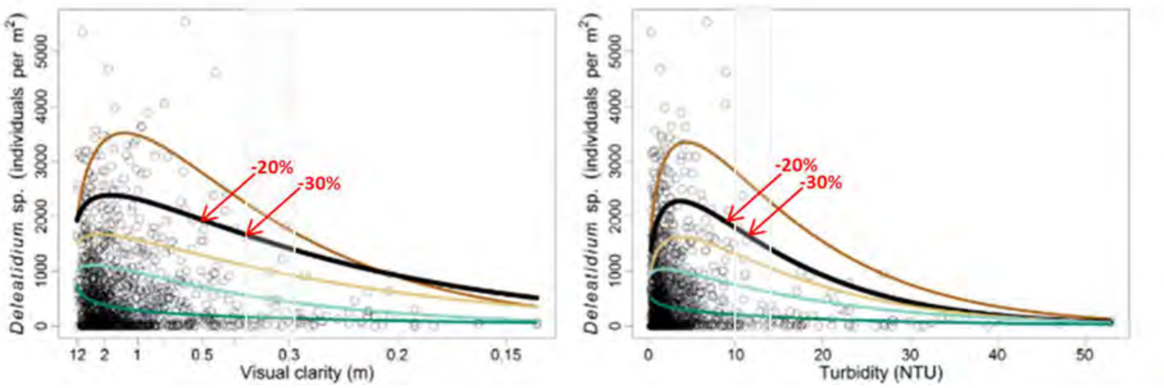


Figure G-4: Quantile regression fits to mayfly (*Deleatidium* spp.) abundance data for visual clarity and turbidity (annual medians) from NWQRN monitoring. Thresholds for 20% and 30% reduction in abundance from maximum shown for 95% quantile. Shaded band indicates range for calculated extirpation XC95 values for various reach classifications. Quantile lines are plotted for the 99% (brown), 95% (black), 90% (light brown), 80% (light turquoise) and 50% (turquoise) percentiles. Extirpation ESVs for *Deleatidium* are shown in Appendix H.

An example of the quantile regressions fits to the NWQRN data for a mayfly (*Deleatidium* spp.) is shown in Figure G-4 for visual clarity and turbidity. *Deleatidium* is generally in the lower 25th percentile of the species sensitivity to suspended sediments (Appendix H). The visual clarity and turbidity values for 20% and 30% reduction from the peak abundance on the 95th percentile quantile are shown, together with the range of extirpation XC95 values calculated for the five classification classes. The extirpation XC95 values for turbidity are generally in the 20-40% reduction below the peak abundance based on the 95th percentile quantile, while the clarity extirpation XC95 values are in the 30-40% reduction range. Notably, the extirpation XC95 values may equate to relatively high abundance of some species depending on the nature of the relationship between the species and the stressor measure.

Discussion

The chosen analytical approach for the macroinvertebrate-suspended sediment ESV responses reflected both the availability of suitable national data that were quantitatively collected using standard methods, and the need to derive quantitative relationships for a multiple stressor environment. The NRWQN dataset satisfied these criteria and provided the additional monitoring data for other potential stressors that could confound establishing a causative relationship with water column suspended sediment (or its surrogates turbidity and black disk visual clarity). The nature of water column suspended sediment in rivers is a combination of particulate organic matter (POM) and inorganic suspended sediment – resulting in a subsidy-stress response for macroinvertebrate communities. A quantile regression approach based on the 95th percentile quantile and a non-linear response function was consistent with the subsidy/stress response relationships.

Our quantile regressions for the macroinvertebrates were based on the 95th percentile quantile for macroinvertebrate responses to annual median values of inverse visual clarity so that the fitted form of the equation was consistent with turbidity. This quantile provided a good bounding of the data cloud without high leveraging, which might occur if the 99th percentile were used for the dataset. Our choice of a 30% reduction in the macroinvertebrate measures is based on this being a substantial reduction. Either lower (e.g., 20% effect) or a higher (e.g., 50% effect) could be used as the basis for the SSD-based guideline derivation. However, we consider that higher values would allow an excessive level of environmental impact and a lower threshold would introduce increased uncertainties relating to the cause-effect nature of the primary stressor (i.e., suspended sediment) and the causation linkage between the annual median visual clarity and turbidity values and the biotic measures. For these reasons, we chose to use the 30% effect measure as a pragmatic threshold for deriving effect-based thresholds from this analysis.

We consider the key advantages of the quantile regression approach are:

- The method is robust for non-linear response relationships.
- Quantitative non-linear relationships can provide numeric effect thresholds.
- The approach is conceptually robust for establishing the maximum community metric or species abundance in relation to the stressor of concern.
- The approach may be used for stressor elimination – providing quantitative information is available on other potential stressors affecting the biotic communities.

We consider some of the main limitations of the quantile regression approach are:

- Suitable quantile regression relationships need to be available to fit non-linear responses to provide numeric derivation of effect thresholds and statistical parameterisation of the relationships.
- Subjective assessments may need to be applied to determine the most appropriate quantile for fitting the data cloud.
- Additional monitoring data needs to be available for other stressors to facilitate a stressor identification/elimination analysis.
- Relatively large numbers of data are required to undertake the analysis and special techniques may be required to appropriately manage the ‘zero’ data in many large datasets.
- The selection of a 30% deviation level for the ‘bottom line’ (i.e., C/D attribute band transition) is arbitrary.

Appendix H Extirpation analyses

Introduction

Here we present the results of an analysis aimed at determining the proportion of macroinvertebrate taxa that may be locally extirpated as sediment ESV state 'worsens'. The units of the deposited fine sediment and turbidity ESVs are percentage coverage (% cover) and nephelometric turbidity units (NTUs), respectively, so their state worsens when their values increase. By contrast, visual clarity—with units of metres (m)—worsens when it decreases. The analysis we undertook is referred to as a biological extirpation analysis (BEA; Cormier et al. (2018)), and is a well-established analytical technique in ecotoxicology (Posthuma et al. 2002; US EPA 2016a).

Broadly, the analysis involves two steps: First, we determine the value of the ESV that will likely result in the local extirpation of the species. In our case, local extirpation means the species has been extirpated from the 'reach' (NZReach), as defined in the New Zealand River Environment Classification (REC; Snelder,Biggs (2002)). Following Cormier et al. (2018), a species is deemed locally extirpated when the probability of its occurrence declines to 5% as a function of a worsening ESV state. The ESV value that corresponds to the 5% probability of occurrence is referred to as that species' XC95. This first step is completed for all species in the assemblage; that is, those species that have passed some data quality assurance criteria (see below for details).

Second, the XC95 values of species in the assemblage are ranked from lowest (most sensitive; least tolerant) to highest (least sensitive; most tolerant), and then that ranked list of XC95 values is transformed to yield a species sensitivity distribution (SSD; Posthuma et al. (2002)). The SSD is essentially a cumulative probability distribution, which is a function whose y-values give the proportion of the species in the analysis that are locally extirpated when the ESV (x-value) reaches that state.

Our BEA algorithm largely follows that outlined by Cormier et al. (2018), but with some modifications to better accommodate smaller datasets and our sediment state classification, hence regionalisation of NOF management bands. The details of our algorithm are provided in the next section.

Methods

Data

The data we used for this BEA came from two sources. The first source was the New Zealand River Water Quality (NRWQN) dataset. These data were sourced from a total of 67 sites coming from 67 unique NZReaches and a total of 1274 samples. The NRWQN was established and managed by NIWA and was characterised by standardised, quantitative sampling protocols across sites and sampling events. Most of the 67 sites contain data spanning 20 years of standardised annual sampling regimes. The NRWQN data was used for the BEA on turbidity and clarity. Within NRWQN sites macroinvertebrate samples were taken once annually, while turbidity and visual clarity measurements were taken monthly. A single macroinvertebrate sample was taken from NRWQN sites each year, and these macroinvertebrate samples were paired with median turbidity and visual clarity estimates, where the medians were calculated over monthly turbidity and visual clarity estimates for the 12 months preceding the macroinvertebrate sampling date.

The NRWQN data were not used for the BEA on deposited fine sediment because it did not contain the variable required for that analysis. The deposited fine sediment variable used for the present BEA was SAM2 % cover instream ('instreamVis'), the proportion of quadrats covered in fine sediment as measured using instream visual inspection within river runs—this variable was included only in the State of the Environment (SOE) dataset which we explain below.

The SOE dataset was the second dataset used for the present BEA. This is the dataset assembled under the New Zealand-wide State of the Environment monitoring undertaken by regional councils. It comprises data from 1311 sites and a total of 8327 individual samples. Macroinvertebrate sampling under the SOE monitoring is not standardised and the method is less quantitative (kick nets as opposed to surber samplers). Moreover, monitoring of the ESVs within the SOE set was less extensive in time, and less consistent. Nevertheless, annual median ESV values were calculated for these data and paired with annual macroinvertebrate samples.

Quality assurance – filtering of global data frames and selection of sediment classification level

Turbidity and visual clarity were measured using similar instruments for both the NRWQN and SOE sets, and they had the same units. Macroinvertebrate presence/absence within a sample was determined for both datasets and the sets were then merged to form a global data frame. To ensure our BEAs were robust, the data had to satisfy numerous conditions, listed throughout this section:

- Filter 1: Remove any rows containing NAs for all three ESVs.
- Filter 2: Remove any rows where turbidity and visual clarity were equal to zero (this was deemed impossible and a measurement error).
- Filter 3: For an NZReach to be included in any BEA it had to contain at least three samples for the ESV of interest, taken across three unique years. This condition was imposed on the analysis to increase the reliability of our assignments of presence or absence of a species to a reach—the less samples within a reach the more likely our assignment of a species' presence/absence to that reach will be incorrect and shaped by chance alone.

For each ESV, suspended/deposited sediment classes (see Appendix D) were a factor in the BEA, towards obtaining NOF management bands that are class- hence region-specific. Data exploration indicated that we had insufficient data to employ the Level 4 sediment classification—relationships between the probability of occurrence of individual taxa and ESVs were too noisy in many of the 12 classes at Level 4, for all ESVs. Level 3 was selected (up to 8 sediment classes) as data exploration indicated it was the finest resolution of classification yielding some convincing relationships between ESVs and a species' probability of occurrence in at least some—or most—of the (Level 3) classes.

For a species to enter the BEA it had to satisfy the following conditions:

- Filter 4: For the turbidity BEA, a species must be present over samples spanning at least 10 NTUs; For the visual clarity BEA, a species must be present over samples spanning at least 1 m; For the deposited fines BEA, a species must be present over samples spanning at least 20% fines.

- Filter 5: When estimating a species' XC95 value it must be present in at least 10 NZReaches, within any given sediment class. Thus, for each ESV, the species composition of the SSD may vary across sediment classes.
- Filter 6: For a species to be retained it must be present in at least 2 of the 8 sediment classes at Level 3.

These conditions were imposed on a species to increase the robustness of the models of probability of occurrence as a function of ESV values, from which XC95 values of species were obtained.

Estimation of XC95s, SSDs and band thresholds

Approximately normal distributions are desirable for BEA as they result in less noisy relationships between probability of occurrence and the ESV (Cormier et al. 2018). Suspended sediment ESVs were strongly log-normal, so they were transformed prior to estimation of XC95s for any species: turbidity was log-transformed (natural logarithm) and visual clarity was inversed, then log-transformed (i.e., $\ln(1/x) = -\ln(x)$). Visual clarity was inverse-transformed for computational simplicity; the BEA algorithm was easiest to encode when all ESVs 'worsened' as values increased.

Estimating the XC95 value for each ESV for each species involved the following steps:

1. The probability of occurrence as a function of ESV was determined by dividing an ESV into 'bins', then calculating the proportion of samples in each bin where the species was present. For each species, 30 bins were defined for each ESV.
2. Filter 7: If the probability of occurrence within an ESV bin was calculated over less than three samples, then the probability of occurrence for that ESV bin was assigned an NA (missing data). This filtering step reduced the frequency of spurious probability functions affecting our BEAs.
3. Once the probability function was estimated for a species, we fitted a binomial linear mixed-effects model (BLMM), where probability of occurrence of an individual species was a function of ESV bin, where sediment class-specific intercepts were allowed to deviate around the global intercept by including sediment class as a random factor (normally-distributed deviations of intercepts around global mean, but only one global slope). The global slope of this model was used to test if a species generally exhibited a negative relationship with the worsening ESV.
4. Filter 8: If a species did not exhibit a negative relationship with the ESV it was precluded from the analysis; its XC95 was undefined, following Cormier et al. (2018).
5. If a global negative slop was returned from the BLMM then we estimated the cumulative probability of occurrence of a species as a function of the ESV. The 95th centile of this cumulative probability function is the XC95; the value of the ESV where the species has reached only a 5% probability of detection and is, therefore, deemed locally extirpated (following Cormier et al. (2018)). A binomial additive mixed-effects model (BAMM) was fitted to a species' cumulative probability function to facilitate estimation of XC95 within each sediment class, within each ESV. This was done to:
    - Improve the accuracy of XC95 estimation, given interpolation was often required to estimate the ESV value at which probability of detection was exactly 5%.

- Facilitate XC95 estimation for certain sediment classes of a species where few data were available. By including sediment class as a random factor in a mixed model the global ‘patterns’ in the data set have an influence on parameter estimation within individual classes containing few data—we borrow the strength of the data set as a whole to estimate class-specific parameters.

To estimate the species sensitivity distribution and the NOF management band thresholds, the following steps were implemented:

1. Once we had the XC95s of species we determined the SSD for each sediment class. This is done by ranking species within classes (least tolerant to most tolerant), then estimating a cumulative percentage function describing the proportion of species extirpated as the ESV worsens.
2. To facilitate estimation of various extirpation thresholds we fitted a binomial additive model whereby percentage ranks of species’ XC95s were a function of their XC95 values, where binomial curves were allowed to vary across sediment classes. Again, by fitting these models we were able to more accurately estimate extirpation thresholds using interpolation. Thus, our SSD for an ESV is now defined by a binomial additive model.
3. The following extirpation thresholds were then estimated from the SSD: ESV values corresponding to 1%, 2.5%, 5%, 10%, 25%, 50% and 75% of species being extirpated within a class.

As is the case with any modelling exercise, assumptions are made. In our case, we have made three important assumptions that may affect our estimation of band thresholds:

The first assumption is that samples taken through time, within any of the monitoring sites, are uncorrelated and/or independent. This assumption could be relaxed by taking a more sophisticated modelling approach, which would require additional analyses for implementation.

The second assumption is that the relationship between a species’ probability of occurrence and an ESV is not confounded by collinearity between the ESV and other drivers of the species’ distribution. As is the case with the first assumption, we could relax this assumption by taking a more sophisticated analytical approach, but this was outside the scope of this project.

The third assumption is that the annual median of turbidity and visual clarity values based on monthly measurements are appropriate physical and statistical measures of the stressor of concern. As each of these measures is a surrogate for the combined effects of particulate organic material and particulate inorganic material, with visual clarity also being affected by dissolved organic matter, these measures must be considered a pragmatic indicator of the stressor. Exposure to flow and suspended sediments is highly time-variable and varies greatly between rivers. The sensitivity of individual species is also expected to differ markedly in their time/concentration tolerance for sediment. Community tolerance may well be set by an annual maximum or by intermittent peaks, however, the data is not available to provide better calibration of the exposure measure. The annual median is the most robust statistical measure and most easily related to catchment characteristics.

Assigning NOF management band thresholds

We recommend the following correspondence between extirpation thresholds and sediment NOF management bands:

1. The threshold between the A and B management bands: 1% extirpation thresholds from BEA.
2. The threshold between the B and C management bands: 2.5% extirpation thresholds from BEA.
3. The threshold between the C and D management bands: 5% extirpation thresholds from BEA.

These recommendations may seem overly environmentally-conservative to certain stakeholders. However, it is important to realise that the extirpation thresholds estimated using our BEAs likely underestimate the true proportion of taxa extirpated as sediment load to New Zealand rivers increases. They are underestimates for the following reasons:

- The process of sampling invertebrates from rivers is imperfect and rarer—potentially more sensitive species—are missed by the sampling process itself.
- The various data filtering steps implemented to ensure the robustness of our BEA removes species with narrow environmental tolerances, a narrow distribution, or low frequency of occurrence in samples (just rare, despite being detected and with a broad tolerance and distribution).

It follows that our BEAs are effectively analyses of how worsening ESV states extirpate the more common macroinvertebrate species from New Zealand river reaches.

Results

Turbidity

After merging the NRWQN and SOE datasets, the global data frame contained a total of 206 taxa. Of these, 62 satisfied filter conditions 1–6, and 44 satisfied conditions 7 and 8. Sufficient data were available to obtain SSDs within 4 of the 8 sediment classes of the Level 3 classification.

The macroinvertebrate taxa comprising the SSDs for the classes for which we had sufficient data, and their XC95 values, are presented in Table H-1. The SSDs that result from the rankings in Table H-1 are presented in Figure H-1. Clear differences in the extirpation thresholds—hence the NOF management bands from this analysis—can also be seen in Figure H-1. The extirpation thresholds themselves are presented in Table H-2; these thresholds are those estimated from data subdivided nationally according to the Level 3 sediment state classification.

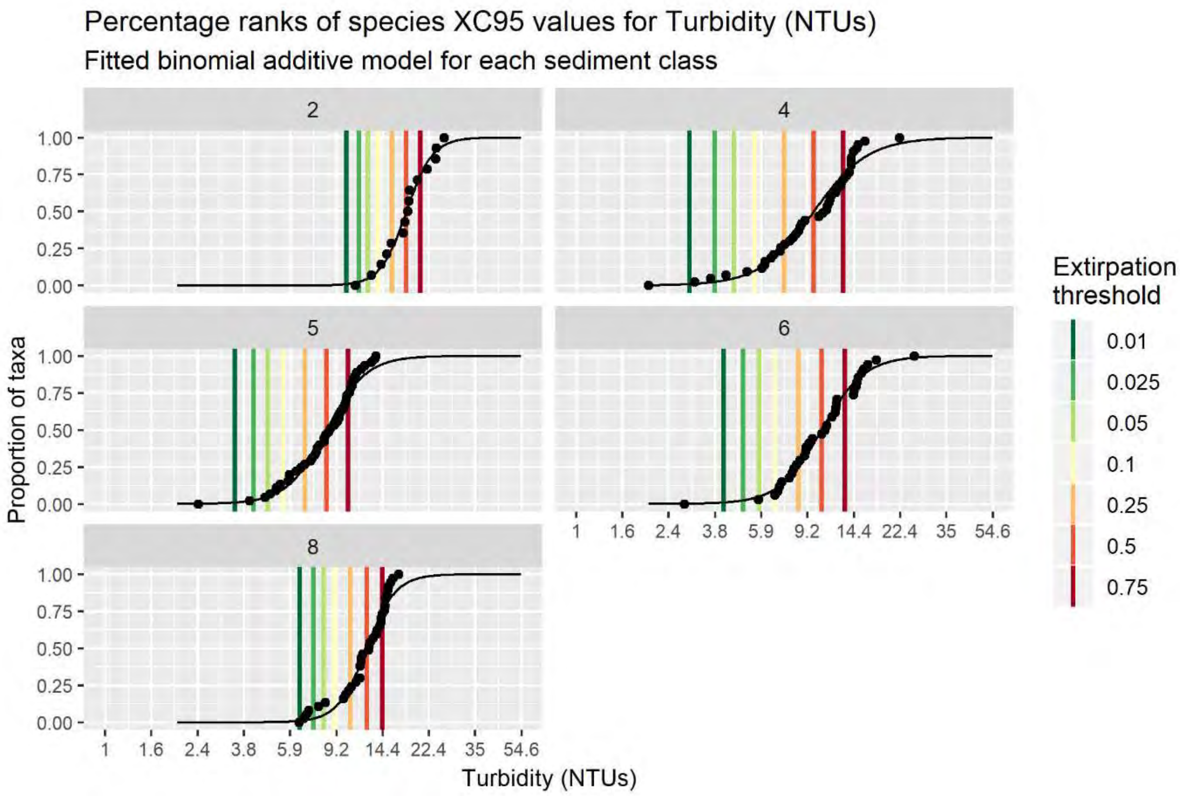


Figure H-1: Proportion of taxa extirpated as a function of annual median turbidity within each of four suspended sediment classes. Classes defined at Level 3 of the suspended sediment classification).

Table H-1: Ranked lists of the values defining the annual median turbidity (NTUs) at which there is a 95% probability of macroinvertebrate taxa being extirpated locally (XC95). Values are provided within suspended sediment classes (Level 3 suspended sediment classification; sufficient data for five of eight classes). Taxa highlighted in light grey are the taxa extirpated at the 25% threshold; while those in darker grey are extirpated at the 5% threshold. Taxon names abbreviated to first 8 characters to save space.

L3.2		L3.4		L3.5		L3.6		L3.8	
Species	**XC95**	**Species**	**XC95**	**Species**	**XC95**	**Species**	**XC95**	**Species**	**XC95**
Olinga	10.76	Stenoper	2.01	Mischode	2.56	Beraeopt	2.99	Empidida	6.56
Psilocho	12.57	Beraeopt	2.81	Physella	3.78	Empidida	5.86	Gyraulus	6.66
Pycnocen	13.91	Hydraeni	2.91	Austrope	3.94	Hydraeni	6.45	Ceratopo	6.68
Pycnocen	14.69	Austrope	3.40	Beraeopt	4.15	Acarina	6.55	Paralimn	6.68
Tanypodi	15.38	Physella	4.69	Copepoda	4.53	Coloburi	6.85	Scirtida	6.94
Deleatid	17.33	Copepoda	5.25	Paralimn	4.89	Aphrophi	7.18	Hydraeni	7.28
Chironom	17.82	Ceratopo	6.07	Megalept	5.02	Austrope	7.28	Zephlebi	9.44
Elmidae	18.15	Coloburi	6.34	Hydraeni	5.19	Megalept	7.47	Tanytars	9.70
Hudsonem	18.15	Acarina	6.43	Gyraulus	5.38	Paralimn	8.08	Austrocl	10.20
Acarina	18.83	Paralimn	6.45	Acarina	5.61	Stenoper	8.08	Muscidae	10.27
Aoteapsy	19.89	Nesamele	6.68	Stenoper	5.66	Diamesin	8.28	Aphrophi	10.96
Hydrobio	22.01	Empidida	6.79	Empidida	5.97	Costacho	8.51	Coloburi	11.03
Aphrophi	23.47	Diamesin	7.15	Coloburi	6.51	Nesamele	8.73	Olinga	11.36
Austrosi	23.68	Gyraulus	7.29	Muscidae	6.80	Muscidae	9.03	Psilocho	11.36
Muscidae	25.72	Tabanida	8.17	Aphrophi	7.25	Olinga	9.12	Austrosi	11.47
		Aphrophi	8.71	Zelandop	7.41	Zelandop	9.69	Pycnocen	11.47
		Megalept	8.88	Scirtida	7.50	Archicha	10.47	Pycnocen	11.47
		Muscidae	8.95	Hudsonem	7.66	Austrosi	10.86	Hexatomi	11.60
		Costacho	10.38	Austrocl	8.01	Hudsonem	10.86	Archicha	12.13
		Zelandop	10.47	Nesamele	8.01	Pycnocen	10.86	Acarina	12.57
		Pycnocen	11.16	Ceratopo	8.15	Austrocl	11.12	Mischode	12.82
		Neurocho	11.22	Diamesin	8.16	Neurocho	11.82	Hudsonem	13.04
		Olinga	11.58	Tabanida	8.17	Psilocho	11.90	Hydrobio	13.16
		Mischode	11.62	Costacho	8.29	Pycnocen	12.01	Nesamele	13.34
		Hudsonem	12.12	Pycnocen	8.47	Deleatid	13.91	Megalept	13.69
		Archicha	12.45	Olinga	9.12	Aoteapsy	14.04	Costacho	13.93
		Pycnocen	12.46	Psilocho	9.29	Elmidae	14.43	Aoteapsy	14.04
		Psilocho	12.92	Pycnocen	9.55	Hydrobio	14.43	Eriopter	14.05
		Austrosi	13.16	Hydrobio	9.72	Hydrobio	15.02	Austrope	14.17
		Aoteapsy	13.41	Austrosi	9.81	Zephlebi	15.06	Deleatid	14.30
		Deleatid	13.53	Aoteapsy	9.90	Tanypodi	15.53	Elmidae	14.30
		Hydrobio	13.65	Elmidae	9.99	Eriopter	16.14	Chironom	14.56
		Tanypodi	13.65	Neurocho	10.03	Chironom	17.82	Neurocho	14.67
		Austrocl	13.81	Deleatid	10.18	Ceratopo	25.53	Stenoper	14.93
		Eriopter	13.93	Tanypodi	10.18			Tanypodi	14.97
		Chironom	14.04	Archicha	10.29			Zelandop	15.73
		Elmidae	14.30	Tanytars	10.53			Neozephl	16.79
		Plectroc	14.37	Chironom	10.56				
		Tanytars	14.97	Eriopter	10.93				

L3.2		L3.4		L3.5		L3.6		L3.8	
Species	**XC95**	**Species**	**XC95**	**Species**	**XC95**	**Species**	**XC95**	**Species**	**XC95**
		Zephlebi	16.42	Hexatomi	11.20				
		Neozephl	21.81	Plectroc	11.62				
				Neozephl	12.20				
				Hydrobio	12.69				
				Zephlebi	12.89				

There were insufficient data for BEA within classes L3.1, L3.3 and L3.7. However, given our sediment state classification is based on hierarchical similarities, we can assign extirpation thresholds to those classes with insufficient data. Such classes are assigned the extirpation thresholds of the next most similar class in the classification. For example, within the suspended sediment state classification, Class L3.3 is most similar to Class L3.2, based on their sediment supply and retention characteristics. So we assign the extirpation thresholds of Class L3.2, for which we had sufficient data, to Class L3.3, for which data were inadequate.

For convenience, we organise the information in Table H-2 according to the Level 4 classification as well. Again, to do this we exploit the hierarchical nature of the sediment state classification and assign extirpation thresholds for Level 4 classes according to the Level 3 classes they are grouped within.

Table H-2: Threshold annual median turbidity values (NTUs) at which 1%, 2.5%,…,75% of macroinvertebrate taxa are extirpated from the community within suspended sediment classes. Classes are at Level 3 of the suspended sediment classification; sufficient data for five of eight classes; classes with insufficient data indicated by an asterisk, and thresholds assigned to those based on method in text.

Potential NOF band threshold		A/B	B/C	C/D				
Sed. Class		1%	2.5%	5%	10%	25%	50%	75%
L3.1*	L4.1	3.89	4.73	5.51	6.44	8.17	10.31	13.00
L3.2	L4.5 L4.8	9.95	11.24	12.27	13.54	15.56	17.89	20.68
L3.3*	L4.6	9.95	11.24	12.27	13.54	15.56	17.89	20.68
L3.4	L4.7 L4.10 L4.11	2.54	3.32	4.09	5.11	7.03	9.67	13.30
L3.5	L4.9	3.03	3.69	4.29	5.02	6.33	7.98	10.01
L3.6	L4.12	3.89	4.73	5.51	6.44	8.17	10.31	13.00
L3.7*	L7.2	3.89	4.73	5.51	6.44	8.17	10.31	13.00
L3.8	L4.3 L4.4	6.11	6.98	7.75	8.66	10.19	11.91	14.02
	Global mean	5.41	6.33	7.15	8.15	9.90	12.03	14.71

Visual clarity

Of the 206 taxa, 63 satisfied filter conditions 1-6, and 44 satisfied conditions 7 and 8. Sufficient data were available to obtain SSDs within 4 of the 8 sediment classes of the Level 3 classification.

Table H-3: Ranked lists of the values defining the annual median visual clarity (m) at which there is a 95% probability of macroinvertebrate taxa being extirpated locally (XC95). Values are provided within suspended sediment classes (Level 3 suspended sediment classification; sufficient data for four of eight classes). Taxa highlighted in light grey are the taxa extirpated at the 25% threshold; while those in darker grey are extirpated at the 5% threshold. Taxon names abbreviated to first 8 characters to save space.

L3.2		L3.4		L3.5		L3.6		L3.8	
Species	XC95	Species	XC95	Species	XC95	Species	XC95	Species	XC95
Olinga	0.54	Plectroc	2.03	Austrope	2.00	Beraeopt	1.37	Scirtida	1.05
Pycnocen	0.44	Mischode	1.51	Mischode	1.77	Austrope	1.33	Tanytars	0.85
Psilocho	0.42	Stenoper	1.32	Stenoper	1.35	Stenoper	1.12	Empidida	0.71
Hudsonem	0.35	Tanytars	1.07	Megalept	1.34	Helicops	0.84	Gyraulus	0.67
Tanypodi	0.33	Hydraeni	1.00	Hydraeni	1.32	Coloburi	0.72	Helicops	0.64
Acarina	0.31	Copepoda	0.89	Tanytars	1.10	Zelandop	0.65	Hexatomi	0.64
Pycnocen	0.31	Zelandop	0.81	Scirtida	1.02	Costacho	0.63	Stenoper	0.60
Deleatid	0.29	Beraeopt	0.80	Copepoda	0.91	Nesamele	0.59	Paralimn	0.59
Chironom	0.28	Helicops	0.80	Neozephl	0.89	Archicha	0.57	Austrocl	0.54
Elmidae	0.26	Physella	0.77	Beraeopt	0.82	Hydraeni	0.56	Zelandop	0.54
Hydrobio	0.24	Costacho	0.70	Helicops	0.79	Acarina	0.55	Costacho	0.53
Aoteapsy	0.23	Austrocl	0.69	Physella	0.79	Austrocl	0.55	Zelandob	0.53
Muscidae	0.23	Coloburi	0.56	Paralimn	0.74	Olinga	0.54	Ceratopo	0.50
Aphrophi	0.21	Gyraulus	0.56	Gyraulus	0.73	Hydrobio	0.53	Acarina	0.49
Austrosi	0.21	Diamesin	0.56	Hexatomi	0.72	Muscidae	0.52	Muscidae	0.46
		Paralimn	0.53	Empidida	0.71	Aphrophi	0.50	Neozephl	0.46
		Nesamele	0.49	Zelandop	0.71	Diamesin	0.45	Coloburi	0.45
		Acarina	0.46	Coloburi	0.67	Hudsonem	0.44	Mischode	0.45
		Archicha	0.44	Costacho	0.62	Pycnocen	0.42	Archicha	0.43
		Neozephl	0.43	Muscidae	0.57	Paralimn	0.41	Psilocho	0.42
		Pycnocen	0.43	Acarina	0.56	Austrosi	0.41	Nesamele	0.42
		Muscidae	0.39	Hydrobio	0.55	Neurocho	0.39	Eriopter	0.41
		Neurocho	0.39	Austrocl	0.53	Psilocho	0.38	Hudsonem	0.40
		Zelandob	0.38	Nesamele	0.51	Zelandob	0.37	Hydraeni	0.40
		Empidida	0.34	Plectroc	0.51	Empidida	0.36	Olinga	0.40
		Hudsonem	0.34	Hudsonem	0.48	Pycnocen	0.33	Hydrobio	0.39
		Psilocho	0.34	Diamesin	0.47	Megalept	0.33	Chironom	0.37
		Austrope	0.31	Pycnocen	0.46	Hydrobio	0.32	Pycnocen	0.37
		Austrosi	0.31	Archicha	0.45	Aoteapsy	0.30	Pycnocen	0.36
		Aphrophi	0.30	Ceratopo	0.44	Deleatid	0.30	Austrosi	0.36
		Olinga	0.30	Aphrophi	0.42	Elmidae	0.30	Aphrophi	0.36
		Hydrobio	0.30	Hydrobio	0.42	Tanypodi	0.30	Aoteapsy	0.35
		Pycnocen	0.30	Psilocho	0.42	Eriopter	0.28	Deleatid	0.35
		Deleatid	0.30	Neurocho	0.41	Ceratopo	0.23	Austrope	0.34
		Aoteapsy	0.29	Austrosi	0.41	Chironom	0.23	Elmidae	0.34
		Elmidae	0.28	Chironom	0.41			Tanypodi	0.33
		Ceratopo	0.27	Zelandob	0.39			Megalept	0.29
		Chironom	0.27	Eriopter	0.39			Neurocho	0.27
		Eriopter	0.25	Olinga	0.38				

L3.2		L3.4		L3.5		L3.6		L3.8	
Species	XC95	Species	XC95	Species	XC95	Species	XC95	Species	XC95
		Megalept	0.24	Deleatid	0.38				
		Tanypodi	0.23	Pycnocen	0.38				
				Elmidae	0.37				
				Aoteapsy	0.36				
				Tanypodi	0.35				

The macroinvertebrate taxa comprising the SSDs for the classes for which we had sufficient data, and their XC95 values, are presented in Table H-3. The SSDs that result from the rankings in Table H-3 are presented in Figure H-2. As was the case for turbidity, clear differences in the extirpation thresholds—hence the potential NOF management bands from this analysis—can also be seen in Figure H-2. The extirpation thresholds themselves are presented in Table H-4; these thresholds are those estimated from data subdivided nationally according to the Level 3 sediment state classification.

There were insufficient data for BEA within classes L3.1, L3.3 and L3.7. However, we assigned extirpation thresholds to those classes with insufficient data, following the method outlined under the Turbidity results.

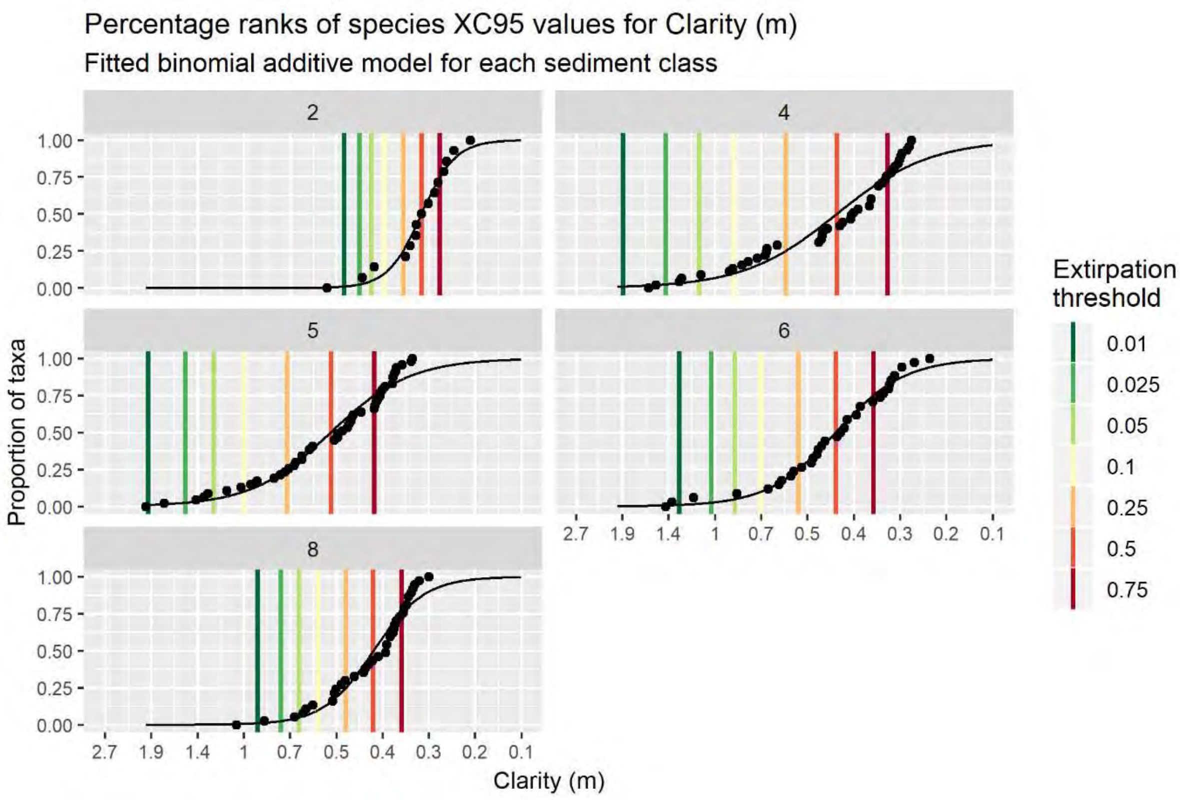


Figure H-2: Proportion of taxa extirpated as a function of annual median clarity within each of five suspended sediment classes. Classes are at Level 3 of the suspended sediment classification

Table H-4: Threshold annual median clarity values (m) at which 1%, 2.5%,…,75% of macroinvertebrate taxa are extirpated from the community within suspended sediment classes. Classes at Level 3 of the suspended sediment classification; sufficient data for four of eight classes; classes with insufficient data indicated by an asterisk, and thresholds assigned to those based on method in text.

Potential NOF band threshold		A/B	B/C	C/D				
Sed. Class		1%	2.5%	5%	10%	25%	50%	75%
L3.1*	L4.1	1.32	1.06	0.90	0.75	0.58	0.45	0.35
L3.2	L4.5 L4.8	0.58	0.50	0.45	0.40	0.34	0.29	0.25
L3.3*	L4.6	0.58	0.50	0.45	0.40	0.34	0.29	0.25
L3.4	L4.7 L4.10 L4.11	1.97	1.47	1.16	0.91	0.64	0.45	0.32
L3.5	L4.9	1.98	1.55	1.27	1.04	0.78	0.58	0.43
L3.6	L4.12	1.32	1.06	0.90	0.75	0.58	0.45	0.35
L3.7*	L7.2	1.32	1.06	0.90	0.75	0.58	0.45	0.35
L3.8	L4.3 L4.4	0.92	0.79	0.71	0.63	0.53	0.44	0.37
	Global mean	1.25	1.00	0.84	0.70	0.55	0.43	0.33

Deposited fine sediment

Of the 206 taxa, 17 satisfied filter conditions 1–6, and none satisfied conditions 7 and 8. Therefore, there were insufficient data for BEA as a function of deposited fine sediment, as assessed using the SAM2 instream visual assessment of fine sediment cover. If the NRWQN data included estimation of deposited fine sediment using the SAM2 instream visual assessment then sufficient data would have been available. Although utilisation of the coarse estimates of deposited fine sediment cover from the NZFFD may be approporiate for animals like fishes that respond to environmental change at relatively coarse scales, use of such data to characterise the response of invertebrates to deposited fine sediment was deemed inapproporiate—invertebrate community spatial patterns respond to much finer scales of change in substrate composition than fishes. There is an urgent need to improve the quality and spatial coverage of monitoring of deposited fine sediment throughout New Zealand.

Appendix I Generalised linear modelling of macroinvertebrate metrics

Introduction

Generalised linear modelling is a form of regression analysis that can be used for estimating relationships among variables. We used generalised linear models (GLMs) to characterise the relationship between several macroinvertebrate metrics and the three sediment ESVs of % cover of deposited fine sediment, turbidity and visual clarity. In contrast to the quantile regression approach, GLMs are used to characterise the 'average' response and typically incorporate further predictor variables to account for the potential influence of factors other than the stressor of interest on the response variable.

Methods

Three macroinvertebrate metrics were used as potential ecological indicators to identify thresholds for each sediment ESV (visual clarity, turbidity and total fines) using this method. The macroinvertebrate metrics were hard-bottom MCI (MCI_hb), a sediment specific macroinvertebrate community index (sed_MCI), and "the proportion of sediment decreases" (Decreaser_abundance) (see 0 for metric description). We applied the same generic analysis method to all combinations of macroinvertebrate metric and ESV.

Let us denote the metric of interest as y, and the sediment variable of interest as ESV. There were several steps to the method applied:

1. Use available observed field data to fit a regression model relating y to each ESV separately. This model should account for landscape-scale variations in climate, physical habitat, and river size.
2. Fit a generalised linear model (glm) of the form y ~ f(ESV, climate&topography, network_position). This glm describes the best-fit relationship between y and ESV for each landscape setting (combination of climate, topography and network position, e.g., small warm-wet-lowland sites). Use the appropriate family within this glm to define the error structure depending on the metric of interest (Table I-1). Apply appropriate weightings to each observation to account for pseudoreplication associated with uneven numbers of observations between sites.
3. Use a model reduction approach to test for the legitimacy of an interaction between ESV and climate&topography by comparing AIC values. Determine if the most parsimonious model accounts for different y~ESV relationships between landscape settings. Note that in landscape settings with few data or a narrow range of observed conditions, inclusion of this interaction may result in non-intuitive predicted y~ESV patterns (where the ESV should decline with y, but is predicted to increase).
4. Quantify quality of model fit using r-squared of observed versus predicted data.
5. For each landscape setting, use the model to quantify the best-fit values of y over the observed range of the ESV (Table I-2). For each landscape setting denote the lowest y value predicted to occur at the observed dirtiest condition (i.e., highest turbidity or lowest visual clarity) as y_{min}.

6. Inspect the look-up between each climate&topography and the appropriate ESV SSC (Appendix D) to find the predicted ESV reference condition for each landscape setting.

7. Obtain the y value corresponding to the ESV reference condition. Denote this value y_{ref}.

8. Obtain the y value corresponding to the ESV worst possible condition (Table I-2). Denote this value y_{worst}. These values represent a hypothetical worst-case macroinvertebrate community comprising 100% Oligochaeta, Chironomus, Psychodidae or Syrphidae.

9. Calculate percent deviations from y_{ref} between y_{ref} and y_{worst} of 6.67%, 13.3%, 20%. These deviations are used to represent, A/B, B/C and C/D thresholds respectively. Denote these y values as y_{AB}, y_{BC}, and y_{CD} respectively. For example, if y_{ref} = 160 and y_{worst} = 20, then y_{CD} = 160 – 0.2(160-20) = 132.

10. Use the best-fit relationship for each landscape setting between y and observed range of ESV to identify ESV values corresponding to y_{AB}, y_{BC}, and y_{CD}. Denote these values ESV_{AB}, ESV_{BC}, and ESV_{CD}. If the required y value is less than y_{min} then the required ESV value is set to the observed dirtiest condition for that ESV (Table I-2).

11. Amalgamate results across different landscape settings to provide single y_{AB}, y_{BC}, and y_{CD} values for each class of the appropriate ESV classification by weighting landscape setting thresholds by the proportions of reaches (defined as RECv2 nzsegments) with that landscape setting contained within each class.

Prior to fitting glms we removed data in the warm-dry-lakefed landscape setting because this setting only contained four observed values.

Table I-1: Macroinvertebrate community metrics and their treatment in the method.

Metric of interest	Symbol	Distribution applied in glm	Worst possible condition
Hard-bottom MCI	MCI_hb	Gamma as appropriate for none-zero right-skewed data.	20
Sediment MCI	sed_MCI	Gamma as appropriate for none-zero right-skewed data.	25
Proportion of sediment decreasers	Decreaser_ abundance	Binomial as appropriate for proportion data.	0

Table I-2: ESV metrics and their observed ranges in the available dataset.

ESV of interest	Observed cleanest condition	Observed dirtiest condition
Turbidity (NTU)	0.1	562
Clarity (m)	31.6	0.018
Total fines (% cover)	0	100

We proposed the following hypotheses prior to viewing the results:

1. The benthic dwelling nature of macroinvertebrate communities would mean that macroinvertebrate metrics would respond more strongly to a deposited fine sediment indicator (total fines) than suspended sediment indicators (turbidity and visual clarity) (H1).

2. Sed_MCI, MCI_hb and Decreaser_abundance should be highest under low total fines, low turbidity and high visual clarity (H2).

3. Invert metrics that respond less strongly to the ESV would yield less environmentally conservative thresholds (H3).

Results

Inclusion of an interaction term between ESV and climate&topography reduced AIC for sed_MCI and MCI_hb, but not Decreaser_abundance regardless of ESV. We therefore, included an interaction term between ESV and climate&topography for all models of sed_MCI and MCI_hb, but a similar interaction was not included for any model of Decreaser_abundance.

Inclusion of each of the three ESVs was able to increase variation explained within the observed data across all three macroinvertebrate metrics in comparison with models that did not include any ESV (Figure I-1). Variation explained for all three macroinvertebrate metrics was greatest for models that included total fines rather than either turbidity or visual clarity. This finding matched well with our hypothesis that macroinvertebrate would respond more strongly to deposited in comparison with suspended sediment (H1). We do note that these results were indicative since datasets were not exactly balanced between various macroinvertebrate metrics-ESV pairs.

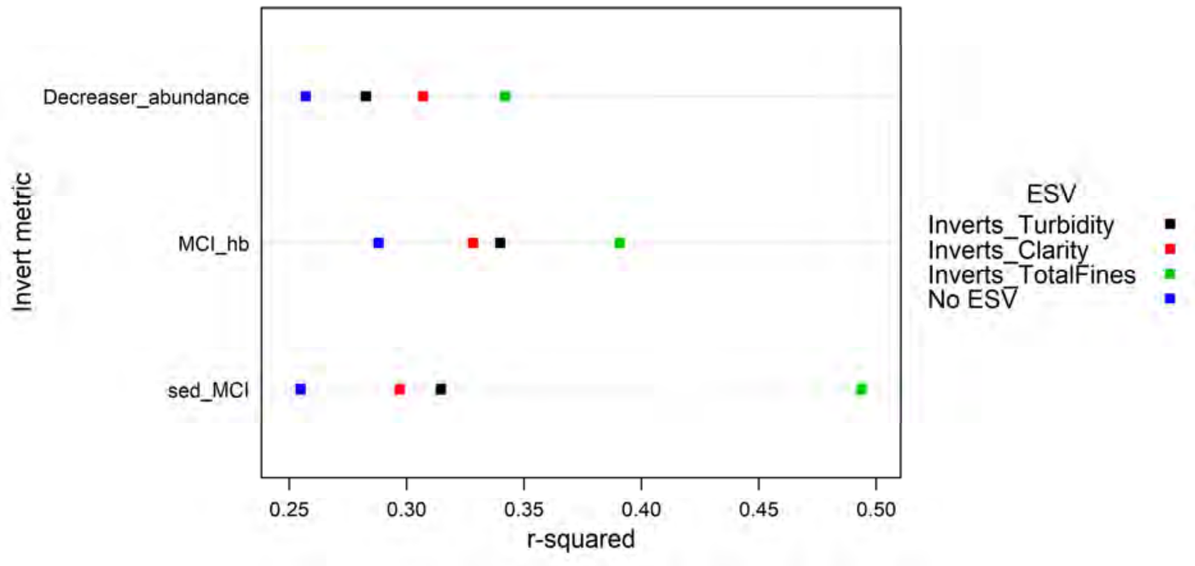


Figure I-1: Variation explained in observed available data for each invert metric and each ESV. "No ESV" represents a reduced model with only landscape setting as predictors.

Patterns predicted by the glm models indicated generally conformed to expected relationships between macroinvertebrate metric and ESV with higher sed_MCI, MCI_hb and Decreaser_abundance in “cleaner” conditions (Figure I-2 to Figure I-4). This supported our hypothesised relationships (H2). However, the predicted relationship was very flat in some landscape settings (e.g., warm-wet-lowland for MCI_hb~total fines; Figure I-2) and did not conform to expected relationships in some landscape settings (e.g., warm-wet-lowland for sed_MCI~total fines; Figure I-3). Scatter of the observed data around the best-fit relationships varied between macroinvertebrate metrics and between landscape settings. This scatter was indicative of measurement uncertainty but also an unbalanced number of observations between landscape settings.

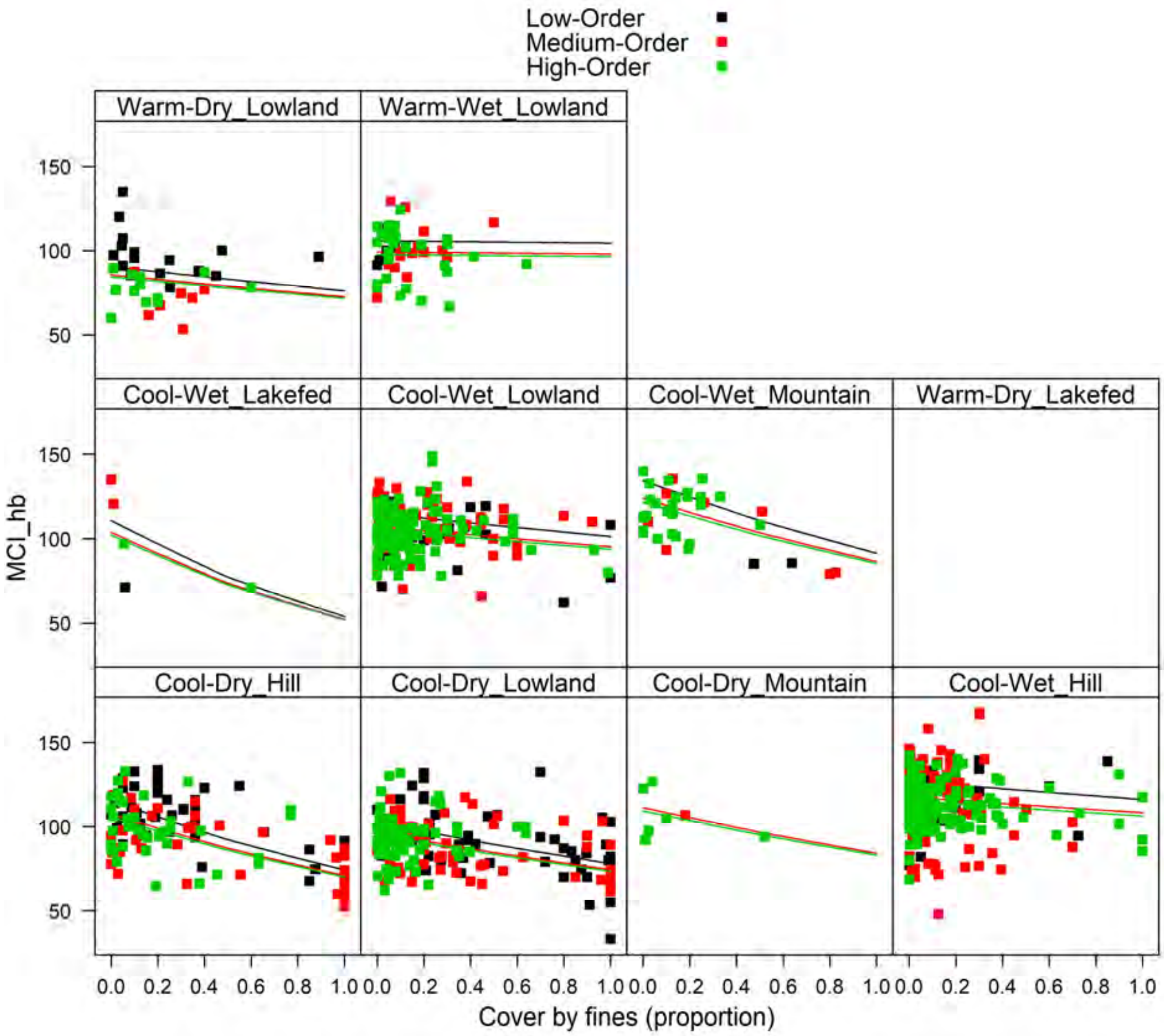


Figure I-2: Available data and glm predicted best-fit patterns for hard-bottom MCI as a function of proportional cover of total fines.

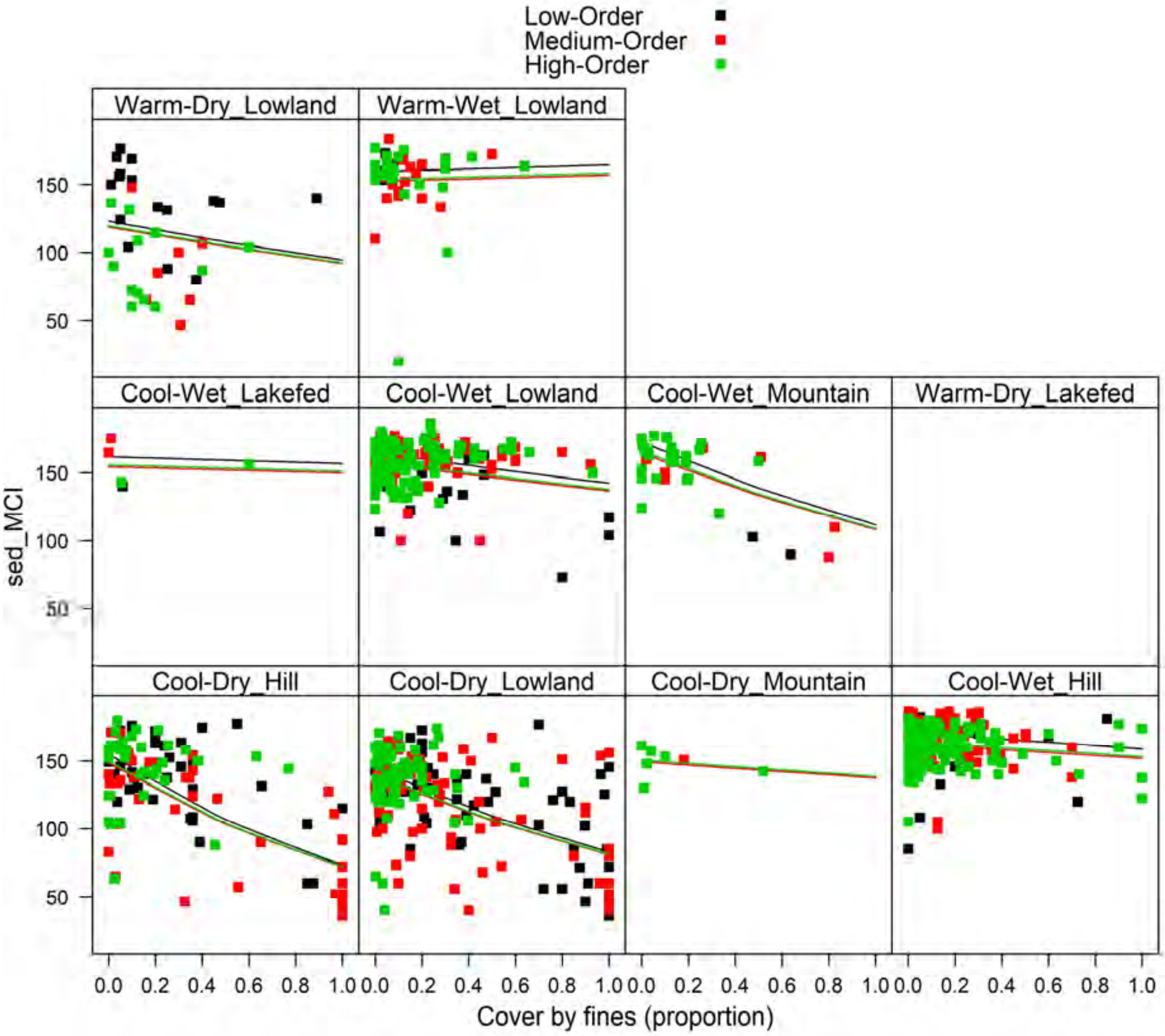


Figure I-3: Available data and glm predicted best-fit patterns for Sediment MCI as a function of proportional cover of total fines.

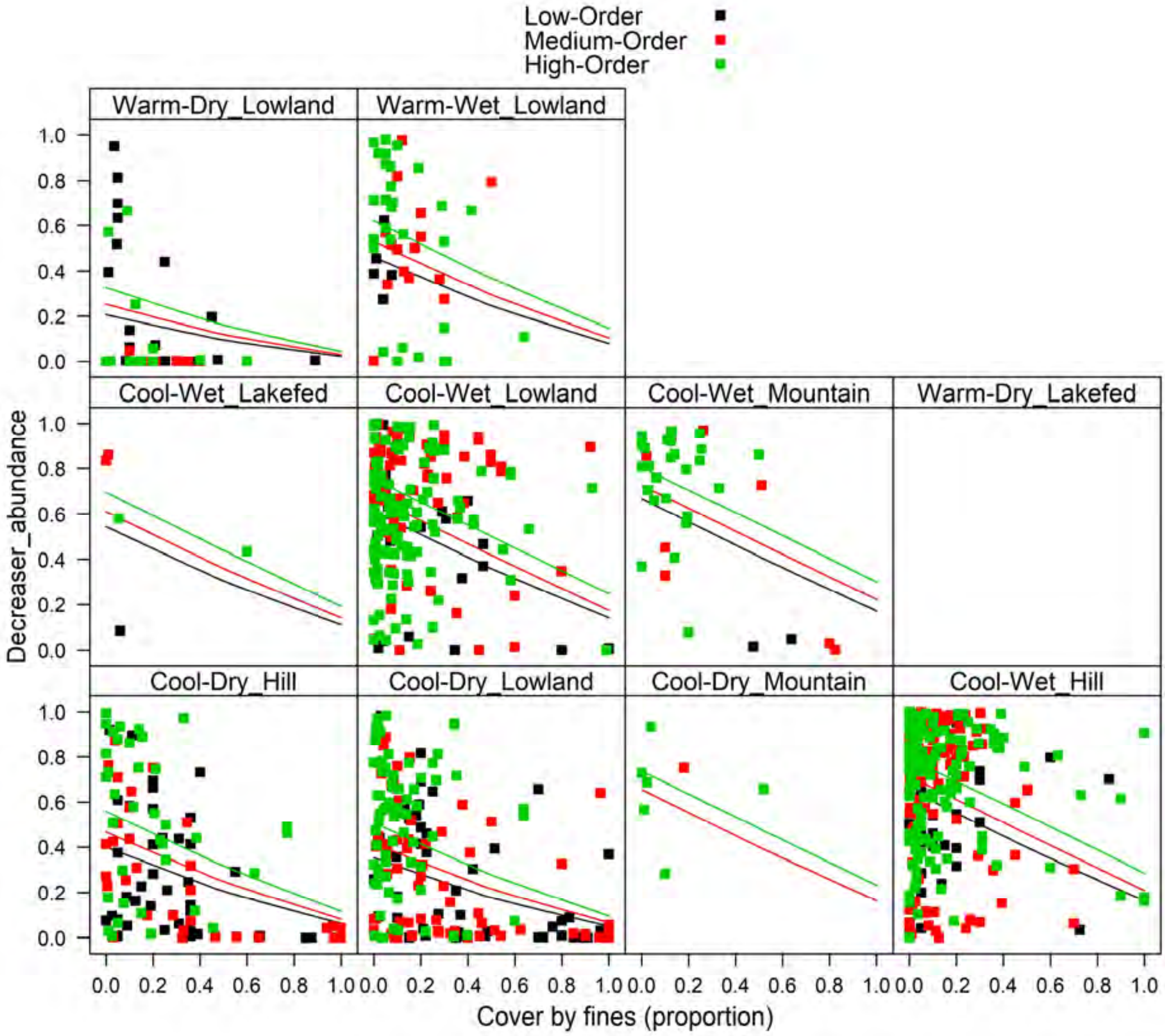


Figure I-4: Available data and glm predicted best-fit patterns for sediment decreasers as a function of proportional cover of total fines.

Completion of the method yielded recommended values for each macroinvertebrate metric-ESV combination. Recommended thresholds for visual clarity and turbidity were not strongly environmentally conservative. This corresponded well with our hypothesis that macroinvertebrate metrics that respond less strongly to the ESV would yield less environmentally conservative thresholds (H3). Some calculated thresholds fell outside the bounds of the observed data. This is reflected in the values shown in Figure I-5 that correspond those in Table I-3, Table I-4 and Table I-5 for deposited sediment, turbidity and visual clarity respectively. The final proposed attribute thresholds are provided in Section 6.

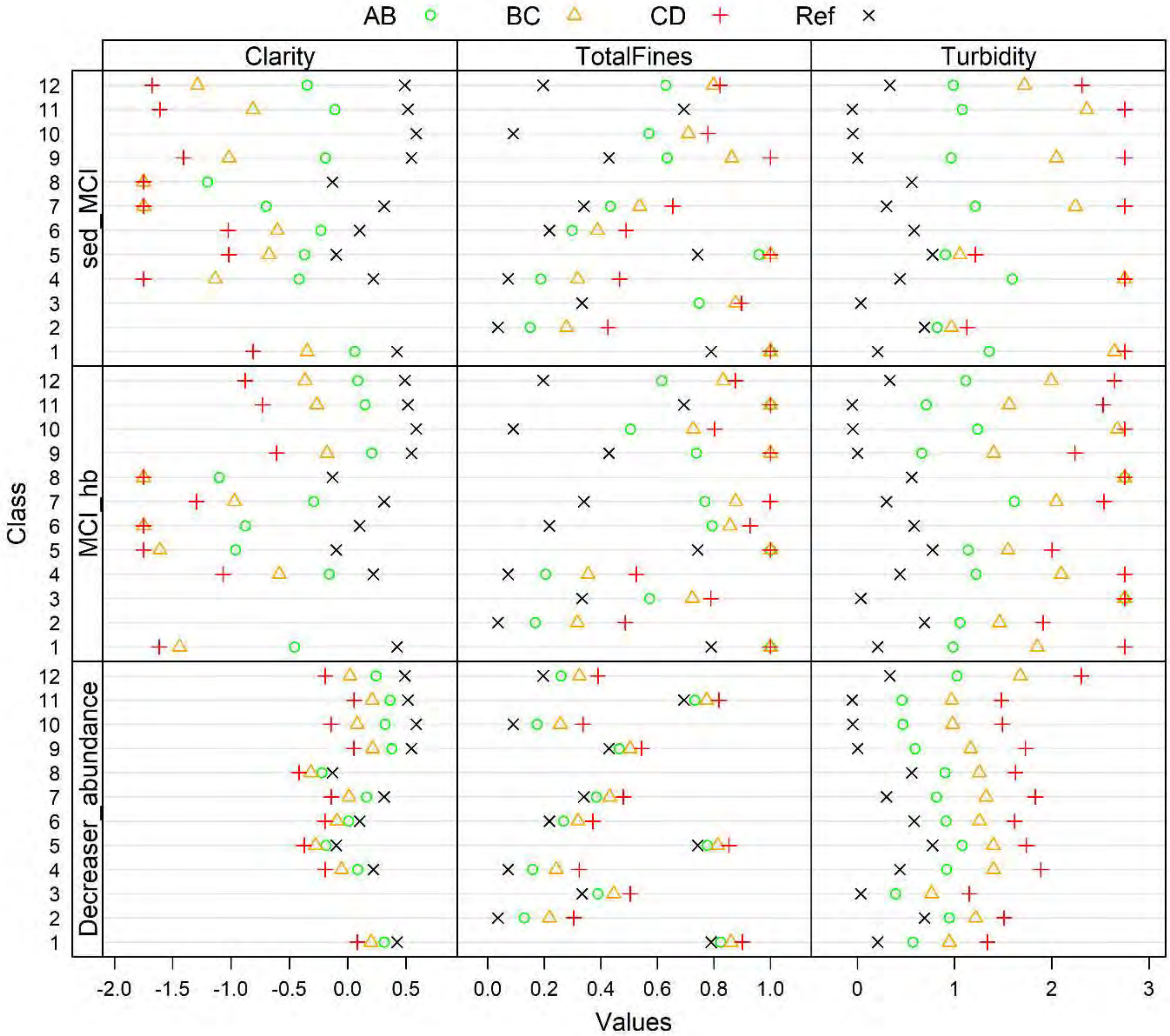


Figure I-5: Potential thresholds calculated from macroinvertebrate metrics for each ESV for the 12 class SSC. Note that in several cases the CD thresholds for the two MCI metrics are at the upper boundary of observed data.

Table I-3: Summary of potential ESV thresholds for total fines (% cover) for the 12 class SSC.

Metric	Class	Predicted reference	A/B threshold	B/C threshold	C/D threshold
MCI	L4.1	79%	100%	100%	100%
MCI	L4.2	4%	17%	32%	49%
MCI	L4.3	33%	57%	72%	79%
MCI	L4.4	7%	21%	36%	53%
MCI	L4.5	74%	100%	100%	100%
MCI	L4.6	22%	79%	86%	93%
MCI	L4.7	34%	77%	88%	100%
MCI	L4.8	13%	NA	NA	NA
MCI	L4.9	43%	74%	100%	100%
MCI	L4.10	9%	50%	73%	80%
MCI	L4.11	69%	100%	100%	100%
MCI	L4.12	20%	62%	83%	88%
Sediment MCI	L4.1	79%	82%	86%	90%
Sediment MCI	L4.2	4%	13%	22%	31%
Sediment MCI	L4.3	33%	39%	45%	50%
Sediment MCI	L4.4	7%	16%	24%	32%
Sediment MCI	L4.5	74%	78%	81%	85%
Sediment MCI	L4.6	22%	27%	32%	37%
Sediment MCI	L4.7	34%	39%	43%	48%
Sediment MCI	L4.8	13%	NA	NA	NA
Sediment MCI	L4.9	43%	47%	50%	55%
Sediment MCI	L4.10	9%	18%	26%	34%
Sediment MCI	L4.11	69%	73%	77%	82%
Sediment MCI	L4.12	20%	26%	32%	39%
Sediment decreasers	L4.1	79%	100%	100%	100%
Sediment decreasers	L4.2	4%	15%	28%	43%
Sediment decreasers	L4.3	33%	75%	88%	90%
Sediment decreasers	L4.4	7%	19%	32%	47%
Sediment decreasers	L4.5	74%	96%	100%	100%
Sediment decreasers	L4.6	22%	30%	39%	49%
Sediment decreasers	L4.7	34%	43%	54%	65%
Sediment decreasers	L4.8	13%	NA	NA	NA
Sediment decreasers	L4.9	43%	63%	86%	100%
Sediment decreasers	L4.10	9%	57%	71%	78%
Sediment decreasers	L4.11	69%	NA	NA	NA
Sediment decreasers	L4.12	20%	63%	80%	82%

Table I-4: Summary of potential ESV thresholds for turbidity (NTU) for the 12 class SSC.

Metric	Class	Predicted reference	A/B threshold	B/C threshold	C/D threshold
MCI	L4.1	1.6	9.6	70.8	562.3
MCI	L4.2	4.9	11.4	28.9	81.6
MCI	L4.3	1.1	562.3	562.3	562.3
MCI	L4.4	2.7	16.6	125.3	562.3
MCI	L4.5	5.9	13.8	35.3	100.5
MCI	L4.6	3.8	NA	NA	NA
MCI	L4.7	2.0	41.1	111.3	343.5
MCI	L4.8	3.6	562.3	562.3	562.3
MCI	L4.9	1.0	4.6	25.2	173.1
MCI	L4.10	0.9	17.2	472.5	562.3
MCI	L4.11	0.9	5.1	36.4	337.3
MCI	L4.12	2.2	13.0	98.6	442.9
Sediment MCI	L4.1	1.6	3.7	8.8	21.7
Sediment MCI	L4.2	4.9	8.8	16.5	32.4
Sediment MCI	L4.3	1.1	2.5	5.8	14.1
Sediment MCI	L4.4	2.7	8.3	25.1	77.2
Sediment MCI	L4.5	5.9	12.0	25.0	54.6
Sediment MCI	L4.6	3.8	8.2	18.0	41.4
Sediment MCI	L4.7	2.0	6.5	21.1	67.7
Sediment MCI	L4.8	3.6	7.9	18.0	42.4
Sediment MCI	L4.9	1.0	3.9	14.6	53.5
Sediment MCI	L4.10	0.9	2.9	9.5	31.0
Sediment MCI	L4.11	0.9	2.9	9.3	30.2
Sediment MCI	L4.12	2.2	10.6	47.3	201.7
Sediment decreasers	L4.1	1.6	22.6	442.5	562.3
Sediment decreasers	L4.2	4.9	6.6	9.3	13.4
Sediment decreasers	L4.3	1.1	NA	NA	NA
Sediment decreasers	L4.4	2.7	39.3	562.3	562.3
Sediment decreasers	L4.5	5.9	8.1	11.3	16.4
Sediment decreasers	L4.6	3.8	NA	NA	NA
Sediment decreasers	L4.7	2.0	16.3	174.6	562.3
Sediment decreasers	L4.8	3.6	NA	NA	NA
Sediment decreasers	L4.9	1.0	9.2	111.5	562.3
Sediment decreasers	L4.10	0.9	NA	NA	NA
Sediment decreasers	L4.11	0.9	12.0	228.7	562.3
Sediment decreasers	L4.12	2.2	9.7	52.1	204.5

Table I-5: Summary of potential ESV thresholds for visual clarity (m) for the 12 class SSC.

Metric	Class	Predicted reference	A/B threshold	B/C threshold	C/D threshold
MCI	L4.1	2.65	0.35	0.04	0.02
MCI	L4.2	NA	NA	NA	NA
MCI	L4.3	NA	NA	NA	NA
MCI	L4.4	1.66	0.69	0.26	0.09
MCI	L4.5	0.80	0.11	0.03	0.02
MCI	L4.6	1.27	0.13	0.02	0.02
MCI	L4.7	2.05	0.51	0.11	0.05
MCI	L4.8	0.74	0.08	0.02	0.02
MCI	L4.9	3.52	1.61	0.67	0.25
MCI	L4.10	3.86	NA	NA	NA
MCI	L4.11	3.28	1.41	0.54	0.19
MCI	L4.12	3.09	1.22	0.43	0.13
Sediment MCI	L4.1	2.65	2.06	1.58	1.21
Sediment MCI	L4.2	NA	NA	NA	NA
Sediment MCI	L4.3	NA	NA	NA	NA
Sediment MCI	L4.4	1.66	1.21	0.89	0.64
Sediment MCI	L4.5	0.80	0.65	0.53	0.42
Sediment MCI	L4.6	1.27	1.02	0.81	0.64
Sediment MCI	L4.7	2.05	1.45	1.02	0.72
Sediment MCI	L4.8	0.74	0.60	0.48	0.38
Sediment MCI	L4.9	3.52	2.38	1.63	1.12
Sediment MCI	L4.10	3.86	2.09	1.20	0.72
Sediment MCI	L4.11	3.28	2.30	1.62	1.14
Sediment MCI	L4.12	3.09	1.75	1.04	0.64
Sediment decreasers	L4.1	2.65	1.15	0.45	0.15
Sediment decreasers	L4.2	NA	NA	NA	NA
Sediment decreasers	L4.3	NA	NA	NA	NA
Sediment decreasers	L4.4	1.66	0.38	0.07	0.02
Sediment decreasers	L4.5	0.80	0.43	0.21	0.10
Sediment decreasers	L4.6	1.27	0.59	0.25	0.09
Sediment decreasers	L4.7	2.05	0.20	0.02	0.02
Sediment decreasers	L4.8	0.74	0.06	0.02	0.02
Sediment decreasers	L4.9	3.52	0.65	0.10	0.04
Sediment decreasers	L4.10	3.86	NA	NA	NA
Sediment decreasers	L4.11	3.28	0.78	0.15	0.02
Sediment decreasers	L4.12	3.09	0.45	0.05	0.02

Discussion and limitations

In fitting universal models to each y~ESV, we must accept all potential influences on y cannot be accounted for within this model. This is a disadvantage of the method because it does not explicitly account for variables such as temperature or nutrient enrichment. This is also an advantage of the method in that spurious findings will not be produced because of co-variance between explanatory variables within the unbalanced observed dataset (e.g., co-variance between an ESV and temperature might attribute less explanatory power to the ESV simply because the ESV is correlated with temperature).

We set the bottom of band A at 6.67% below reference condition. To some extent this offset below reference accounted for natural variations in ESV state within landscape settings, and the fact that our observed data were one-off observations rather than medians over time. An alternative approach would be to set the bottom of band A to be one standard error below the estimated reference conditions. This method was not applied because it would result in recommended bands that would be a function of data availability rather than ecological effect.

Appendix J Community deviation analyses

Introduction

This section sets out the full details of the technical methods used in the process of characterising fish-sediment ESV and macroinvertebrate-sediments ESV responses using the community deviation method. The objective of this component of the project was to test for, and characterise, relationships between fish and macroinvertebrates and sediment ESVs that could be used to inform the development of a sediment NOF attribute for the protection of ecosystem health.

The main steps involved in this stage of the project were to:

1. determine the availability of suitable datasets
2. model probability of capture as a function of sediment ESVs within landscape settings
3. evaluate community change in response to deviation of ESV state from reference conditions, and
4. derive potential sediment ESV thresholds.

Data availability

Fish

The New Zealand Freshwater Fish Database (NZFFD) contains 42,154 unique observations of fish from across the country. The data in the NZFFD can be extremely useful for the study of fish community changes, but there are some key limitations to using the data effectively that must be accounted for prior to analysis.

In general, the methods applied by Crow et al. (2016) and Crow et al. (2014) were used to extract and organise data from the NZFFD for analysis. This included removing records from prior to 1970, only selecting records with an identified NZ reach number (i.e., the unique reach number from the RECv1), removing reaches that were not from rivers, eliminating records observed from angling or with an unknown fishing method, and collapsing fishing method into six categories.

Records observed prior to 1970 were removed from the analysis as these are generally considered less reliable than more recent records. Furthermore, only records with an identified NZ reach number relating it to the national river network (RECv1) were included in the analysis. This allowed more effective pairing of fish and sediment ESV observations later in the analyses.

NZFFD records can be entered for any location where a fish may be observed. This includes lakes, wetlands, ponds and water raceways. Only records identified as being from rivers were desirable for this analysis. The “locality” field from each record was, therefore, used to remove records containing observations that were not from sites on rivers. This included wetlands, estuaries, tarns, ponds, and water races. In addition, localities associated with lakes, dams, harbours, lagoons, canals, swamps, and reservoirs were removed from the analysis unless they were also associated with tributaries, streams, rivers, creeks, or brooks. For example, records with locality descriptions similar to “reservoir tributary” or “tributary to large lake” were included in the analysis, whereas localities similar to “isolated pond”, “large raceway” or “small wetland” were removed from the analysis.

Following the method of Crow et al. (2014), fishing method was collapsed into six categories (visual, netting, trapping, combinations of methods and electric fishing). Visual methods included daytime observation, diving and spotlighting. Trapping methods included Gee minnow traps, box traps, and bait traps. Netting methods included fyke nets, seine nets and set nets. Electric fishing included backpack and mains set methods. Combinations included combinations of electric fishing and nets, combinations of nets and traps, and combinations of nets, traps and electric fishing. Records observed from angling or unknown fishing methods were removed from the analysis.

Whilst a proportion of NZFFD data records contain data on observed abundances, fish abundance was not used in the analysis for two reasons. First, abundance is strongly related to fishing effort and area fished, which are often not available or imprecisely measured for many records. Also, fishing effort may not be transferable between sites due to differences in physical conditions (size of river, substrate size, presence of vegetation etc.). Second, the locations at which abundances have been observed are biased towards certain catchment and regions of the country. Fish distributions are strongly related to landscape setting such as distance from sea and altitude. These characteristics may also be related to sediment characteristics. Therefore, to best characterise the relationships between fish and sediment, this landscape-scale information must first be accounted for. This is best achieved by utilising fish observations spread across the entire range of catchment conditions. Analyses were, therefore, carried out using presence-absence data from a total of 34,364 NZFFD records remaining after data sorting.

Macroinvertebrates

The macroinvertebrate dataset collated for the BRT analyses was also used for these analyses (Appendix F). Invertebrate taxa within that dataset were identified to the MCI level. Hereafter these taxa will be referred to as species. All observations were converted to presence-absence of individual species for the purposes of this analysis.

Deposited sediment data

Many NZFFD records also contain observations of substrate cover recorded by instream visual estimate over the sampling reach at the time of the survey. The proportional areal cover of fine sediments (mud/silt <1 mm and sand <2 mm categories) was available for 22,946 of the NZFFD records.

Because the NZFFD data had been observed over a long time period (47 years), a brief investigation was undertaken to assess the strength of any temporal trend in the deposited sediment data. A generalised linear model was applied using a binomial distribution as is appropriate for proportion data. Explanatory variables were year of record, network position (a proxy for river size), fishing method, climate class and topography class. Results indicated that, once other variables have been accounted for, there was no significant relationships between year and deposited total fine sediment (Table J-1). All other variables were significant. This indicated that it was legitimate to employ all available NZFFD deposited sediment data in the analyses of fish-deposited sediment stressor-response analyses.

For the macroinvertebrate analyses, the paired observations of SAM2 % cover instream data collated as part of the macroinvertebrate dataset for the BRT analyses were used (Appendix F).

It should be noted that the observations of % cover of total fines from the NZFFD (NZFFD % total fines) were not confined to individual habitat types (e.g., runs). They may, therefore, not be directly comparable with % cover of total fines that have been observed only in particular habitat types (e.g., runs as is the case for SAM2). See Appendix N for further discussion of the relationships between different deposited sediment ESV measures.

Table J-1: Results from a GLM of deposited total fine sediment using data from the NZFFD (n = 22,946).

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-0.261	0.197	-1.325	0.1850
Year	0.001	0.002	0.664	0.5069
NET_POSN Medium-Order	-0.332	0.036	-9.256	0.0000
NET_POSN High-Order	-0.503	0.059	-8.593	0.0000
Fishmethod Electric	-1.377	0.076	-18.158	0.0000
Fishmethod Net	-0.262	0.092	-2.859	0.0043
Fishmethod Trap	0.165	0.096	1.722	0.0850
Fishmethod Unknown	-0.356	0.120	-2.979	0.0029
Fishmethod Visual	-1.194	0.090	-13.228	0.0000
CLIMATE Cool-Wet	-0.642	0.051	-12.712	0.0000
CLIMATE Cool-ExtremelyWet	-1.250	0.059	-21.180	0.0000
CLIMATE Warm-Dry	0.775	0.079	9.866	0.0000
CLIMATE Warm-Wet	0.013	0.055	0.237	0.8129
CLIMATE Warm-ExtremelyWet	-0.601	0.127	-4.737	0.0000
TOPOGRAPHY Hill	0.237	0.074	3.177	0.0015
TOPOGRAPHY Lowland	0.714	0.077	9.244	0.0000
TOPOGRAPHY Lakefed	1.311	0.138	9.501	0.0000
GEOLOGY SS	0.602	0.048	12.521	0.0000
GEOLOGY Al	0.771	0.053	14.548	0.0000
GEOLOGY VA	0.419	0.051	8.171	0.0000

Suspended sediment data

Visual clarity and turbidity are not routinely collected from the same places as fish are sampled. For the purposes of this project predicted median visual clarity or turbidity for each NZFFD observation were in-filled using recently developed statistical models trained using available state-of-the-environment monitoring data. These statistical models were random forest models with a suit of landscape-scale predictors. The model took the same training data and predictors as that of Whitehead (2019), but also included sediment yield estimated by Hicks et al. (2019) as a predictor. One outcome of the inclusion of sediment yield as a predictor was a decrease in predicted visual clarity and an increase in predicted turbidity for some rivers located in the Southern Alps and the West Coast of the South Island.

Paired median visual clarity and turbidity data were available for many of the macroinvertebrate observations collated for the BRT analyses and so were used for these analyses.

Environmental data

The NZ reach identified for each NZFFD record was used to obtain various landscape and reach-scale information from databases that have previously been linked to the national river network. Available catchment characteristics included a range of categorical and continuous variables including a hierarchical classification of New Zealand's rivers called the River Environment Classification (REC;Snelder,Biggs (2002)). Snelder et al. (2005) showed that grouping river segments by nested categorical subdivisions provided an a priori hydrological regionalisation at various levels of detail and spatial resolution. The first three levels of this hierarchical classification are: 1) climate categories; 2) the joining of climate and topography categories; and 3) the joining of climate, topography and geology categories. These are known as the first, second and third levels of the REC classification system. The second level is also referred to as the Source-of-Flow grouping factor.

Information on proportions of landcover in the catchment upstream of each observation were obtained from LCDB3. Several LCDB3 categories were lumped together to calculate the proportion of the upstream catchment that could be described as heavy pasture, exotic vegetation and urban landcover (see Depree et al. (2018)).

Matching fish data with observed ESV data

To evaluate sediment ESV – fish responses it was necessary to try and pair sediment ESV observations with fish observations by matching them spatially and temporally. Spatial matches were evaluated using NZ reach numbers associated with both the NZFFD records and the independent sediment ESV observations.

Several spatial matches between independent ESV observations and NZFFD records on the same reach, but on different dates, were found (SAM2 % cover instream = 260, SAM1 % cover bankside = 440, turbidity = 133, TSS = 143, RHA100 = 283, visual clarity = 158). However, the duration between fish observations and independent ESV observations at the same site were frequently long (>5 years apart), and in many cases fish observations were recorded many years before the ESV observations (Figure J-1). Only three of the paired sediment and fish observations also coincided by sampling date.

To increase the number of spatial matches, and hence increase the probability of obtaining combined spatial and temporal matches, the rules for spatial matching were relaxed. Upstream-downstream searches were conducted to match any independent ESV observations with any NZFFD records that were not on the same NZ reach, but were located within the same catchment. Many matches spatial were found, especially in larger catchments, but few were in adjacent or nearby reaches (Figure J-2), and there were very few time-series of paired observations in the same catchment.

In summary, there was a lack of combined spatial and temporal matches where fish observations were made in the same location on the same dates as independent sediment ESV observations making this approach unsuitable for this project. In the case of deposited sediment, it was decided to use the deposited sediment data (NZFFD % total fines) associated with the NZFFD records. To advance the analyses for suspended sediment, modelled median visual clarity and turbidity derived were used as substitutes for observed sediment ESV data. These modelled values are available for all locations on the NZ river network allowing pairing with all 34,364 NZFFD records. Because the summary statistic of the suspended sediment ESVs was the long-term median, results would be compatible with existing state of the environment monitoring strategies for these variables.

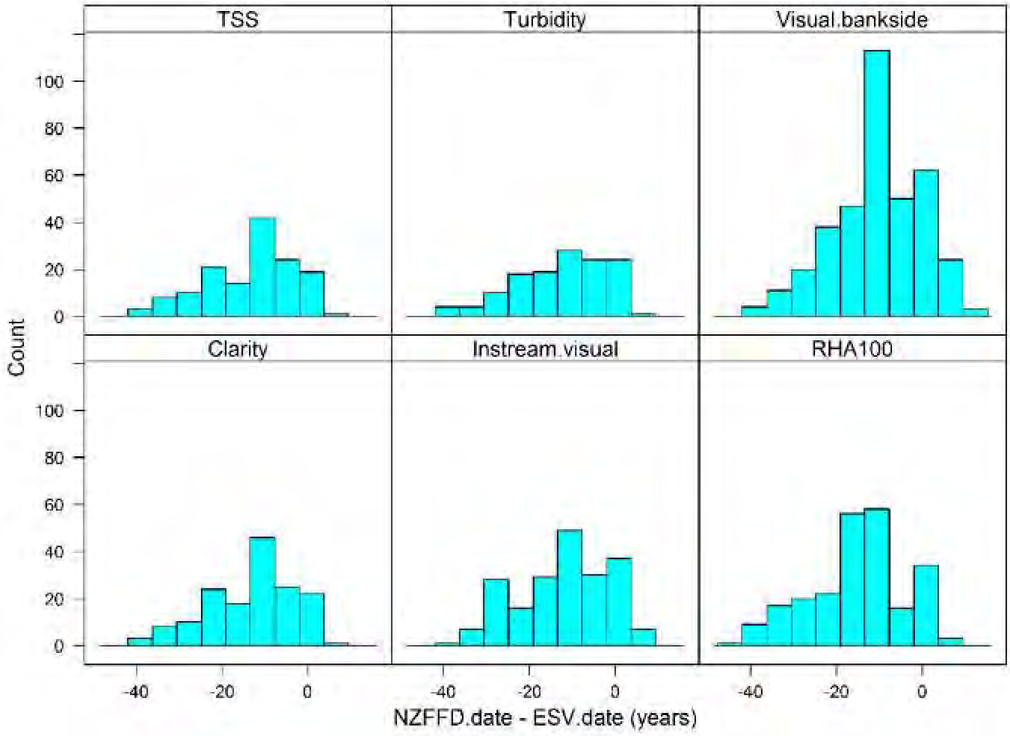


Figure J-1: Summary of temporal separation between spatially paired fish and sediment ESV observations. Negative numbers mean that the fish observation was prior to the sediment observation. Positive numbers mean that the sediment observation was prior to the fish observation. TSS, turbidity and visual clarity data refer to suspended sediment dataset from Unwin,Larned (2013). Visual bankside, instream visual and RHA100 refer to the sediment metrics in Clapcott et al. (2011) for deposited sediment.

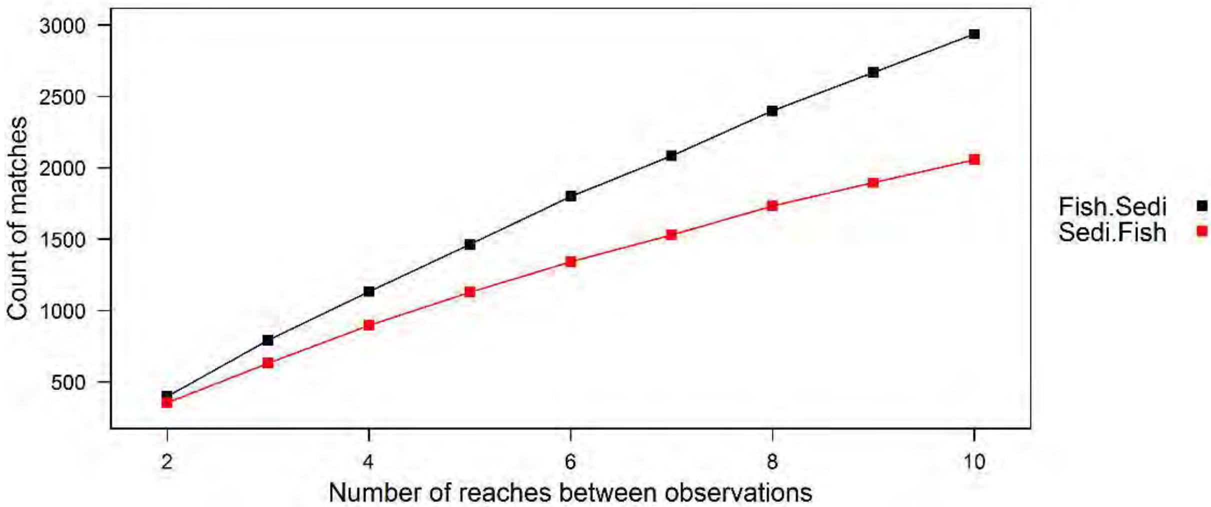


Figure J-2: Count of spatial matches between NZFFD records located upstream (Fish.Sedi) or downstream (Sedi.Fish) of an independent sediment ESV observation.

Probability of capture as a function of sediment ESV within landscape settings

The community deviation method was applied separately to the fish community and to the macroinvertebrate community for each of visual clarity, turbidity, and deposited fine sediment. The fish species included in the analyses comprised nine native, non-geographically restricted species, plus brown trout, that were present in at least 5% of sites. Brown trout were included because they are a highly valued species who are known to respond to the ESVs and because their habitat is protected under the RMA. Freshwater crayfish (kōura) were also included in the analysis at the request of MfE because of their biodiversity value and due to the possibility that this species may show a response to the ESVs. Presence and absence of each species was obtained for each record (a set of observations from the same location and date) within the NZFFD (Figure J-3). The macroinvertebrate community comprised 25 of 31 species (taxon identified at the MCI-level) present at more than 5% of observed sites (Figure J-4 to Figure J-6).

The community deviation method required that probability of capture of each species be related to the ESV of interest in addition to a set of other variables representing influential landscape-scale conditions; climate, hydrology and physical conditions. Each taxon was, therefore, modelled as a function of an ESV (visual clarity, turbidity or total fines) and a set of other predictors:

- Classes of the second level of the REC as defined by the joining of Climate and Topography categories (e.g., warm-dry-lowland or cool-wet-mountain). These classes were included to represent spatial variations in hydrological conditions and physical habitat known to influence fish and inverts.
- Network position (large, medium or small rivers defined using stream order). These classes were included to represent variations in physical habitat and are also related to dispersal of individuals across river networks.
- Distance to the sea (Log to the base 10 transformed) for fish (but not for invertebrates) because it is an important factor determining the distribution of species that migrate to or from the sea for some part of their live history. This is a common species trait amongst New Zealand's native fish species.
- Fishing method for fish (but not for invertebrates) because it is a factor influencing probability of capture (Figure J-11).
- Competition between salmonids and each of inanga, banded kōkopu and kōaro.

For each invert and fish taxon, an interaction between the ESV and REC climate class plus topography class was tested. The interaction term did not provide systematic improvement in predictive performance (defined by Area Under Curve; AUC) and did not improve model parsimony (AIC) across all analyses (Figure J-7 to Figure J-10). The interaction term was, therefore, not included in the final models.

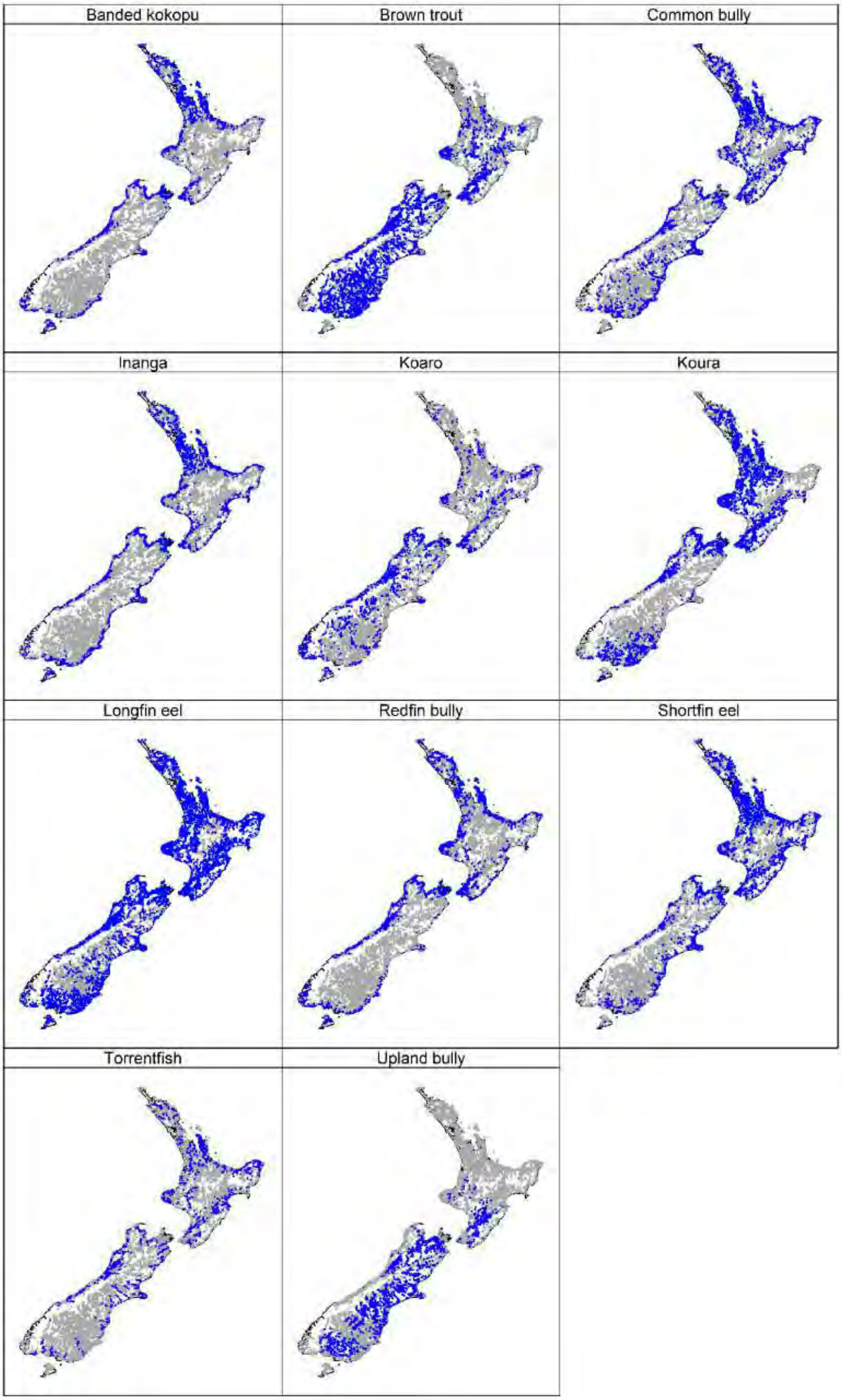


Figure J-3: Maps of presence (blue) and absence (grey) in the NZFFD records for the eleven species used in these analyses.

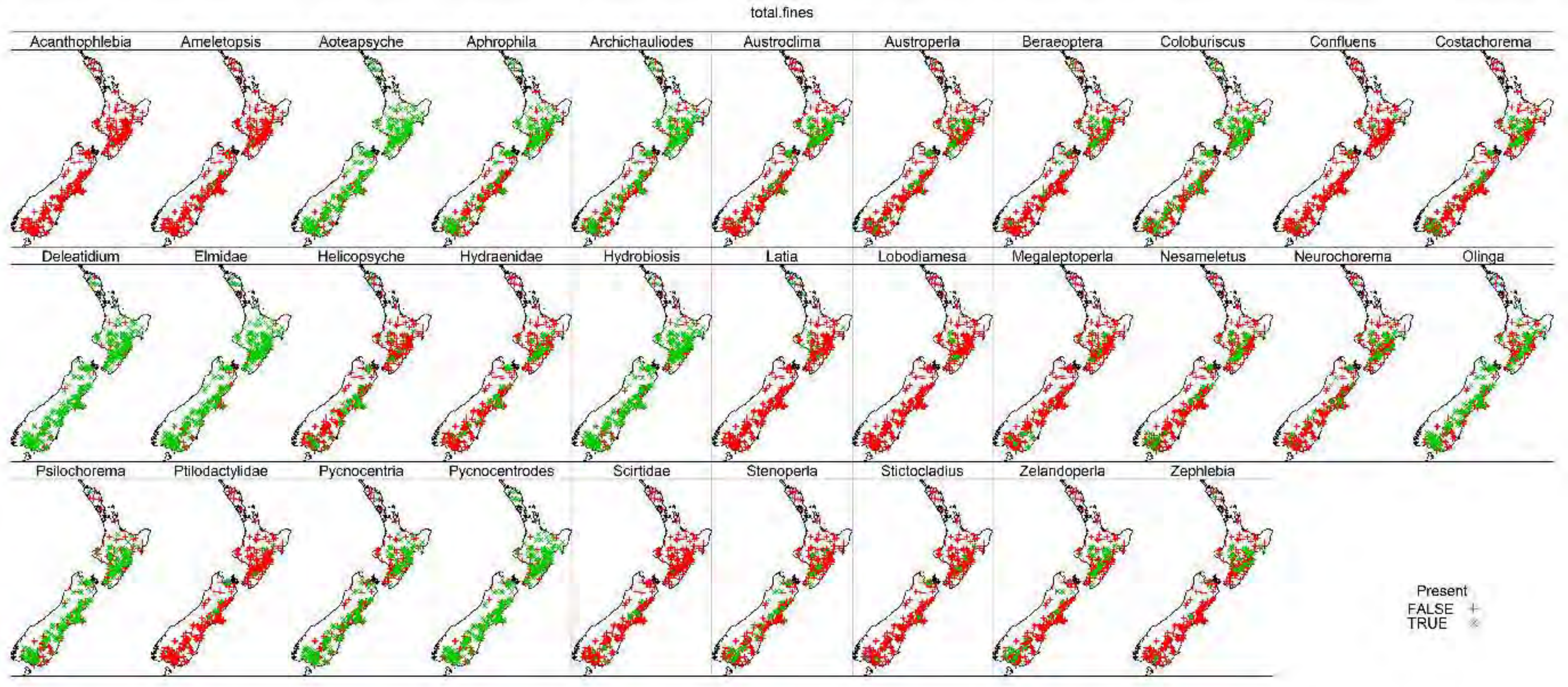


Figure J-4: Distribution maps for the macroinvertebrate species included in the analyses for the deposited sediment ESV. Data are from SOE monitoring sites. Green indicates the species was present at the survey site at the time of the survey. Red indicates the species was not captured at the survey site at the time of the survey.

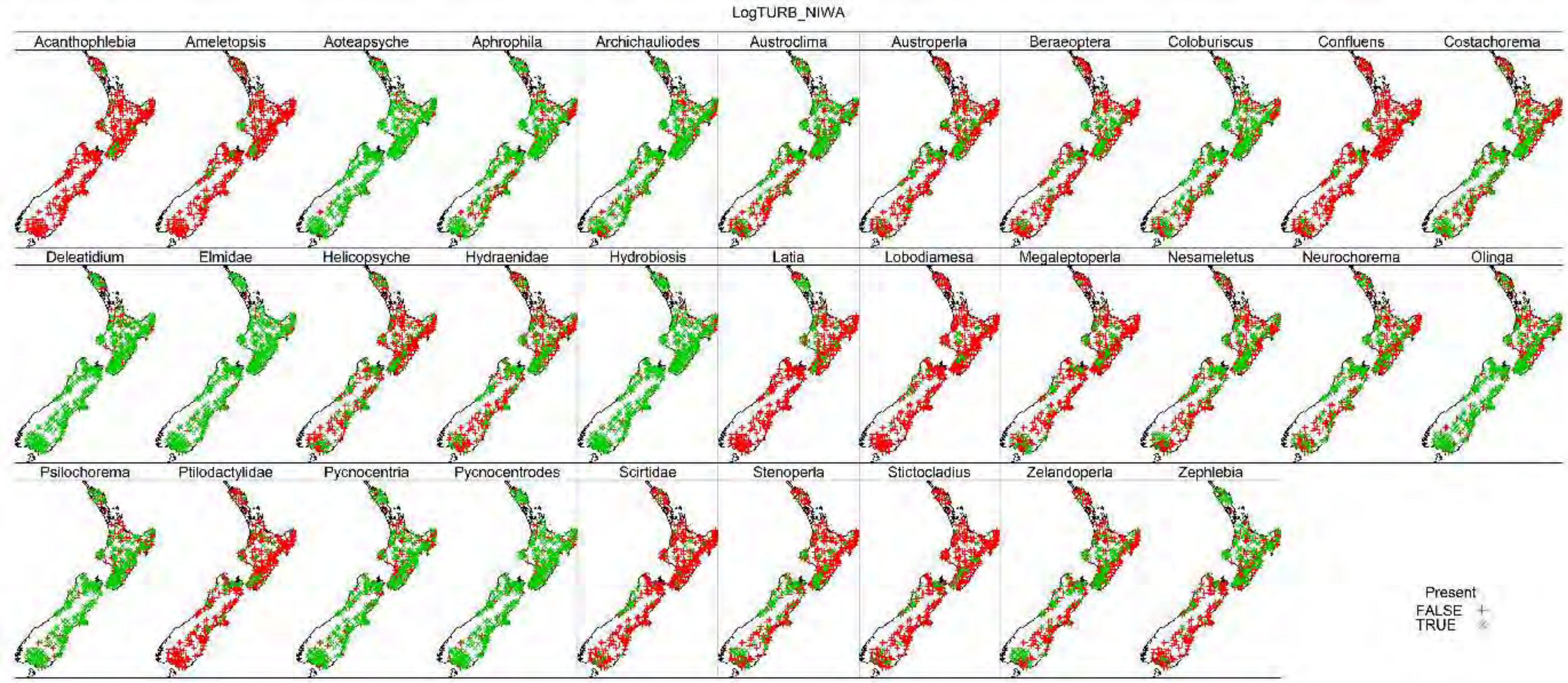


Figure J-5: Distribution maps for the macroinvertebrate species included in the analyses for the turbidity ESV. Data are from SOE monitoring sites. Green indicates the species was present at the survey site at the time of the survey. Red indicates the species was not captured at the survey site at the time of the survey.

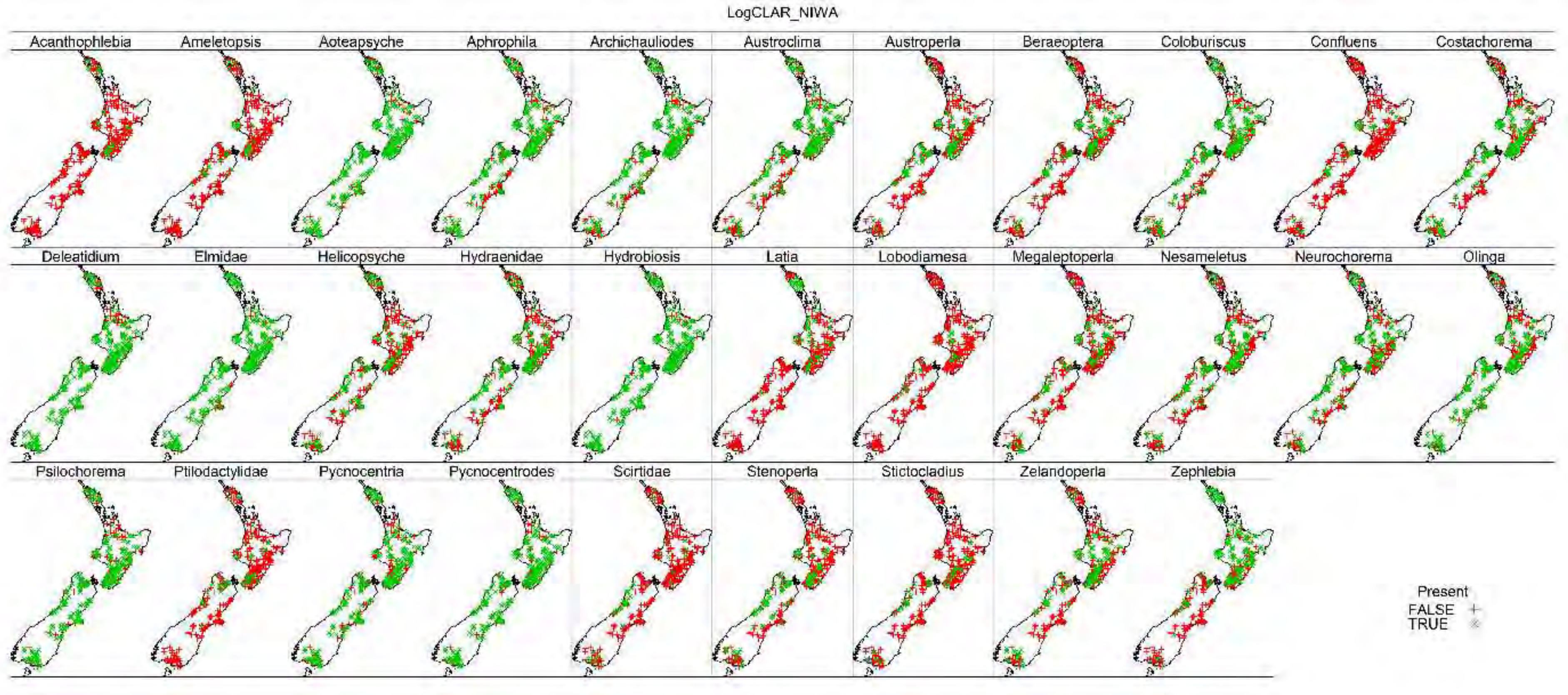


Figure J-6: Distribution maps for the macroinvertebrate species included in the analyses for the visual clarity ESV. Data are from SOE monitoring sites. Green indicates the species was present at the survey site at the time of the survey. Red indicates the species was not captured at the survey site at the time of the survey.

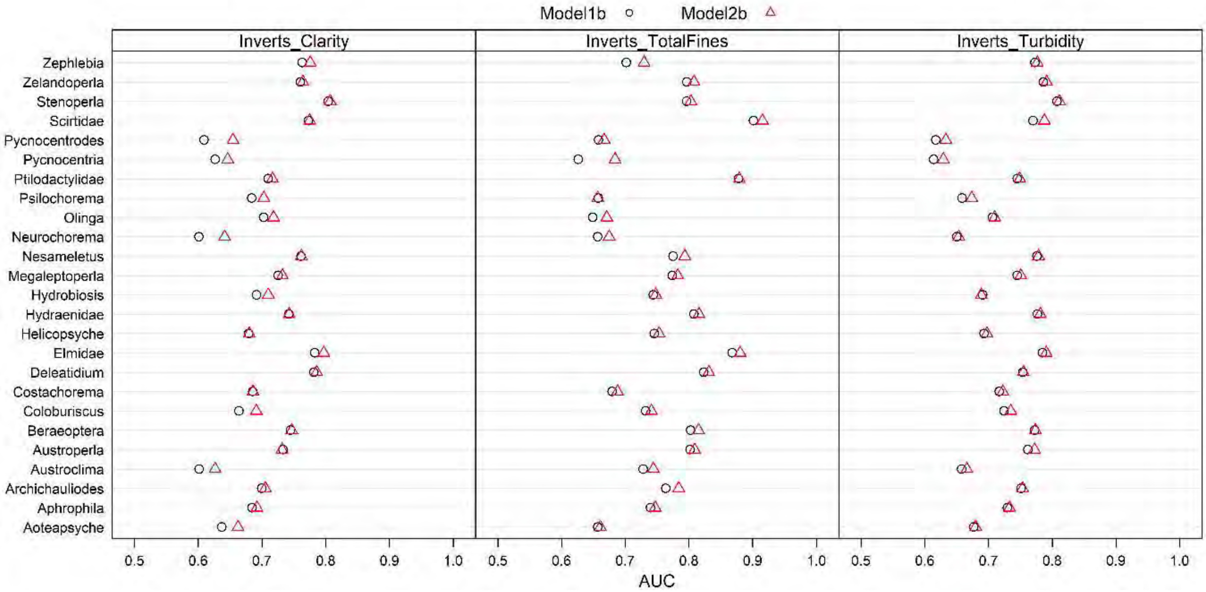


Figure J-7: Area under curve (AUC) results for glm models of invertebrate taxa against each ESV. Model 1b has no interaction between the ESV and climate plus topography class and Model 2b includes an interaction between the ESV and climate plus topography class.

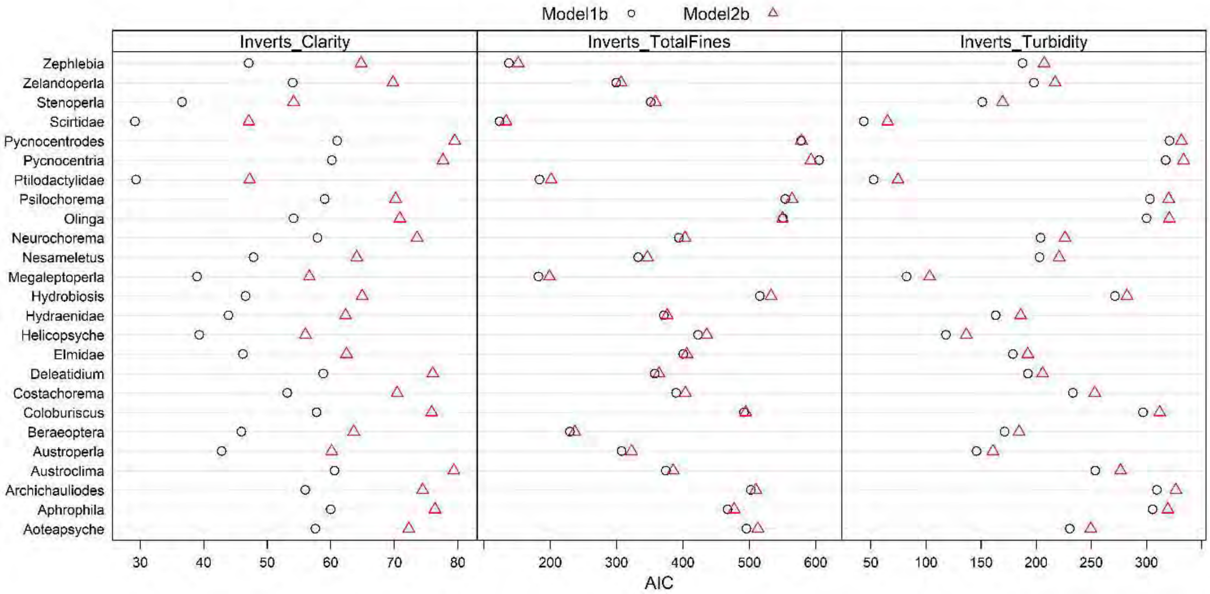


Figure J-8: Akaike Information Criteria (AIC) results for glm models of invertebrate taxa against each ESV. Model 1b has no interaction between the ESV and climate plus topography class and Model 2b includes an interaction between the ESV and climate plus topography class.

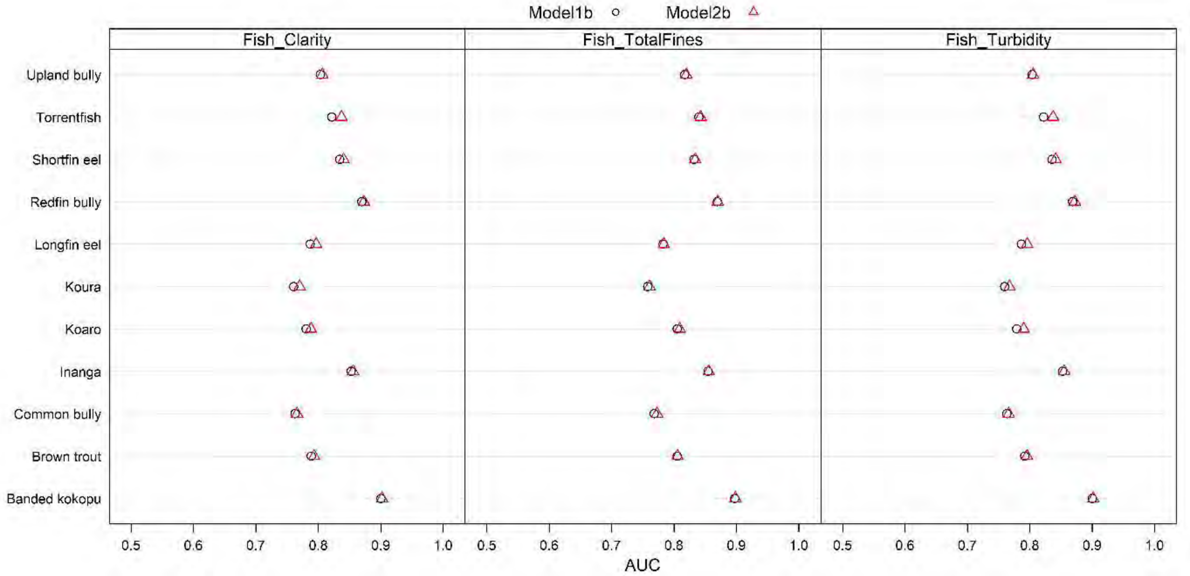


Figure J-9: Area under curve (AUC) results for glm models of fish species against each ESV. Model 1b has no interaction between the ESV and climate plus topography class and Model 2b includes an interaction between the ESV and climate plus topography class.

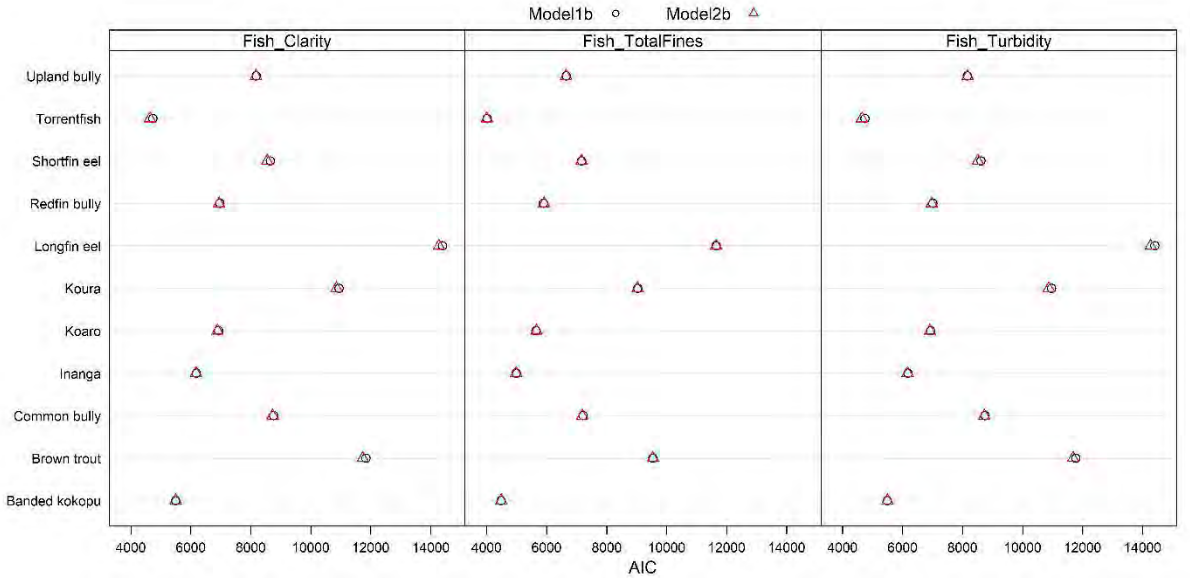


Figure J-10: Akaike Information Criteria (AIC) results for glm models of fish species against each ESV. Model 1b has no interaction between the ESV and climate plus topography class and Model 2b includes an interaction between the ESV and climate plus topography class.

Interactions between ESV and fishing method were also tested for evidence that clearer water or less fines eventuated in different slopes of the relationship between FPC and ESV (e.g., stronger relationships for electric fishing than for visual fishing methods). Inspection of the models showed that whilst some interactions were statistically significant for some species, the effect of this interaction was negligible due to very low coefficients on the interaction terms. Inclusion of this interaction showed only very small increases in model performance as assessed using the Area Under Curve (AUC) method applied by Crow et al. (2014).

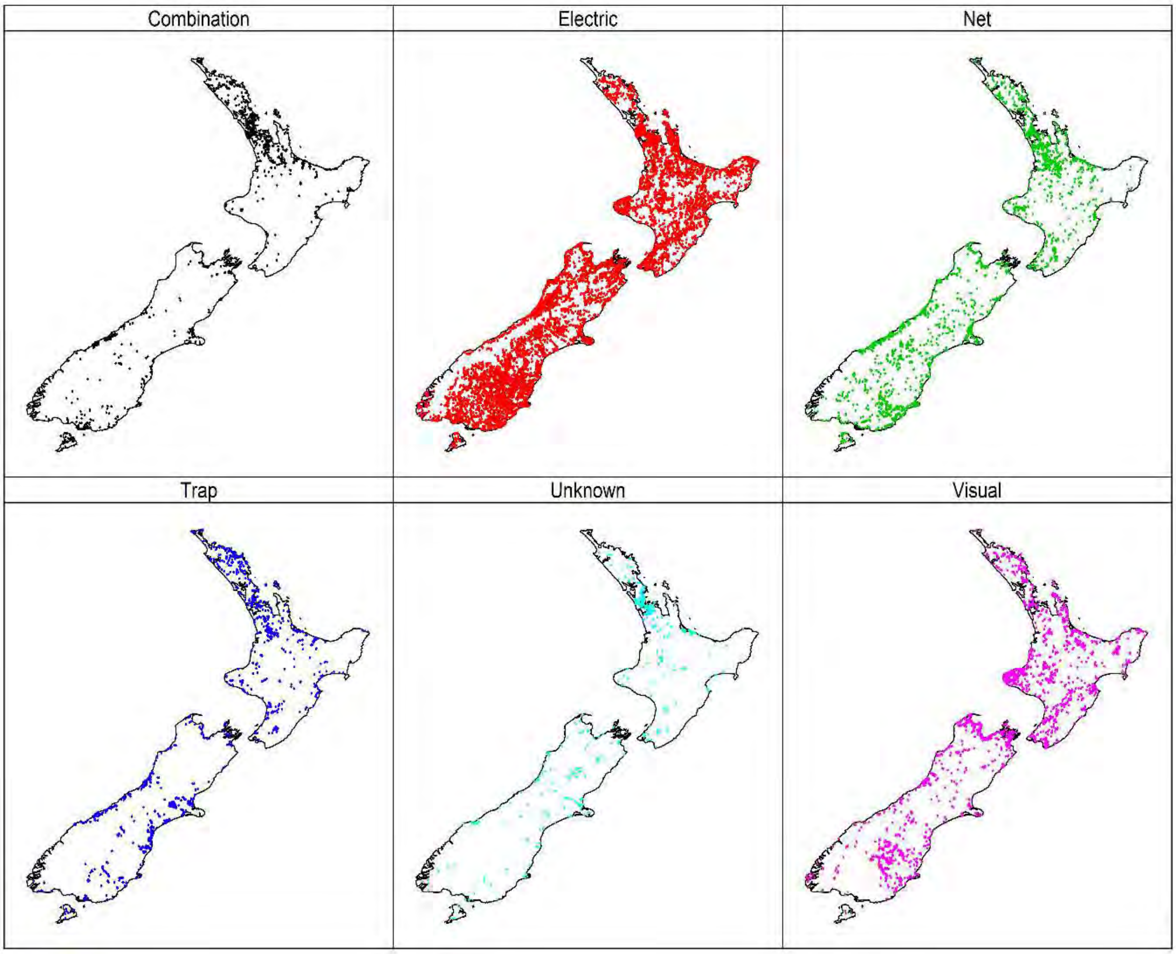


Figure J-11: Map showing NZFFD sample locations by fishing method. Combinations included combinations of electric fishing and nets, combinations of nets and traps, and combinations of nets, traps and electric fishing; electric fishing included backpack and mains set methods; netting methods included fyke nets, Seine nets and set nets; trapping method included Gee minnow traps, box traps, and bait traps and visual methods included daytime observation, diving and spotlighting. Unknown capture records were excluded from the analysis.

The presence of inanga, banded kōkopu and kōaro are potentially influenced by salmonid competitors, particularly brown trout and rainbow trout. The influence of salmonid competitors (possibly interacting with the ESV) was, therefore, tested for inclusion in models of inanga, banded kōkopu and kōaro that would subsequently be used in the fish community analysis. Models with no competition, the addition of presence of a salmonid competitor, and interaction between competition and the presence of a salmonid competitor were compared. Akaike Information Criteria (AIC) values were inspected, and log-likelihood ratio tests were applied to inform on the legitimacy of including salmonid competitors for inanga, banded kōkopu and kōaro. Regardless of ESV, the most parsimonious models for inanga, banded kōkopu and kōaro included no competition, the addition of competition, and an interaction between competition and the ESV respectively.

Probability of capture for each of the species was modelled as a function of each sediment ESV using a generalised linear mixed-effects model (Figure J-12). The response of each FPC was modelled as a function of each ESV by including the ESV as a fixed-effect (Equations (1) and (2)). The proportion of the stream bed observed to be covered by total fines in the NZFFD records (NZFFD % total fines) was used as the deposited sediment ESV for fish and SAM2 % cover instream for macroinvertebrates. The modelled median clarity and turbidity values of were used to relate fish probability of capture to these ESVs and observed medians for macroinvertebrates. No competition from salmonids was assumed as this gave the steepest relationship with the ESV for kōaro and banded kōkopu (the most environmentally conservative outcome). The assumption of no competition from salmonids slightly increased the predicted probability of capture, but made very little difference to final calculated thresholds since difference in predicted prevalence between species is controlled for in the community deviation calculation. Observations within the same NZReach of the RECv2 were down-weighted to avoid pseudo-replication within both the fish and macroinvertebrate datasets.

The following model was selected as the most appropriate for describing the response of fish probability of capture to changes in sediment ESVs:

$$\text{FPC} \sim \text{ESV} + \text{fishing method} + \text{distance to sea} + \text{network position} + \text{Climate\_topography} \quad (1)$$

The following model was selected as the most appropriate for describing the response of macroinvertebrate probability of capture to changes in sediment ESVs:

$$\text{FPC} \sim \text{ESV} + \text{network position} + \text{Climate\_topography} \quad (2)$$

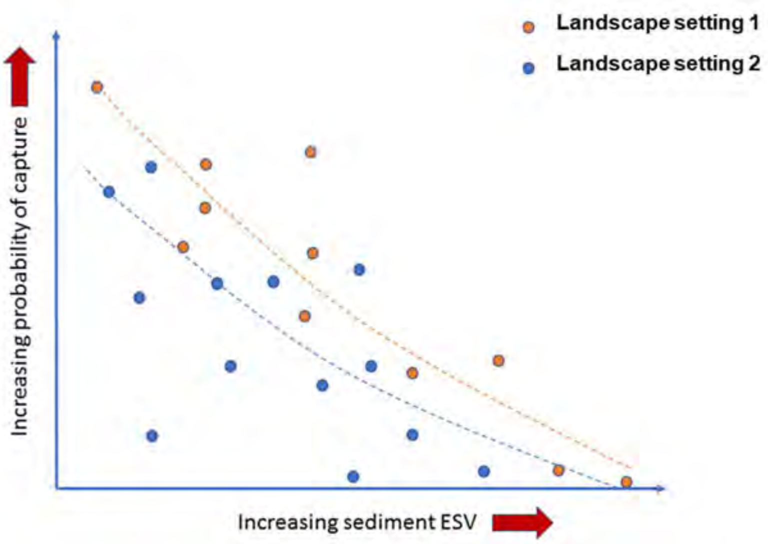


Figure J-12: Simplified example of how variations in fish FPC with increasing sediment ESV are modelled across different landscape settings. The fitted curves derived from this step of the analysis (Equation (1) and (2)) are subsequently used to develop the metrics of community change. Settings refer to landscape settings (e.g., Warm-wet hill).

The probability of capture models (Equations (1) and (2)) provide an estimate of the probability of capturing a species in a particular landscape setting (climate/topography/network position/distance inland) at a given sediment ESV value. These probabilities can be converted to presence/absence data using a threshold probability (Manel et al. 2001) and used to inform interpretation of the expected consequences of changing ESV state for fish community composition.

A range of metrics are available for determining the optimal threshold probability for species distribution models (Manel et al. 2001; Wilson et al. 2005). Cohen's kappa (Cohen 1960) is a measure of the proportion of all possible cases of presence or absence that are predicted correctly after accounting for chance effects, and has been identified as an effective statistic for evaluating presence-absence models, while also being relatively unaffected by prevalence (i.e., the frequency of occurrence of an organism) (Manel et al. 2001). The probability of capture threshold at which Cohen's kappa was maximised (maxKappa) was calculated in R using the 'PresenceAbsence' package for each species. In effect, if probability of capture > maxKappa the species is more likely present than absent, and if probability of capture < maxKappa the species is more likely absent than present (Figure J-13).

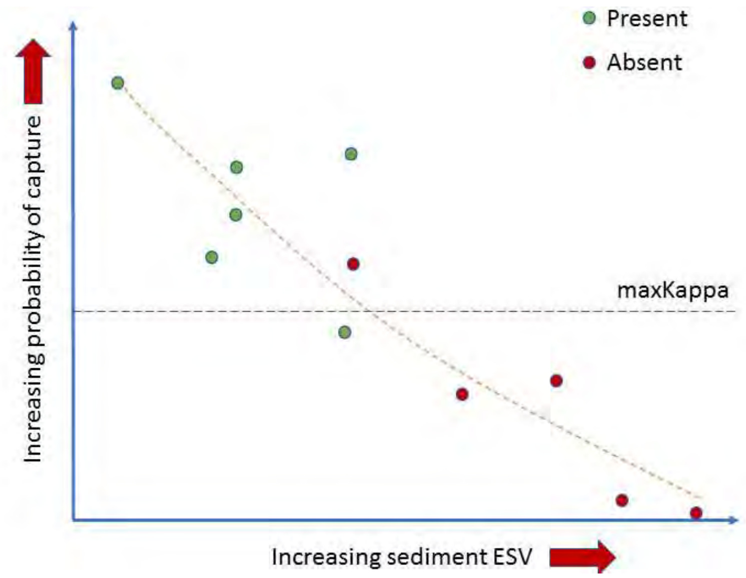


Figure J-13: Illustration of how maxKappa is derived relative to the observed fish data (presence-absence) and the probability of capture for a species. In effect, when probability of capture > maxKappa (above the purple dashed line) a species is most likely to be predicted as present. When probability of capture < maxKappa (below the purple dashed line) a species is most likely to be predicted as absent. However, note that it is possible to get false positives (i.e., a red dot above the maxKappa line) and false negatives (i.e., a green dot below the maxKappa line).

Assessing community change resulting from changes in ESV state

The information on predicted ESV reference state for each landscape setting (see Appendix D), and probability of capture for each species in each landscape setting, were subsequently combined to evaluate the potential impacts on community composition resulting from changes in ESV state. In all cases, the fish probability of capture predictions were made for electric fishing only as this gave the highest probability of capture for all species except kōura, banded kōkopu and inānga.

Species predicted to prefer greater turbidity, less visual clarity, or greater coverage of deposited sediment were first removed from the subsequent calculations of community deviation (Figure J-14). This means that the response to the ESV is consistent between species, avoiding negative consequences for some species being 'cancelled out' by positive changes for other species.

Several steps were required to translate the predicted probability of capture (PC) ESV responses for individual species into a metric of expected community change at different ESV states (Figure J-15). In simple terms this first involved determining the PC at reference ESV state and an array of different ESV states for each individual species in each landscape setting. These values were then combined into a metric (ΔC) describing the overall expected change in community relative to the community that might be expected at reference ESV state.

The first step was to calculate PC under predicted reference ESV state (PC_{ref}) for each landscape setting. Predicted sediment ESV reference states were determined from Level 4 (i.e., 12 class) of the sediment state classification (SSC). Because the landscape setting of the PC models (Equations (1) and (2)) is defined by the REC climate and topography classes (plus network position and distance inland), but the SSC classes are an amalgamation of joined REC climate, topography and geology classes, there were up to four (i.e., the number of different geology classes) reference ESV states for each PC climate-topography landscape setting. PC_{ref} and subsequently ΔC were, therefore, calculated using the four different reference ESV states for each distance inland-network position-climate-topography landscape setting.

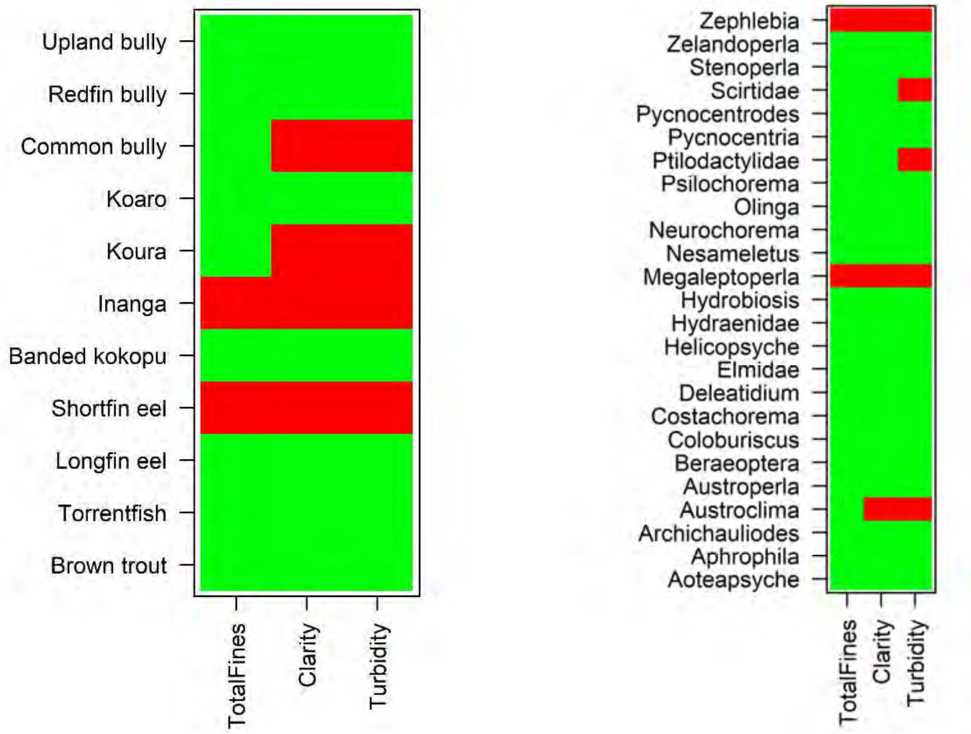


Figure J-14: Taxa included in the fish and macroinvertebrate community analysis. Red = excluded. Green = included.

PC was then also calculated at different ESV values (PC_{ESV}) for each setting. The ratio of these PCs to maxKappa was then calculated:

$$P_{ESV} = PC_{ESV} / \text{maxKappa} \quad (3)$$

$$P_{ref} = PC_{ref} / \text{maxKappa} \quad (4)$$

Where P_{ref} or P_{ESV} >1 the species can be considered more likely to be present than absent. Where P_{ref} or P_{ESV} <1 the species can be considered more likely to be absent than present. The difference between P_{ref} and P_{ESV} represents the deviation away from reference condition (with respect to a particular ESV) for a species at a particular ESV state:

$$\Delta P_{ESV} = P_{ESV} - P_{ref} \quad (5)$$

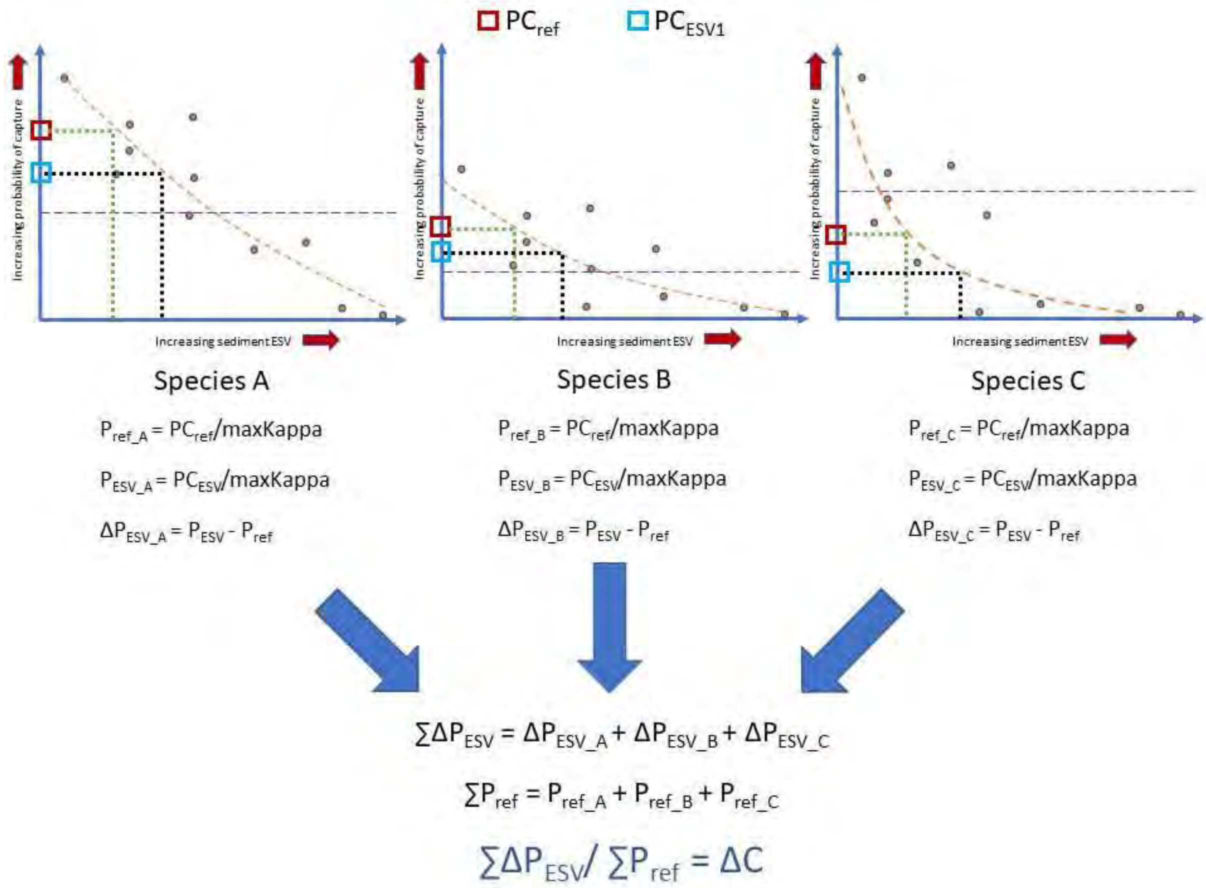


Figure J-15: Illustration of how ΔC is derived from the probability of captures for each species for different sediment ESV states. The probability of capture (PC) at reference ESV state (green dashed line links the predicted reference ESV, for zero heavy pasture, with the corresponding predicted reference PC) is first derived for each species. PC at a different ESV state (dashed black line) is then calculated for each species. Subsequently, the difference in expected probability between the reference ESV state and the alternative ESV state is derived for each species (ΔP_{ESV}). These metrics are then combined from each species to calculate overall expected community change (ΔC).

Positive values of ΔP_{ESV} can be interpreted to mean a species is more likely to occur than at reference condition. Negative values of ΔP_{ESV} indicate that a species is less likely to occur than at reference condition. The value of ΔP_{ESV} was calculated for all species.

For each fish setting, for each ESV value, these deviations from reference condition were then summed over all species ($\sum \Delta P_{ESV}$) and standardised by the sum of P_{ref} over all species ($\sum P_{ref}$).

$$\sum \Delta P_{ESV} / \sum P_{ref} = \Delta C \qquad (6)$$

Standardising by $\sum P_{ref}$ ensures that changes are quantified relative to those expected under reference conditions for the considered ESV. This avoids the situation where communities with more species expected to be present at the reference ESV state (i.e., PC_{ref} > maxKappa for more species), would always show more change in the expected community under different ESV states, relative to their reference state.

ΔC is always zero at the reference ESV state. Negative values in ΔC represent a net loss in the community composition relative to reference conditions. Positive values in ΔC represent net gains in community composition across species relative to reference conditions. ΔC, therefore, represents a deviation in community integrity relative to reference conditions.

ESV band derivation

The calculations of ΔC were used as the basis of deriving ESV bands that could potentially inform the development of the sediment NOF attribute. Because ΔC is a gradient response, as opposed to a threshold response, a risk-based approach was utilised to evaluate band thresholds. The greater the reduction in ΔC from reference, the greater the risk to community integrity. Consequently, increasing departure from reference state was considered to increase the risk of negative outcomes for ecological communities.

For the purposes of this study a 20% departure in community integrity relative to average reference condition (i.e., $\Delta C = -0.20$) was set as the C/D bottom-line threshold. Potential A/B and B/C band limits were subsequently set at equal intervals ($\Delta C = -0.07$ and -0.13 respectively) between the reference condition and the C/D threshold.

The absolute values for the A/B, B/C and C/D thresholds were calculated for each sediment ESV (% fines, turbidity and visual clarity) for each landscape setting (i.e., distance inland-network position-climate-topography-geology combination). Because there are multiple landscape settings in each SSC class and, therefore, multiple ESV thresholds within an SSC class, a single ESV threshold for a class was derived by weighted averaging. The weightings were derived by determining the proportion of segments within the national river network in each landscape setting that existed in each SSC class. The multiple ESV thresholds within an SSC class were then averaged after having applied the weightings derived for each landscape setting to result in a single ESV threshold for an SSC class.

Results

Deposited sediment

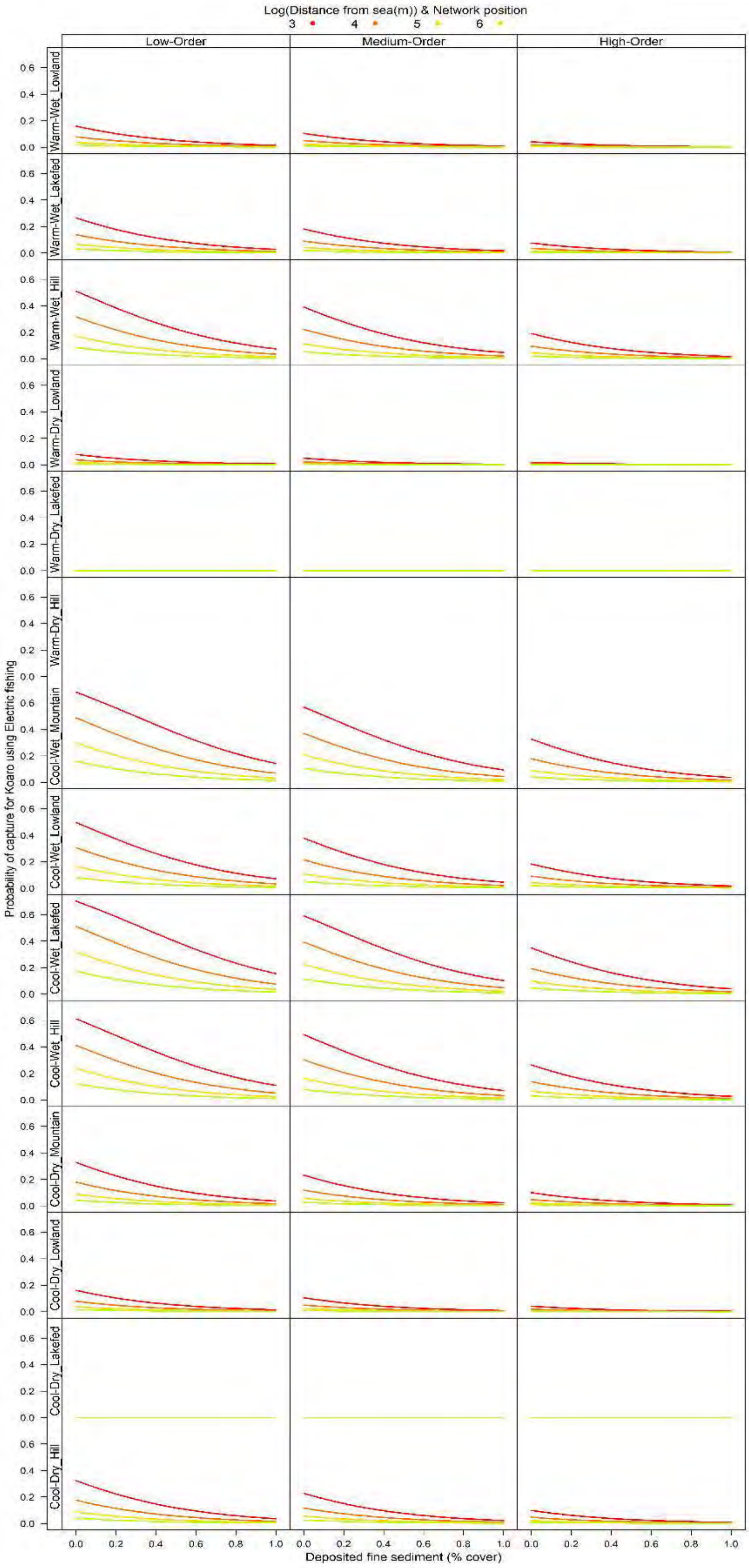


Figure J-16: Example of how kōaro probability of capture varies with increasing deposited fine sediment in different landscape settings. Different lines within landscape setting (Climate-Topography-Network position) reflect different distances inland.

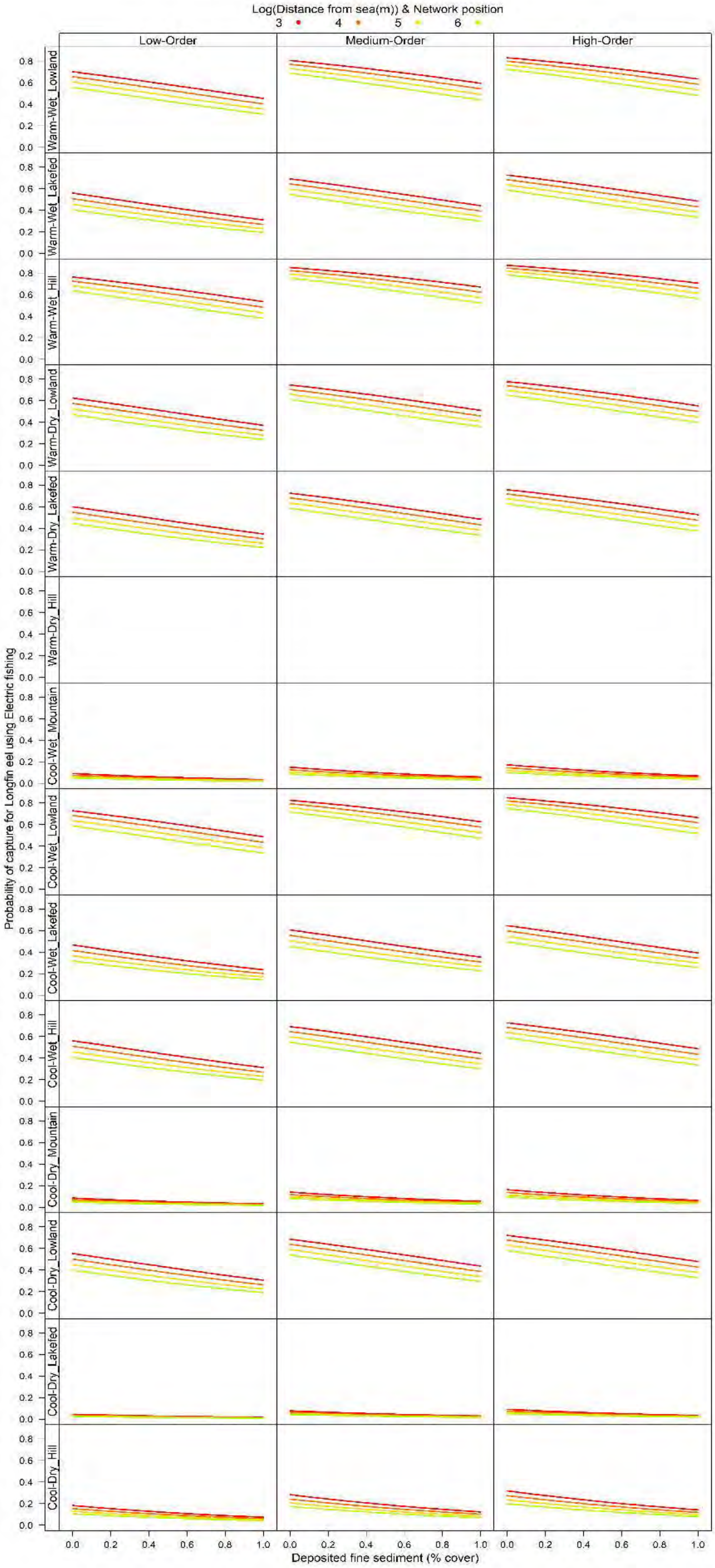


Figure J-17: Example of how longfin eel probability of capture varies with increasing deposited fine sediment in different landscape settings. Different lines within landscape setting (Climate-Topography-Network position) reflect different distances inland.

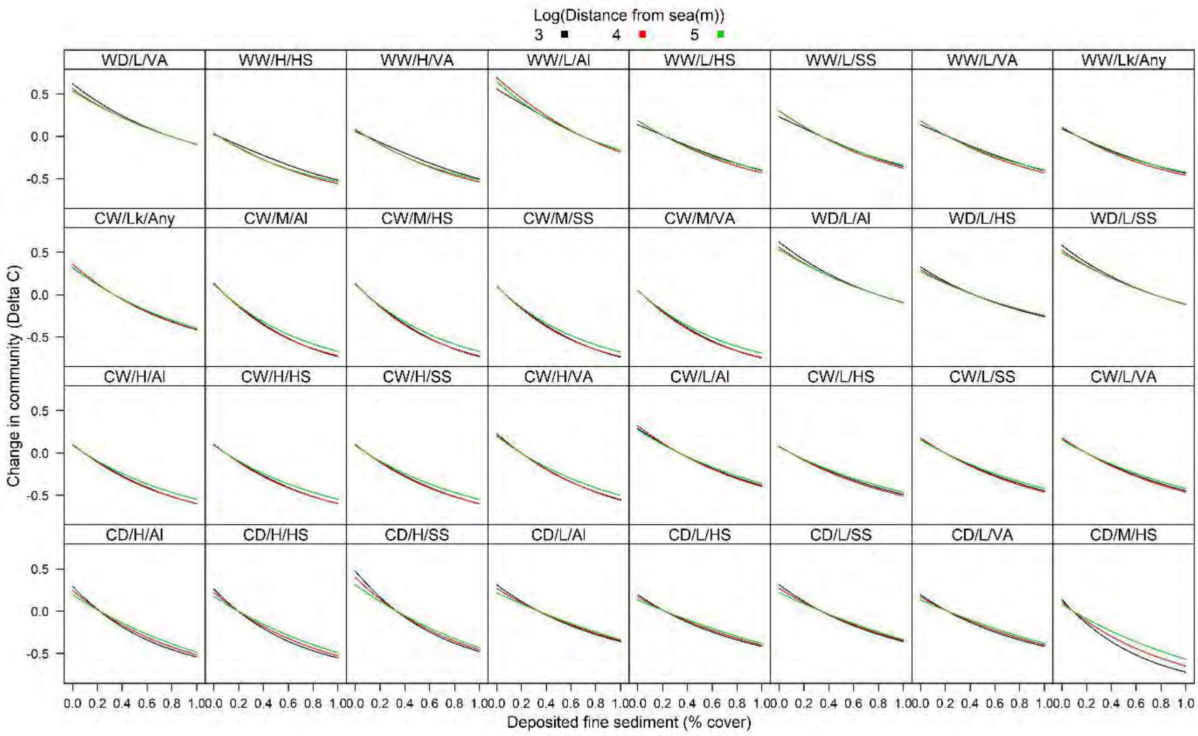


Figure J-18: Change in fish community (ΔC) with increasing cover of deposited fine sediment across Climate-Topography-Geology (CTG) groups for medium order streams. Different coloured lines within landscape settings (CTG-Network Position) represent different distances from the sea.

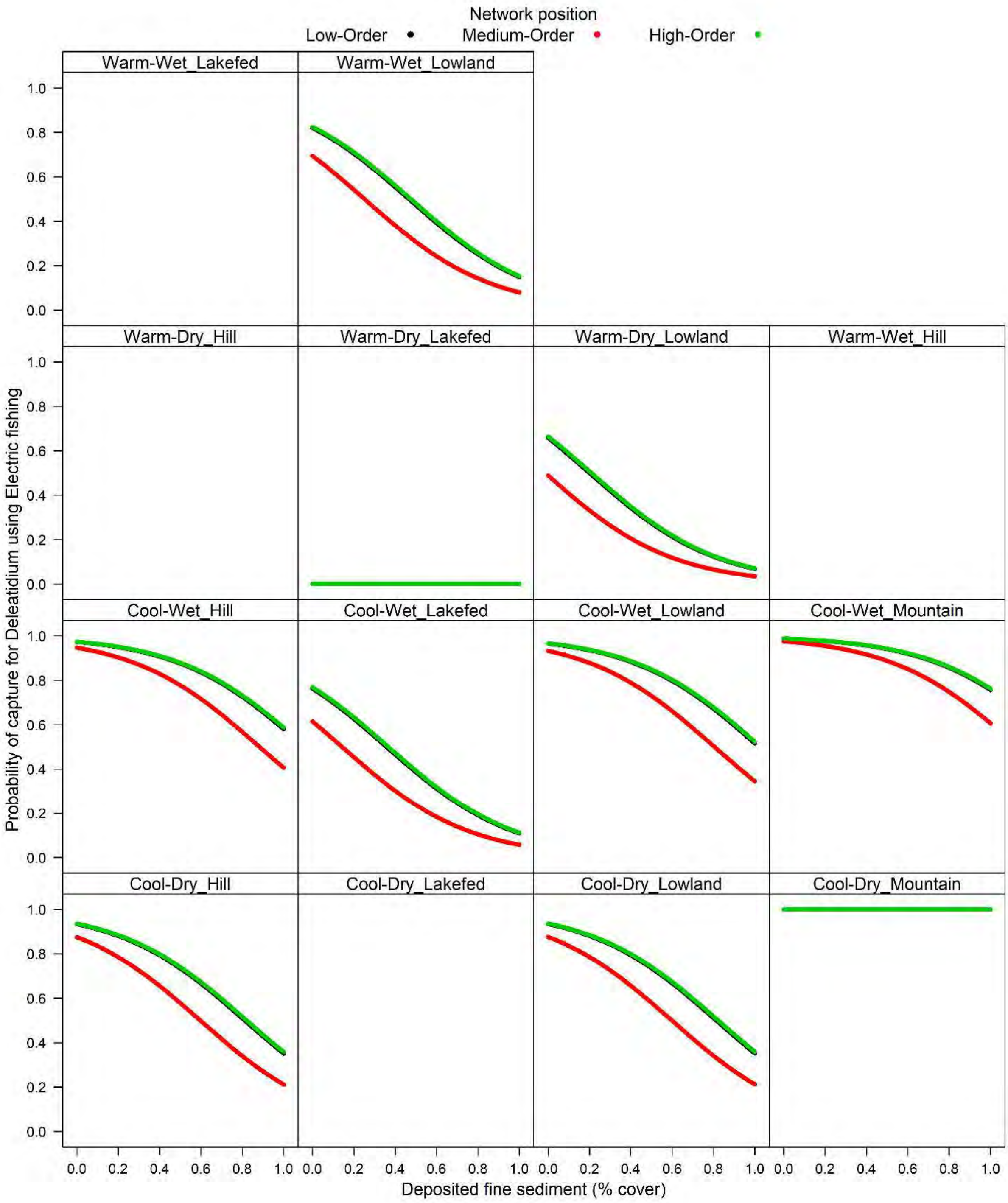


Figure J-19: Example of how *Deleatidium* probability of capture varies with increasing deposited fine sediment in different landscape settings. Different lines within landscape setting (Climate-Topography) reflect different network positions.

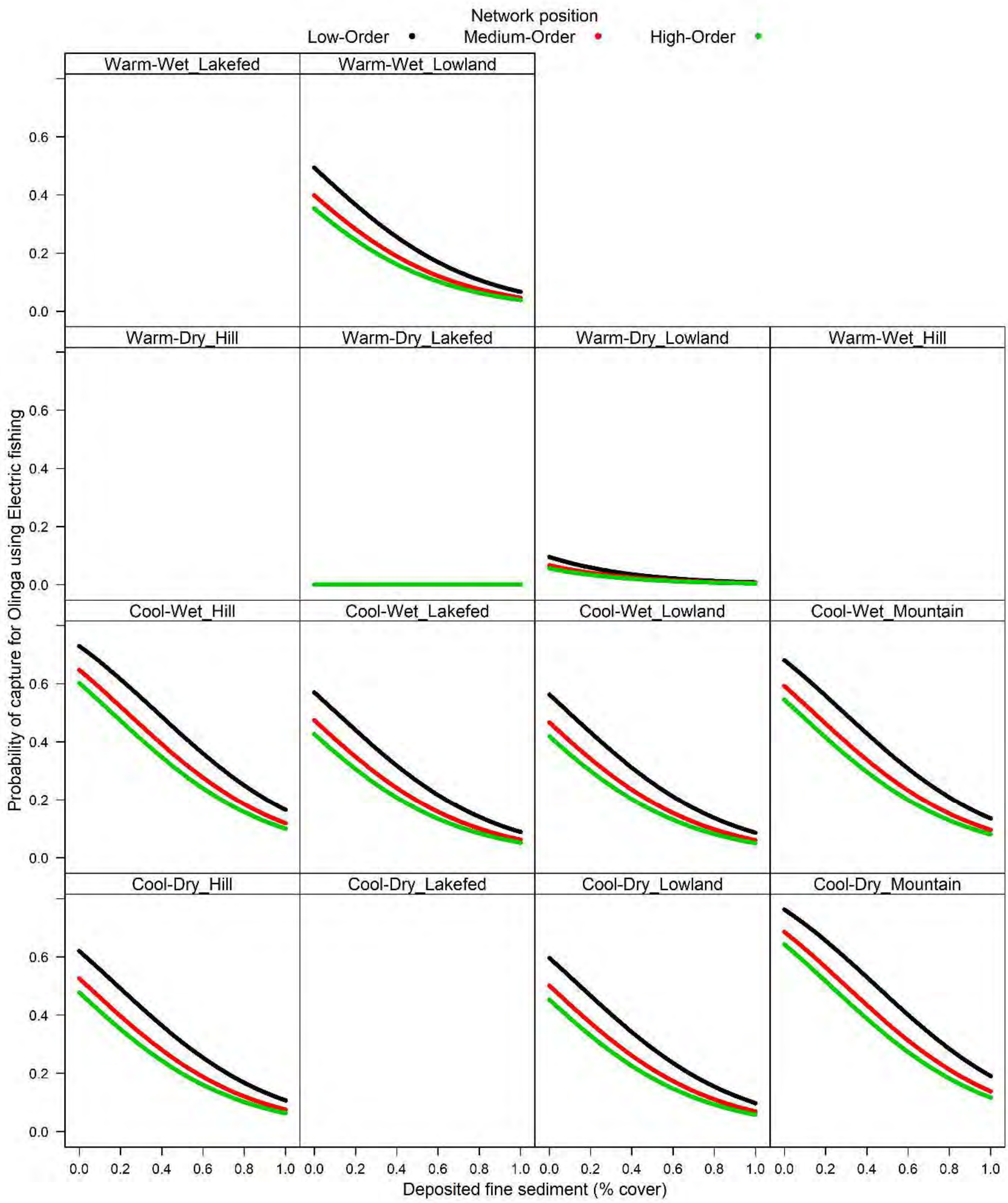


Figure J-20: Example of how *Olinga* probability of capture varies with increasing deposited fine sediment in different landscape settings. Different lines within landscape setting (Climate-Topography) reflect different network positions.

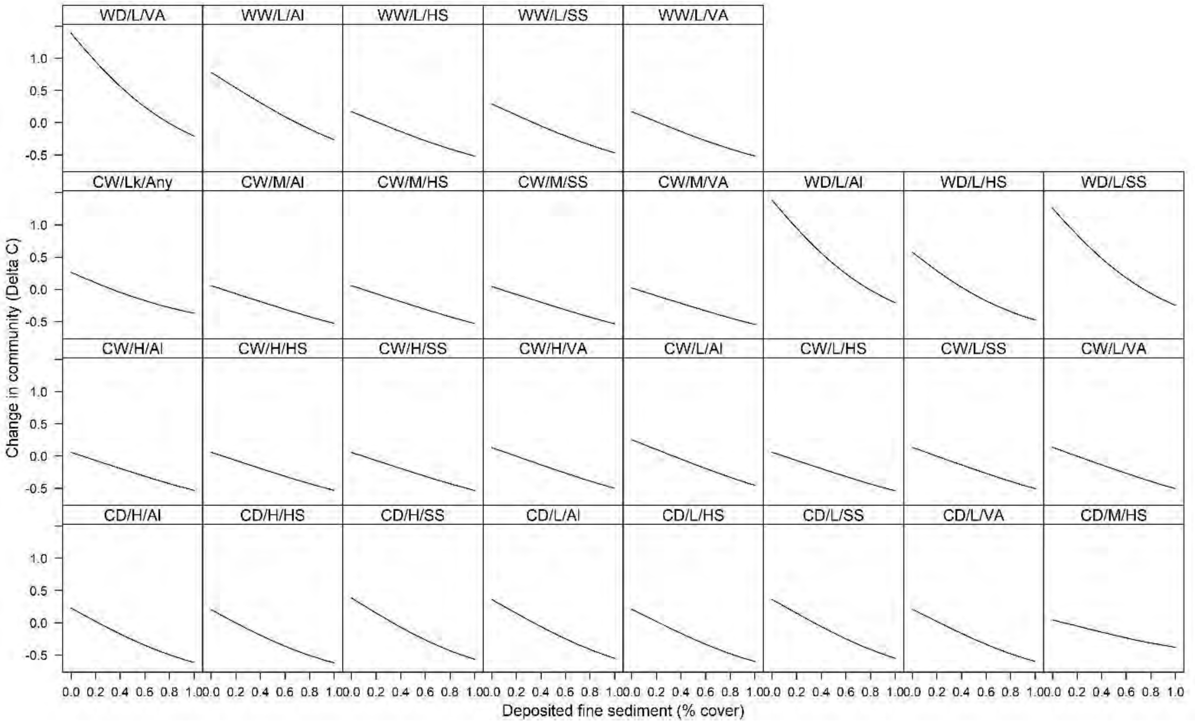


Figure J-21: Change in macroinvertebrate community (ΔC) with increasing cover of deposited fine sediment across Climate-Topography-Geology (CTG) classes for medium size rivers. Distance to the sea is not predictor variable in the macroinvertebrate models meaning there is only one line per CTG class.

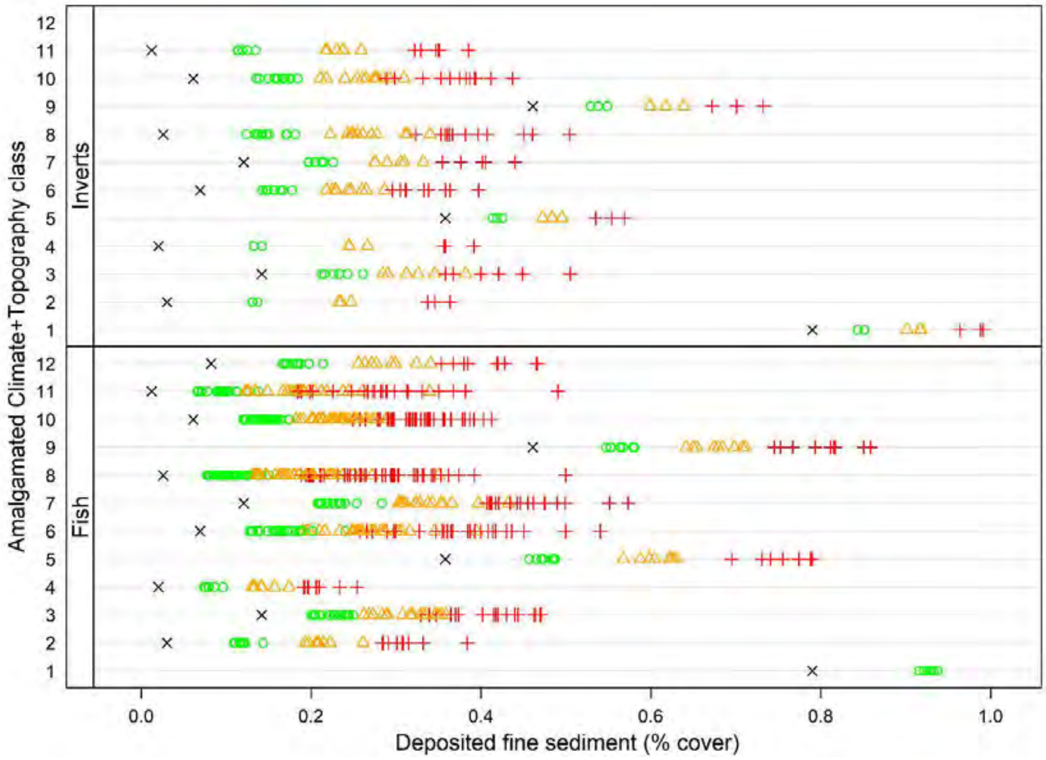


Figure J-22: Deposited sediment band thresholds for all landscape settings within SSC classes for fish and macroinvertebrates using the community deviation method. Black crosses indicate the reference sediment state. Green circles represent A/B thresholds (ΔC of -0.066). Orange triangles represent B/C thresholds (ΔC of -0.133). Red crosses represent C/D thresholds (ΔC of -0.20).

Table J-2: Potential band thresholds for deposited sediment based on the fish community deviation method. Thresholds are presented as proportions of the bed covered by fine sediment for the 12 classes at the fourth level of aggregation in the SSC. NA indicates that the thresholds exceed the maximum value of 1.0. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	A/B threshold	B/C threshold	C/D threshold
L4.1	0.79	0.92	NA	NA
L4.2	0.04	0.09	0.15	0.21
L4.3	0.33	0.42	0.51	0.61
L4.4	0.07	0.12	0.17	0.23
L4.5	0.74	0.88	NA	NA
L4.6	0.22	0.32	0.42	0.54
L4.7	0.34	0.44	0.55	0.67
L4.8	0.13	0.22	0.33	0.45
L4.9	0.43	0.56	0.70	0.85
L4.10	0.09	0.15	0.22	0.29
L4.11	0.69	0.81	0.94	NA
L4.12	0.20	0.27	0.36	0.45

Table J-3: Potential band thresholds for deposited sediment based on the macroinvertebrate community deviation method. Thresholds are presented as proportions of the bed covered by fine sediment for the 12 classes at the fourth level of aggregation in the SSC. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	A/B threshold	B/C threshold	C/D threshold
L4.1	0.79	0.84	0.90	0.97
L4.2	0.04	0.09	0.15	0.21
L4.3	0.33	0.42	0.50	0.60
L4.4	0.07	0.12	0.17	0.23
L4.5	0.74	0.80	0.86	0.92
L4.6	0.22	0.30	0.38	0.46
L4.7	0.34	0.41	0.48	0.56
L4.8	0.13	0.22	0.33	0.45
L4.9	0.43	0.48	0.54	0.61
L4.10	0.09	0.15	0.22	0.29
L4.11	0.69	0.76	0.82	0.89
L4.12	0.20	0.27	0.36	0.45

Turbidity

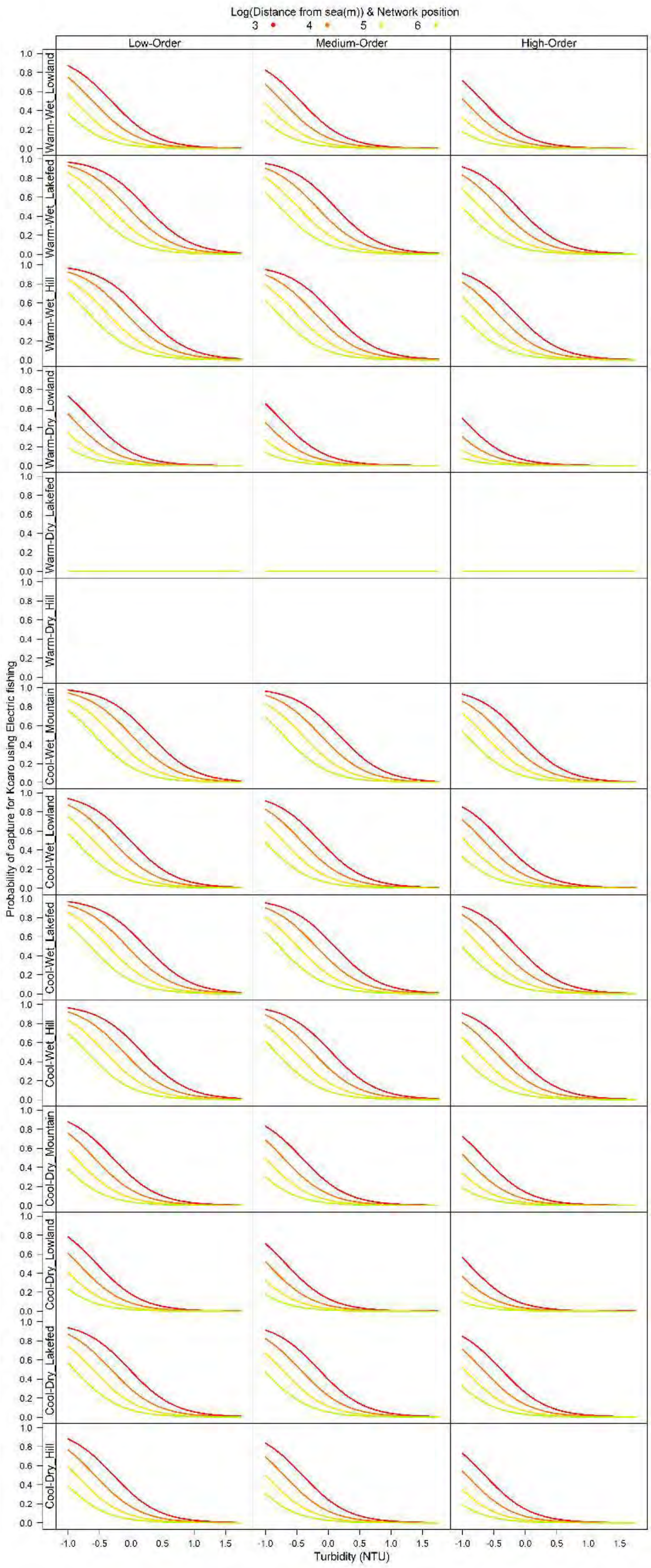


Figure J-23: Example of how kōaro probability of capture varies with increasing turbidity in different landscape settings. Different lines within landscape setting (Climate-Topography-Network position) reflect different distances inland.

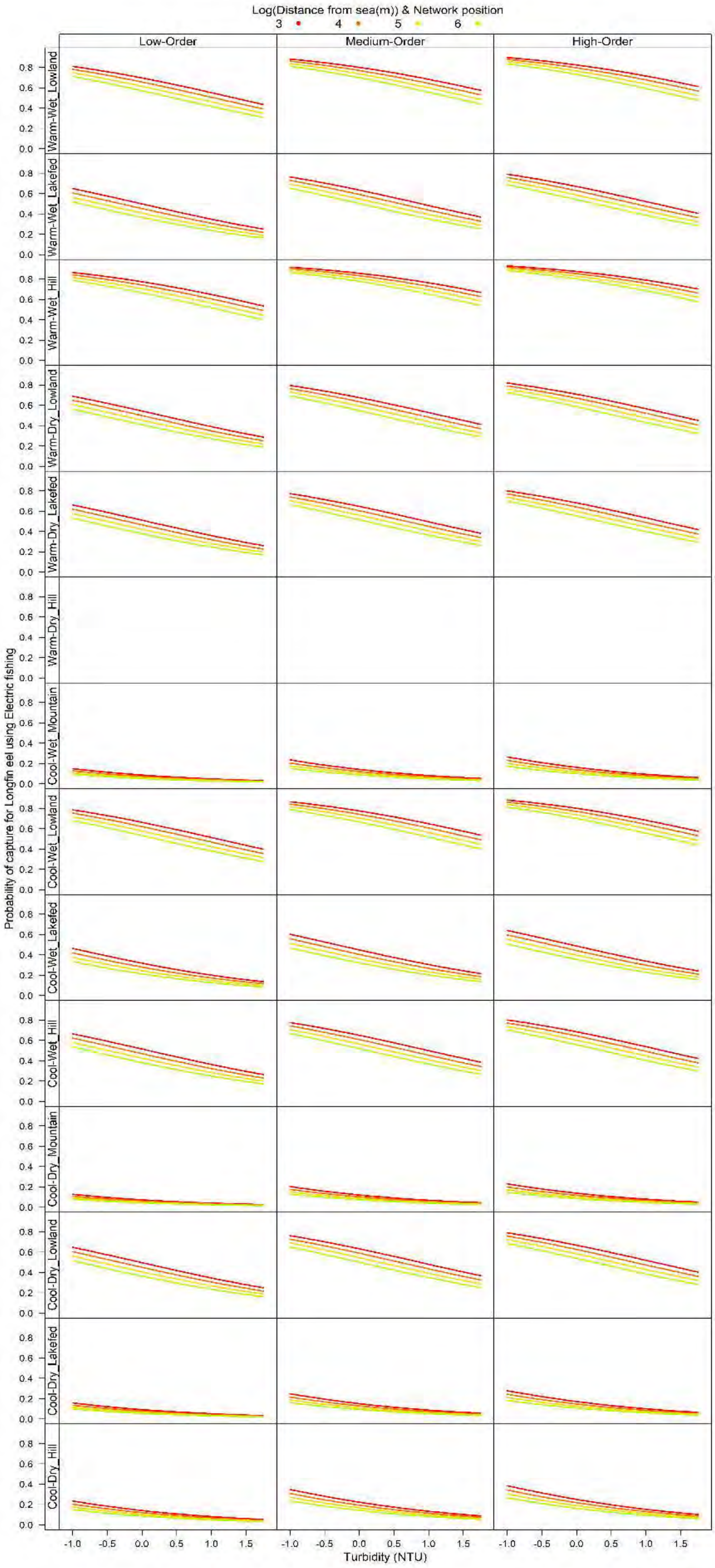


Figure J-24: Example of how longfin eel probability of capture varies with increasing turbidity in different landscape settings. Different lines within landscape setting (Climate-Topography-Network position) reflect different distances inland.

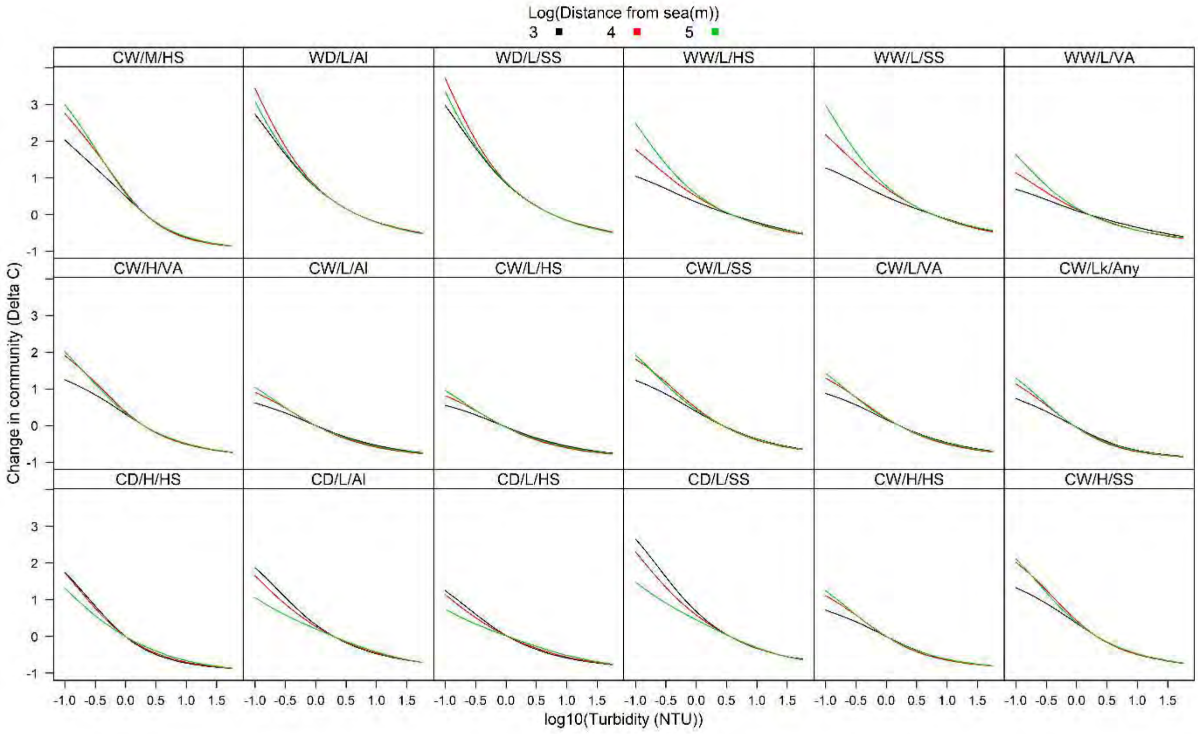


Figure J-25: Change in fish community (ΔC) with increasing cover of turbidity across Climate-Topography-Geology (CTG) groups for medium order streams. Different coloured lines within landscape settings (CTG-Network Position) represent different distances from the sea.

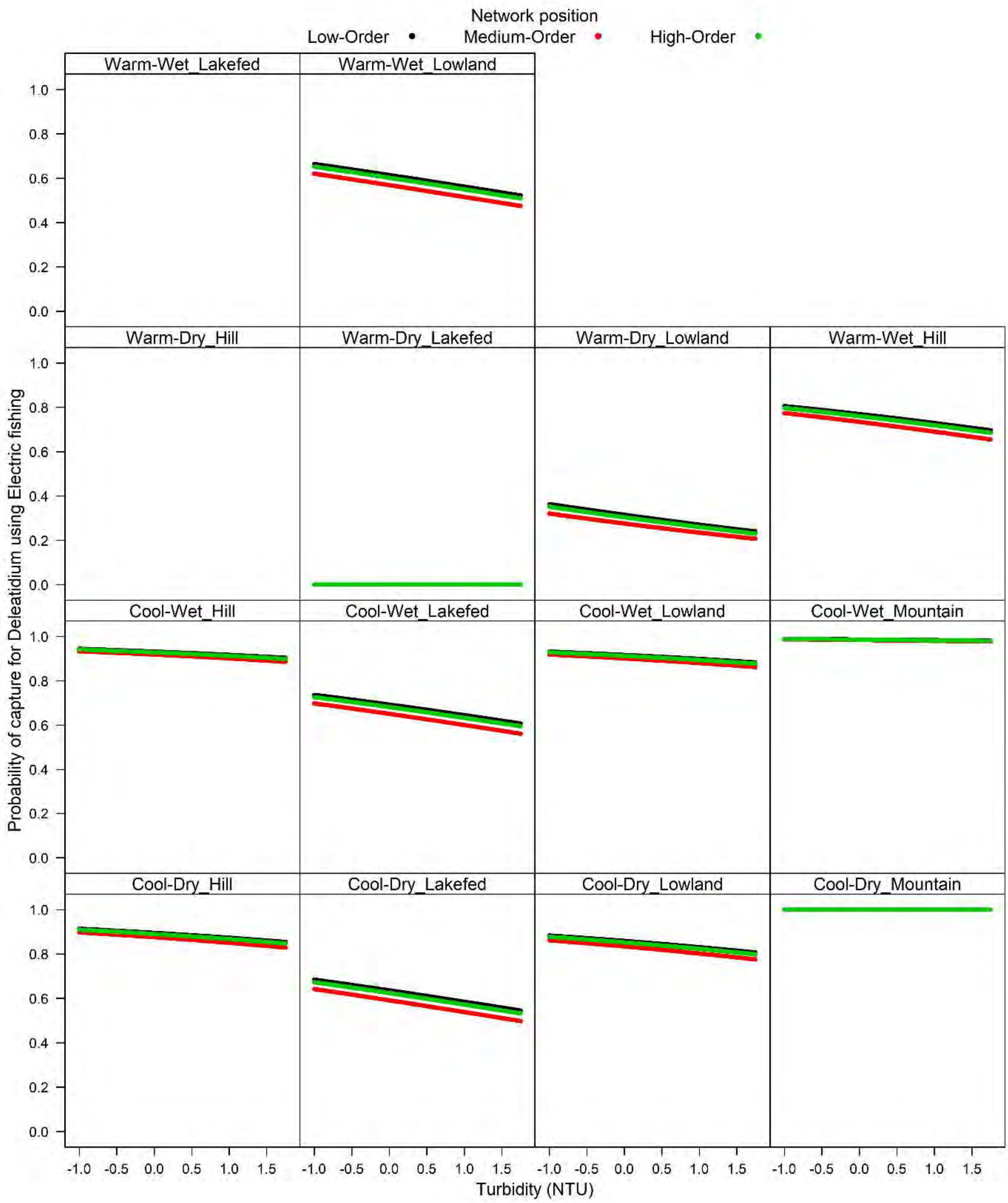


Figure J-26: Example of how *Deleatidium* probability of capture varies with increasing turbidity in different landscape settings. Different lines within landscape setting (Climate-Topography) reflect different network positions.

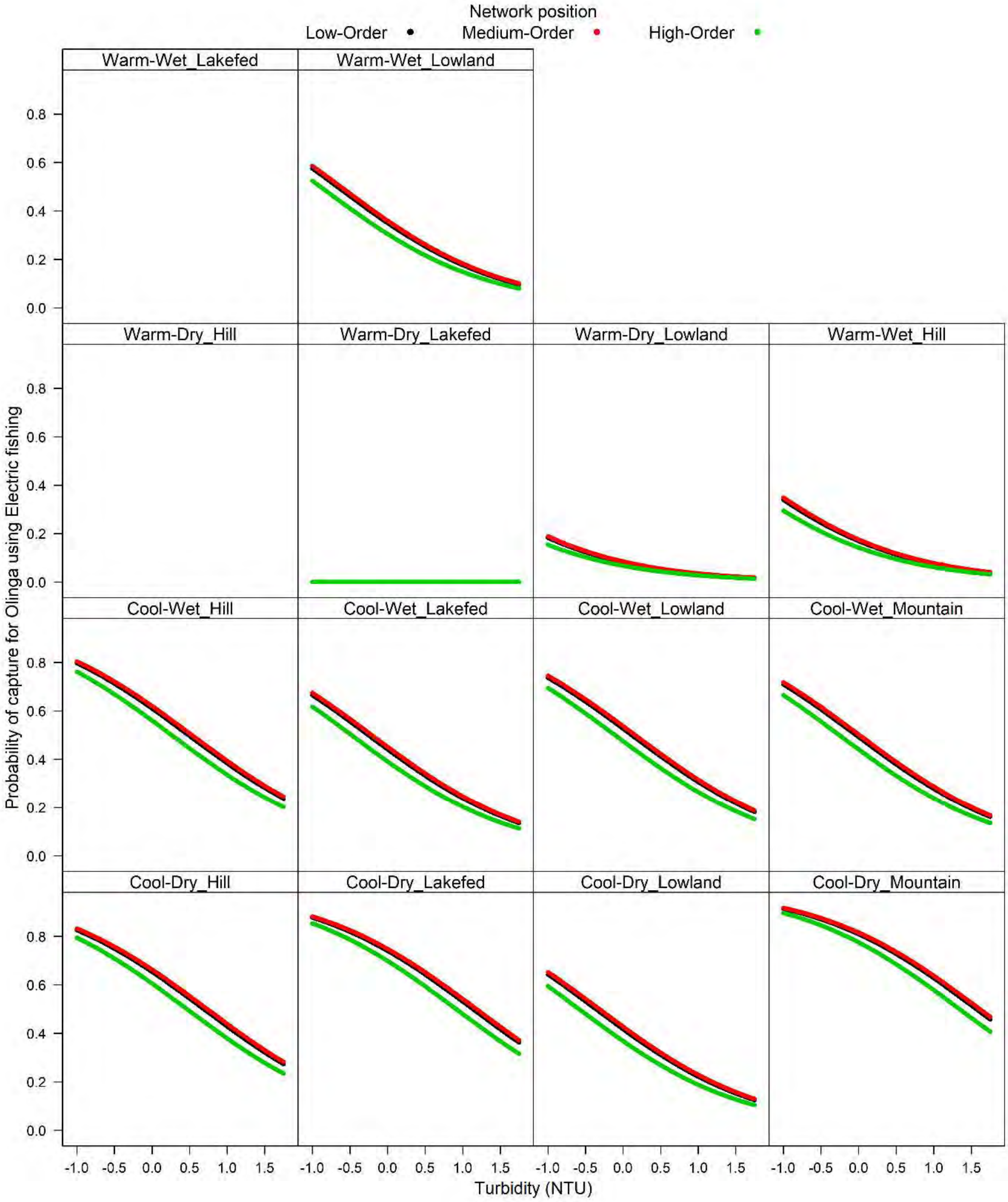


Figure J-27: Example of how *Olinga* probability of capture varies with increasing turbidity in different landscape settings. Different lines within landscape setting (Climate-Topography) reflect different network positions.

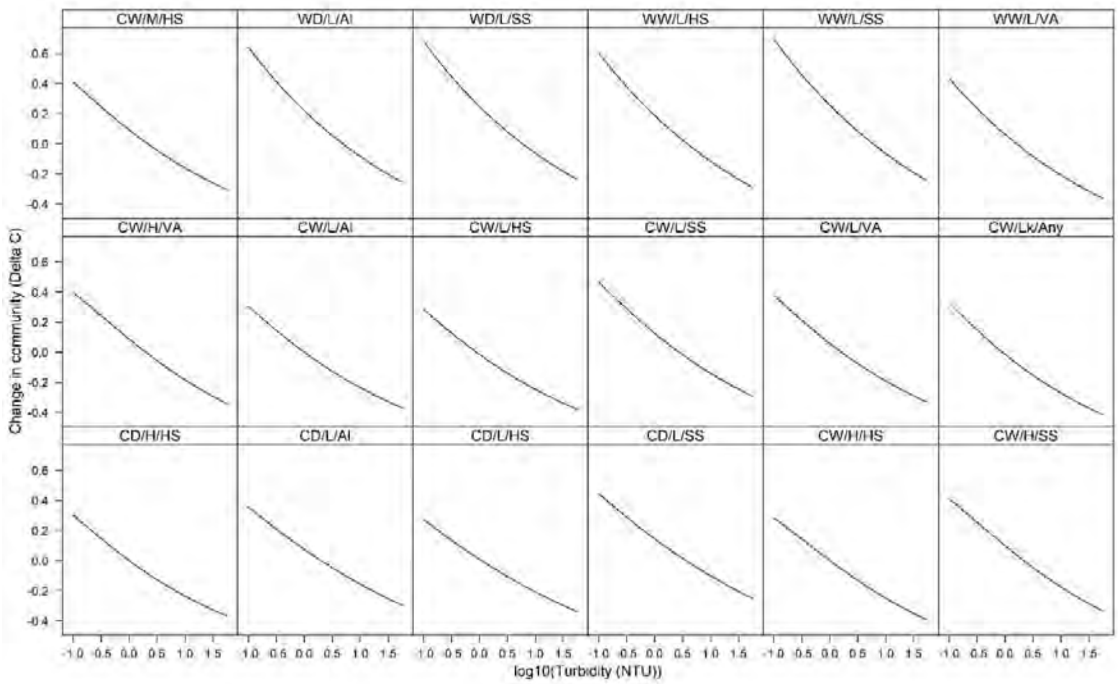


Figure J-28: Change in macroinvertebrate community (ΔC) with increasing cover of turbidity across Climate-Topography-Geology (CTG) classes for medium size rivers. Distance to the sea is not a predictor variable in the macroinvertebrate models meaning there is only one line per CTG class.

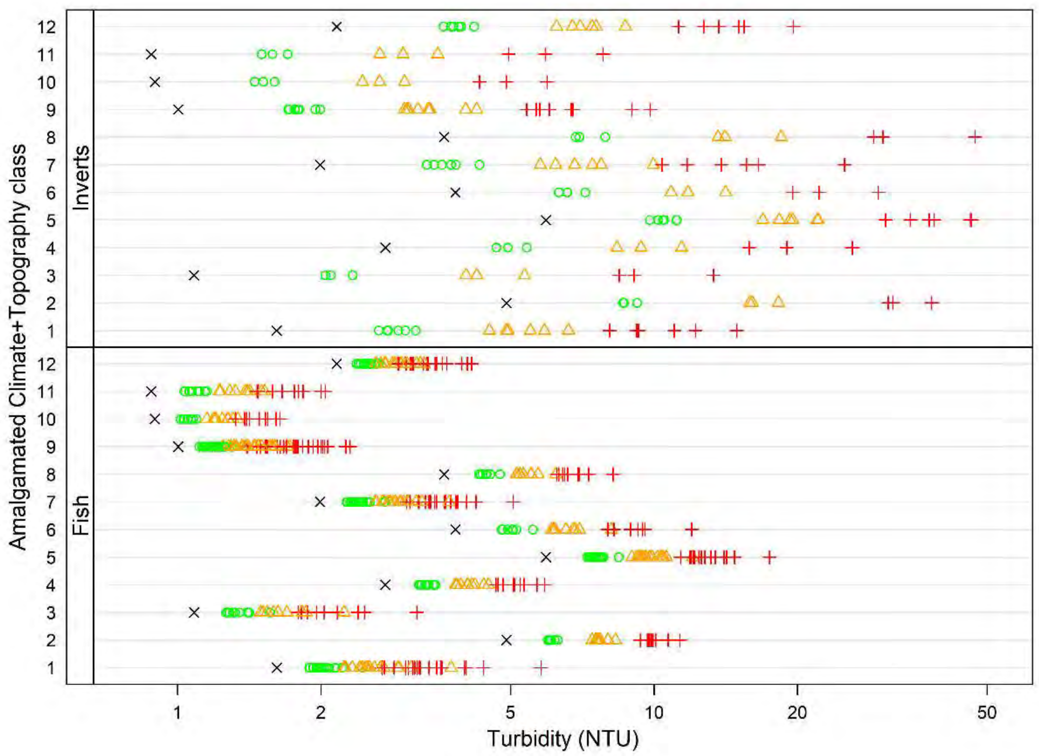


Figure J-29: Turbidity band thresholds for all landscape settings within SSC classes for fish and macroinvertebrates using the community deviation method. Black crosses indicate the reference sediment state. Green circles represent A/B thresholds (ΔC of -0.066). Orange triangles represent B/C thresholds (ΔC of -0.133). Red crosses represent C/D thresholds (ΔC of -0.20).

Table J-4: Potential band thresholds for turbidity based on the fish community deviation method. Thresholds are presented turbidity (NTU) for the 12 classes at the fourth level of aggregation in the SSC. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	A/B threshold	B/C threshold	C/D threshold
L4.1	1.6	2.0	2.5	3.2
L4.2	4.9	6.2	7.9	10.5
L4.3	1.1	1.3	1.6	2.0
L4.4	2.7	3.3	3.9	4.8
L4.5	5.9	7.5	9.8	13.1
L4.6	3.8	4.8	6.3	8.3
L4.7	2.0	2.3	2.8	3.3
L4.8	3.6	4.3	5.2	6.4
L4.9	1.0	1.2	1.4	1.6
L4.10	0.9	1.1	1.3	1.5
L4.11	0.9	1.1	1.3	1.6
L4.12	2.2	2.4	2.7	3.1

Table J-5: Potential band thresholds for turbidity based on the macroinvertebrate community deviation method. Thresholds are presented turbidity (NTU) for the 12 classes at the fourth level of aggregation in the SSC. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	A/B threshold	B/C threshold	C/D threshold
L4.1	1.6	2.7	4.7	8.6
L4.2	4.9	8.7	16.1	31.7
L4.3	1.1	2.1	4.1	8.7
L4.4	2.7	4.8	8.7	16.9
L4.5	5.9	10.1	18.0	33.7
L4.6	3.8	6.4	11.1	20.3
L4.7	2.0	3.5	6.5	12.6
L4.8	3.6	6.9	13.9	29.6
L4.9	1.0	1.7	3.1	5.8
L4.10	0.9	1.5	2.8	5.3
L4.11	0.9	1.5	2.7	5.2
L4.12	2.2	3.8	6.9	13.5

Visual clarity

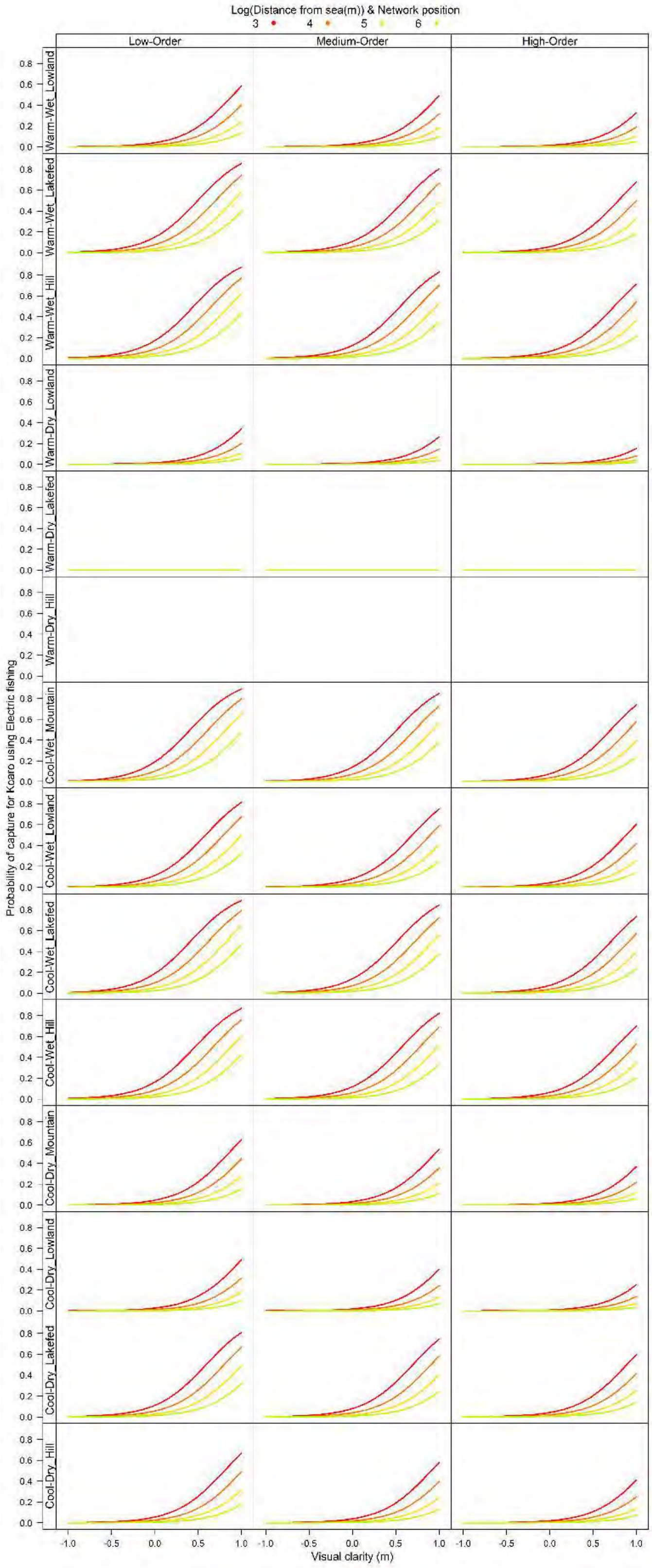


Figure J-30: Example of how kōaro probability of capture varies with increasing visual clarity in different landscape settings. Different lines within landscape setting (Climate-Topography-Network position) reflect different distances inland.

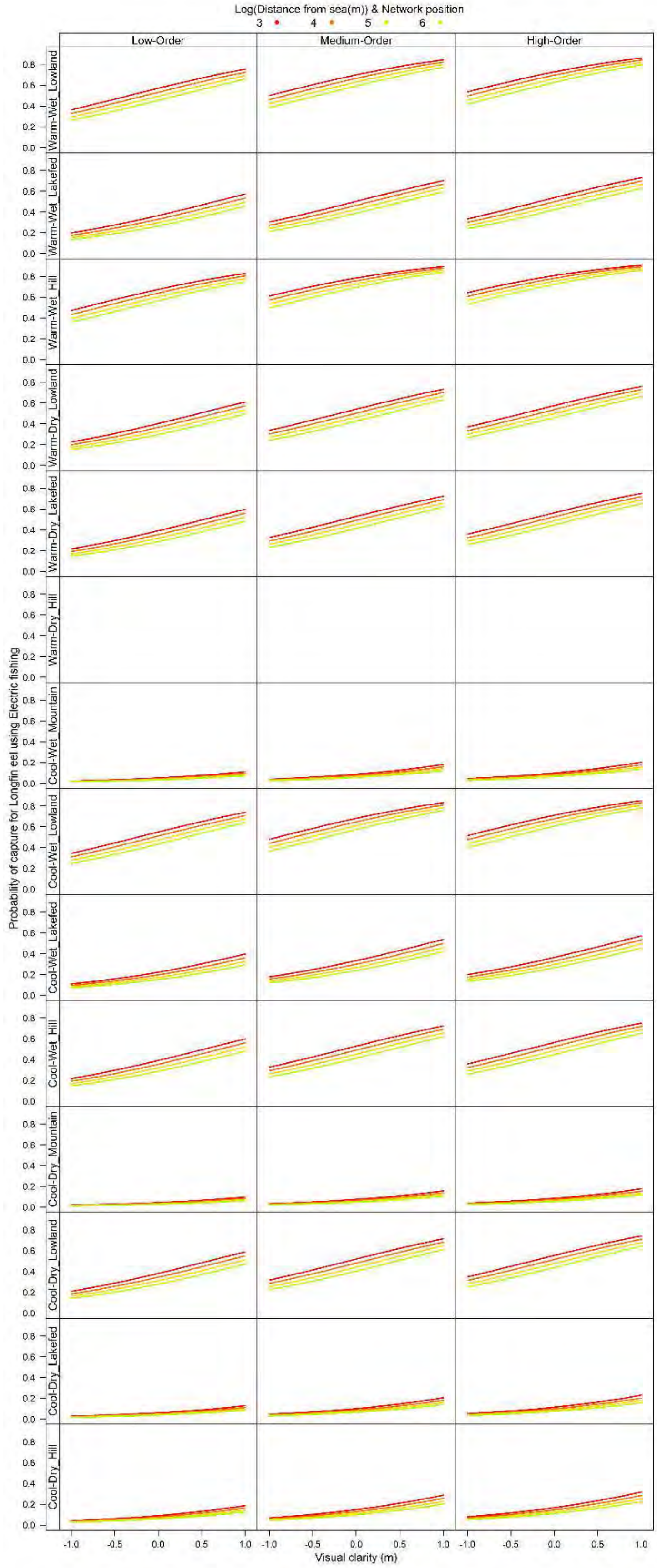


Figure J-31: Example of how longfin eel probability of capture varies with increasing visual clarity in different landscape settings. Different lines within landscape setting (Climate-Topography-Network position) reflect different distances inland.

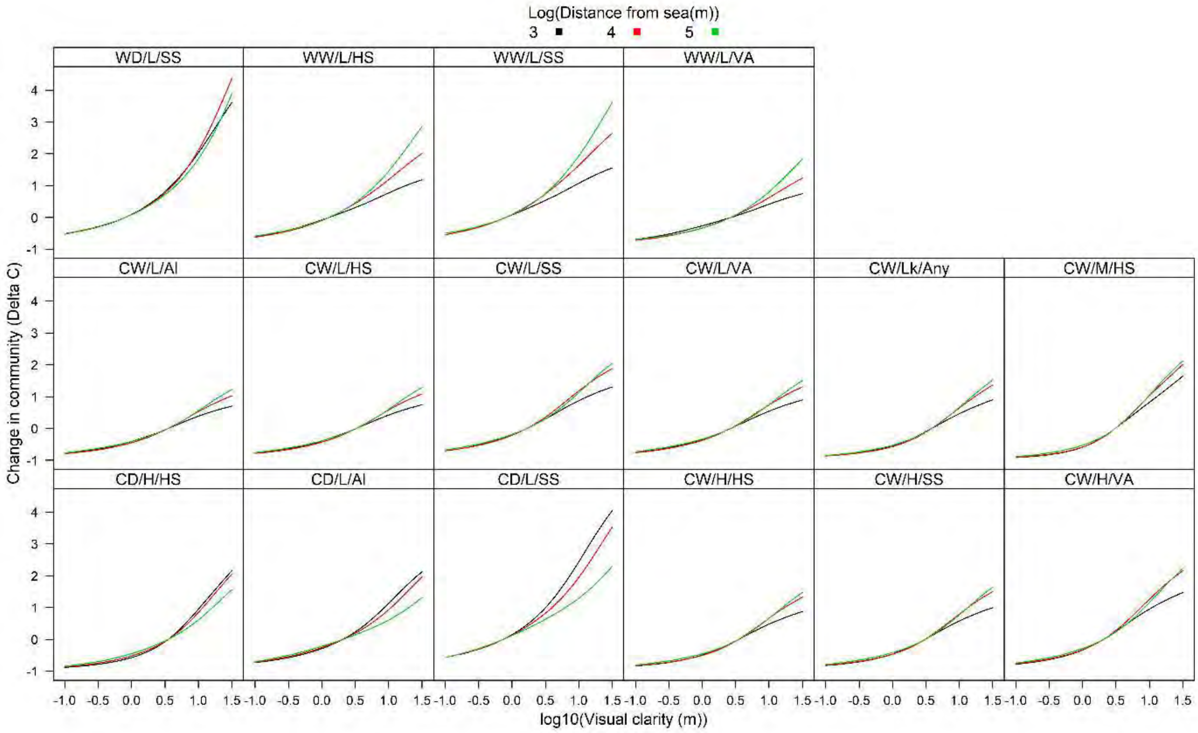


Figure J-32: Change in fish community (ΔC) with increasing cover of turbidity across Climate-Topography-Geology (CTG) groups for medium order streams. Different coloured lines within landscape settings (CTG-Network Position) represent different distances from the sea.

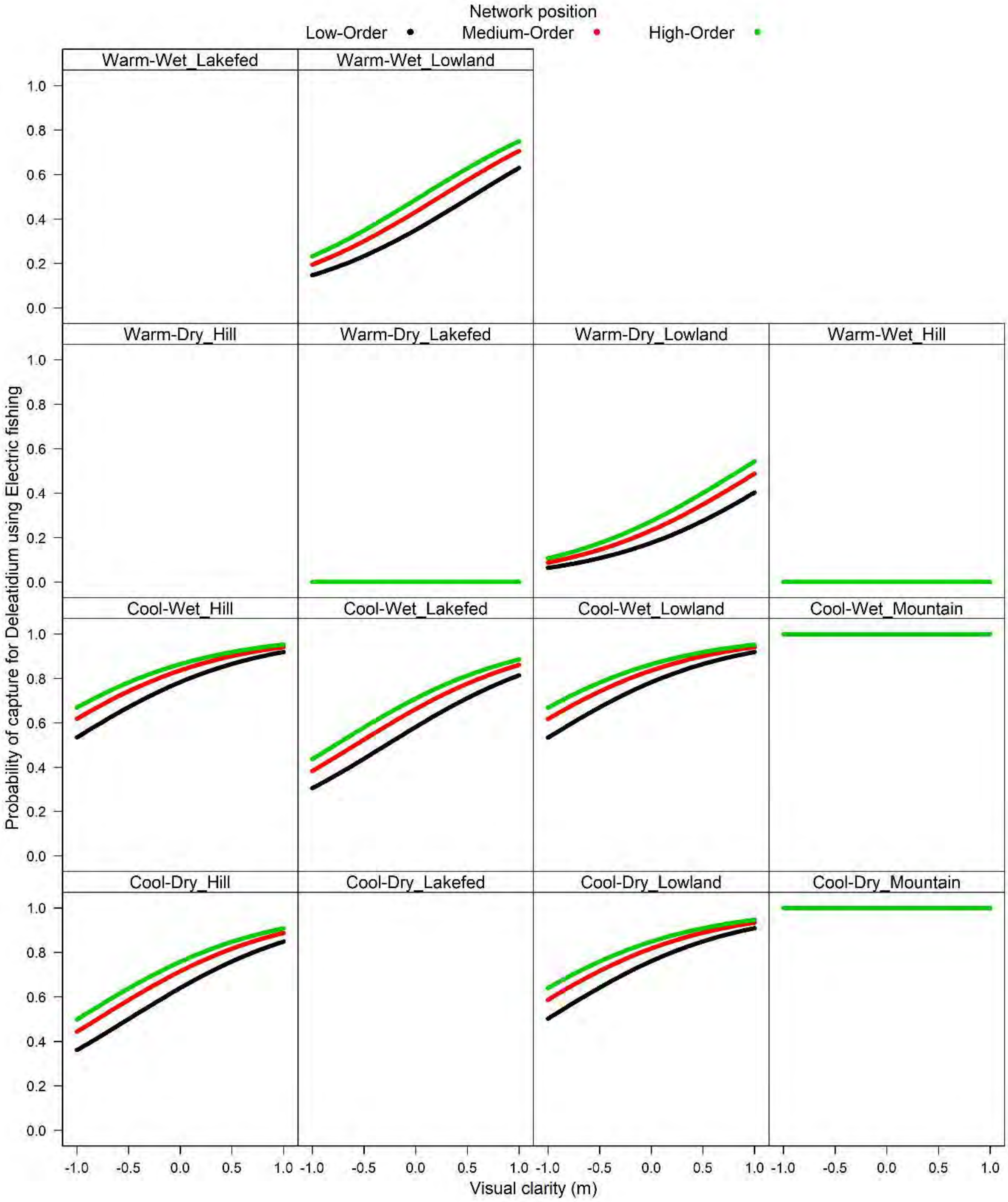


Figure J-33: Example of how *Deleatidium* probability of capture varies with increasing visual clarity in different landscape settings. Different lines within landscape setting (Climate-Topography) reflect different network positions.

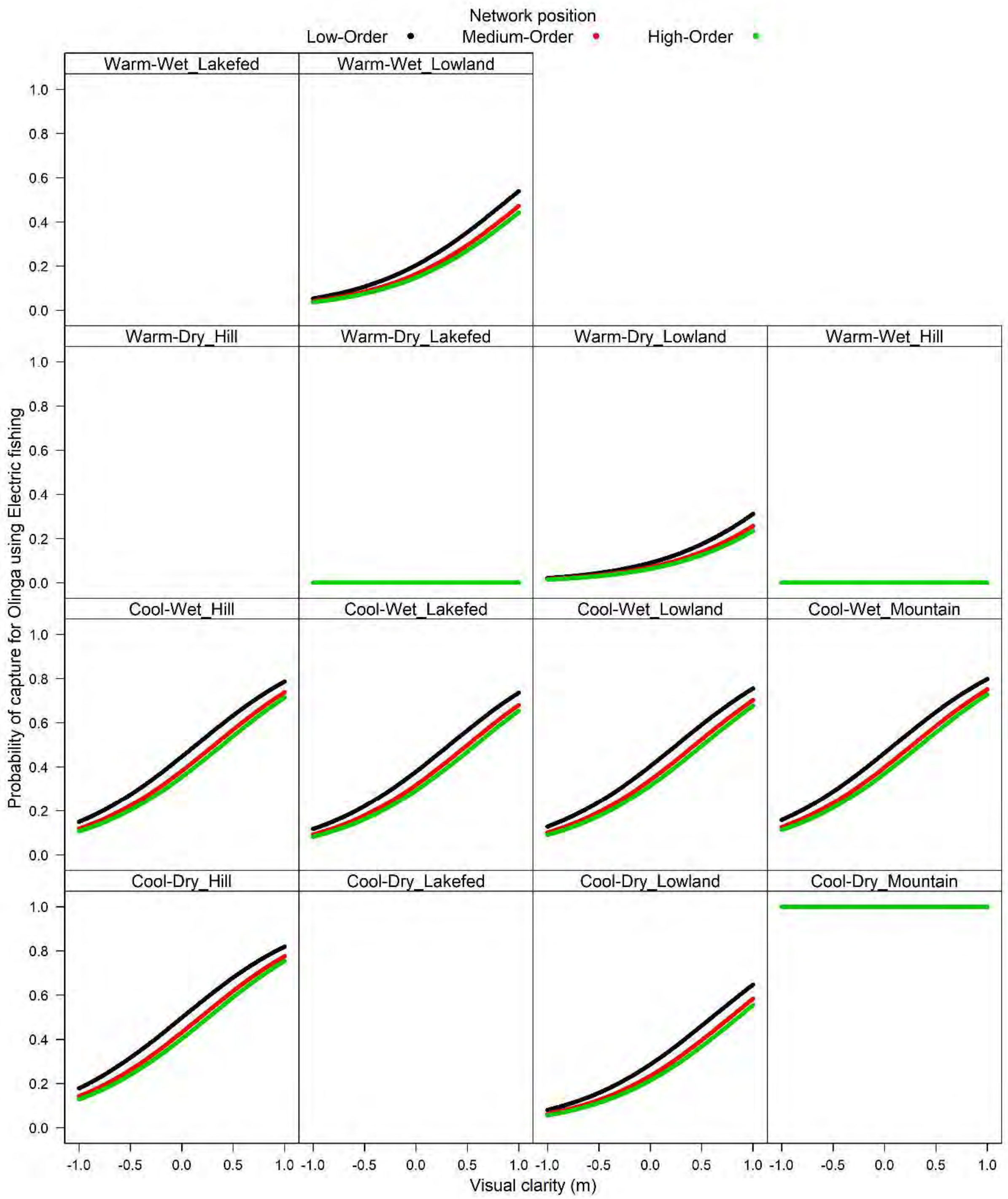


Figure J-34: Example of how *Olinga* probability of capture varies with increasing visual clarity in different landscape settings. Different lines within landscape setting (Climate-Topography) reflect different network positions.

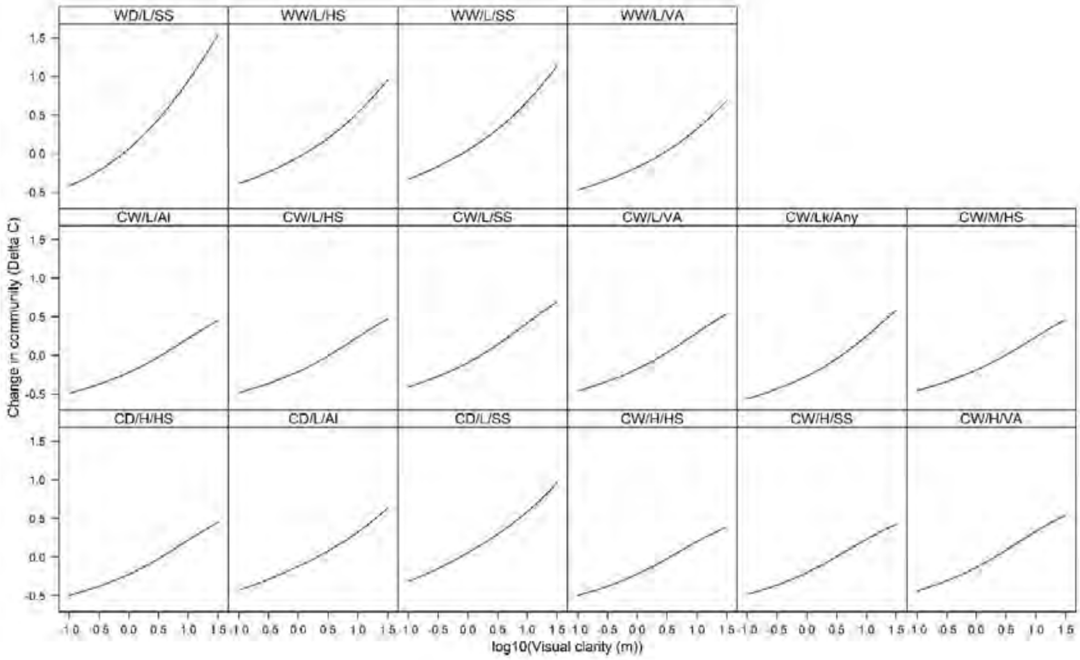


Figure J-35: Change in macroinvertebrate community (ΔC) with increasing cover of visual clarity across Climate-Topography-Geology (CTG) classes for medium size rivers. Distance to the sea is not a predictor variable in the macroinvertebrate models meaning there is only one line per CTG class.

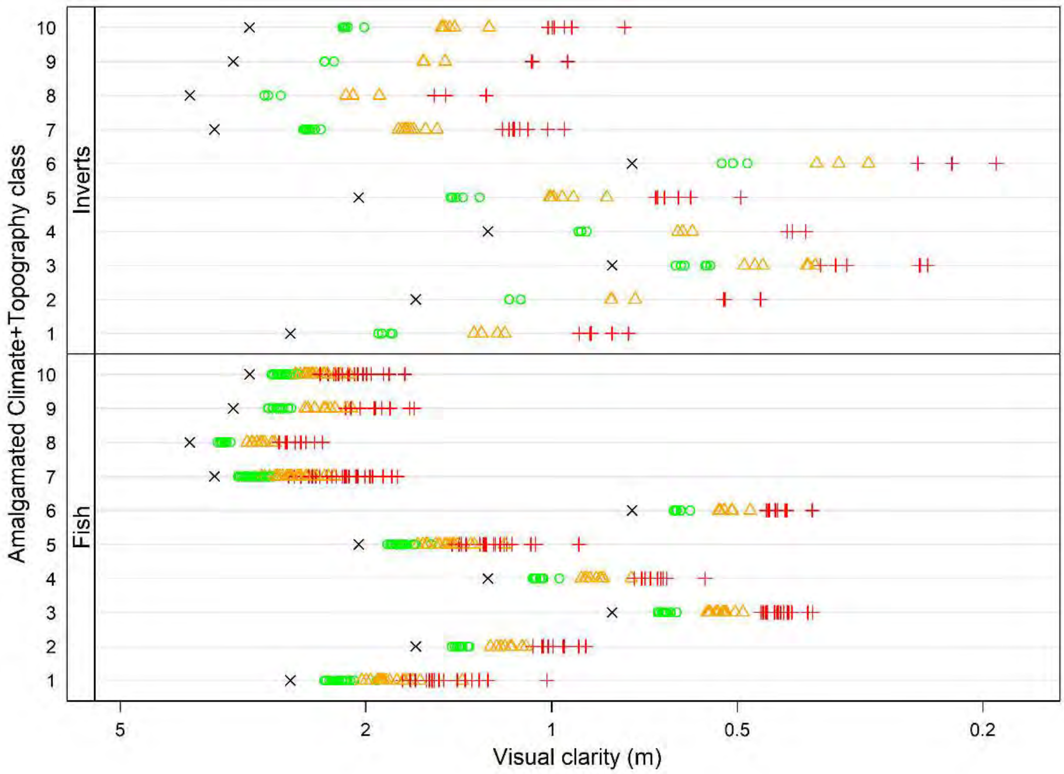


Figure J-36: Visual clarity band thresholds for all landscape settings within SSC classes for fish and macroinvertebrates using the community deviation method. Black crosses indicate the reference sediment state. Green circles represent A/B thresholds (ΔC of -0.066). Orange triangles represent B/C thresholds (ΔC of -0.133). Red crosses represent C/D thresholds (ΔC of -0.20).

Table J-6: Potential band thresholds for visual clarity based on the fish community deviation method. Thresholds are presented visual clarity (m) for the 12 classes at the fourth level of aggregation in the SSC. NAs occur where insufficient data were available within the class to predict reference state. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	A/B threshold	B/C threshold	C/D threshold
L4.1	2.65	2.25	1.88	1.55
L4.2	2.86	2.43	2.02	1.65
L4.3	1.72	1.45	1.21	1.00
L4.4	1.66	1.43	1.22	1.02
L4.5	0.80	0.66	0.53	0.42
L4.6	1.27	1.06	0.87	0.70
L4.7	2.05	1.78	1.53	1.30
L4.8	0.74	0.63	0.53	0.44
L4.9	3.52	3.10	2.71	2.35
L4.10	3.86	3.38	2.93	2.51
L4.11	3.28	2.84	2.43	2.06
L4.12	3.09	2.79	2.51	2.23

Table J-7: Potential band thresholds for visual clarity based on the macroinvertebrate community deviation method. Thresholds are presented visual clarity (m) for the 12 classes at the fourth level of aggregation in the SSC. NAs occur where insufficient data were available within the class to predict reference state. A/B threshold = ΔC of -0.066; B/C threshold = ΔC of -0.133; C/D threshold = ΔC of -0.20.

SSC	Predicted reference state	A/B threshold	B/C threshold	C/D threshold
L4.1	2.65	1.90	1.32	0.89
L4.2	2.86	2.26	1.76	1.34
L4.3	1.72	1.20	0.83	0.55
L4.4	1.66	1.17	0.80	0.52
L4.5	0.80	0.58	0.41	0.28
L4.6	1.27	0.90	0.62	0.41
L4.7	2.05	1.44	0.99	0.66
L4.8	0.74	0.52	0.36	0.25
L4.9	3.52	2.48	1.72	1.15
L4.10	3.86	2.82	2.00	1.38
L4.11	3.28	2.33	1.61	1.07
L4.12	3.09	2.16	1.48	0.97

Appendix K Community deviation thresholds for different levels of acceptable deviation from reference

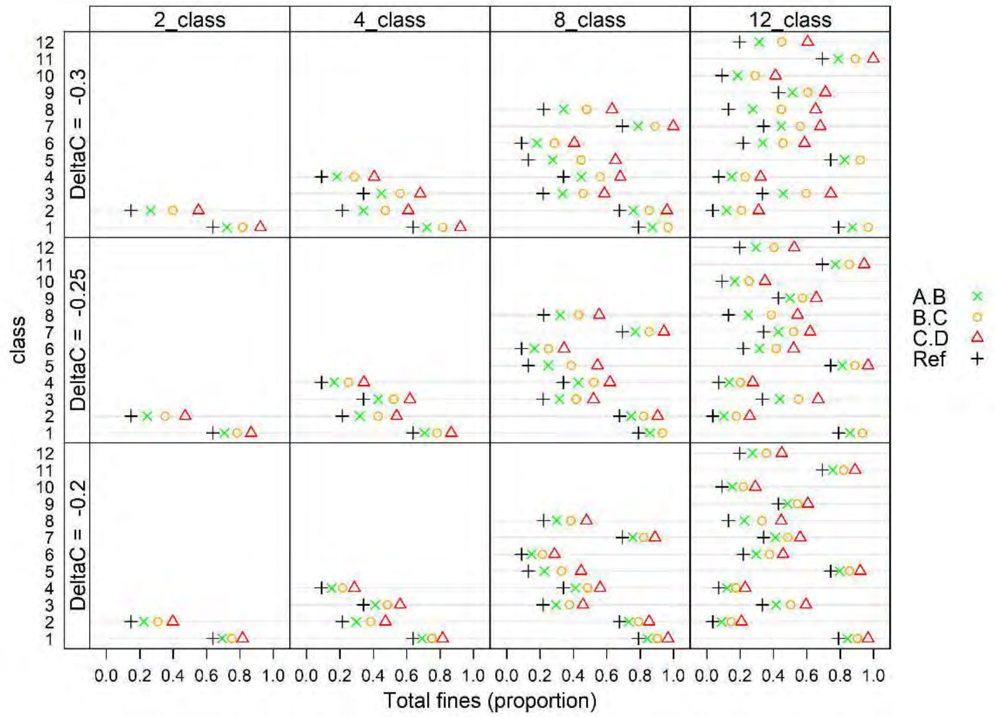


Figure K-1: Illustration of the consequences of choosing different community deviation values for deposited sediment thresholds at each aggregation level of the SSC.

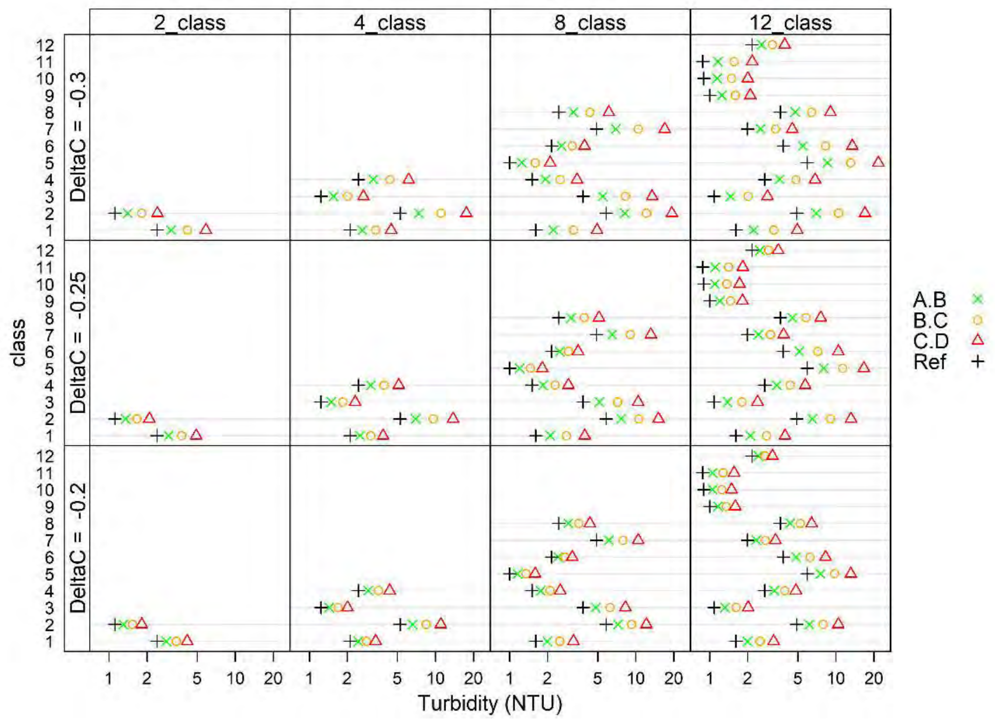


Figure K-2: Illustration of the consequences of choosing different community deviation values for turbidity thresholds at each aggregation level of the SSC.

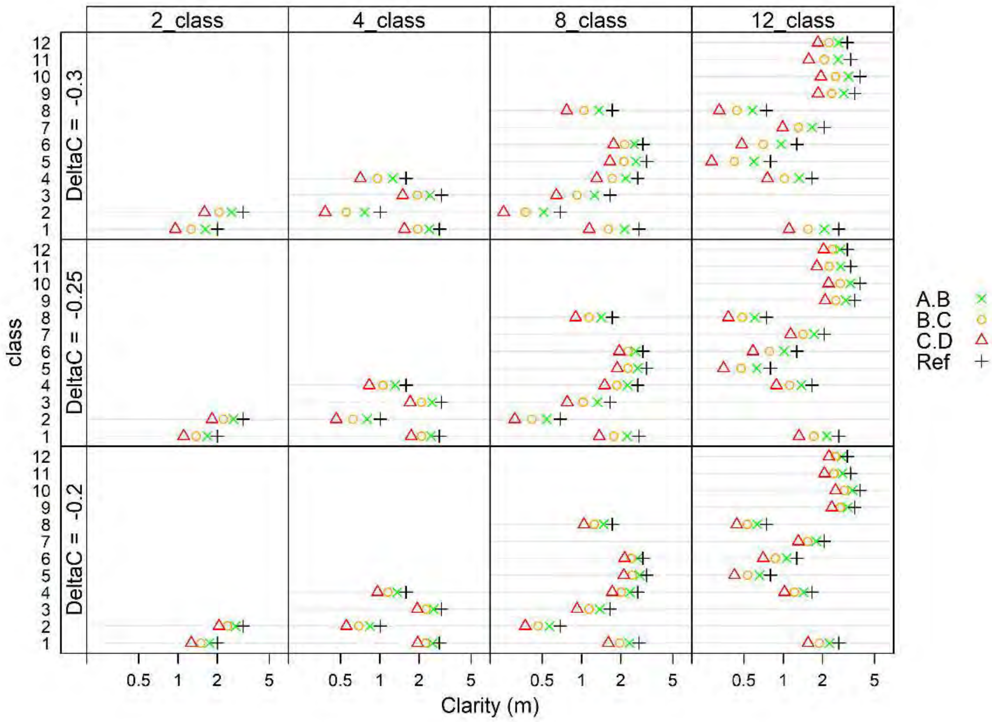


Figure K-3: Illustration of the consequences of choosing different community deviation values for visual clarity thresholds at each aggregation level of the SSC.

Appendix L Weight of evidence scoring tables

Deposited sediment

Analytical method:	**Boosted Regression Trees (BRT)**
Response variables: Macroinvertebrate Community Index (MCI) Sediment Macroinvertebrate Community Index (Sediment MCI) % EPT richness % sediment decreasers	
ESV measure:	SAM2 % cover instream
Relevance score:	+/++
Relevance description: **Biological:** + Primary and secondary consumers; mid food web and so transfer of energy and nutrients in the food web. + Close link to ecosystem function (e.g., organic matter and nutrient cycling). + Relatively long lived and integrate annual impacts. + Wide range in community composition so higher potential to include different sensitivities. + Less mobile and mostly restricted to benthos so will be responding to local conditions. + Recognised indicator of ecosystem health by public. + Mechanistically shown to respond to elevated deposited sediment. +- Includes all taxa (MCI). + Focused on sediment sensitive species (Sediment MCI; % sediment decreasers). **Physical:** -+ 'Run' scale measure spatially disconnected from predominantly riffle sampled macroinvertebrates, but both selected to represent reach-scale. - Only reflects one mechanistic pathway of effect. - Usually one-off observation and so doesn't account for temporal variability. + Is the metric for proposed attribute. **Environmental:** + Dataset is national and relatively broad spatial coverage.	
Reliability score:	+/++
Reliability description: **Design and execution:** + QC for invert processing. + Pseudo-replication accounted for. +- Correlative field survey makes the most of existing data. + Moderate sample size. + Relatively high spatial coverage. - Not all paired observations in time.. **Sample size:** + Moderate number of sites (hundreds) + Relatively high spatial coverage. **Minimise confounding:** + Used landscape variables as surrogates for potential confounding variables. + Used chl-*a* to account for nutrient stressor pathway. - Covariance and interactions not directly accounted for in interpretation of output.	

Specificity:
- Field survey data.
+ Community response to sediment.

Potential or bias:
- Dominated by SOE network which is biased towards wadeable streams in impacted areas.
+ Used SAM2 % cover instream..
- Doesn't account for temporal variability

Standardisation:
+ Standardised data collection and processing protocols for inverts and sediment.
- Sample method not controlled for in the analysis.

Corroboration:
+ Published method.

Transparency:
+- Informative model but thresholds strongly influenced by dataset.
- Subjective visual assignment of thresholds.
- Model approach not widely used.

Peer review:
+ Published method.

Consistency:
+ Standard sampling and processing methods.
- Do not account for temporal variation (space for time).

Consilience:
+ Output makes sense.
+ Aligns to theory: invert metrics go down as sediment goes up.

Suitability score: +

Suitability description:

Bottom-line:
+ Can be used to identify impact initiation and cessation thresholds.

Bands:
+- May be possible to use to inform definition of band thresholds.

Global v classes:
+ Global analysis possible.
- Analysis within classes limited by need for larger sample size (n>100).

Reproducible:
- Reliant on subjective interpretation of impact initiation.

Departure from reference:
+- Could potentially be used for calculating departure from reference.

Analytical method:	Generalised linear modelling (GLM)
Response variables: Macroinvertebrate Community Index (MCI) Sediment Macroinvertebrate Community Index (Sediment MCI) % sediment decreasers	
ESV measure:	SAM2 % cover instream
Relevance score:	+/++
Relevance description: **Biological:** + Primary and secondary consumers; mid food web and so transfer of energy and nutrients in the food web; + close link to ecosystem function (e.g., organic matter and nutrient cycling) + Relatively long lived and integrate annual impacts + Wide range in community composition so higher potential to include different sensitivities + Less mobile and mostly restricted to benthos so will be responding to local conditions + Recognised indicator of ecosystem health by public + Mechanistically shown to respond to elevated deposited sediment +- Includes all taxa (MCI) + Focused on sediment sensitive species (Sediment MCI; % sediment decreasers) **Physical:** -+ 'Run' scale measure spatially disconnected from predominantly riffle sampled inverts, but both selected to represent reach-scale EH - Only reflects one mechanistic pathway of effect - Usually one-off observation and so doesn't account for temporal variability + Is the metric for proposed attribute **Environmental:** + Dataset is national and relatively broad spatial coverage	
Reliability score:	+/++
Reliability description:	
Design and execution: + QC for sediment measure standardised + Pseudo-replication accounted for in analytical method - Field survey data +- Medium sample size +- Reasonable spatial coverage - Not always paired observations in time at site **Sample size:** +- Medium number of samples +- Reasonably broad spatial coverage **Minimise confounding:** + Used landscape variables (REC) as surrogate for possible confounding variables - Doesn't explicitly include confounding variables **Specificity:** - Field survey data + Uses SAM2 % cover instream **Potential for bias:** +- Reasonably broad spatial coverage of both response and driver data - Doesn't use an average over time	

- Limited to wadeable rivers **Standardisation:** - Sampling method not controlled for in analysis method **Transparency:** + It's logical and objective **Consistency:** + Get consistent relationship between ESV and invert community change **Consilience:** + MCI goes down as ESV goes up
Suitability score: ++/+++
Suitability description: **Bottom-line:** +- Can be used to identify threshold but reliant on normative decision on acceptable deviation **Bands:** +- Can be used to identify threshold but reliant on normative decision on acceptable deviation **Global v classes:** + Global analysis possible + Analysis within classes possible **Reproducible:** + Thresholds derived based on numeric criteria **Departure from reference:** + Suited to calculating departure from reference

Analytical method:	**Community deviation method**
Response variables: Macroinvertebrate taxa presence/absence Fish taxa presence/absence	
ESV measure:	SAM2 % cover instream/NZFFD % fines
Relevance score:	++
Relevance description:	
Macroinvertebrates **Biological:** + Middle of food web and so integrate impacts at lower trophic levels + Food source for higher trophic levels - Shorter lived and so don't integrate impacts over longer period - Immobile and so less able to avoid conditions they don't like - Immobile and so less able to find refuge during extreme events and so observed may reflect disturbance regime more + Immobile and so responding to local conditions + Mechanistically shown to respond to elevated deposited sediment (impacts on	**Fish** **Biological:** + Top end of food web and so integrate impacts at lower trophic levels +- Longer lived and so integrate impacts over longer period + Mobile and so more likely to avoid conditions they don't like + Mobile and so more able to find refuge during extreme events - Mobile and so may not be responding to local conditions + Recognised indicator of ecosystem health by public + Mechanistically shown to respond to elevated deposited sediment (impacts on

reproduction/microhabitat suitability/food availability etc.) - Only decreasers used in deltaC calculations - Presence/absence less representative/sensitive end-point **Physical:** - Run scale measure and so doesn't reflect sediment conditions across all mesohabitats - Doesn't reflect 'embededness' or depth of sediment, which may functionally be more direct cause - One off observation and so doesn't account for temporal variability + Metric proposed to be used for implementing attribute **Environment:** +- Accounts for some landscape influences that may structure invert communities + Dataset has reasonably broad - Dataset has some biases in spatial coverage in NZ river network	reproduction/microhabitat suitability/food availability etc.) - Only decreasers used in deltaC calculations - Presence/absence less representative/sensitive end-point **Physical:** + 'Reach' scale measure and so reflects sediment conditions across all mesohabitats - Doesn't reflect 'embededness' or depth of sediment, which may functionally be more direct cause - One off observation and so doesn't account for temporal variability + Equivalence with metric that you're proposing to be used for monitoring **Environment:** + Accounts for landscape influences that structure fish communities + Dataset has broad and unbiased spatial coverage in NZ river network
Reliability score: ++	
Reliability description:	
Macroinvertebrates **Design and execution:** + QC for sediment measure standardised + Pseudo-replication accounted for in analytical method - Field survey data +- Medium sample size +- Reasonable spatial coverage - Not always paired observations in time at site **Sample size:** +- Medium number of samples +- Reasonably broad spatial coverage **Minimise confounding:** + Used landscape variables (REC) as surrogate for possible confounding variables - Doesn't explicitly include confounding variables **Specificity:** - Field survey data + Focused on decreasers + Uses instreanVis **Potential for bias:** +- Reasonably broad spatial coverage of both response and driver data	**Fish** **Design and execution:** - QC lower for sediment measures + Pseudo-replication accounted for - Field survey data + High sample size + High spatial coverage + Paired observations in time at site **Sample size:** + Large number of samples + Broad spatial coverage **Minimise confounding:** + Used landscape variables (REC) as surrogate for possible confounding variables - Doesn't explicitly include confounding variables **Specificity:** - Field survey data + Focused on decreasers **Potential for bias:** + Broad spatial coverage of both response and driver data - Doesn't use instreamVis - Doesn't use an average over time

- Doesn't use an average over time - Limited to wadeable rivers **Standardisation:** + Use of presence/absence - Sampling method not controlled for in analysis method **Corroboration:** - Bespoke method that has not been used elsewhere **Transparency:** + It's logical and objective - It can be less intuitive to understand - deltaC hard to define qualitatively **Peer review:** - It's a bespoke method that has not been formally peer-reviewed and published + Peer reviewed within team **Consistency:** + Get consistent relationship between ESV and invert community change **Consilience:** + invert decreasers go down as ESV goes up	- Potential under-representation of large rivers **Standardisation:** + Use of presence/absence + Sampling method controlled for in analysis method **Corroboration:** - Bespoke method that has not been used elsewhere **Transparency:** + It's logical and objective - It can be less intuitive to understand - deltaC hard to define qualitatively **Peer review:** - It's a bespoke method that has not been formally peer-reviewed and published + Peer reviewed within team **Consistency:** + Get consistent relationship between ESV and fish community change **Consilience:** + Fish go down as ESV goes up
Suitability score: ++/+++	
Suitability description: **Bottom-line:** +- Can be used to identify threshold but reliant on normative decision on acceptable deviation **Bands:** +- Can be used to identify threshold but reliant on normative decision on acceptable deviation **Global v classes:** + Global analysis possible + Analysis within classes possible **Reproducible:** + Thresholds derived based on numeric criteria **Departure from reference:** + Suited to calculating departure from reference	

Analytical method:	**Quantile regression**[14]
Response variables: Macroinvertebrate Community Index (MCI) Sediment Macroinvertebrate Community Index (Sediment MCI) % EPT richness % sediment decreasers	
ESV measure:	SAM2 % cover instream
Relevance score:	++
Relevance description: **Biological:** + Primary and secondary consumers; mid food web and so transfer of energy and nutrients in the food web + Close link to ecosystem function (e.g., organic matter and nutrient cycling) + Relatively long lived and integrate annual impacts + Wide range in community composition so higher potential to include different sensitivities + Less mobile and mostly restricted to benthos so will be responding to local conditions + Recognised indicator of ecosystem health by public + Mechanistically shown to respond to elevated deposited sediment +- Includes all taxa (MCI) + Focused on sediment sensitive species (Sediment MCI; % sediment decreasers) **Physical:** -+ 'Run' scale measure spatially disconnected from predominantly riffle sampled macroinvertebrates, but both selected to represent reach-scale - Only reflects one mechanistic pathway of effect - Usually one-off observation and so doesn't account for temporal variability + Is the metric for proposed attribute **Environmental:** + Dataset is national and relatively broad spatial coverage	
Reliability score:	+
Reliability description:	
Design and execution: + QC for invert processing - Pseudo-replication not accounted for +- Correlative field survey makes the most of existing data + Moderate sample size + Relatively high spatial coverage - Not all paired observations in time **Sample size:** + Moderate number of sites (hundreds) + Relatively high spatial coverage **Minimise confounding:** + Use of quantiles incorporates impacts of confounding variables	

[14] Reported in Depree et al. (2018)

- Covariance and interactions not directly accounted for **Specificity:** - Field survey data + Community response to sediment **Potential or bias:** - Dominated by SOE network which is biased towards wadeable streams in impacted areas + Used instreamVis - Doesn't account for temporal variability **Standardisation:** + Standardised data collection and processing protocols for inverts and sediment - Sample method not controlled for in the analysis **Corroboration:** + Published method **Transparency:** - Derived thresholds dependent on subjective choice of acceptable deviation and choice of benchmark for change + Reproducible and choice of acceptable deviation can be balanced with other values **Peer review:** + Published method **Consistency:** + Standard sampling and processing methods - Do not account for temporal variation (space for time) **Consilience:** + Output makes sense + Aligns to theory: invert metrics go down as sediment goes up
Suitability score: +
Suitability description: **Bottom-line:** +- Can be used to identify threshold but reliant on normative decision on acceptable deviation **Bands:** +- Can be used to identify threshold but reliant on normative decision on acceptable deviation **Global v classes:** + Global analysis possible +- Analysis within classes limited by data availability **Reproducible:** + Thresholds derived based on numeric criteria **Departure from reference:** + Suited to calculating departure from reference

Analytical method:	**Gradient Forest (GF)**[15]
Response variables: Macroinvertebrate taxa proportional relative abundance	
ESV measure:	SAM2 % cover instream
Relevance score:	+
Relevance description: **Biological:** + Primary and secondary consumers; mid food web and so transfer of energy and nutrients in the food web; + close link to ecosystem function (e.g., organic matter and nutrient cycling) + Relatively long lived and integrate annual impacts + Wide range in community composition so higher potential to include different sensitivities + Less mobile and mostly restricted to benthos so will be responding to local conditions - Includes only a subset of the taxa; rare often excluded **Physical:** -+ 'Run' scale measure spatially disconnected from predominantly riffle sampled inverts, but both selected to represent reach-scale EH - Only reflects one mechanistic pathway of effect - Usually one-off observation and so doesn't account for temporal variability + Is the metric for proposed attribute **Environmental:** + Dataset is national and relatively broad spatial coverage	
Reliability score:	+
Reliability description:	
Design and execution: + QC for invert processing - Pseudo-replication not accounted for - Correlative field survey makes the most of existing data + Moderate sample size + Relatively high spatial coverage - Not all paired observations in time **Sample size:** + Moderate number of sites (hundreds) + Relatively high spatial coverage - Smaller dataset than BRT because CHLA required **Minimise confounding:** + Used landscape variables as surrogates for potential confounding variables + Used chl *a* to account for nutrient stressor pathway - Covariance and interactions not directly accounted for in interpretation of output **Specificity:** - Field survey data + Combined output based on individual taxa response to sediment	

[15] Reported in Depree et al. (2018)

Potential or bias: - Dominated by SOE network which is biased towards wadeable streams in impacted areas + Used instreamVis - Doesn't account for temporal variability **Standardisation:** + Standardised data collection and processing protocols for inverts and sediment - Sample method not controlled for in the analysis **Corroboration:** + Published method **Transparency:** +- Informative model but thresholds strongly influenced by dataset + Model assigned taxa tolerance/sensitivity thresholds - Model always identifies a turnover function, which may include non-informative/sensible taxa responses - Model approach not widely used **Peer review:** + Published method **Consistency:** + Standard sampling and processing methods - Do not account for temporal variation (space for time) **Consilience:** + Output makes sense + Aligns to theory: invert taxa respond to sediment
Suitability score: -
Suitability description: **Bottom-line:** + Can be used to identify impact initiation and cessation thresholds **Bands:** +- May be possible to use to inform definition of band thresholds **Global v classes:** + Global analysis possible - Analysis within classes limited by need for larger sample size (n>100) **Reproducible:** - Reliant on subjective interpretation of impact initiation/cessation **Departure from reference:** - Not well suited for calculating departure from reference

Suspended sediment

Analytical method:	**Boosted Regression Trees (BRT)**
Response variables: Macroinvertebrate Community Index (MCI) Sediment Macroinvertebrate Community Index (Sediment MCI) % EPT richness % sediment decreasers	
ESV measure:	Turbidity
Relevance score:	+
Relevance description: **Biological:** + Primary and secondary consumers; mid food web and so transfer of energy and nutrients in the food web + Close link to ecosystem function (e.g., organic matter and nutrient cycling) + Relatively long lived and integrate annual impacts + Wide range in community composition so higher potential to include different sensitivities + Less mobile and mostly restricted to benthos so will be responding to local conditions - Mostly restricted to benthos + Recognised indicator of ecosystem health by public +- Includes all taxa (MCI) + Focused on sediment sensitive species (Sediment MCI; % sediment decreasers) **Physical:** -+ Annual median and so accounts for long-term temporal variation - Annual median and so does not account for shorter term temporal variations that may be mechanistically more relevant + Is the metric for proposed attribute **Environmental:** + Dataset is national and relatively broad spatial coverage	
Reliability score:	+/++
Reliability description:	
Design and execution: + QC for invert processing + Pseudo-replication accounted for +- Correlative field survey makes the most of existing data + Moderate sample size + Relatively high spatial coverage - Observations paired in space **Sample size:** + Moderate number of sites (hundreds) + Relatively high spatial coverage **Minimise confounding:** + Used landscape variables as surrogates for potential confounding variables + Used chl-*a* to account for nutrient stressor pathway	

- Covariance and interactions not directly accounted for in interpretation of output **Specificity:** - Field survey data + Community response to sediment **Potential or bias:** - Dominated by SOE network which is biased towards wadeable streams in impacted areas + Used annual median turbidity - Doesn't account for short-term temporal variability **Standardisation:** + Standardised data collection and processing protocols for inverts and sediment - Sample method not controlled for in the analysis **Corroboration:** + Published method **Transparency:** +- Informative model but thresholds strongly influenced by dataset - Subjective visual assignment of thresholds - Model approach not widely used **Peer review:** + Published method **Consistency:** + Standard sampling and processing methods - Do not account for short-term temporal variation **Consilience:** + Output makes sense + Aligns to theory: invert metrics go down as sediment goes up
Suitability score: +
Suitability description: **Bottom-line:** + Can be used to identify impact initiation and cessation thresholds **Bands:** +- May be possible to use to inform definition of band thresholds **Global v classes:** + Global analysis possible - Analysis within classes limited by need for larger sample size (n>100) **Reproducible:** - Reliant on subjective interpretation of impact initiation **Departure from reference:** +- Could potentially be used for calculating departure from reference

Analytical method:	**Quantile regression**
Response variables: Macroinvertebrate Community Index (MCI) % EPT richness Taxa absolute abundance Taxa richness	
ESV measure:	Turbidity/Visual clarity
Relevance score:	+
Relevance description: **Biological:** + Primary and secondary consumers; mid food web and so transfer of energy and nutrients in the food web; + close link to ecosystem function (e.g., organic matter and nutrient cycling) + Relatively long lived and integrate annual impacts + Wide range in community composition so higher potential to include different sensitivities + Less mobile and so will be responding to local conditions - Mostly restricted to benthos + Recognised indicator of ecosystem health by public +- Includes many taxa - Some key species groups (e.g., mussels and crayfish) not included **Physical:** -+ Annual median and so accounts for long-term temporal variation - Annual median and so does not account for shorter term temporal variations that may be mechanistically more relevant + Is the metric for proposed attribute **Environmental:** + Dataset is national and relatively broad spatial coverage	
Reliability score:	++
Reliability description:	
Design and execution: + QC for invert processing - Pseudo-replication not accounted for +- Correlative field survey makes the most of existing data + Moderate sample size + Relatively high spatial coverage - Not all paired observations in time - QR sensitive to outliers and 0 counts - Uncertainty associated with mathematical model used to fit QR curves **Sample size:** + Moderate number of sites (hundreds) + Relatively high spatial coverage **Minimise confounding:** + Use of quantiles incorporates impacts of confounding variables - Covariance and interactions not directly accounted for	

Specificity:

- Field survey data

+ Individual taxa responses to sediment used

- May be a more variable measure than taxa richness or 'sensitive' species measures.

Potential or bias:

- Dominated by SOE network which is biased towards wadeable streams in impacted areas

+ Used annual median turbidity

- Doesn’t account for short-term temporal variability

Standardisation:

+ Standardised data collection and processing protocols for inverts and sediment

- Sample method not controlled for in the analysis

+ NRWQN data are quantitative

Corroboration:

+ Published method

Transparency:

- Arbitrary basis for 'adverse effect' threshold

- Derived thresholds dependent on subjective choice of acceptable deviation and choice of benchmark for change

+ Reproducible and choice of acceptable deviation can be balanced with other values

Peer review:

+ Published method

Consistency:

+ Standard sampling and processing methods

- Do not account for short-term temporal variation

Consilience:

+ Output makes sense

+ Aligns to theory: invert metrics go down as sediment goes up

Suitability score: ++

Suitability description:

Bottom-line:

+- Can be used to identify threshold but reliant on normative decision on acceptable deviation

Bands:

+- Can be used to identify threshold but reliant on normative decision on acceptable deviation

Global v classes:

+ Global analysis possible

+- Analysis within classes limited by data availability

Reproducible:

+ Thresholds derived based on numeric criteria

Departure from reference:

+ Suited to calculating departure from reference

Analytical method:	**Generalised linear modelling (GLM)**
Response variables: Macroinvertebrate Community Index (MCI) Sediment Macroinvertebrate Community Index (Sediment MCI) % sediment decreasers	
ESV measure:	Turbidity/Visual clarity
Relevance score:	+/++
Relevance description: **Biological:** + Middle of food web and so integrate impacts at lower trophic levels + Food source for higher trophic levels - Shorter lived and so don't integrate impacts over longer period - Immobile and so less able to avoid conditions they don't like - Immobile and so less able to find refuge during extreme events and so observed may reflect disturbance regime more + Immobile and so responding to local conditions - Mechanistically shown to respond to elevated deposited sediment (impacts on reproduction/microhabitat suitability/food availability etc.) + Community level response **Physical:** + Widely measured ESVs + Majority of literature uses turbidity, hence more equivalence with other lines of evidence - Mechanistically visual clarity stronger than turbidity + Representative of TSS which has causal impacts on inverts (e.g., gill clogging) - Does not reflect impact of short-term variations in turbidity/visual clarity that may be functionally more significant **Environment:** +- Accounts for some landscape influences that may structure invert communities + Dataset has reasonably broad - Dataset has some biases in spatial coverage in NZ river network	
Reliability score:	+/++
Reliability description: **Design and execution:** + ESV representative of long-term median which is what is proposed to be used for implementation + Pseudo-replication accounted for - Field survey data + High sample size + High spatial coverage + Paired observations in time at site **Sample size:** + Reasonably large number of samples + Broad spatial coverage **Minimise confounding:**	

+ Used landscape variables (REC) as surrogate for possible confounding variables
- Doesn’t explicitly include confounding variables

Specificity:

- Field survey data
+ Focused on decreasers

Potential for bias:

+ Broad spatial coverage of both response and driver data
+ Uses an average over time
- Potential under-representation of large rivers

Standardisation:

- Sampling method not controlled for in analysis method

Corroboration:

- Widely used statistical method

Transparency:

+ It’s logical and objective
- It can be less intuitive to understand

Peer review:

+ Peer reviewed method

Consistency:

+ Get consistent relationship between ESV and invert community change

Consilience:

+ Inverts go down as ESV gets worse

Suitability score: ++/+++

Suitability description:

Bottom-line:

+- Can be used to identify threshold but reliant on normative decision on acceptable deviation

Bands:

+- Can be used to identify threshold but reliant on normative decision on acceptable deviation

Global v classes:

+ Global analysis possible
+ Analysis within classes possible

Reproducible:

+ Thresholds derived based on numeric criteria

Departure from reference:

+ Suited to calculating departure from reference

Analytical method:	Community deviation method
Response variables: Macroinvertebrate taxa presence/absence Fish taxa presence/absence	
ESV measure:	Turbidity/Visual clarity
Relevance score:	++
Relevance description:	
Macroinvertebrates **Biological:** + Middle of food web and so integrate impacts at lower trophic levels + Food source for higher trophic levels - Shorter lived and so don't integrate impacts over longer period - Immobile and so less able to avoid conditions they don't like - Immobile and so less able to find refuge during extreme events and so observed may reflect disturbance regime more + Immobile and so responding to local conditions + Mechanistically shown to respond to elevated deposited sediment (impacts on reproduction/microhabitat suitability/food availability etc.) - Only decreasers used in deltaC calculations - Presence/absence less representative/sensitive end-point **Physical:** +- Less widely measured ESV - Majority of literature uses turbidity, hence less equivalence with other lines of evidence + Mechanistically may be stronger driver than turbidity +- Modelled data - Less representative of TSS which have causal impacts on inverts (e.g., gill clogging) + Annual median may be more appropriate for visual clarity than turbidity because functionally more relevant **Environment:** +- Accounts for some landscape influences that may structure invert communities + Dataset has reasonably broad - Dataset has some biases in spatial coverage in NZ river network	**Fish** **Biological:** + Top end of food web and so integrate impacts at lower trophic levels + Longer lived and so integrate impacts over longer period + Mobile and so more likely to avoid conditions they don't like + Mobile and so more able to find refuge during extreme events - Mobile and so may not be responding to local conditions + Recognised indicator of ecosystem health by public + Mechanistically shown to respond to elevated suspended sediment (impacts on feeding success/migration etc.) - Only decreasers used in deltaC calculations **Physical:** - Less widely measured ESV - Majority of literature uses turbidity, hence less equivalence with other lines of evidence + Mechanistically more direct driver for visual feeders + Annual median may be more appropriate for visual clarity than turbidity because functionally more relevant **Environment:** + Accounts for landscape influences that structure fish communities + Dataset has fairly broad spatial coverage in environment of interest - Modelled so available everywhere

Reliability score: ++

Reliability description:

Macroinvertebrates

Design and execution:

+ ESV representative of long-term median which is what is proposed to be used for implementation

+ Pseudo-replication accounted for

- Field survey data

+ High sample size

+ High spatial coverage

+ Paired observations in time at site

Sample size:

+ Reasonably large number of samples

+ Broad spatial coverage

Minimise confounding:

+ Used landscape variables (REC) as surrogate for possible confounding variables

- Doesn’t explicitly include confounding variables

Specificity:

- Field survey data

+ Focused on decreasers

Potential for bias:

+ Broad spatial coverage of both response and driver data

+ Uses an average over time

- Potential under-representation of large rivers

Standardisation:

+ Use of presence/absence

- Sampling method not controlled for in analysis method

Corroboration:

- Bespoke method that has not been used elsewhere

Transparency:

+ It’s logical and objective

- It can be less intuitive to understand

- deltaC hard to define qualitatively

Peer review:

- It’s a bespoke method that has not been formally peer-reviewed and published

+ Peer reviewed within team

Consistency:

+ Get consistent relationship between ESV and fish community change

Consilience:

+ Fish go down as ESV gets worse

Fish

Design and execution:

- Modelled data for ESV

+ ESV representative of long-term median which is what is proposed to be used for implementation

+ Pseudo-replication accounted for

- Field survey data

+ High sample size

+ High spatial coverage

- Not paired observations in time at site

Sample size:

+ Large number of samples

+ Broad spatial coverage

Minimise confounding:

+ Used landscape variables (REC) as surrogate for possible confounding variables

- Doesn’t explicitly include confounding variables

Specificity:

- Field survey data

+ Focused on decreasers

Potential for bias:

+ Broad spatial coverage of both response and driver data

+ Uses an average over time

- Potential under-representation of large rivers

Standardisation:

+ Use of presence/absence

+ Sampling method controlled for in analysis method

Corroboration:

- Bespoke method that has not been used elsewhere

Transparency:

+ It’s logical and objective

- It can be less intuitive to understand

- deltaC hard to define qualitatively

Peer review:

- It’s a bespoke method that has not been formally peer-reviewed and published

+ Peer reviewed within team

Consistency:

+ Get consistent relationship between ESV and fish community change

Consilience:

	+ Fish go down as ESV gets worse

Suitability score:	+++

Suitability description:

Bottom-line:

+- Can be used to identify threshold but reliant on normative decision on acceptable deviation

Bands:

+- Can be used to identify threshold but reliant on normative decision on acceptable deviation

Global v classes:

+ Global analysis possible

+ Analysis within classes possible

Reproducible:

+ Thresholds derived based on numeric criteria

Departure from reference:

+ Suited to calculating departure from reference

Analytical method:	**Extirpation analysis and SSD**
Response variables:	
Macroinvertebrate taxa	
ESV measure:	Turbidity/Visual clarity
Relevance score:	++

Relevance description:

Biological:

+ Middle of food web and so integrate impacts at lower trophic levels

+ Food source for higher trophic levels

- Shorter lived and so don't integrate impacts over longer period

- Immobile and so less able to avoid conditions they don't like

- Immobile and so less able to find refuge during extreme events and so observed may reflect disturbance regime more

+ Immobile and so responding to local conditions

- Mechanistically shown to respond to elevated deposited sediment (impacts on reproduction/microhabitat suitability/food availability etc.)

+ Reflects individual level response

Physical:

+ Widely measured ESVs

+ Majority of literature uses turbidity, hence more equivalence with other lines of evidence

- Mechanistically visual clarity stronger than turbidity

+ Representative of TSS which has causal impacts on inverts (e.g., gill clogging)

- Does not reflect impact of short-term variations in turbidity/visual clarity that may be functionally more significant

Environment:

- Does not account for some landscape influences that may structure invert communities

+ Dataset has reasonably broad coverage

Reliability score:	++

Reliability description:

Design and execution:

+ ESV representative of long-term median which is what is proposed to be used for implementation

- Pseudo-replication not accounted for

- Field survey data

+ High sample size

+ High spatial coverage

+ Paired observations in time at site

+ Robust, model-free technique

Sample size:

+ Reasonably large number of samples

+ Broad spatial coverage

Minimise confounding:

+ Used landscape variables (REC) as surrogate for possible confounding variables

- Doesn't explicitly include confounding variables

Specificity:

- Field survey data

+ Focused on decreasers

Potential for bias:

+ Broad spatial coverage of both response and driver data

+ Uses an average over time

- Potential under-representation of large rivers

Standardisation:

- Sampling method not controlled for in analysis method

Corroboration:

- Widely used method for deriving numeric criteria with standard rules for application

Transparency:

+ It's logical and objective

- It can be less intuitive to understand

Peer review:

+ Peer reviewed method

Consistency:

+ Get consistent relationship between ESV and invert community change

Consilience:

+ Inverts go down as ESV gets worse

Suitability score:	++/+++

Suitability description:

Bottom-line:

+- Can be used to identify threshold but reliant on normative decision on acceptable deviation

Bands:

+- Can be used to identify threshold but reliant on normative decision on acceptable deviation

Global v classes:

+ Global analysis possible

+ Analysis within classes possible

Reproducible:

+ Thresholds derived based on numeric criteria

Departure from reference:

- Not typically used to calculate departure from reference

Appendix M Analysis of temporal variability in deposited sediment measurements

Introduction

We explored temporal variation in the % cover of deposited sediment metrics to determine the minimum/recommended number of sampling events required to assess compliance against proposed deposited sediment attribute thresholds.

Methods

We identified sites from the collated database where the % cover of fine sediment had been repeatedly measured over time using either the bankside (SAM 1) or instream (SAM2) visual assessment methods from Clapcott et al. (2011). For both SAM1 and SAM2 methods we calculated the variance for all sites with ≥4 temporal samples. We then used the relationship between mean and standard deviation to estimate the number of samples required to estimate the mean within an absolute 10% fine sediment cover margin of error (i.e., +/- 5%). We used this approach to account for observer error; visual assessments can only assess % deposited sediment cover in 5% intervals at best (more realistically at 10% intervals).

Results

Temporal variation in the standard deviation of % sediment cover ranged from 0.6 to 44% when measured using the visual bankside method (SAM1). For SAM2 data (instream), the standard deviation varied from 0% to 50%. There was a significant difference in the variance observed at reference sites compared to non-reference sites for SAM1 (Welch's t-test, $p < 0.001$; Figure M-1) and SAM2 data (Welch's t-test $p = 0.007$).

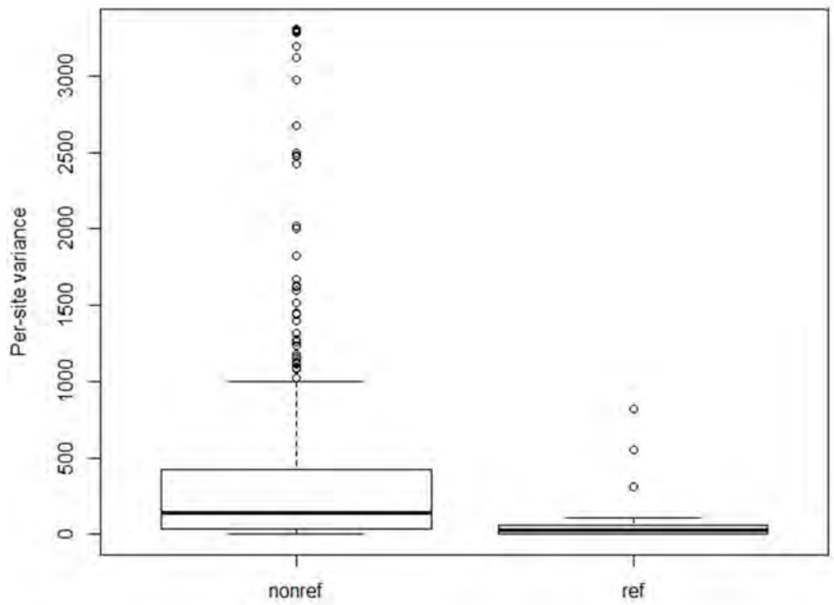


Figure M-1: Average temporal variation in % cover of deposited fine sediment measured using the visual bankside (SAM1) method at reference sites and non-reference sites.

The variance was strongly dependent on the mean sediment cover: sites with mean sediment closer to 50% had higher variance (Figure M-2). Based on this relationship, the number of samples required to estimate up to 30% sediment cover within +/-5% was 24 samples (i.e., two years of monthly samples. At most, 37 samples were needed to accurately estimate mean values around 50% cover. However, the loss of precision following collection of 24 samples when the mean sediment cover is 50% was only 6.2%, which is not noticeably different from our selected 5% error. As such, it appears 24 monthly samples would enable estimation of the mean sediment cover using the SAM2 method sufficiently accurately.

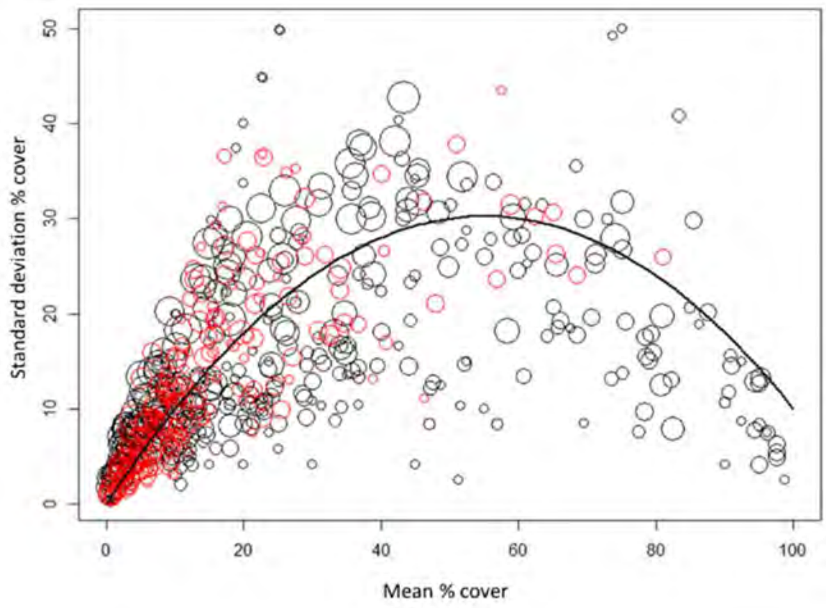


Figure M-2: The linear relationship (with a quadratic term; R^2 = 0.85) between mean and standard deviation in % sediment cover. Red and black circles show % cover bankside and % cover instream data respectively. Circle size is relative to sample number for each site.

Appendix N Comparison between deposited sediment measures

We have used two main measures of deposited sediment for the analyses in this project; NZFFD % fines and SAM2 % cover instream. The NZFFD % fines data are determined from an instream visual assessment of the percentage cover of different substrate sizes across the reach (i.e., all habitats) surveyed for fish. The SAM2 % cover instream measurement is a more systematic assessment of the cover of fine sediments in run habitats using an instream viewer. We had concerns over the equivalence of the two measures and so carried out exploratory analyses to understand their comparability.

We firstly compared where data using the two different measures had been collected (Figure N-1 to Figure N-4). Visual inspection of the histograms suggested that patterns across landscape settings in these two sets of observations were broadly similar, but that greater proportions of fine sediment cover were observed in the NZFFD data, especially for warmer and lowland settings (Figure N-1 and Figure N-2). However, further investigation indicated a large discrepancy between the size of rivers within the two data sets (Figure N-3 and Figure N-4). NZFFD % fines observations were located across a broader range of river sizes (as represented by Strahler stream order) and, therefore, included observations from many more smaller rivers in comparison with the independent SAM2 % cover instream observations.

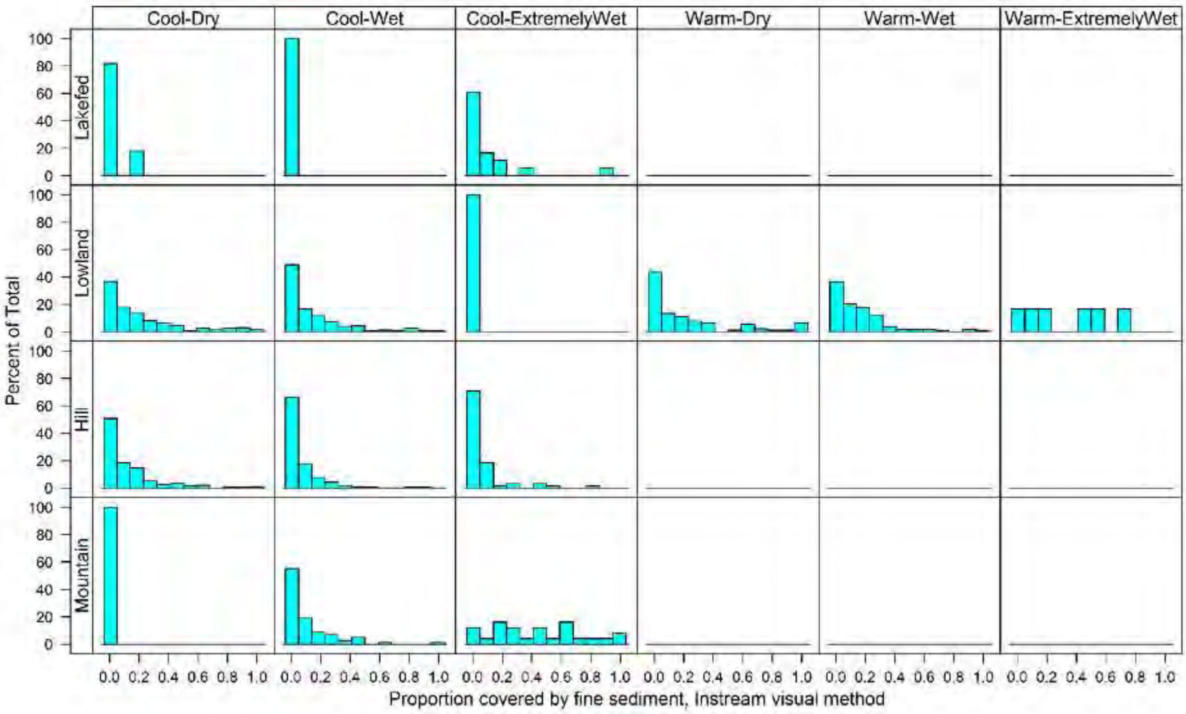


Figure N-1: Histograms of areal cover of deposited fine sediment observed using the instream visual method (SAM2) by REC climate and topography classes. These data are from the deposited sediment dataset assembled by Depree et al. (2018).

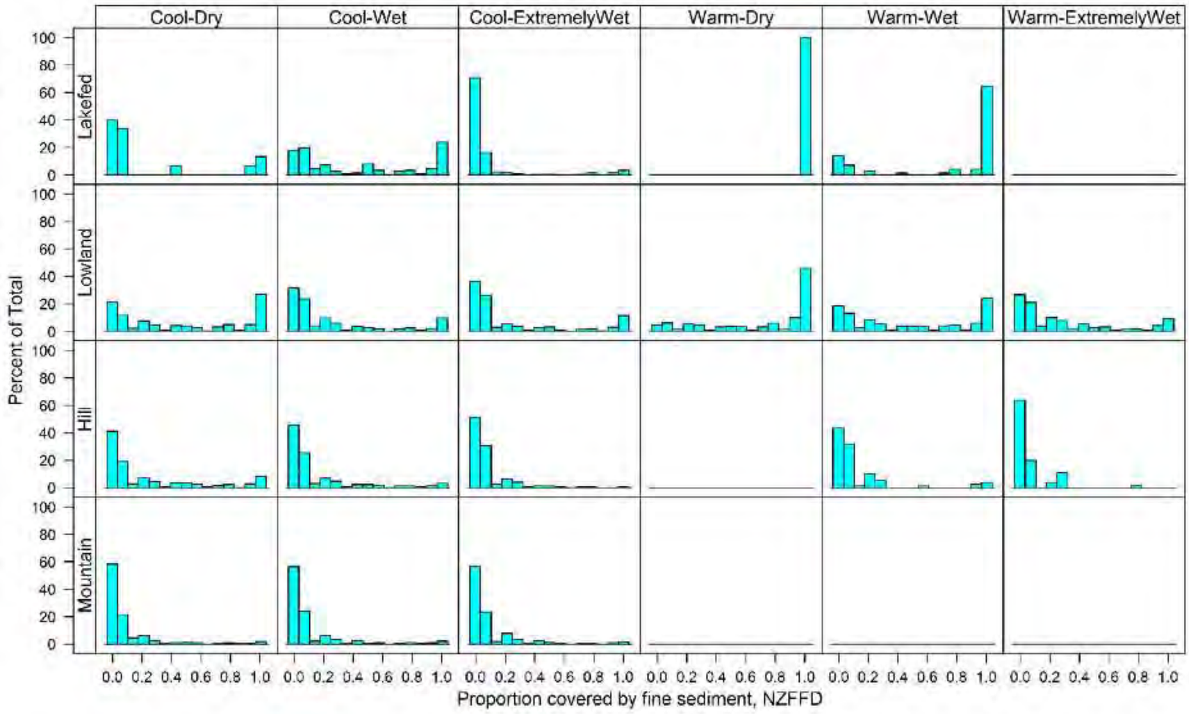


Figure N-2: Histograms of areal cover of deposited fine sediment recorded in the NZFFD by REC climate and topography classes.

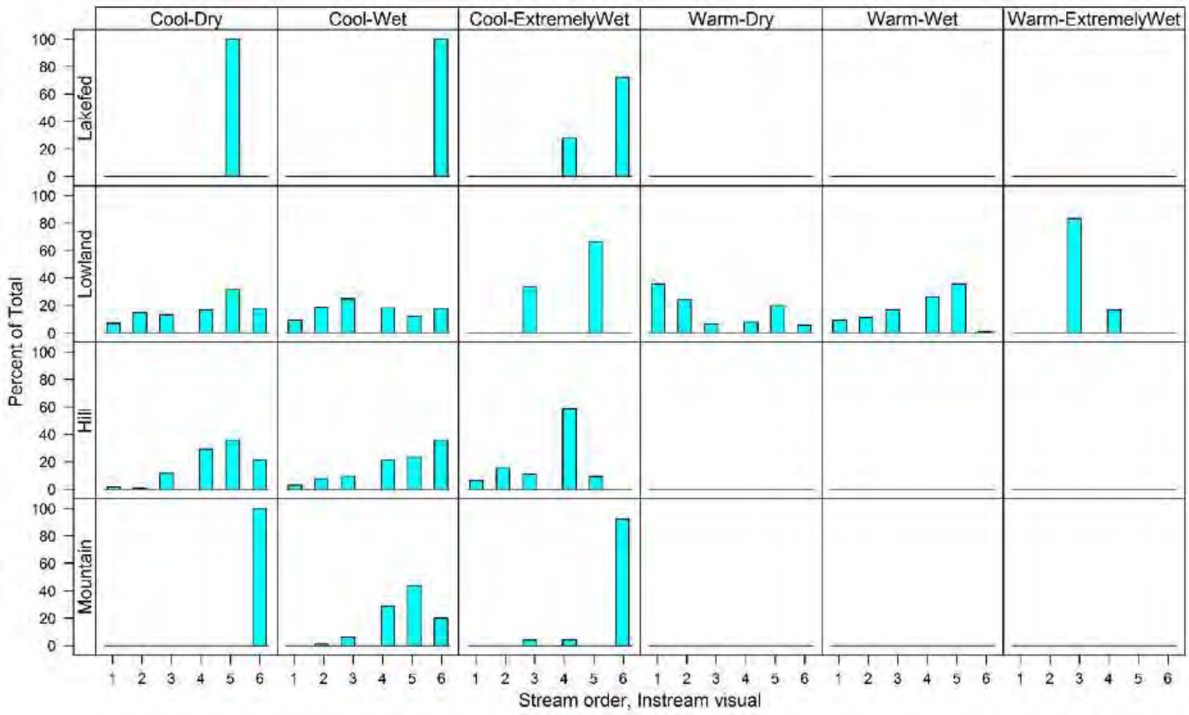


Figure N-3: Histogram of stream orders from which deposited fine sediment cover has been observed independently using the SAM2 instream visual method. Stream order 6 represents orders 6 and above.

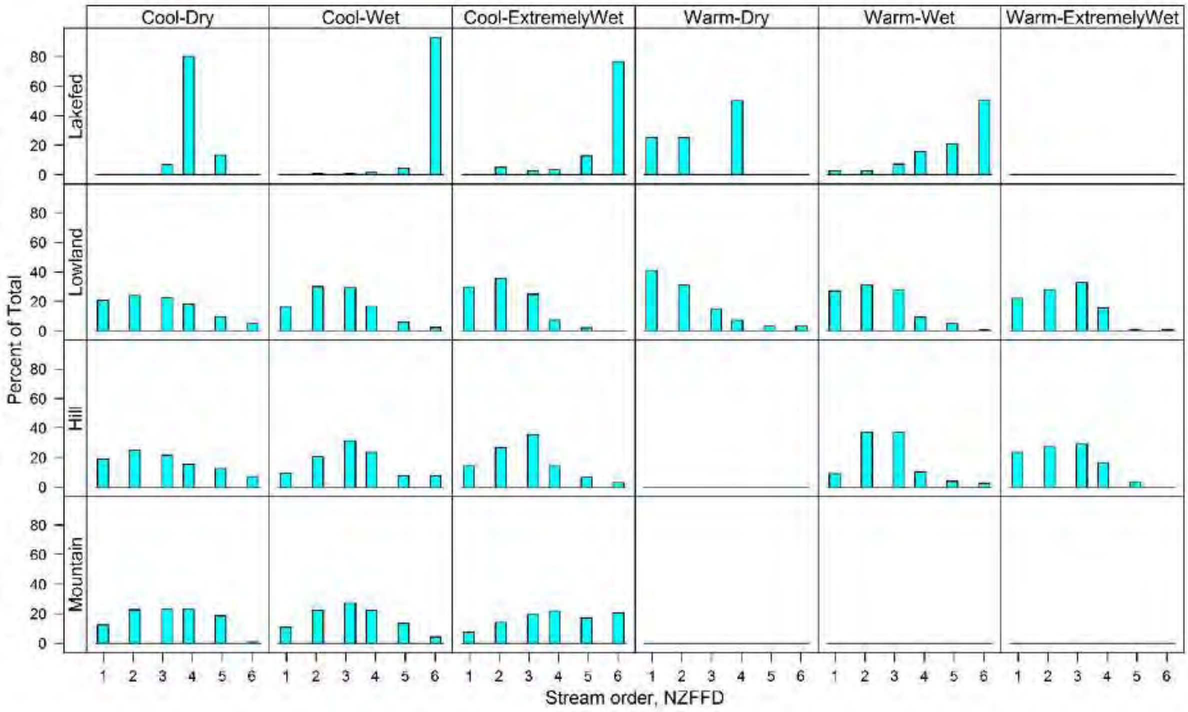


Figure N-4: Histogram of stream orders from which deposited fine sediment cover has been observed in the NZFFD records. Stream order 6 represent orders 6 and above.

Comparison between values of NZFFD % fines and SAM2 % cover instream independently observed on the same NZ reach, but at different times, showed very weak patterns (Figure N-5). This indicated that there could be great variation in % cover of total fines either: a) over time; b) within NZ reaches; or c) between techniques of observing % cover of total fines. In our view it is likely that all three factors are driving the lack of observed concordance. In the absence of observations of both NZFFD % fines and SAM2 % cover instream collected at the same place at the same time, we looked at comparisons between the SAM1 % cover bankside method and SAM2 % cover instream method (Figure N-6). The SAM1 % cover bankside method is relatively similar to the NZFFD % fine method in that it determines the proportional cover of different substrate sizes in runs, riffles and pools (i.e., across a reach) by visual assessment. This comparison indicated a strong (r^2 = 0.92) correlation between the SAM1 % cover bankside reach and SAM2 % cover instream run variables, and that the relationship was close to 1:1. This gave us confidence to assume an equivalence between the NZFFD % fines and SAM1 % cover instream variables used for the analyses for deriving exposure-response relationships.

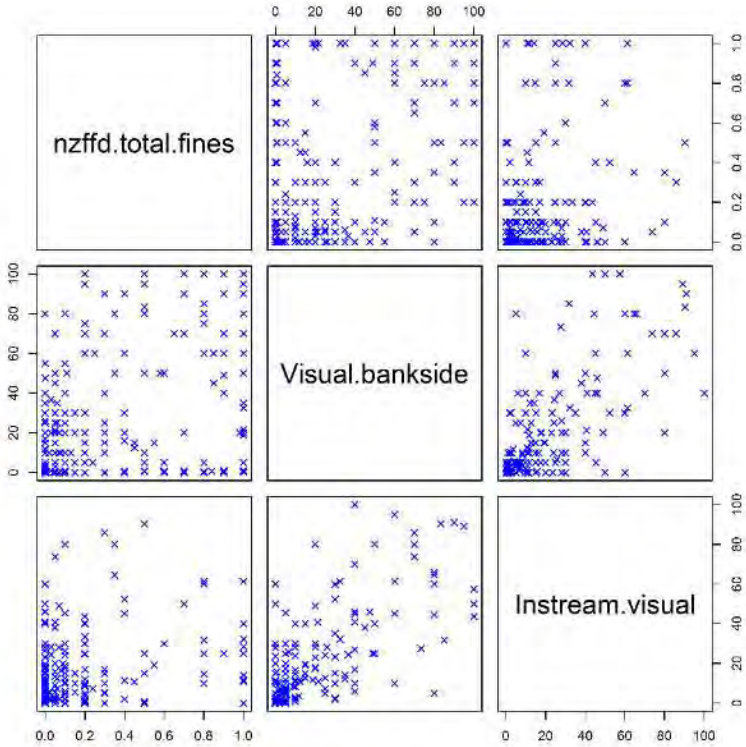


Figure N-5: Comparison between non-synchronously observed sediment ESVs. Comparison of % cover of total fines measurements from the NZFFD (nzffd.total.fines) at the same NZ reaches, using visual bankside method (Visual.bankside, SAM1) and using the instream visual method (Instream.visual, SAM2).

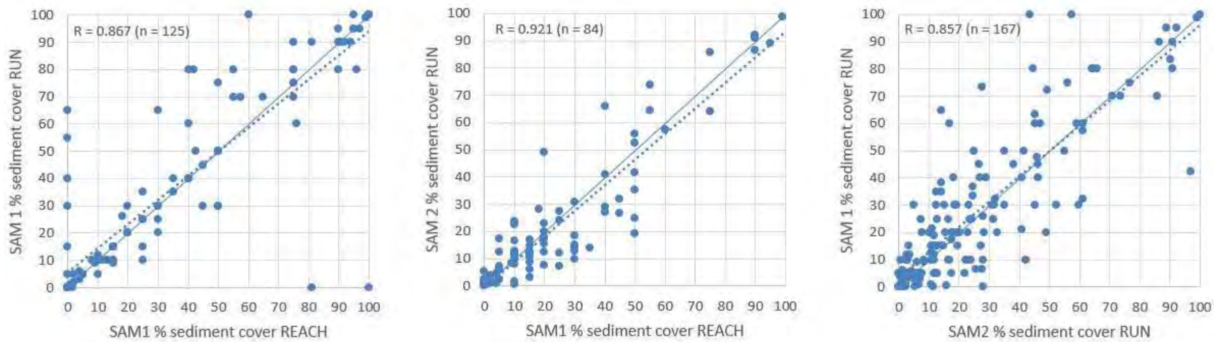


Figure N-6: Comparison of SAM1 and SAM2 deposited sediment measurement methods. The 'SAM1 % sediment cover reach' variable is very similar to the NZFFD % fines variable and so is used here as a surrogate for understanding the equivalence with SAM1 % cover instream.

Appendix O Converting between turbidity and visual clarity

Visual clarity and turbidity are generally correlated with each other, but the strength of the correlation is often site-specific, and 'one size fits all' regressions may not be sufficiently robust to allow interconversions to be carried out with confidence (Davies-Colley and Smith 2001). This limitation indicates that when going from a single site, to regional/organisational, or to national datasets, it should be anticipated that regressions between suspended sediment metrics will be less robust.

Paired turbidity and visual clarity data (derived from regional SOE and NRWQN monitoring sites; n=722) demonstrate a strong correlation (r^2 = 0.81) when data from all sites are considered (Figure O-1). This relationship could be used for interconversion between turbidity and visual clarity measurements or thresholds where locally calibrated interconversions (i.e., from paired samples) are not available.

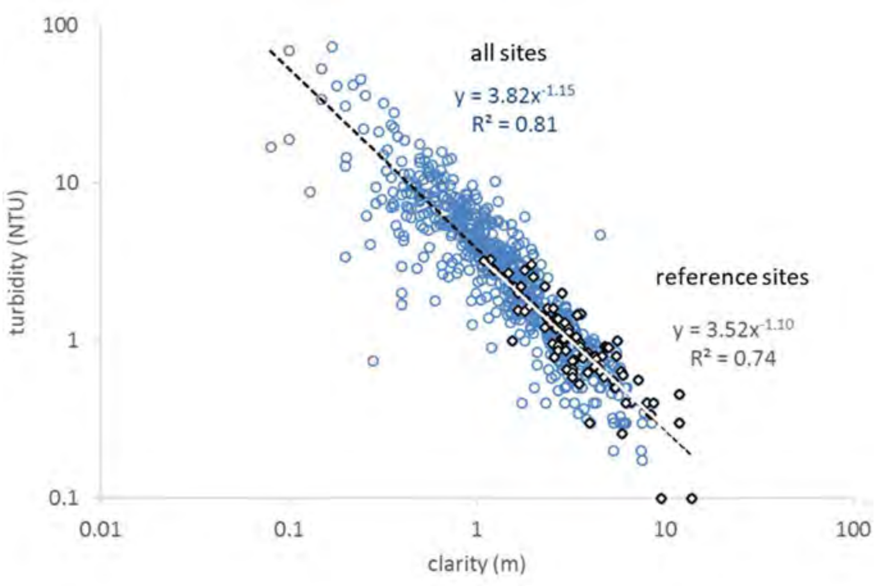


Figure O-1: Regression of turbidity and visual clarity using long-term site medians. For all sites n=722 (blue circles) and for reference sites n=83 (black diamonds).