Development of ecosystem health *bottom-line* thresholds for suspended and deposited sediment in New Zealand rivers and streams

Prepared for Ministry for the Environment

May 2018

Prepared by:
Craig Depree*
Joanne Clapcott**
Doug Booker*
Paul Franklin*
Chris Hickey*
Annika Wagenhoff**
Fleur Matheson*
James Shelley*
Martin Unwin*
Sanjay Wadhwa*
Eric Goodwin**
James Mackman**
Hayden Rabel**

*NIWA
** Cawthron Institute

For any information regarding this report please contact:

Craig Depree
Group Manager
Chemistry & Ecotoxicity
+64-7-756 7026

National Institute of Water & Atmospheric Research Ltd
PO Box 11 115
Hamilton 3251

NIWA CLIENT REPORT No: 2017076HN
Report date: May 2018
NIWA Project: MFE16211

Quality Assurance Statement		
	Reviewed by:	Dr John Quinn
	Formatting checked by:	Alison Bartley
	Approved for release by:	Dr David Roper

Contents

Tables

Figures

Executive summary

Background

Increased sediment input arising from human modification of the landscape and ongoing land use activities is a major stressor on freshwater ecosystems (and estuarine and coastal receiving waters) in New Zealand and worldwide. Increased sediment inputs adversely impact freshwaters in several ways.

The National Policy Statement for Freshwater Management (NPS-FM) requires regional councils, through their regional plans, to set freshwater objectives that provide for freshwater values, and to set limits and management actions to achieve those objectives. The NPS-FM includes the National Objectives Framework (NOF), which defines attributes that assist regional councils to set freshwater (i.e., numeric) objectives and justifiable policies (including limits) for achieving these. For Ecosystem Health, a compulsory national freshwater value, existing NOF attributes specify four ecosystem states defined by numeric thresholds that separate the four management bands (A, B, C and D); A/B band threshold is the most stringent while the C/D threshold is the national bottom line that defines the minimum ecosystem state required for all streams and rivers in New Zealand. Ministry for the Environment (MfE) requires science to inform the definition of NOF attributes aimed at protecting ecosystem health.

The NOF does not currently include attributes for sediment despite it being a major stressor on freshwater ecosystems. The difficulties associated with defining nationally applicable freshwater objectives and attributes for sediment were not satisfactorily resolved when the 2014 NPS-FM was released.

MfE commissioned a project aimed at providing science to inform definition of deposited and suspended fine sediment attributes for inclusion in the NOF to protect ecosystem health. The main objective was to investigate the relationship between measures of deposited and suspended fine sediment and indicators of the ecosystem health (fishes and macroinvertebrates) across different stream and river types. The project was divided into two stages:

- *Stage 1*: Dataset collation and classification system (Stage 1A and 1B reports): Stage 1A report summarised deposited and suspended sediment datasets, and ecological datasets (fishes and macroinvertebrates) for subsequent investigation of ecological responses to sediment gradients. Stage 1B report characterised (or modelled) natural state variation in suspended and deposited sediment and proposed a national classification system to account for this variation. Note that the key components of the sediment classification reports are included in Chapter 2 of this report. For both suspended (section 2.3) and deposited (section 2.4) sediment, this report includes revisions/refinements to the classifications presented in the respective 1B reports (Depree 2017; Clapcott and Goodwin 2017).

- *Stage 2*: Investigate biological responses to gradients of different measures of deposited and suspended sediment environment state variables (ESVs) and proposed bottom-line (i.e., band C/D) thresholds for definition of NOF attributes for deposited and suspended fine sediment (this report).

Stage 2 Objectives

This work involved data analyses for the purpose of:

- developing a sediment (suspended and deposited) classification system for New Zealand streams
- identifying ecological indicators of sediment effects
- quantifying the stressor-response relationship between sediment ESVs and ecological indicators, and
- identifying thresholds in stressor-response relationships to help inform development of NOF sediment attributes, for the Ecosystem Health value, which involves definition of management thresholds (in particular the C/D band threshold).

Summary of main points

Suspended and deposited sediment classification systems for New Zealand streams (Chapter 2)

- The purpose of a classification system is to account for the natural variation in deposited and suspended sediment measures along rivers and across the country due to factors such as geology, soils, channel and catchment slope and climate. Some of the key drivers of natural state variation differ between deposited sediment and suspended sediment. Deposited sediment cover of the streambed can vary substantially between reaches along rivers with changes in local (reach scale) river characteristics (e.g., stream slope and power), whereas suspended sediment is relatively insensitive to these factors. This required different approaches to the classification systems: the deposited sediment classification required a modelling approach that accounted for local river influences to predict natural levels of deposited fine sediment percentage cover at every stream reach, whereas suspended sediment classification was derived from simpler analysis of data to characterise natural state variation.
- 80-90 suspended sediment reference sites where identified from approximately 800 water quality sites, using newly developed criteria and catchment landcover areas (LCDB4), and used to characterise natural state variation of three sediment measures – namely total suspended sediment (TSS), visual clarity and turbidity.
- The absolute levels and variation in median natural state suspended sediment measures were relatively low (e.g., turbidity at natural state sites varied 6-fold, most sites ranging between 0.4 and 2.2 NTU. Furthermore, because the focus of the project was limited to defining bottom-line thresholds (C/D bands), the observed variation natural state of suspended sediment was much lower than anticipated bottom-line threshold values informed by a literature search. For turbidity, we expected the lowest C/D band threshold would be medians between 5 and 7 NTU, compared to natural state variation of site median values (0.4-2.2 NTU).
- Differentiation reference site suspended sediment measures by climate class suggested that aggregate 'warm' River Environment Classification (REC) climate class, on average, have up to 1 NTU higher median turbidity values compared to reference sites from 'cool' REC climates. We therefore recommend a potential +1 NTU offset for warm climate classes. This is largely a

pragmatic recommendation based on few reference sites, but is also consistent with modelled reference state measures of suspended sediment (McDowell et al. 2013).

- A boosted regression tree (BRT) model, trained with deposited sediment data from 2,022 reference sites spread across New Zealand, was used to predict natural state (i.e., sediment state without or little human influence) for all stream segments of the national river network. Across all stream segments, the model predicted a 'national' median deposited sediment cover of 13% for the hypothetical natural state of streams in New Zealand. It also predicted that 75% of all stream segments in New Zealand would have <30% sediment cover in their hypothetical natural state while the remaining 25% of stream segments have >30% sediment cover.

- We explored the grouping of sediment natural state predictions using Classification and Regression Trees (CART) and River Environment Classification classes. The CART identified 9 sediment classes with a statistically significant difference in mean sediment values. These classes were pragmatically combined into three classes: 1) <30% sediment cover ('hard-bottom streams with low-medium sediment levels'), 2) 30-60% sediment cover ('hard-bottom streams with high sediment levels') and 3) >60% sediment cover ('soft-bottom streams'). The three classes are recommended for attribute band application, and the BRT model predictions are recommended for the assignment of stream segments into each class.

Development of deposited sediment thresholds based on macroinvertebrate responses (Chapter 3 & 4)

- The development of stressor-specific macroinvertebrate metrics, including metrics for deposited sediment, was explored in the parallel MfE-funded macroinvertebrate project (Contract No. 21630, Clapcott et al. 2017) using a national dataset consisting of several research datasets. Macroinvertebrate metrics, including macroinvertebrate community index (MCI) and numbers of sediment-sensitive taxa (i.e., 'No. of decreasers'), showed the strongest responses (among a range of candidate metrics) to a gradient in deposited sediment in the training data, and were used in this project for development of sediment thresholds.

- Three analytical approaches were used to model stressor-response relationships between deposited sediment and benthic macroinvertebrate communities from which effects-based sediment thresholds were derived. These approaches included: 1) linear quantile regression; 2) boosted regression tree (BRT) analysis, both of which were used to model the responses of macroinvertebrate metrics; and 3) gradient forest (GF) analysis, which investigates points across the sediment gradient where macroinvertebrate communities change more dramatically.

- Analyses were performed on a national dataset consisting of data from state-of-the-environment monitoring, National River Water Quality Network (NRWQN) sites and several research studies. Three measures of deposited sediment were used:

    i. the percentage of sediment cover on the streambed assessed by standing in the stream (% sediment cover instream);

    ii. the percentage of sediment cover on the streambed assessed from standing on the stream bank (% sediment cover bankside); and

    iii. suspendable inorganic sediment (SIS) assessed using the 'Quorer method'.

- The response curves from boosted regression tree (BRT) models for four macroinvertebrate community metrics (*MCI, EPT richness, Sediment MCI, No. of decreasers*) across the gradient of % sediment cover instream were most informative and showed consistent and meaningful thresholds. The GF analysis was limited by the number of taxa for which models could be built, however the derived thresholds were largely consistent with the BRT model output and provided complementary information for defining proposed C/D band threshold values.
- Based on BRT and GF analyses of macroinvertebrate response to deposited fine sediment, we recommend C/D band threshold (national bottom line) values of:
  - **30% fine sediment cover**, for streams classified as 'hard-bottom streams with low-medium sediment cover' – defined as streams with predicted reference state fine-sediment cover of <30% (assessed via instream method)
  - **60% fine sediment cover** for streams classified as 'hard-bottom streams with high sediment cover' – defined as streams with predicted reference state fine-sediment cover of 30 to 60% (assessed via instream method).

These thresholds were found to be adequately protective of fish as described in Section 6. We consider streams with predicted reference state of >60% fine sediment cover as 'soft-bottom' and do not recommend sediment attributes are applied here.

Development of suspended sediment thresholds informed by macroinvertebrate responses (Chapter 5)

- The analytical approach to derivation of effects-based sediment thresholds (for both visual clarity and turbidity) involved two steps:
  i. Step 1 involved modelling of the stressor-response relationships between suspended sediment measures and seven benthic macroinvertebrate community metrics as well as seven selected macroinvertebrate taxa using linear and non-linear quantile regression models. Non-linear models were selected to fit expected subsidy/stress relationships between suspended sediment and macroinvertebrate indicators. From the response curves, indicative sediment thresholds were derived at the point at which there was a 30% reduction in the macroinvertebrate metric or taxon density, compared to the maximum value.
  ii. Step 2 involved generating a 'species' sensitivity distribution (SSD) plot from the thresholds identified in step 1 from which then thresholds were defined that corresponded to different levels of protection. The analyses were performed for two surrogate measures (or ESV) of suspended sediment; turbidity and visual clarity (black disc).
- Analyses were performed on the NRWQN dataset consisting of monthly measurements of turbidity and visual clarity and annual macroinvertebrate samples taken from 67 river sites spread across New Zealand. For modelling, each macroinvertebrate sample collected in the period from 1990 to 2013 was matched with the respective annual median visual clarity and turbidity calculated from the preceding 12 monthly sediment measurements.
- The suspended sediment ESV corresponding to an 80% protection level is considered consistent with the narratives generally applied to C/D band thresholds for other NPS-FM

attributes (i.e., point defining significant adverse ecological effects). The indicative visual clarity and turbidity thresholds based on an 80% protection level of macroinvertebrates were 1.0 m and 4.3 NTU, respectively.

Development of deposited and suspended sediment thresholds informed by fish responses (Chapter 6)

Theanalyses were based on fish presence-absence data retrieved from the New Zealand Freshwater Fish Database (NZFFD). For many fish records, the NZFFD contains data on deposited fine sediment which is a visual estimate of the percent cover of fine sediment on the streambed (% sediment cover) within the fish sampling reach assessed instream on the same day. The NZFFD does not hold data on suspended sediment, hence we used model-predicted values for suspended sediment ESVs. These were derived from existing models that can predict long-term median visual clarity and turbidity for each NZ reach in the national river network.

- Eleven fish species, ten natives and brown trout, were selected for the analysis. Each was hypothesised to be influenced by sediment levels, found across New Zealand, and present in a reasonable proportion of samples in the NZFFD.
- A generalised linear mixed model was used to relate likelihood of capture of each fish species to a sediment ESV (both suspended and deposited) whilst also taking account of other variables that influence fish distribution at the landscape scale such as distance from sea, climate, topography and size of river.
- Overall, a higher likelihood of capture was found for most species in conditions with less deposited fine sediment, higher visual clarity or less turbidity. In contrast, shortfin eel was an example where its likelihood of capture was favoured by high deposited sediment, high turbidity and low visual clarity.
- The predicted likelihood of capture for individual fish species across a sediment ESV were translated into a metric that describes the overall expected change in fish community relative to the community that might be expected at reference ESV state. Reference sediment ESV states were predicted using a separate analysis. A 20% departure in fish community integrity from reference state was selected as the threshold for defining C/D band thresholds values. There was substantial variation in ESV reference state and the resulting C/D threshold values across REC classes. Averages (and range) of reference state and C/D thresholds were as follows: clarity reference = 2.5 m (1.3 – 3.9 m), clarity C/D = 1.5 m (0.6 – 2.4 m), turbidity reference = 1.5 NTU (0.7 – 2.9 NTU), turbidity C/D = 3.2 NTU (1.2 – 6.7 NTU), sediment cover reference = 19% (6 – 46%), sediment cover C/D = 50% (31 – 80%).

Synthesis and final proposed thresholds for deposited and suspended sediment based on multiple lines of evidence (Chapter 7)

- Macroinvertebrates were more sensitive to deposited fine sediment than fishes, and so these were used for the proposed bottom-lines for the deposited fine sediment attribute. We propose thresholds of 30% (where reference state is <30% of the stream bed covered in fine sediment) or 60% (where the reference state is between 30 and 60% cover). Where predicted reference state is >60% cover, the sites are classified as naturally soft-bottom and are considered exempt from the attribute.

- The proposed NOF attribute (ecosystem health) table for deposited fine sediment in wadeable NZ stream and rivers is provided in Table ES-1.
- For suspended fine sediment (measured as median turbidity or visual clarity), macroinvertebrates and fish yielded comparable bottom-line thresholds for warm climate classes, but the results indicated that fish in cold climate areas were more sensitive to suspended sediment than macroinvertebrates.
- A classification system for suspended fine sediment is proposed which is based on aggregated river environmental classification (REC) climate classes: 'cool' (cool dry, cool wet and cool extremely wet) and 'warm' (warm dry, warm wet and warm extremely wet). Median turbidity at reference 'cool' and 'warm' classes was 1.0 and 2.0 NTU, respectively. For the national bottom-line we propose a median turbidity of 5.0 and 6.0 NTU in 'cool' and 'warm' climate classes, respectively (corresponding to respectively median visual clarity values of around 0.85 m and 0.7 m.
- The proposed NOF attribute (ecosystem health) table for suspended fine sediment in wadeable NZ stream and rivers is provided in Table ES-2.

Table ES-1: Proposed NOF attribute table for assessing deposited fine sediment in wadeable rivers

Value	Ecosystem Health		
Freshwater Body Type	Rivers (wadeable only)		
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	**Numeric Attribute State ‘Low-to-medium’ level (<30%)[1] of natural sediment**	**Numeric Attribute State ‘high’ level (30-60%)[1] of natural sediment**	**Narrative Attribute State**
	Annual mean[2]	Annual mean[2]	
A	NA	NA	
B	NA	NA	
C	<30%	<60%	Low to moderate cover relative to reference state providing excellent to fair habitat for biota. Risk of sensitive macroinvertebrate species being lost and change in community composition.
National Bottom Line	**30%[3]**	**60%[3]**	
D	>30%	>60%	High likelihood of sediment cover exceeding reference state providing poor habitat for biota. High probability of loss of sensitive macroinvertebrate species.

NA = not applicable

1) Classes are streams and rivers defined according to predicted reference state for deposited sediment, currently this is based on predicted reference state from the BRT REF model. Streams with greater than 60% fine sediment cover are classified as naturally soft-bottomed streams and are exempt. Based on a monthly monitoring regime.
2) The minimum record length for grading a site based on an instream visual assessment of % fine sediment cover (SAM2) is 2 years.
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 integrity index*), however they may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations (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 ES-2: Proposed NOF attribute table for assessing suspended fine sediment in streams and rivers

Value	Ecosystem Health				
Freshwater Body Type	Rivers				
Attribute	Suspended fine sediment quantity (Surrogate measures: visual clarity or turbidity)				
Attribute Unit	Visual clarity, m (metres); turbidity, NTU (Nephelometric Turbidity Units)				
Attribute State	**Numeric Attribute State: Visual clarity (m)** [1]		**Numeric Attribute State: Turbidity (NTU)** [1]		**Narrative Attribute State**
	Annual median[2,3]		Annual median[2,3]		
	'cool'[4]	'warm'[4]	'cool'[4]	'warm'[4]	
A	NA	NA	NA	NA	
B	NA	NA	NA	NA	
C	>0.85	>0.7	<5	<6	Protects biodiversity measures from >30% impact.
National Bottom Line[5]	0.85	0.7	5	6	
D	<0.85	<0.7	>5	>6	High likelihood of loss of sensitive species and marked reduction in biodiversity. High probability of extirpation of sensitive macroinvertebrate species.

NA = not applicable

1. Classes are for all wadeable streams and rivers with the following exclusions: (i) 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 inorganic sediment from the catchment).
2. Based on a monthly monitoring regime. The minimum number of samples is 20, this will generally mean that assessment again the thresholds will require 2 years of monthly data, or 5 years of quarterly data.
3. Interconversion of visual clarity and turbidity is acceptable as derivation based on database of annual median data for these parameters (i.e., not concurrent instantaneous measurements). The more sensitive of the visual clarity or turbidity measures will determine the site grading. Visual clarity will be a more sensitive measure of changes in river particulate organic material and inorganic SS in high quality (i.e., low turbidity) waters.
4. Aggregated REC climate classes: 'cool' consists of cool dry (CD), cool wet (CW) and cool extremely wet (CX); 'warm' consists of warm dry (WD), warm wet (WW) and warm extremely wet (WX). Suspended sediments summary statistics from minimally disturbed condition sites (Depree 2017) and predicted reference states (McDowell et al. 2013) indicate that 'cool' and 'warm' sites have respective median turbidity values of 1 and 2 NTU; 'cool' and 'warm' median visual clarity values for reference sites are around 3.5 and 2.5 m respectively. Insufficient data was available in the macroinvertebrate database to derive distinct thresholds for the cool and warm classes. Differentiation based on differences in reference conditions.
5. Visual clarity values based on average of 2 regression equations between visual clarity and turbidity. 1) all NRWQN data (n=76), 2) all SoE monitoring data (n=722) (refer to Stage 1B, Depree 2017).

1 Introduction

1.1 Background

The National Policy Statement for Freshwater Management (NPS-FM) requires regional councils, through their regional plans, to set freshwater objectives that provide for freshwater values, and to set limits and management actions to achieve those objectives. The NPS-FM includes the National Objectives Framework (NOF), which defines attributes that assist regional councils to set freshwater (i.e., numeric) objectives and justifiable policies (including limits) for achieving these.

For Ecosystem Health, a compulsory national freshwater value, existing NOF attributes specify four ecosystem states defined by numeric thresholds that separate the four management bands (A, B, C and D); A/B band threshold is the most stringent while the C/D threshold is the national bottom line that defines the minimum ecosystem state required for all streams and rivers in New Zealand. Ministry for the Environment (MfE) requires science to inform the definition of NOF attributes aimed at protecting ecosystem health.

The NOF does not currently include attributes for fine sediment, despite it being a major stressor on freshwater ecosystems (and coastal waters). Sediment affects ecosystem health through various modes of impact but can be coarsely categorised into suspended sediment and deposited fine sediment[1]. Suspended sediment is typically quantified by four environment state variables (ESVs) suspended sediment concentration (also known as total suspended solids), visual clarity (which also for the purposes of this report includes the correlate, turbidity), light penetration (LP), and deposited fine sediment[2]. The difficulties associated with defining nationally applicable freshwater objectives and attributes for sediment were not satisfactorily resolved when the 2014 NPS-FM was released, hence sediment has been identified as a priority area for the development of attributes and bottom lines for future revisions of the NPS-FM.

The Ministry for the Environment (MfE) and the Ministry for Primary Industries (MPI) convened a process that considered how NOF attributes for sediment can be developed. The process concluded (MfE 2015) that research and development is required to solve two key problems associated with sediment NOF attributes: (i) the transformation of catchment sediment loads into ESVs (Hicks et al. 2016); and (ii) defining numerical thresholds for sediment ESVs that relate to effects on ecosystem values. Consequently, MfE commissioned a project in two stages. Stage 2 is the content of this report.

1.2 Aim and objectives

As detailed in the Statement of Work (Appendix A), the main research aim of Stage 2 of the large project is to use existing data to define numerical thresholds for sediment ESVs (for deposited and suspended sediment) that are based on effects of sediment on the ecosystem health value (i.e., define effects-based sediment thresholds) and can be used for definition of NOF sediment attributes.

The research tasks to address this aim were:

1) Scoping workshop.

[1] In this report we define deposited sediment as sediment of grain size finer than 2 mm (i.e., includes sand and silt fractions). Hereafter, whenever deposited sediment is referred to it concerns deposited fine sediment.

2) Compile and QC ecological datasets.

3) Compile sediment ESV datasets.

4) Field study to investigate the relationship between deposited sediment and macroinvertebrate abundances and diversity.

5) Development of a classification system to differentiate NZ rivers according to variation in ‘reference state’ sediment ESV characteristics.

6) Summary of sediment ESV characteristics (e.g., frequency criteria) within stream classes for both reference and current conditions.

7) Summarise the relationship between different suspended sediment ESVs (within the classification system developed in Task 5) – namely suspended sediment concentration, visual clarity and turbidity to identify whether any of the ESVs are redundant (i.e., are 3 suspended sediment ESVs required to manage the effects of suspended sediment?).

8) Development of ecological indicators for sediment stressor-specific macroinvertebrate metrics – existing, commonly used, macroinvertebrate community metrics (e.g., MCI) were developed for organic enrichment in streams, and therefore are potentially not suitable for assessing the effects of sediment gradients.

9) Analyse ecological responses to sediment ESV gradients and collate published information on relevant ecological responses to ESV gradients.

10) Integration of response of ecological indicators to sediment ESV within the recommended classification system and recommend ESV thresholds.

11) Identification of data gaps and recommendations to address these.

1.2.1 Project structure

The project (Contract No. 21511) was comprised of three stages, with Stage 1A and 1B represented ‘hold points’ in the contract. Proceeding beyond a hold-point was conditional on approval from the Ministry.

Stage 1A: Data collation (tasks 1-4)

Research tasks 1 to 4 have been reported in the following companion reports:

- Stage 1A: summary of data (Depree and Clapcott 2016)[3]
- Effects of the Waihi Dam failure on the benthic invertebrates of the Waiau River (Clapcott 2016)[4]

Stage 1B: National classification system for suspended and deposited sediment (tasks 5-7)

Research tasks 5 to 7 have been report in the following companion reports:

[3] Depree C, Clapcott J 2016. Sediment Stage 1A report: summary of data. Prepared for Ministry for the Environment. NIWA Client Report No: 2016124HN. 27 p.

[4] Clapcott J 2016. Effects of the Waihi Dam failure on the benthic invertebrates of the Waiau River. Prepared for Hawkes Bay Regional Council and Ministry for the Environment. Cawthron Report No. 2926. 22 p.

- Stage 1B: Deposited sediment report (Clapcott and Goodwin 2017)[5]
    - executive summary provided in Appendix A
- Stage 1B: Suspended sediment report (Depree 2017)[6]
    - executive summary provided in Appendix A

Stage 2: Investigate ecological responses to sediment ESV gradients and propose thresholds for definition of NOF attributes (at least a C/D band threshold) for deposited and suspended sediment aimed at protecting ecosystem health of rivers and streams (task 8-11)

- Reporting of research tasks 8-11 forms the content of this report

1.3 Stage 2 report layout

Stage 2 report presents the complete output from research tasks 8-11 that were concerned with development of management thresholds for definition of NOF attributes for deposited and suspended sediment aimed at protecting ecosystem health of rivers and streams., We also included relevant information on NOF attribute development from tasks reported in Stage 1A and 1B into the Stage 2 report to provide a stand-alone report presenting the major outcomes of the MfE Sediment project (Contract No. 21511). Chapters 2-7 report on the research tasks and outputs summarised below. Each of the Chapters 2-6 presents an independent and comprehensive body of research so Chapter summaries have been provided. Note that Chapters 4-6 propose effects thresholds defined to be consistent with NOF bottom-line threshold values. The derived thresholds (based on macroinvertebrate and fish responses) contribute to multiple lines of evidence (including literature thresholds and expert opinion) that are assessed in Chapter 7. The output of this synthesis of evidence are proposed NOF C/D band thresholds for suspended and deposited sediment attribute tables.

The structure of the report and connectivity between chapters is summarised in Figure 1-1.

[5] Clapcott J, Goodwin E 2017. Technical report on developing a deposited sediment classification for New Zealand streams. Prepared for Ministry for the Environment. Cawthron Report No. 2994. 36 p.

[6] Depree C 2017. Sediment Attribute Stage 1B. Proposed classification system. Prepared for Ministry for the Environment. NIWA Client Report No. 2017076HN. 76 p.

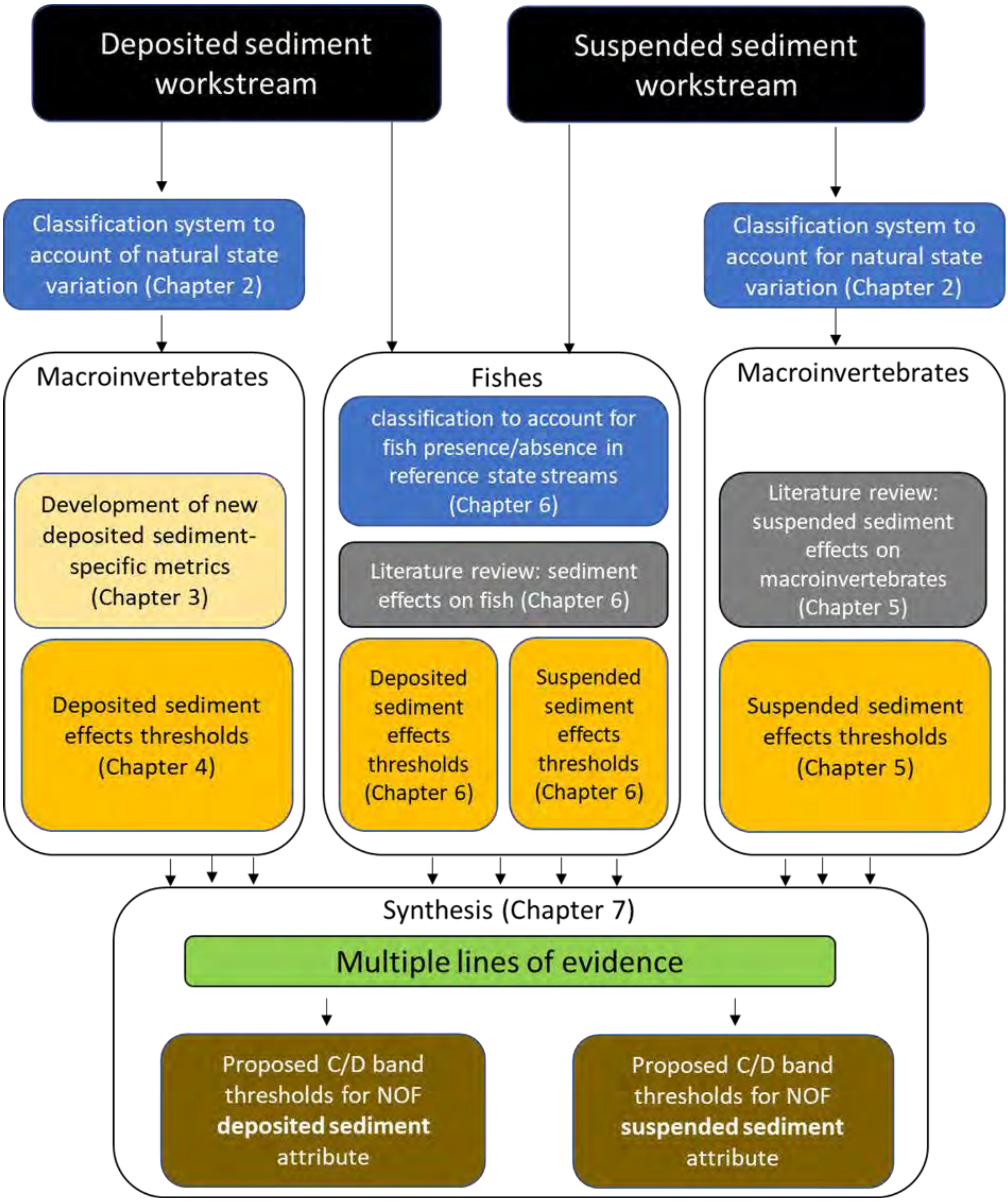


Figure 1-1: Summary of the workflow of the Stage 2 sediment threshold project, illustrating the major components of the chapters, and how these contributed to the final proposed C/D band thresholds for suspended and deposited sediment attributes. Blue boxes refer to development of classification systems to account for natural state variation in sediment ESVs – note the fish threshold work required a separate classification to account for the strong influence of landscape setting on fish presence/absence. Grey boxes indicate threshold literature reviews. Orange boxes represent analytical approaches for deriving threshold from sediment ESVs and biological data (pale orange box Chapter 3, metric development) was a prerequisite for threshold derivation in Chapter 4). Multiple lines of evidence from the 3 workflows were synthesised (green box) to generate the final proposed NOF C/D band thresholds for the suspended and deposited attribute tables (brown boxes).

Brief summaries of the main components of work contained within each chapter are provide below.

Chapter 2 – Suspended and deposited sediment classification

- Suspended sediment:
  - Argument for including turbidity-based thresholds in the suspended sediment attribute table
  - Brief summary, and update, of the proposed 'single' classification system reported in Stage 1B to include aggregate REC 'cool' and 'warm' climate classes
- Deposited sediment:
  - Detailed summary of proposed classification system reported in 1B – based on predicted reference state condition and Classification and Regression Tree (CART) methods to aggregate these into suitable classes
  - Update on stage 1B classification – 3 classes based on predicted reference state of deposited fine sediment with class boundaries informed by ecological responses (Chapter 4)

Chapter 3 – Development of macroinvertebrate metrics as ecological indicators of deposited sediment effects

- Introduction of sediment-specific invertebrate metrics developed overseas
- Field study designed to test whether macroinvertebrate metrics respond to a broad gradient in specific sediment yield by relating deposited sediment ESVs to specific sediment yield and secondly relating macroinvertebrate metrics to deposited sediment ESVs
- Systematic literature review using a formal causal criteria analysis and the Eco Evidence software to identify what ecological evidence exists to inform sediment-specific metric development
- Development of sediment-specific metrics using gradient forest analysis and expert opinion; use of a large national dataset specifically compiled for this task and consisting of several research studies.

Chapter 4 – Development of deposited sediment thresholds: 1st line of evidence –macroinvertebrate responses

- Introduction of analytical approaches to defining effects-based thresholds
- Analysis of macroinvertebrate responses to deposited sediment using three selected approaches; the analysis included new stressor-specific macroinvertebrate metrics developed in the parallel MfE macroinvertebrate project (see Chapter 3); use of a national dataset specifically compiled for this and the parallel MfE project and consisting of SoE monitoring data and research data

- Analysis of temporal variation of deposited sediment ESVs
- Proposed attribute bands for deposited sediment based on the analysis presented in this chapter; attribute bands are stream-type specific and stream types were defined by a sediment classification system developed in this project (see Chapter 2)
- Recommendation for future work.

Chapter 5 – Development of suspended sediment thresholds based on macroinvertebrate responses

- Literature review of effects of suspended sediment on macroinvertebrates and analytical approaches to defining numeric standards; includes New Zealand and international studies
- Analysis of macroinvertebrate responses to suspended sediment using a single approach involving two distinct steps; use of the National River Water Quality Network (NRWQN) monitoring dataset
- Proposed attribute bands for suspended sediment based on the analysis presented in this chapter; attribute bands are not stream-type specific
- Recommendation for future work.

Chapter 6 – Development of deposited and suspended sediment thresholds based on fish responses.

- Literature review of effects of deposited and suspended sediment on fish to inform development of management thresholds in addition to thresholds identified analytically in this chapter; includes New Zealand and international studies
- Use of expert knowledge to derive expected sensitivity of New Zealand's main fish species to elevated fine sediment inputs and hypothesised mechanisms
- Prediction of sediment ESV reference state, and classification system that accounted for influences of 'landscape setting' on the presence/absence of fishes
- Analysis of fish responses to deposited and suspended sediment using a single approach involving two distinct steps; use of data retrieved from the New Zealand Freshwater Fish Database (NZFFD)
- Proposed attribute bands for deposited and suspended sediment based on the analysis presented in this chapter; attribute bands are stream-type specific, the stream types are second-level 'Source-of-Flow' REC classes
- Recommendation for future work.

Chapter 7 – Synthesis of multiple lines of evidence and proposed NOF C/D thresholds for suspended and deposited sediment attributes.

- Synthesis of data from all workstreams; comparing sediment thresholds based on fish and macroinvertebrates, and reconciling these against literature value and reference state
- Population of NOF attribute tables, including proposed C/D band thresholds for the management of deposited and suspended sediment in NZ rivers and streams.

2 Sediment classification

2.1 Chapter summary

This chapter presents the development of classification systems for suspended and deposited sediment to account for the natural state variation of relevant sediment ESVs across the country within which NOF sediment attribute band thresholds (limited in this report to the C/D band threshold values only) would be defined. Sediment classification is desirable if natural sediment conditions vary significantly across flowing waters. The work has been reported on in detail in the Stage 1B reports (Clapcott & Goodwin 2017, Depree 2017). In this report only the main and relevant outcomes are presented, although some updates and refinement are included.

For suspended sediment, the dataset available comprised 514, 722 and 832 water quality sites for TSS, visual clarity and turbidity data, respectively. Catchment areas of different land cover were calculated for all sites using the latest database (LCDB4) and criteria developed to define a subset of reference condition sites. For TSS, visual clarity and turbidity, the number of reference sites were 52, 85, and 92, respectively. These were used to characterise the natural state variation of suspended sediment. A major difference to deposited sediment was the finding that, using long-term site medians as the statistic, suspended sediment levels did not vary as much as we assumed. For example, for turbidity, 80% of reference sites had values between 0.5 and 2.5 NTU compared to 0-100% for deposited sediment. Moreover, deposited sediment along rivers due to influences of channel physical characteristics (e.g., slope and flow), which required a different (modelled) approach to characterising reference condition at a national scale.

For deriving ecosystem health bottom-line thresholds for suspended sediment ESVs, we concluded that natural state variation (e.g., 0.4-2.2 NTU for turbidity) was not significant compared to the likely range of a C/D band threshold (e.g., 5-7 NTU).[7] Accordingly, we determined that a classification system was not required for suspended sediment. Model reference state (McDowell et al. 2013) and groupings by REC aggregate climate, did however, suggest applying a +1 NTU 'offset' to the 'general' C/D band threshold for aggregated 'warm' REC climate classes.

For deposited sediment, sediment data (% sediment cover) collected using four different field assessment methods (SAM1, SAM2, RHA100 and NZFFD data) were compiled for all sites (n = 15,281) and for a subset of reference sites defined by land cover rules (n = 2,022). The data were used to develop boosted regression tree (BRT) models to predict the natural state of deposited sediment for all stream segments in the national river network. The BRT model based on reference data (BRT REF) had better predictive performance and more realistic spatial patterns than the all-sites data model (BRT ALL), although data limitation prevented model validation for large areas of eastern New Zealand. A classification and regression tree (CART) model illustrated that there were nine distinct groupings of stream segments at the national level with different natural sediment levels predicted by the BRT REF model. Post hoc grouping of predicted natural sediment levels by REC classes did not show a strong delineation of streams.

Subsequently, site-specific predictions of reference state deposited sediment levels from the BRT REF model were recommended to identify sites as having low (<30% cover), medium (30-60% cover) and high (>60% cover) for attribute band allocation. The classes were respectively called 'hard-bottom

[7] Informed by review of suspended sediment effects literature and focusing on the lowest values, and those related to chronic exposures.

streams with low-medium sediment levels', 'hard-bottom streams with high sediment levels' and 'soft-bottom streams'.

2.2 Environmental classification approaches for suspended and deposited sediment

Natural spatial variations in environmental conditions often necessitates the development of classification systems for summarising and grouping similar environment types for planning, monitoring and reporting purposes. For example, in New Zealand the River Environment Classification (REC) is a hierarchical classification system based on spatial variation in climate, geology, source of flow and land use (Snelder & Biggs 2002, Snelder 2004). The REC facilitates grouping of stream segments in the national river network by similarity in these factors (categorical) and has been used to summarise the water quality and ecological condition of streams for regional and national State of the Environment reporting. For example, the REC is used in the NPS-FM to identify 'Productive' and 'Default' classes of streams for the application of the periphyton attribute. These are defined by the combination of REC 'Dry' Climate categories and REC Geology categories that characteristically have naturally high levels of nutrient enrichment.

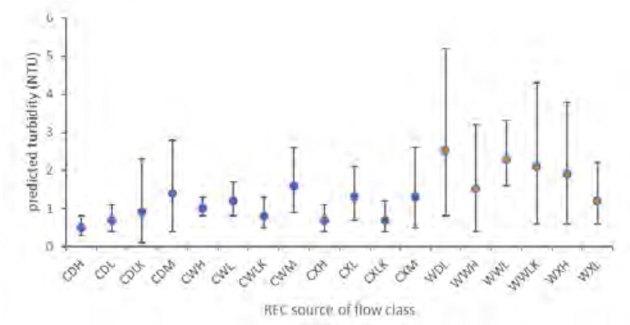
Ideally, measured values, rather than predicted values, should be used to develop stream classes for sediment (or any other) NOF attribute implementation and sediment data from streams with minimal or no human impact ('reference sites') across stream 'types' in New Zealand i.e., would be available to develop a classification system. These data would allow us to determine whether sediment ESV measures (based on the selected statistic, i.e., median) vary sufficiently to justify the use of a classification system.

In the case of suspended and deposited sediment, it was assumed that measures of relevant sediment environmental state variables (ESVs) would vary naturally, given their strong relationship with primary drivers of geology, slope and climatic factors (Hicks et al. 2016). Hence the project aim was to explore the spatial and temporal variation of suspended and deposited sediment at reference sites. However, the spatial representativeness and quality of the data available varied for deposited and suspended sediment measures and necessitated different data exploration approaches (Table 2-1).

A series of data exploration approaches were undertaken as part of Stage 1B of this project (Clapcott & Goodwin 2017, Depree 2017). The following section summarises the output of data exploration of both suspended (section 2.3) and deposited (section 2.4) sediment and provides further detail, and justification, on the recommended stream classifications for sediment attribute implementation.[8]

[8] Note, a classification system for suspendable inorganic solids (SIS) was developed (see Clapcott and Goodwin 2017) but the system developed for '% sediment cover' is recommended as suitable for both of these deposited sediment measures, based on the strong correlation between variables at low-to-medium sediment levels.

Table 2-1: Summary of data availability and approaches taken to explore natural variation in sediment environmental state variables (ESVs) for determining stream classifications. Reference sites are defined as stream sites with no or minimal human impact; for deposited sediment (LCDB3: >90% native vegetation, <5% exotic vegetation, <10% high production pasture, 0% urban) and for suspended sediment (LCDB4: >70% natural landcover including >40% indigenous forest, <10% high production pasture, <5% urban).

<table>
<tr><th>ESV / sediment measure</th><th>Number of reference sites</th><th>Approach</th></tr>
<tr><td>Turbidity</td><td>92 with 7-10 years of mainly monthly, but some regions with quarterly data</td><td rowspan="3">Reference sites were used to define natural state variation at reference which were relatively well distributed throughout the country. TSS was unsuitable for this purpose. Measurement reference state variation was relatively low, with reference turbidity ranging between 0.5 and 2.5 NTU. This was in good agreement with previous model reference state predictions (see graph below -reproduced for data in McDowell et al. 2013).


The ability to adequately characterise natural state variation in suspended sediment (at a national-scale) using measured data from SoE sites, combined with its relatively low variation was the fundamental difference between suspended sediment and deposited sediment. The next step involved assessing the ecological relevance of natural state suspended sediment levels. For C/D band thresholds, the natural state variation was considered minor, and so a classification system was not justifiable.</td></tr>
<tr><td>Clarity</td><td>85 with > 12 samples, mainly monthly, but some regions with quarterly data</td></tr>
<tr><td>TSS</td><td>52 samples with an average of c. 40 data points</td></tr>
<tr><td>% Sediment cover</td><td>2,022 with ≥ 1 sample</td><td>Explore spatial coverage of data. Use boosted regression tree model to predict reference state for all stream reaches in the national river network. Use effects-based thresholds to group streams into low, medium and high sediment classes.</td></tr>
<tr><td>Suspendable inorganic sediment (SIS; g/m^2)</td><td>27 with ≥ 1 sample</td><td>Explore spatial coverage of data. Use flexible regression to predict reference state for all stream reaches in the national river network. Use effects-based thresholds to group streams into low, medium and high sediment classes.</td></tr>
</table>

2.3 Approach to classification system to account for natural state variation in suspended sediment

The development of the classification system is described in detail in the Stage 1B report (Depree 2017). The main components of this report are covered in this section to allow a standalone Stage 2 report.

The Request for Proposals (RFP) document (MfE 2015) required the development of a classification system that differentiates New Zealand river according to “reference state” variation in ESV characteristics. The RfP document highlighted the complexities of this task, with suspended and deposited sediment measurements exhibiting very high temporal and spatial variability. For example, at a given site, suspended sediment can vary over 3 to 4 orders-of-magnitude. The importance of these large event-based suspended sediment variations (within site) when defining a summary statistic (e.g., median) for assessing natural state variation was uncertain. Accordingly, the basic steps we implemented for assessing the requirements and justification for a suspended sediment classification system were:

1. From all available data (NRWQN and SoE monitoring), identify a subset of suitable reference sites.
2. Based on the number and quality of reference site data available, determine which measures of suspended sediment are more suitable for characterising natural state variation.
3. Using an appropriate summary statistic, characterise natural state variation in suspended sediment measures (informed by 2).
4. Using established landscape settings (i.e., river environment classification, REC), determine whether existing classification schemes allow for better differentiation of the reference state measures of suspended sediment.
5. Assess the potential ‘ecological relevance’ of reference state variation in suspended sediment to provide justification for any REC-based ‘classification’ (from 4).

Before presenting results and key findings under each of these 5 steps, it is useful to briefly discuss the different suspended sediment ESVs and measurements.

2.3.1 Suspended sediment ESVs and other measurements: relevance to assessing natural state variation

Turbidity as a useful measurement of suspended sediment and ‘proxy ESV’

Although not initially proposed as one of three suspended sediment ESVs, turbidity is a widely used proxy measure of suspended sediment concentration. Previous studies have shown strong correlations between turbidity and TSS (e.g., Holliday et al. 2003, Davies-Colley et al. 2014), supporting inclusion of turbidity for data exploration and guiding development of the classification. Additional reasons for its inclusion as a sediment measure for both assessment of natural state variation and for enumerating thresholds include:

1. Turbidity is one of four water column metrics described as ‘sediment targets for informing river catchment management’ in Collins et al. (2011, Figure 1), the framework on which the New Zealand sediment ESVs are based. The water column metrics included ‘light penetration’,

‘turbidity’, ‘suspended sediment concentration’ and ‘sediment regimes’. The New Zealand adaptation of this sediment management framework substituted turbidity with visual clarity.

2. All 16 regional councils (RCs) and unitary authorities in New Zealand current monitor turbidity as part of their state of the environment (SoE) monitoring. By comparison, only 12 of the 16 regional authorities currently monitoring visual clarity – those regions that do not monitor visual clarity include Otago, Gisborne, Marlborough and Auckland.

3. New Zealand currently has close to 1,000 water quality monitoring sites which have turbidity data; in the data set used for this report (complete to 2013), there were 833 water quality sites with long-term (>10 years) of turbidity data.

4. Turbidity is used internationally for water quality assessments, and, as such, suspended sediment effects literature and numerous regulatory guideline values are expressed in units of turbidity. These resources provide *multiple lines of evidence* (in addition to the effects thresholds derived in chapter 4-6 of this report) for enumerating proposed thresholds for the suspended sediment attribute.

5. Assuming best practice methods are followed, turbidity measurements have sufficiently low detection levels to characterise natural state variations in suspended sediment.

6. Assuming best practice methods are followed, turbidity can be used to characterise suspended sediment gradients from reference state, up to the highest values for the most impacted sites.

7. National environmental monitoring standards (NEMS) include guidance for measurement of discrete turbidity measurements (NEMS 2016).

8. A New Zealand study (Barter and Deas 2003) that compared the performance of 5 portable nephelometric turbidimeters reported that for turbidity values in the range 0.5 to >100 NTU had a coefficient of variation (CV) of between 6 and 12%. For formazin standards, the CV ranged between 1.5 and 6.8%. Except for very clear headwater streams, a power analysis of the variance suggested that single replicate samples are usually sufficient to detect changes in visual clarity recommended in MfE (1994) guidelines. This study involved turbidimeters that conformed to different standards (EPA180.1 or ISO7027), and therefore, with modern turbidimeters and a single standard recommended for New Zealand (ISO7027), turbidity measurement will become more comparable/reliable across national datasets.

Although there may be limitations regarding turbidity measurements (summarised in Davies-Colley and Smith (2001)), with modern instruments, national standards prescribing best practice, and for the numerous reasons mentioned above, we made the decision to include turbidity for both assessing natural state variation and for enumerating proposed thresholds for the suspended sediment attribute.

Euphotic depth as suspended sediment ESV for streams and rivers

Justification for excluding euphotic depth as a suspended sediment ESV in flowing water is provided in Appendix B. The key findings of this technical discussion and additional reasoning are presented below. It is emphasised that, if euphotic depth was not excluded on a technical basis, it could not be readily included as a suspended sediment ESV for either defining natural state or threshold derivation because there is insufficient available from national or regional monitoring programmes.

Davies-Colley and Nagels (2008) found that, although suspended sediment was the main controller of light penetration into NZ rivers, coloured dissolved organic matter (CDOM) also limited light penetration, as illustrated in Table 2.2. In NZ's generally shallow rivers, light 'shading' by the water column is seldom likely to prevent on plants occuring if suspended sediment is managed to protect other values, although there is potential to reduce primary production rates. Large (deep) rivers with persistent loads of suspended materials (leading to high light-attenuation) may be severely light-limiting (e.g., Julian et al. 2008). A study of fine sediment (clay) discharge impacts in West Coast, NZ streams (depth c. 0.3 m) showed that increases in median turbidity from 2.4 NTU upstream to 15 NTU downstream, reduced the average daytime (12 h) light at the bed in these naturally brown-water streams (i.e., sensitive to light further light reduction) by a median of 45% (from 340 $\mu E/m^2/s^1$) upstream) and primary production showed proportional reductions (Davies-Colley et al. 1992). However, the recommended limit for increases in suspended sediment concentration (5 $g\ m^{-3}$) and turbidity (5 NTU) to protect macroinvertebrates in these sensitive West Coast streams (Quinn et al. 1992) should also protect the river bed light climate for primary production. Any suspended sediment attributes to protect instream primary production in rivers would apply more to baseflow when other conditions are more suitable for plant growth (e.g., suitable current velocities and lack of scour/abrasion by bed load and sand particles) than at high flows.

The model framework from Davies-Colley and Nagels (2008) was used to estimate benthic irradiance (as a fraction of incident radiation) for 'average' and 'dirty' rivers at depths of 1 and 2 m (Table 2-2).

Table 2-2: Irradiance (PAR) at the bed of a NZ river as a fraction of incident irradiance.

River state scenario (i.e., SS and CDOM)	Visual clarity (m)	CDOM (g_{340}, 1/m)	K_d (PAR) (1/m)	Bed irradiance as a % of incident irradiance E_{bed}/E_o	
				1 m depth	2 m depth
average SS; average CDOM	1.28	4.10	1.03	38	15
High SS; average CDOM	0.36	4.10	1.95	16	2.7
High SS; high CDOM	0.36	12.2	2.46	10	1.0

For a NZ river 1 m deep (average) exhibiting average light-attenuation (median clarity = 1.28 m), 38% of the incident light (i.e., light at the surface) reaches the stream bed. These conditions would not be light-limiting to most benthic plant communities, but could reduce the rate of production (as illustrated in the West Coast stream study discussed above). In a very 'dirty' and coloured NZ river (95^{th} %ile clarity and CDOM), at 1 m depth 10% of incident light will reach the bed. Under these conditions the growth of some light-demanding benthic plants may start to be constrained (Duarte 1991). If the water was deeper, light limitation would be more severe. For example, if the depth of water in the previous example was increased to 2m, the proportion of incident light reaching the bed would decrease to 1% (the level commonly defined as the euphotic depth in lakes), extinguishing most benthic plants.

These simple calculations suggest that light limitation by water shading in NZ rivers is likely to be protected by attribute states for related ESVs making specific protection of light penetration in rivers unnecessary.

2.3.2 Suspended sediment data

Suspended sediment data (TSS, visual clarity and turbidity) was taken from the water quality data set used by Larned et al. (2015) for the national state and trends report for New Zealand rivers and lakes. This dataset had summary statistics for all NRWQN and regional SoE monitoring sites that met certain data requirements. At most sites, the summary statistics (in this case, medians) for turbidity and visual clarity were calculated for the 10-year period 2004 to 2013. The median number of data for TSS sites was 44, and sites with <18 observations were not included. The total number of sites with TSS, visual clarity and turbidity data was 514, 722 and 832, respectively. The respective medians (and 25th and 75th percentile values in parentheses) for TSS, visual clarity and turbidity were 3.4 g/m^3 (2.4-6.0 g/m^3), 1.5 m (0.9-2.7 m), and 2.5 NTU (1.1-5.1 NTU).

Table 2-3 provides a summary of the number of sites with suspended sediment measures, grouped by region.

Table 2-3: Summary of the number of water quality (WQ) sites with suspended sediment measures (TSS, visual clarity or turbidity). Turbidity and visual clarity data from collation of Larned et al. (2015) and TSS from Hicks et al. 2016)

Regional authority[1]		Number of WQ sites by variable		
		TSS[2]	Visual clarity	Turbidity
AC[3]	Auckland Council	29	26	29
BOP	Bay of Plenty Regional Council	19	32	35
ECAN	Environment Canterbury	79	89	91
ES	Environment Southland	67	67	67
GDC	Gisborne District Council	17		18
GWRC	Greater Wellington Regional Council	52	52	52
HBRC	Hawkes Bay Regional Council	44	49	49
HRC	Horizons Regional Council	79	82	82
MDC	Marlborough District Council	30		32
NCC	Nelson City Council	27	28	28
NRWQN	National River Water Quality Network		77	77
NRC	Northland Regional Council	20	31	31
ORC	Otago Regional Council	41	1	44
TDC	Tasman District Council	10	43	43
TRC	Taranaki Regional Council		10	10
WCRC	West Coast Regional Council		37	37
WRC	Waikato Regional Council		98	107
Total		**514**	**722**	**832**

[1] NRWQN data set includes sites from the 16 regions. [2] excludes sites with <18 data. [3] Auckland Council no longer measure visual clarity at their water quality monitoring sites.

2.3.3 Characterising natural state variation in suspended sediment measures

Calculating land cover areas for catchments upstream of suspended sediment sites

Land-cover was determined using the latest version of the New Zealand Land Cover Database (LCDB v4.1). Each site has a reach identifier (NZReach) assigned to it that links it to River Environment Classification (REC). The REC reach data has nodes that link the reaches to form a drainage network. For a given site, all upstream reaches were selected using 'from' and 'to' node information available in the REC stream network dataset. Watershed polygons for the selected reaches from the REC database were then combined to make the upstream catchment polygon. Using a Geographic Information System, this catchment polygon was then intersected with the LCDB v4.1 dataset (layer) to generate the area of each land-cover class (a total of 35) for the upstream catchment of all 832 water quality sites with turbidity and/or visual clarity data. Validation of this method is provided in Appendix C.

Definition of reference sites (i.e., minimally disturbed condition) for suspended sediment measures

A new set of criteria for defining reference sites, specifically related to a suspended sediment (using turbidity as a proxy), was developed for this project. We focussed on the turbidity data when testing and refining the criteria because this measure of suspended sediment had the largest number of sites (i.e., 832, compared to 722 and 514 for clarity and TSS, respectively).

The criteria used to define the upper bounds of suspended sediment reference (i.e., minimally disturbed) sites are summarised in Table 2-4.

The results of applying the LCDB v4.1 land-cover thresholds in Table 2-4 to suspended sediment data sets are summarised in Table 2-5. The numbers of reference sites for TSS, clarity and turbidity were 51, 85 and 92, respectively (approximately 10% of the total number of sites). The coverage of turbidity reference sites (N=92) is shown in Figure 2-1.

Table 2-4: LCDB v4.1 land-cover type thresholds for upper bound of 'reference' sites.

Landcover type (LCDB v4.1 class)	Threshold	Comment/notes
Natural Combination of 15 land-cover types (refer to comments/notes column)	>75%	Lower than Clapcott et al. 2016 of 85%, however this definition excludes short tussock and non-native types of scrub/shrublands. Accordingly, this threshold is difficult to compare with those derived using older versions of LCDB. Natural land-cover types included: Indigenous Forest (69); Broadleaved Indigenous Hardwoods (54); Manuka and/or Kanuka (52); Matagouri or Grey Scrub (58); Sub Alpine Shrubland (55); Tall Tussock Grassland (43); Fernland (50); Flaxland (47); Permanent Snow and Ice (14); Sand or Gravel (10); Alpine Grass/Herbfield (15); Lake or Pond (20); River (21); Estuarine Open Water (22); Herbaceous Freshwater Vegetation (45);
Heavy pastoral High producing exotic grassland (40) + short-rotation cropland (30)	<10%	Lower than REC threshold for pastoral land-cover (including horticultural cropping) of 25%. Greater than Clapcott et al. (2016) reference site definition of 5%. Settled on 10% as turbidity showed no apparent correlation with this combined land-cover type This is consistent with deposited sediment reference state definition, where a 10% heavy pasture threshold is used.
Light and heavy pastoral High producing exotic grassland (40) + short-rotation cropland (30) + Low Producing Grassland (41)	<15%	Somewhat arbitrary threshold. Introduced largely because of the relatively high heavy pastoral threshold used (10%). This combined 15% threshold attempts to recognise the additive pressure that different agricultural land-cover types may have on suspended sediment ESV's.
Regenerating native Fernland (50) + Manuka and/or Kanuka (52) + Broadleaved Indigenous Hardwoods (54)	<40%	Somewhat arbitrary threshold, but aims to take into account that largely native catchments comprised of regenerating land-cover types (to ultimately form indigenous forest cover) are not natural – i.e., most likely reflects the slow reversion of 'disturbed land' (cleared for pasture) back to native forest. During this reversion it is assumed that suspended sediment ESV's will be significantly higher than for indigenous forest. As such, it was considered appropriate to have a maximum threshold for regenerating native land-types.
Urban Built-up area (settlement) (1) + transport infrastructure (5)	<5%	More conservative than the REC threshold value of 15%, although the REC class included urban parks/open spaces – whereas LCDB4 v4.1 class *urban parkland/open space* (2) was excluded from the urban threshold
Mines/quarries Surface mine or dump (6)	0%	Potential point source discharges having significant impact on downstream water quality monitoring site. Easiest way to manage uncertainty from this land-cover type (which represents very small areas) is to set zero threshold
Wetland Herbaceous Freshwater Vegetation (45)	<50%	The classification system applies to flowing waters – wetlands are a special class and therefore sites with upstream catchments dominated by wetland land-cover (>50%) have been excluded.
Permanent snow and ice (14)	<10%	Attempt to eliminate rivers that have naturally high suspended sediment ESV values due to glacial flour. The intention is that the classification system for suspended sediment ESVs will not include glacial flour-impacted rivers. In Otago, these types of rivers (glacial source) are excluded from turbidity standards in the Regional Plan.
Indigenous forest (69)	na	Natural land-cover of upstream catchments vary depending on altitude and latitude. It was therefore considered too arbitrary to set a threshold for indigenous forest cover.

Table 2-5: Summary of reference (i.e., minimally disturbed condition) sites for different suspended sediment ESV measures.

SS variable	Total no. of sites	No. of reference site	% reference sites
TSS	514	51	10
Clarity	722	85	12
Turbidity	832	92	11

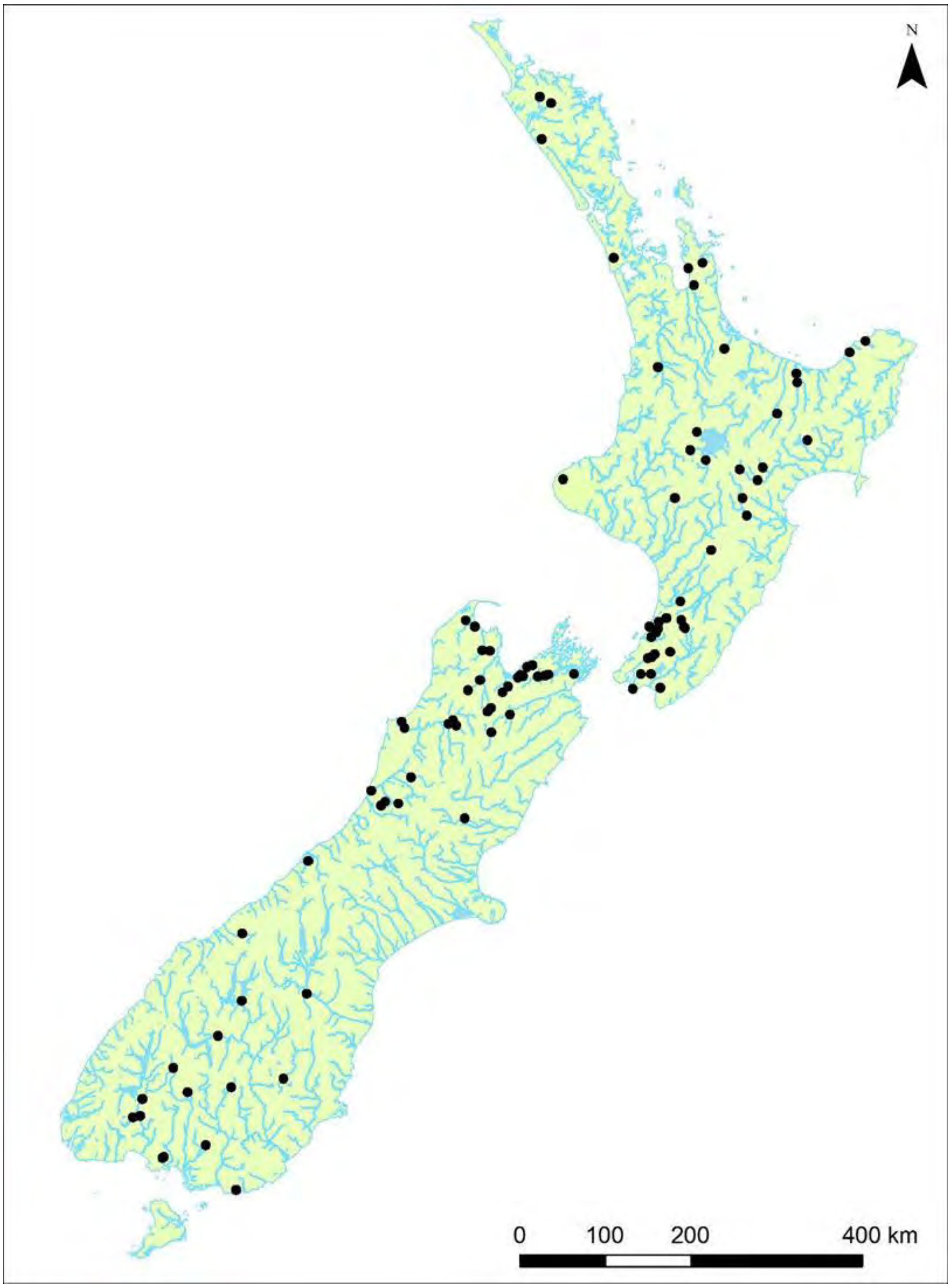


Figure 2-1: Distribution of the 92 reference sites with turbidity data. Reference sites were defined by applying catchment land cover (LCDB4) criteria in Table 2-4.

2.3.4 Natural (i.e., reference) state variation in suspended sediment measures (TSS, visual clarity and turbidity)

Summary statistics of long-term site medians for all data and defined reference sites for TSS, visual clarity and turbidity are provided in Table 2-6.

TSS values for reference sites (n=51) ranged from 0.3 to 6 g/m^3. The median, lower and upper quartile TSS values were 2.0, 2.0 and 3.0 g/m^3, respectively (reflecting the large proportion of data recorded as below detection limit). The method detection limits for quantifying TSS are relatively high (i.e., typically 2-3 g/m^3 for most regions), and therefore it is not a suitable metric for assessing natural state variation in suspended sediment. For the 51 TSS reference sites, 80% of the natural variation (i.e., 10^{th}-90^{th} %ile) was between 0.6 and 3.0 g/m^3 (i.e., <5-fold variation).

Visual clarity values for reference sites (n=85) ranged from 1.1 to 13.9 m. The median, lower and upper quartile clarity values were 2.6, 3.4 and 4.7 m, respectively. For the 85 visual clarity reference sites, 80% of the values (i.e., 10^{th}-90^{th} %ile) varied between 1.7 and 6.9 m (i.e., 4-fold variation).

Turbidity values reference sites (n=92) ranged from 0.1 to 3.8 NTU. The median, lower and upper quartile clarity values were 0.9, 0.6 and 1.3 NTU, respectively. For the 92 turbidity reference sites, 80% of values (i.e., 10^{th}-90^{th} %ile) varied between 0.4 and 2.2 g/m^3 (i.e., <6-fold variation).

Summary plots showing the distribution of reference sites against 'all data' for each suspended sediment measure are shown in Figure 2-2.

Table 2-6: Summary statistics for long-term site medians of suspended sediment measures at all sites and reference sites. Criteria used to defined reference sites are given in Table 2-4.

Statistic	TSS (g/m^3)		Clarity (m)[4]		Turbidity (NTU[3])	
	Reference n=51	All data n=514	Reference n= 85	All data n=722	Reference n=92	All data n=832
minimum	0.3	0.3	1.1	0.1	0.1	0.1
10^{th} %ile	**0.6**	**1.5**	**1.7**	**0.5**	**0.4**	**0.6**
25^{th} %ile	2.0	2.4	2.6	0.9	0.6	1.1
Median	**2.0**	**3.4**	**3.4**	**1.5**	**0.9**	**2.5**
75^{th} %ile	3.0	6.0	4.8	2.7	1.3	5.1
90^{th} %ile	**3.0**	**10.2**	**6.9**	**4.3**	**2.2**	**8.6**
maximum	6.0	99.0	13.9	13.9	3.8	80.0
IQR[1]	1.0	3.6	2.2		0.7	4.0
variation between 10^{th} & 90^{th} %iles	4.8-fold	6.8-fold	4.1-fold	8.5-fold	5.7-fold	14.3-fold
MAD[2]	1.0		1.0	0.8	0.3	1.6

[1] interquartile range. [2] Median absolute deviation (from the median). [3] In this report we are using NTU as a generic unit for nephelometric turbidimeters, but in practice (NEMS 2016) it is likely that turbidity would be reported in FNU (assuming ISO7027 turbidimeters are used). [4] note visual clarity is inversely related to TSS, hence 10^{th} and 25^{th} %ile visual clarity statistics are equivalent to the 90^{th} and 75^{th} %ile statistics (respectively) for TSS/turbidity measures.

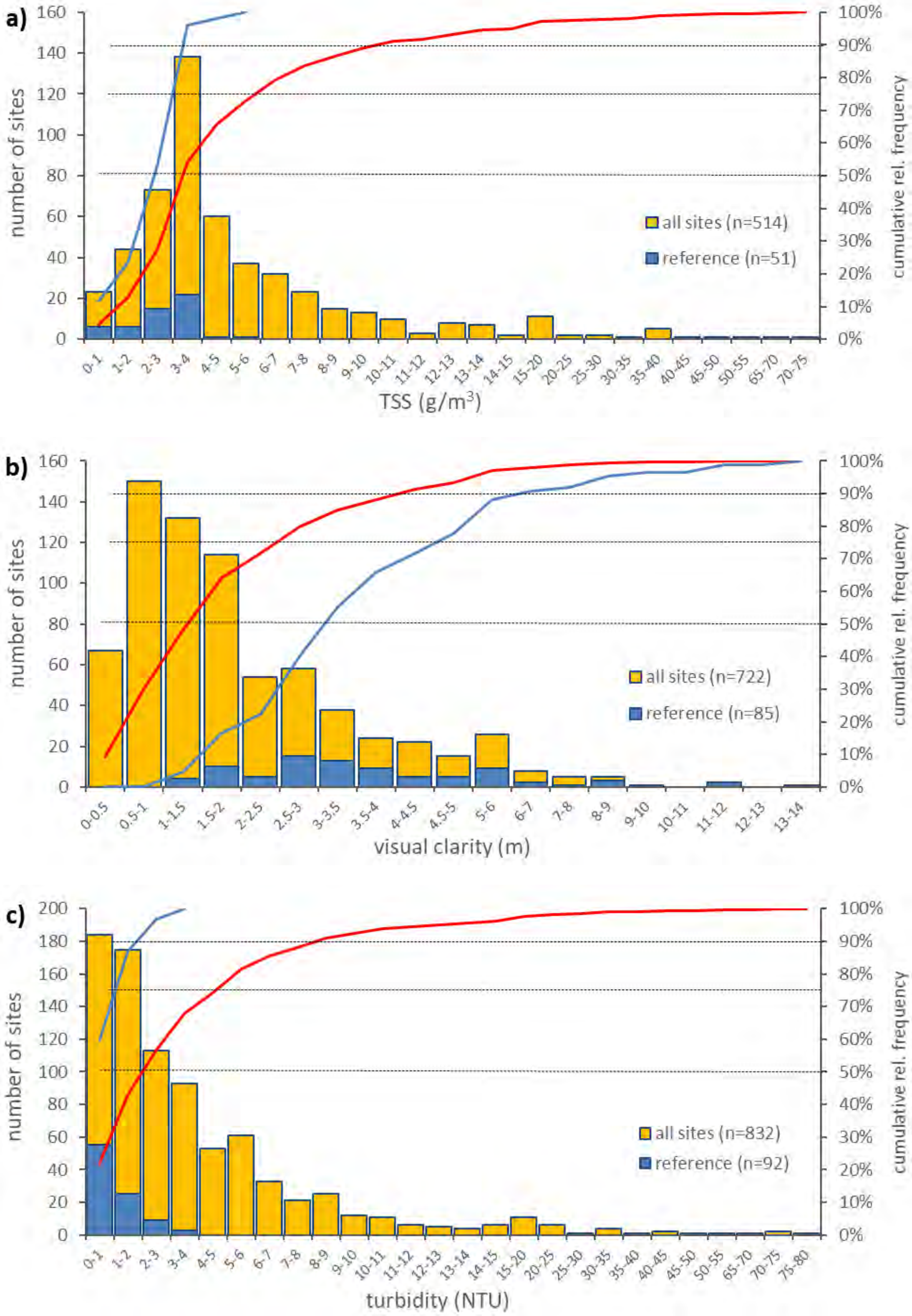


Figure 2-2: Distribution of suspended sediment measures (TSS, visual clarity and turbidity) across all sites (orange bars, red line) and reference sites (blue bars, blue line). Cumulative frequency (relative) distributions are plotted as curves with value corresponding to the right y-axis.

Unsuitability of total suspended sediment (TSS) concentration for defining natural state variation

TSS was not suitable for characterising natural state variation (as it is currently measured), and for reasons briefly discussed below, we did not include this sediment ESV in the threshold derivation workflow (limited to visual clarity and turbidity, refer to Chapters 5 (macroinvertebrates) and 6 (fishes).

Firstly, TSS had a lower number of reference sites – 51, compared to 85 for visual clarity and 92 for turbidity; furthermore, it does not appear to be routinely monitored by TDC, WCRC, WRC or TRC (based on 2004-2012 data). Secondly, in many regions, the method detection limits for TSS is too high to adequately characterise natural state levels of TSS – typically the detection limit is around 2 to 3 g/m^3. Across the 10 regions with TSS data, there were a total 2,483 data for the 51 reference sites, and of these, 1,337 (54%) were below the method detection limit. Twenty reference sites had greater than 50% of the data flagged as being below the detection limit; and at reference sites, 100% of the reported TSS data was <DL. Accordingly, for the purpose of defining reference state variation, TSS data, as collected across the country, was not suitable for defining reference state variation of the suspended sediment. As such, efforts were focus on proxy measures turbidity and visual clarity.

2.3.5 Ecological relevance of natural state variation of suspended sediment:

For reasons of brevity, the keys points presented in this section are made using turbidity as the suspended sediment measure.

For the 92 turbidity reference sites, 80% of values (i.e., 10th-90th %ile) varied between 0.4 and 2.2 g/m^3, which was less than a 6-fold variation. These values were consistent with predicted reference state values reported by McDowell et al. (2013), which were derived using a generalised linear mixed model (GLMM) (same method used for reference state predictions for fish threshold derivation -Chapter 6). When grouped at the 2nd level of the REC (topography nested in climate), the predicted reference state values for turbidity ranged from 0.5 to 2.5 NTU, with approximate medians for aggregate ‘cool’ and ‘warm’ climate classes of 1.0 ± 0.3 NTU and 2.0 ± 0.4 NTU, respectively (± median absolute deviation, MAD, from the median).

Following characterising the natural state variation, in the next step we had to determine whether a classification system to differentiate the observed reference state variation (i.e., 6-fold) was justified for the purposes of defining attribute thresholds. The answer to this depends on the thresholds being derived. The higher the ‘environmental standard’ (i.e., A/B band compared to lower C/D band) the lower the turbidity threshold values will be. The lower the threshold values, the more likely these thresholds will overlap with natural state levels of turbidity.

At a relatively early stage in the project, we made the decision to focus on deriving suspended sediment thresholds for ecosystem health that were consistent with the national bottom-line, or the C/D band threshold. A draft guidance document on implementation of the NPS-FM described an ecosystem health C-band as a state that *generally represents a minimum safe level before an ecological tipping point* (MfE 2014) – which corresponds to a significant adverse effect level. A brief review of relevant literature on effects of suspended sediment (refer to literature reviews in Chapters 5 and 6), indicated that adverse effects in the field were typical observed to occur around 7 to 10 NTU, with the lowest values being around 5 NTU (e.g., De Robertis et al. 2003; Lloyd 1987; Quinn et al. 1992).

Figure 2-3a shows all 92 turbidity reference site values, relative to the turbidity effects range (informed by literature review) that we considered represented a probably C/D band threshold (i.e., 5-7 NTU, red shaded area). The lower plot (Figure 2-3b) shows predicted reference state turbidity values (McDowell et al. 2013). Both figures show that for the purposes of defining a C/D threshold for suspended sediment, it is, arguably, not necessary to classify or group sites, simply because the observed natural state levels of suspended sediment (turbidity in this example) are markedly lower than published adverse effects levels that were anticipated to be consistent with C/D band thresholds.

It is emphasised that this 'single' suspended sediment class is only applicable to C/D band threshold derivation. Higher environmental objectives (A/B and B/C) threshold would require a more comprehensive classification system to account for natural state variation (which for this bands would be significant). Importantly, this approach was not valid for deposited sediment because in contrast to suspended sediment, deposited sediment naturally varies between 0 and 100% sediment cover and is influenced by reach scale variations in channel characteristics (see section 2.4).

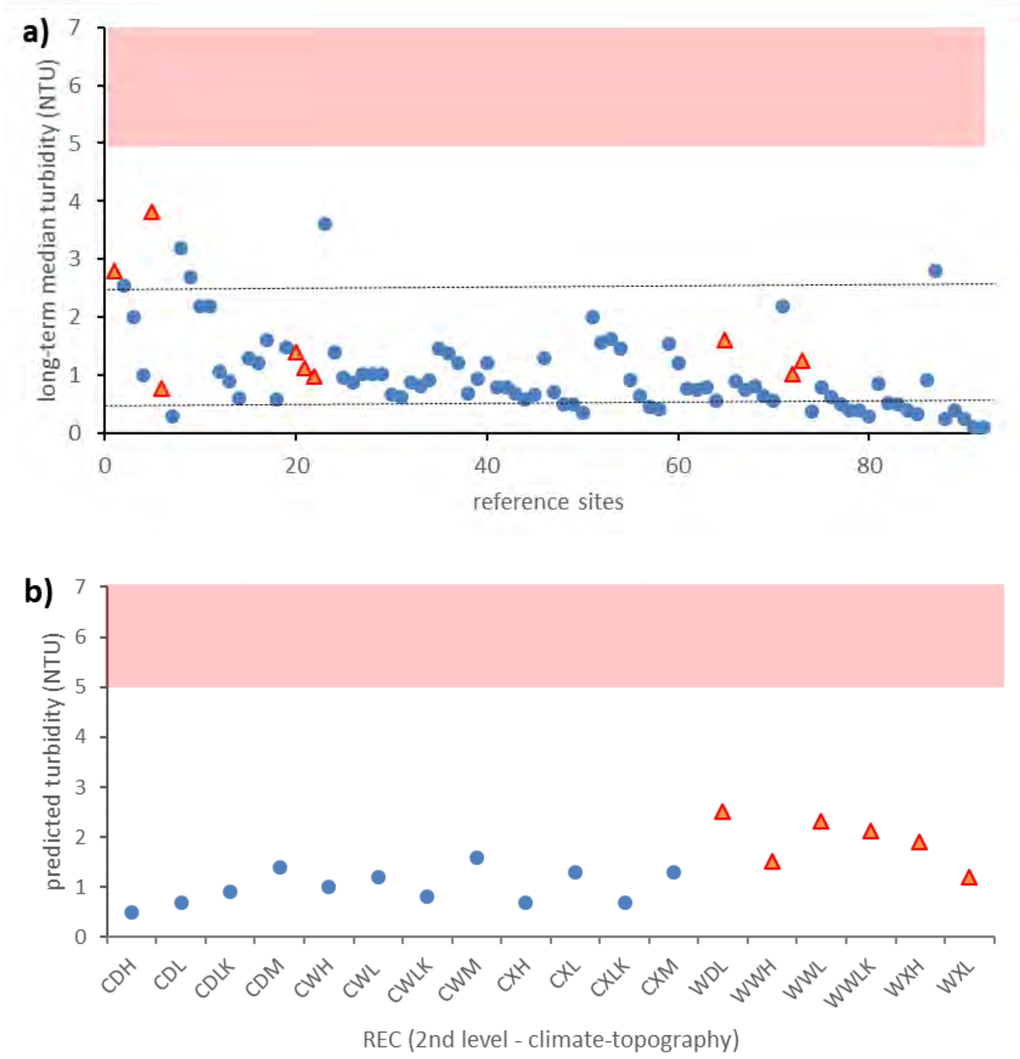


Figure 2-3: Measured (a) and predicted (b) estimates of natural state variation of median turbidity values (as median values) relative to an estimated adverse effect ranges informed from literature studies (red region, 5-7 NTU). 'Cool' and 'warm' REC classes are indicated by blue circles and orange triangle, respectively. The dashed black horizontal lines (a) represent the 10^{th} and 90^{th} %ile turbidity values of 0.5 and 2.5 NTUs, respectively.

2.3.6 Potentially useful REC-based classifications for suspended sediment

The previous section showed that for the purposes of deriving C/D band thresholds, it was unnecessary to differentiate suspended sediment sites into different classes. This was considered beneficial for analysing effects thresholds (macroinvertebrates - Chapter 5) as it allowed all available data to be analysed as one data set. Differentiation of suspended sediment into classes would resulted in ecological effects to be analysed on smaller data sets 'binned' according to the classification system.

It is, however, of interest to understanding how we might better describe the natural state variation of suspended sediments. A couple of reasons include:

- Future efforts develop higher environmental band thresholds (A/B and B/C) will require a relatively comprehensive classification system
- Although a single classification was justifiable for the analysis of ecological effects (e.g., macroinvertebrate responses;), a better understanding of reference state variation could allow 'reference offsets' to be applied to a 'general' threshold value.

Of the 92 turbidity reference sites, only 9 were associated with 'warm' climate classes. Figure 2-3(a/b) suggests that 'cool' climates have, on average, lower reference state turbidity values than 'warm' REC climate classes. Predicted reference values (McDowell et al. 2013) for 12 'cool' source-of-flow (topography nested in climate) REC classes ranged from 0.5 to 1.4 NTU with a median of 1.0 NTU. For 6 'warm' source-of-flow classes, the predicted turbidity values ranged from 1.2 to 2.5 NTU, with a median of 2.0 NTU. Climate is clearly not a driver of suspended sediment, and so assuming the predicted values are correct, the climate and/or topography class must incorporate landscape settings/drivers that influence suspended sediment. For example, warmer climates may be dominated by geologies/particle size distributions (i.e., % clayey soils) that are conducive higher levels of suspended sediment.

For the 83 'cool' REC climate reference sites, the measured median and 75th percentile turbidity values were 0.8 and 1.2 NTU respectively. For 'warm' climates (only n=9) the median and 75th values were 0.6 and 1.3 NTUs higher at 1.4 and 2.5 NTUs, respectively. A combination of predicted and measured estimates of reference state suspended sediment (using turbidity as an example) suggest that there it may be justified to apply a +1 NTU offset to the 'general' C/D band threshold derived or proposed for management of suspended sediment in NZ flowing waters. This offset, derived in turbidity units can be converted to visual clarity equivalents using a combination of 'turbidity-visual clarity' regression equations based on all paired data, reference paired data and NRWQN paired data (Appendix D).

This type of approach was broadly supported when we looked at the relationship of turbidity for all 832 grouped by REC climate (1st level) and plotted along an anthropogenic landscape setting gradient (high productivity grassland + exotic forest – LCDB4) (Figure 2-4). Simple exponential line fits gave Y-intercept values of 1.3 NTU for warm and 0.8 NTU for 'cool' climates – indicating at least a 0.5 NTU 'offset'. Interestingly, the distance between the 'warm' and 'cool' turbidity curves increases with increasing anthropogenic landcover. For example, above 0.6 proportional "anthropogenic land" coverage, the CD (n=168) and WD (n=47) climate classes have median turbidity values of 2.6 and 5.4 NTU, respectively – a difference of almost 3 NTU.

We realise that this is an overly simplistic approach, as Figure 2-5 shows the influence of 2nd level (topography) and selected major source-of-flow classes (i.e., WDL, WWL and CWL), and the importance of geology (3rd level - climate-topography-geology). Further analysis was beyond the scope of the project, but we have included this data (Figure 2-4 and Figure 2-5) to stimulate further discussion/development of more detailed suspended sediment classifications (if required).

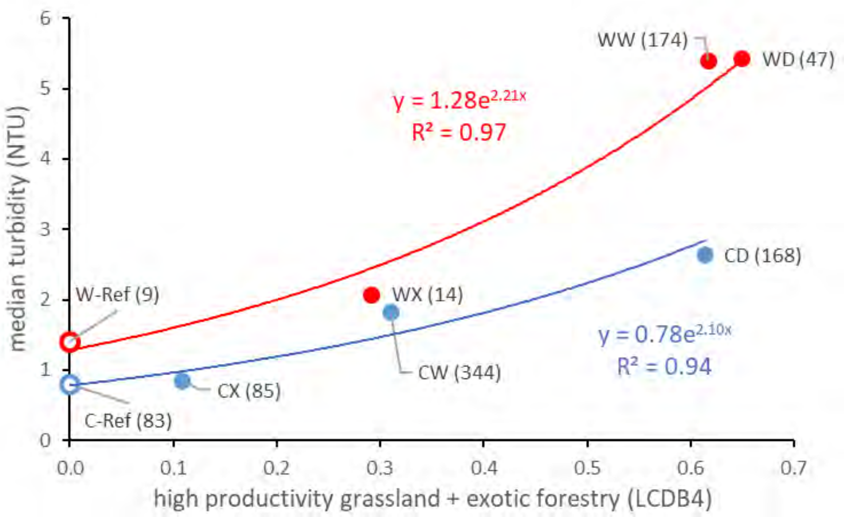


Figure 2-4: Median turbidity values group by REC climate classes plotted against anthropogenic landcover gradient (high producing grassland + exotic forestry; LCDB4).

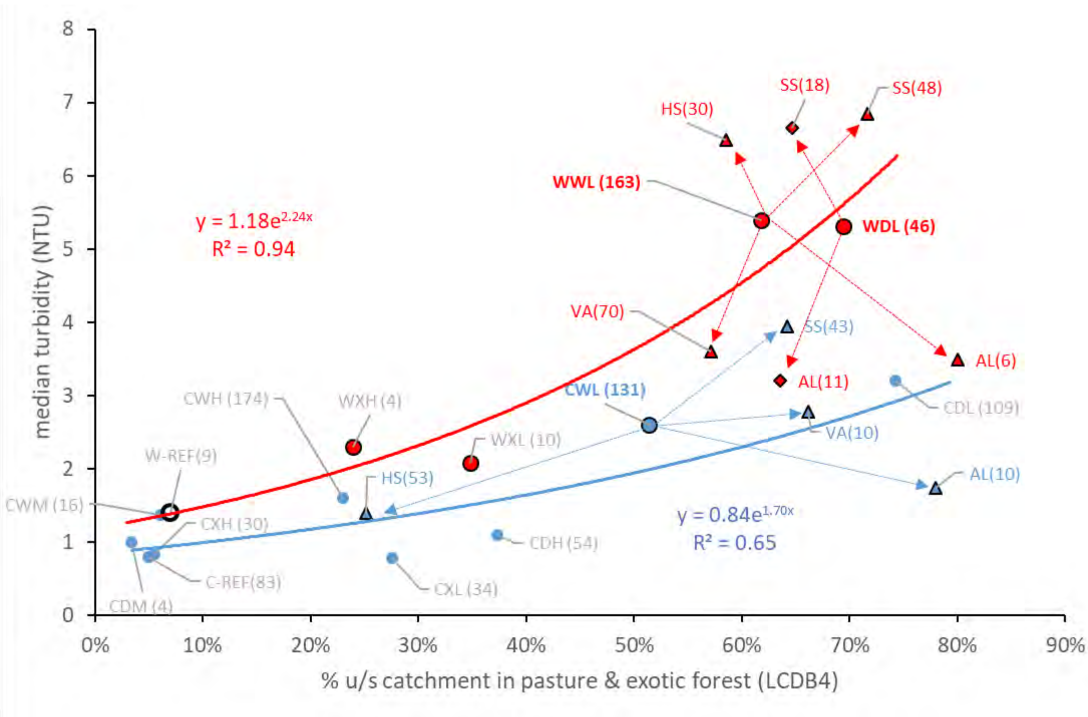


Figure 2-5: Median turbidity values group by REC source-of-flow classes (topography nested in climate) plotted against anthropogenic landcover gradient (high producing grassland + exotic forestry; LCDB4). The influence of the 3rd REC level (geology) is highlighted for selected major source of flow classes (WWL, CWL and WDL).

2.4 Deposited sediment

2.4.1 Background and methods

One step towards the development of a sediment NOF attribute is provision of knowledge of where attribute management bands should be applicable. Streams with naturally high deposited sediment volumes should not unnecessarily be categorised as degraded. Degraded streams should be identified as those where human activity is responsible for increasing deposited sediment from expected natural levels, to a degree that impacts the stream's ecological value. As such, knowledge of the natural or reference states is needed for any given stream. Reference state of a test site can be established from investigating a site or sites of a similar environmental condition but with minimal human impact (reference site) or predicted from a model relating sediment and land use, which requires sediment data from stream sites that cover a gradient of human impact (typically a land-use gradient).

Reference state will vary across the country due to natural environmental gradients such as the source and nature of the fine sediment (e.g., geology, soil), the delivery of the sediment (e.g., erosion, rainfall, elevation) and the ability of the stream to retain sediment (e.g., slope, flow). Understanding and classifying this variation is needed to determine where sediment attribute bands should be applied. For example, it would not be appropriate to impose a low-medium threshold level of 20% deposited fine sediment cover to a site where the natural level is 50% cover.

Classification systems provide a way to group sites based on their natural environmental characteristics (e.g., in this case, the sediment level that would be there if humans had not influenced it). Classes are defined by similarity within groups and dissimilarity among groups. Classification systems provide a framework for freshwater assessment, ensuring that appropriate management bands, or guidelines, are applied to appropriate stream classes. A classification system could be developed at the habitat, site, segment or catchment scale. It is important to understand the variation within each of these scales to assess the robustness of any given classification and its application. For example, the catchment scale has been proposed as the most suitable scale for the management of sediment (Owens 2008), but sediment deposition and erosion can occur at the habitat scale because they are controlled by shear stress, roughness and turbulence at the stream boundary layer. The River Environment Classification (Snelder & Biggs 2002) and following Freshwater Environments of New Zealand (FENZ, Leathwick et al. 2010) provide a segment scale framework for the development of a sediment classification. The FENZ database provides a wide selection of environmental measurements (calculated or predicted) for every stream segment in the national river network (the length of stream between tributary confluences, typically several hundred metres long).

In this section, we explore the variation in the natural state of deposited sediment to assign stream classes for the application of sediment NOF attribute bands. We used two datasets to do this. First, we examined spatial variation of deposited sediment levels within a dataset consisting of reference sites spread throughout New Zealand; reference sites were identified based on a set of land cover rules. Secondly, we built predictive models using sediment data collected from a range of sites across New Zealand and spanning a wide land use gradient. Models were used to predict the natural state for all stream segments in the national river network which we then used to also examine spatial variation.

2.4.2 Compilation of deposited sediment data

Existing data

An existing database compiled for the Sediment Stage 1 (Hicks et al. 2016) project included 628 unique sites where deposited sediment had been measured using standardised methods to derive a range of environmental state variables (ESV) of deposited sediment (Table 2-7).

Table 2-7: Description of assessment methods for six different measures of deposited sediment.

Deposited sediment measure	Description of the assessment method
Bankside visual assessment ('% cover bankside')	A rapid qualitative visual assessment of the % of fine (2< mm) sediment covering the streambed in a run habitat. Also known as SAM1.
Instream visual assessment ('% cover instream')	The average % cover of fine sediment covering the streambed in a run habitat calculated from a minimum of 20 stratified views using an underwater viewer. Also known as SAM2.
Wolman count (% fine)	The proportion of particles less than 2 mm in diameter recorded from a Wolman walk, or the measurement of a minimum of 100 particles picked up throughout a run habitat. Also known as SAM3.
Suspendable inorganic sediment (SIS; g/m^2)	The average amount of inorganic fine sediment entrapped and covering the streambed in a run habitat calculated from a minimum of 5 'Quorer' samples in a run habitat. Also known as the Quorer method and SAM4.
Suspendable benthic sediment volume (SBSV, L/m^2)	Same as SIS but sediment volume rather than weight is recorded.
Shuffle score (0-5)	An average qualitative assessment of the size and duration of a sediment plume resuspended when disturbing the streambed at 3 sites within a run habitat. Also known as SAM5.

This Stage 1 database was manually checked for accuracy of NZReach numbers (the ID for each segment in the national river network) and a significant number of errors were corrected. The errors were mainly due to an incorrect spatial link conducted during the Stage 1 project. Some NZReach assignments from a previous project, Sediment Assessment Methods (Clapcott et al. 2011), were also corrected. The SIS values for up to 50 sites were incorrectly assigned and this was also corrected during this data checking stage.

In addition, the proportion of fine sediment cover (mud, silt, sand) from 22,946 unique records were sourced from the New Zealand Freshwater Fish Database (NZFFD) as described in section 6.

New data

A request to regional councils in July 2016 resulted in the collation of new data which included 1364 unique sites where deposited sediment had been evaluated using the methods described above. Additionally, data collected using a new rapid habitat assessment method was compiled (Table 2-8). New data was assigned an NZReach based on matching site names or site ID numbers with the Stage 1 (Hicks et al. 2016) corrected dataset.

Table 2-8: Details of the rapid habitat assessment (RHA100) method.

Sediment method (metric)	Description
Rapid habitat assessment component 1 (RHA100)	A rapid qualitative visual assessment of the % of fine (<2 mm) sediment covering the streambed in a run habitat scored in the field on a scale of 1-10 and converted to % cover using guidelines provided on field sheets: 1 = 75%, 2 = 60%, 3 = 50%, 4 = 40%, 5 = 30%, 6 = 20%, 7 = 15%, 8 = 10%, 9 = 5%, 10 = 0%.

The Waihi Dam field study (Clapcott 2016) and an additional field survey undertaken in February 2017 (see section 3.3) resulted in the collection of deposited sediment data measured using a visual bankside assessment of percent cover of fine sediment (% cover bankside), as well as suspendable inorganic sediment (SIS). A total of 23 unique sites were added to the dataset.

A parallel MfE-funded macroinvertebrate project (Contract No. 21630) collated research datasets which were used to develop sediment-specific macroinvertebrate metrics. Some of these data were made available for our analysis in the sediment project. This additional research data included the deposited sediment measures of '% cover bankside', '% cover instream' and SIS (see section 3.5)

Summary of deposited sediment data

Deposited sediment data collated as part of this project were combined and summarised at two levels – site and samples per site. Level 1 summarises the number of unique sites (by NZReach) sampled within the preceding 5 years (2012-2017 inclusive), and Level 2 provides the number of unique sites with 3+ or 12+ replicate samples per site (by NZReach) collected within the preceding 5-year period (Table 2-9). The same information was given for all data available (not just the preceding 5-year period).

Table 2-9: Summary of deposited sediment data - by site (Level 1) and by site and number of samples per site (Level 2). Each sediment ESV is assessed with a different sediment assessment method (SAM), see Clapcott et al. (2011) for details.

	Level 1		Level 2			
Sediment assessment method (and ESV)	Number of sites	Number of sites in last 5 years	Number of sites with 3+ samples	Number of sites with 3+ samples in last 5 years	Number of sites with 12+ samples	Number of sites with 12+ in last 5 years
RHA100	661	659	64	62	24	23
'% cover bankside' (SAM1)	803	326	443	185	167	145
'% cover instream' (SAM2)	522	313	269	268	65	58
SAM3 (% fines)	741	461	236	147	62	56
SAM4 (SIS)[1]	438	95	63	3	0	0
SAM 4 (SBSV)[1]	75	24	0	0	0	0
SAM5 (Shuffle score)	167	25	4	0	0	0

[1]SIS = suspendable inorganic sediment (Quorer method); SBSV = suspendable benthic solids by volume.

2.4.3 Defining reference condition

Source data

We used the following upstream catchment land cover rules to define reference sites:

- > 90% native vegetation
- < 5% exotic vegetation
- < 10% pastoral heavy
- 0% urban.

This selection was based on *a priori* expectations of the relationship between land use and sediment delivery and retention in streams, and validated by the response of deposited fine sediment ('% sediment cover instream') to these land cover predictors, explored with a 4-predictor boosted regression tree (BRT) model (Figure 2-6).

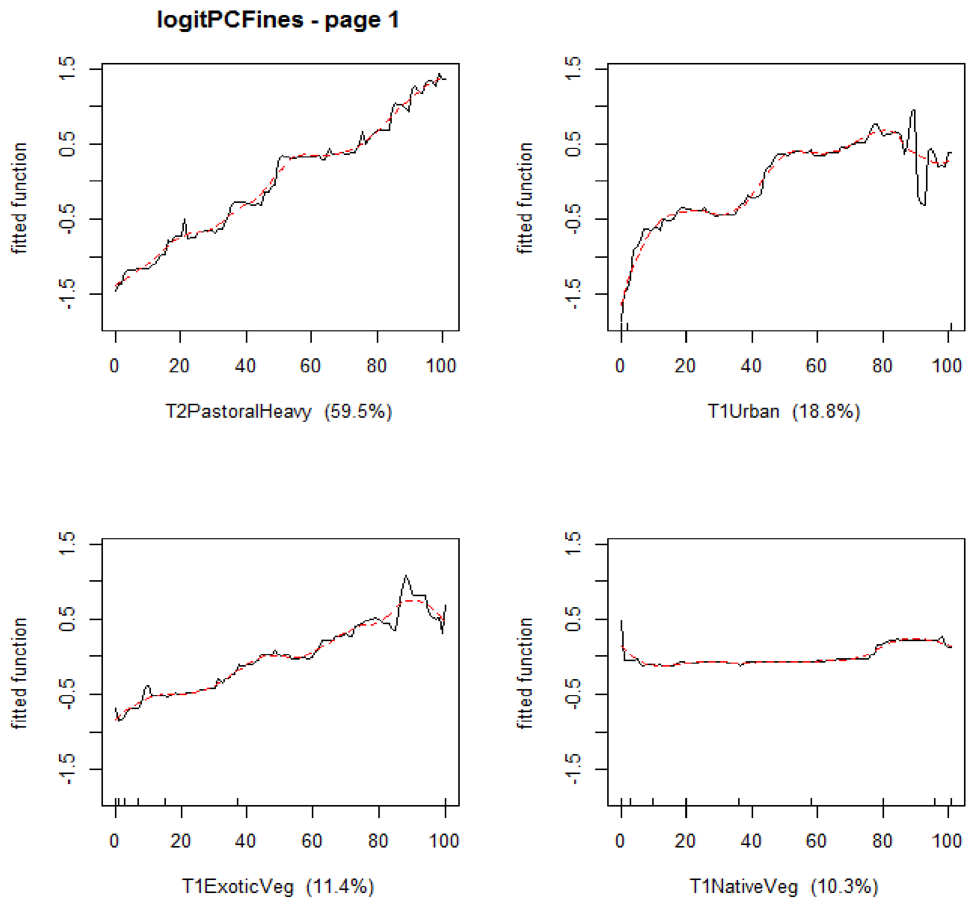


Figure 2-6: Shape of the relationship (fitted functions of the BRT) between deposited sediment cover (logit transformed) and land cover variables. Plots show distribution of data as rug plots on the x axis, the marginal contribution of each predictor to the mean sediment value on the y-axis, and the proportion of total deviance explained by each variable in parentheses. The black line is the fitted function and the red line is the smoothed fit.

We calculated the number of samples available to explore spatial variation in deposited sediment in three combined datasets (Table 2-10). Methods RHA100, SAM1 and SAM2 provide equivalent measures of the sediment cover of the streambed surface within a run habitat (Hicks et al. 2016). Hence, these data were combined and averaged for each NZReach to provide a *'% cover A'*. In addition, sediment data from 22,947 unique records of the percentage of mud, silt or sand were sourced from the New Zealand Freshwater Fish Database (NZFFD) as described in section 6. These were added to '% cover A' data to provide a '% cover B' dataset. Finally, SIS (SAM4) provides a measure of sediment entrapped within the top 10 cm of the streambed. Compared to surface cover of sediment, this measure has shown to better relate to sediment loads (Hicks et al. 2016) and may potentially also relate better to ecological indicators. Multiple replicate observations per site and sampling date were averaged for each NZReach.

Table 2-10: Number of samples at reference sites and total number of sites for each deposited sediment measure.

Deposited sediment measure	Sediment assessment method	Number of samples	Number of unique NZReaches	
			Reference sites	All sites
'% cover A'	SAM1, SAM2 and RHA100	13,391	93	1,452
'% cover B'	% cover A and NZFFD	36,504	2,022	15,281
SIS	SAM4	268	27	257

NZReach identifiers were used to compile environmental data for regression analyses. The primary environmental gradients we identified a priori included both reach-scale and catchment-scale stream descriptors (Table 2-11). Variables were chosen based on their likely mechanistic relationship to sediment delivery and deposition in streams. This excluded variables such as eastings, northings, and temperature, which may correlate with sediment but for which there is no mechanistic relationship to sediment delivery or retention in streams. Extensive exploratory analysis was also conducted of the relationship between environmental variables and sediment measures to further inform variable selection.

Variables were transformed where necessary to meet the assumptions of normality for linear regression including:

- Logit[9] transformation of '% cover A', '% cover B'.
- Log transformation of *SIS*, *Catchment sediment load*, *Stream power*, *Segment slope*, *USDaysRain*, *Specific mean flow*, *Specific low flow*.
- Square-root transformation of *Elevation*.

Table 2-11: Mean and range of three deposited sediment measures (response variables) and 22 predictor variables used in regression analysis. N = 15,281 sites from the '% cover B' dataset.

Variable	Description	Mean (range)	Source of data
Response variable			
% cover A	The average cover of sediment on the streambed in a run habitat (%); logit transformed.	25 (0–100)	this project
% cover B	The average cover of sediment on the streambed in a run habitat or at a reach scale (%); logit transformed.	30 (0.5–100)	this project
SIS	The average amount of inorganic fine sediment entrapped and covering the streambed in a run habitat (g/m^2); log transformed.	540 (1.1–11000)	this project

[9] A logit transformation is applied to proportional data to spread out upper and lower data bounds and approximate a normal distribution.

Variable	Description	Mean (range)	Source of data
Land cover predictor variables (n=6)			
Exotic cover	The cover of exotic vegetation in the upstream catchment including exotic forest, deciduous hardwoods, forest-harvested, gorse and mixed exotic shrubs (%).	8.6 (0–100)	LCDB3
Pastoral heavy cover	The cover of pastoral vegetation in the upstream catchment including high producing exotic grassland, short rotation crops, orchards, vineyards or other perennial crops (%).	31 (0–100)	LCDB3
Pastoral light cover	The cover of pastoral vegetation in the upstream catchment including low producing grassland, tussock and depleted grass (%).	19 (0–100)	LCDB3
Urban cover	The cover in the upstream catchment of urban parkland, built up areas, transport infrastructure, mines or dumps (%).	0.61 (0–100)	LCDB3
Surface water allocation	The low flow remaining after the upstream consented daily water allocation (not including groundwater abstractions or flow restrictions on allocations) is deducted (proportion).	0.03 (0–1)	Clapcott & Goodwin 2010
Environmental predictor variables (n=16)			
Catchment sediment load	Predicted sediment load for the total upstream catchment (t/y) as a function of rainfall, lithology and slope; log transformed.	5.7 (-3.9–17)	Hicks et al. 2011
Specific stream power	Product of the density of water (1000 kg m^3), acceleration due to gravity (9.8 m/s^2), mean flow (m^3/s), and segment slope (degrees), per unit width at mean annual low flow (m); log transformed.	-2.8 (-13–3.6)	Current project using Booker 2010
Elevation	Average segment elevation (masl); square-root transformed.	19 (0–53)	REC1
USSlope	Average slope in the catchment (degrees).	17 (0–55)	REC1
Segment slope	Average segment slope (degrees); log transformed.	0.31 (-5.3–4.1)	REC1
USPhosphorus	Average phosphorus content of rocks in the catchment, 1 = very low to 5 = very high.	2.4 (1–5)	FENZ; Leathwick et al. 2003
USHardness	Average hardness of rocks in the catchment, 1 = very low to 5 = very high.	3.3 (1–5)	FENZ; Leathwick et al. 2003
USCalcium	Average calcium content of rocks in the catchment, 1 = very low to 4 = very high.	1.5 (1–4)	FENZ; Leathwick et al. 2003
USDaysRain	Days ⁄ year with rainfall in the catchment > 25 mm; log transformed.	2.5 (0–4.9)	FENZ

Variable	Description	Mean (range)	Source of data
Specific mean flow	Mean annual flow divided by catchment area ($m^3/s/km^2$); log transformed.	-3.6 (-7–0)	Woods et al. 2006
Specific low flow	Mean annual 7-day low flow divided by catchment area ($m^3/s/km^2$); log transformed.	-5.6 (-12–1)	FENZ; Pearson 1995
Flow stability	Mean annual low flow ⁄ mean annual flow (ratio).	0.19 (0–0.63)	FENZ; Pearson 1995
FRE3	Average number of floods per year (based on the mean daily flow) exceeding three times the median flow.	15 (1.8–41)	Booker 2013
Geology	Categorical REC classification at the geology level.	NA	REC1
Order	Strahler stream order.	1 (1-8)	REC1
DSDistCoast	Distance to coast (km), from mid-point of each river segment; log transformed.	3.9 (-4.6–6.1)	FENZ

We explored the spatial distribution of sites in the *% cover B* dataset (including RHA100, SAM1, SAM2 and NZFFD data) across the country (Figure 2-7) and the distribution of sites across continuous environmental gradients (Figure 2-8) to assess their representativeness of stream types across New Zealand.

Spatially, there was a lack of reference sites for the eastern seaboard, especially for Canterbury and the Southern Alps (Figure 2-7). Despite visual similarity, there was also a significant difference between distributions for most environmental gradients when tested with a Kolmogorov-Smirnov test (Conover 1971); the subset of reference sites differed compared to the national network (summarised by relative number of segments not stream length) for all gradients except *USHardness* and *USCalcium* (Figure 2-8). These explorations suggest the *% cover B* dataset, despite being 10-fold larger than the *% cover A* dataset, does not represent the full range in potential reference states in New Zealand (Figure 2-8). Based on this, we chose to predict reference state for all stream segments.

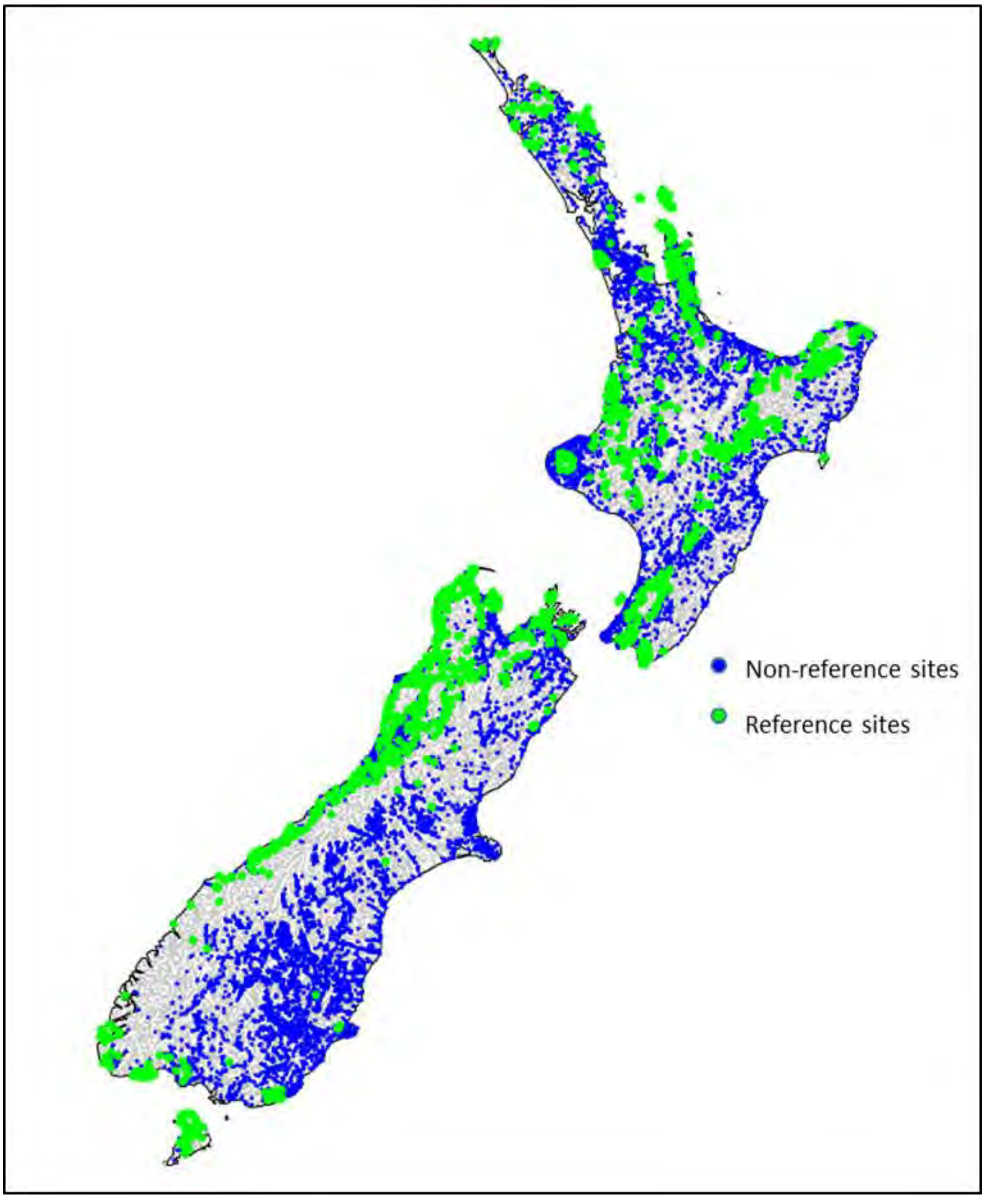


Figure 2-7: Distribution of reference sites (green; n = 2,022) and non-reference sites (blue; n = 13,259) where % cover B data was collected.

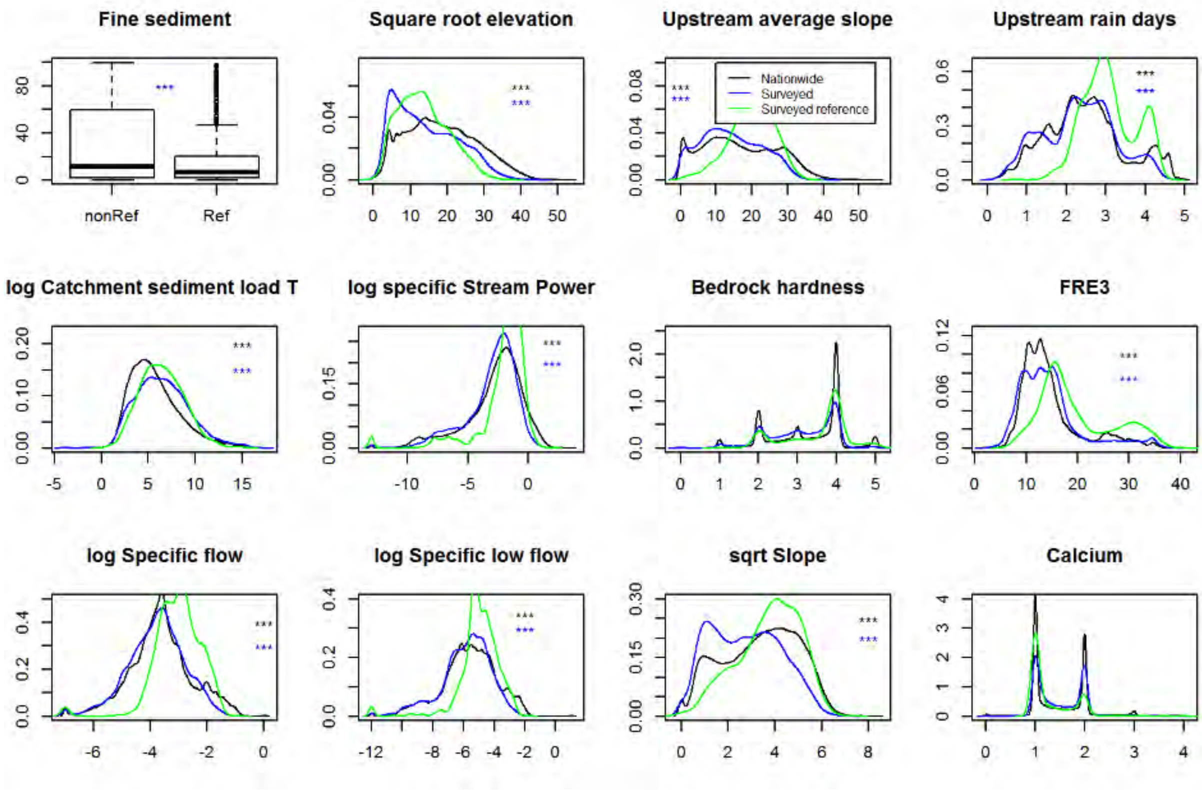


Figure 2-8: The distribution of reference sites and all sites in the % cover B dataset, relative to all stream segments in New Zealand, across continuous gradient of environmental descriptors. Stars in the top right indicate significance of the difference between the distributions of the reference sites (green, n = 2,022) and non-reference (blue, n = 13,259) and nationwide stream segments (black, n = 576,688), according to a Kolmogorov Smirnov test: *** = 0.001, ** = 0.01, * = 0.05.

2.4.4 Predicting reference state for all NZReach segments

We chose to estimate deposited sediment reference state for all stream segments using a flexible spatial regression approach, namely boosted regression tree models (BRT) (Elith et al. 2008). This machine learning method fits relationships of complex shapes, i.e., the relationship is not constrained to being either linear, quadratic, logistic or geometric, but may rise and fall to best fit the training data. Overfitting (i.e., model complexity) is avoided by internal cross-validation during model building.

We developed two BRT models using the '% cover B' dataset (SAM1, SAM2, RHA100 and NZFFD data). The first model (BRT REF) used data from reference sites only (n = 2,022) and the second model (BRT ALL) used data from all sites (n = 15,281) as training data.

Predictive model built on reference site data only (BRT REF)

For the reference sites only, we modelled sediment as a function of environmental variation. We tested the performance of the model by plotting the observed (measured) data against the predicted (modelled) data and calculating the following model performance indicators:

- the Nash-Sutcliffe efficiency (NSE) statistic which indicates how well the plot of observed versus predicted fits the 1:1 line, where values greater than 0 are satisfactory but values greater than 0.5 indicate good model performance (Nash & Sutcliffe 1970)

- root mean squared deviation (RMSD) is an estimate of model uncertainty (overall departure between observed and predicted values), where smaller values indicate lower uncertainty than large values (Piñeiro et al. 2008)
- bias, which measures the average tendency of the predicted values to be larger or smaller than the observed, where positive values indicate model underestimation and negative values indicate overestimation bias.

The sixteen environmental predictor variables explained 53% of the deviance in the reference site deposited sediment data (Figure 2-9) and the model had a cross-validation correlation (CV) coefficient of 0.55. The most important predictors of deposited fine sediment (% cover), contributing to almost half of the total deviance explained at reference sites were elevation, upstream slope, geology and specific stream power. Fitted functions showed predominantly meaningful response curves. For example, sediment cover decreased with increasing elevation, upstream slope, segment slope, specific stream power, and flow stability. The NSE statistic suggested good model performance (0.53) and there was effectively no model bias (Figure 2-10).

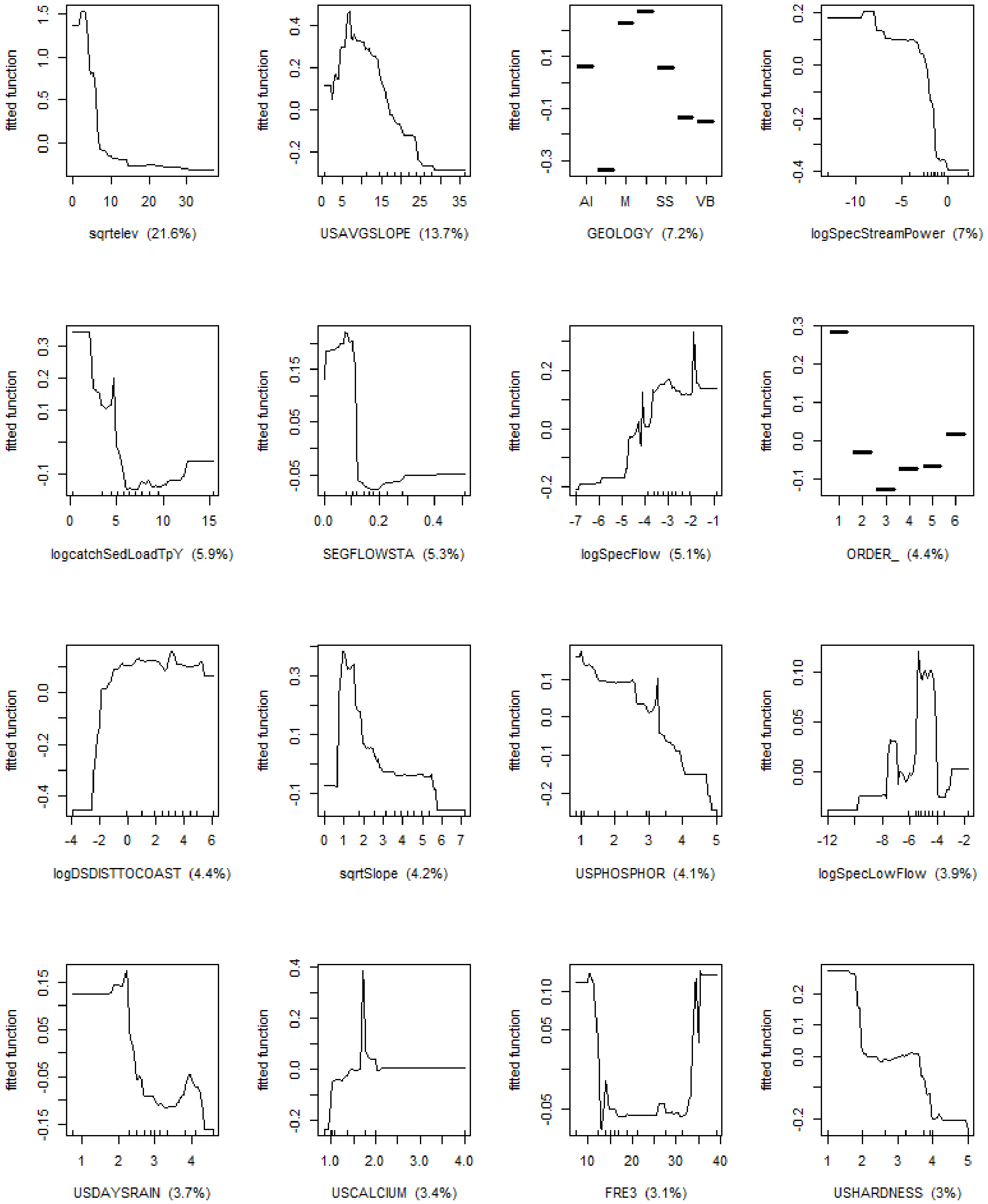


Figure 2-9: Shape of the relationship (BRT fitted functions) between deposited fine sediment cover (logit transformed) and environmental variables. Relationship applied at 2,022 reference sites in the % cover B dataset. Plots show distribution of data as rug plots on the x-axis, the marginal contribution of each predictor to the mean sediment value on the y-axis, and the proportion of total deviance explained by each variable in parentheses.

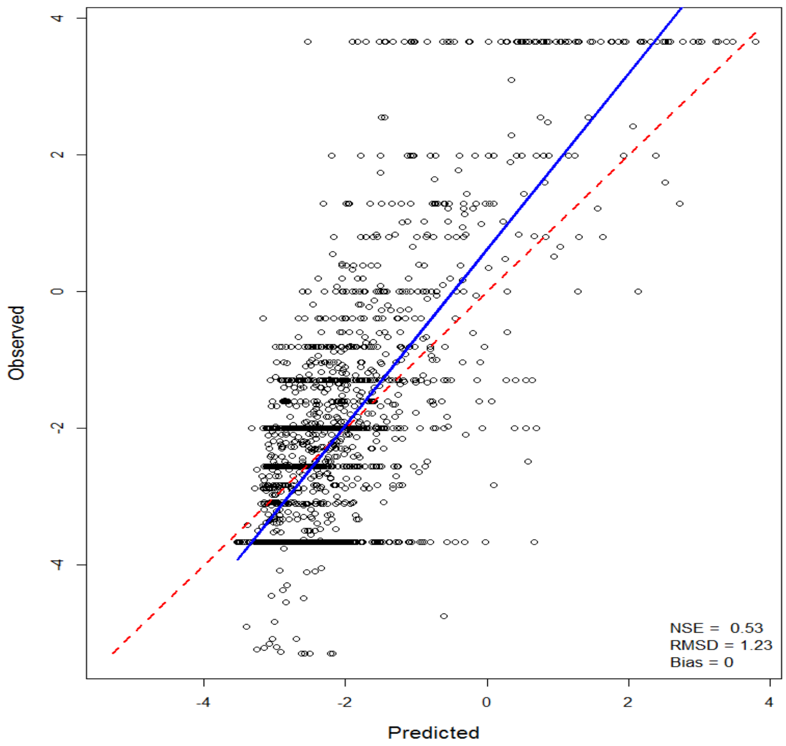


Figure 2-10: Scatterplot of the relationship between observed and predicted (logit transformed) sediment cover values from the BRT REF model (n = 2,022). The dashed red line is the 1:1 line and the blue line is the line of best fit. Model performance statistics presented at the bottom right are explained in text.

We used the BRT REF model to predict natural deposited sediment state for all stream segments in New Zealand and viewed the output map (Figure 2-11). Based on expert opinion, the distribution of deposited fine sediment at a national scale seemed reasonable. For example, higher sediment values were predicted for parts of Stewart Island and Northland where higher sediment is observed at reference sites. On the other hand, in other parts of the country where overall high '% sediment cover' (>10-30) values are predicted but few reference sites exist (e.g., for large areas of the central plateau, in Canterbury and Otago), it is difficult to determine how realistic these predictions are.

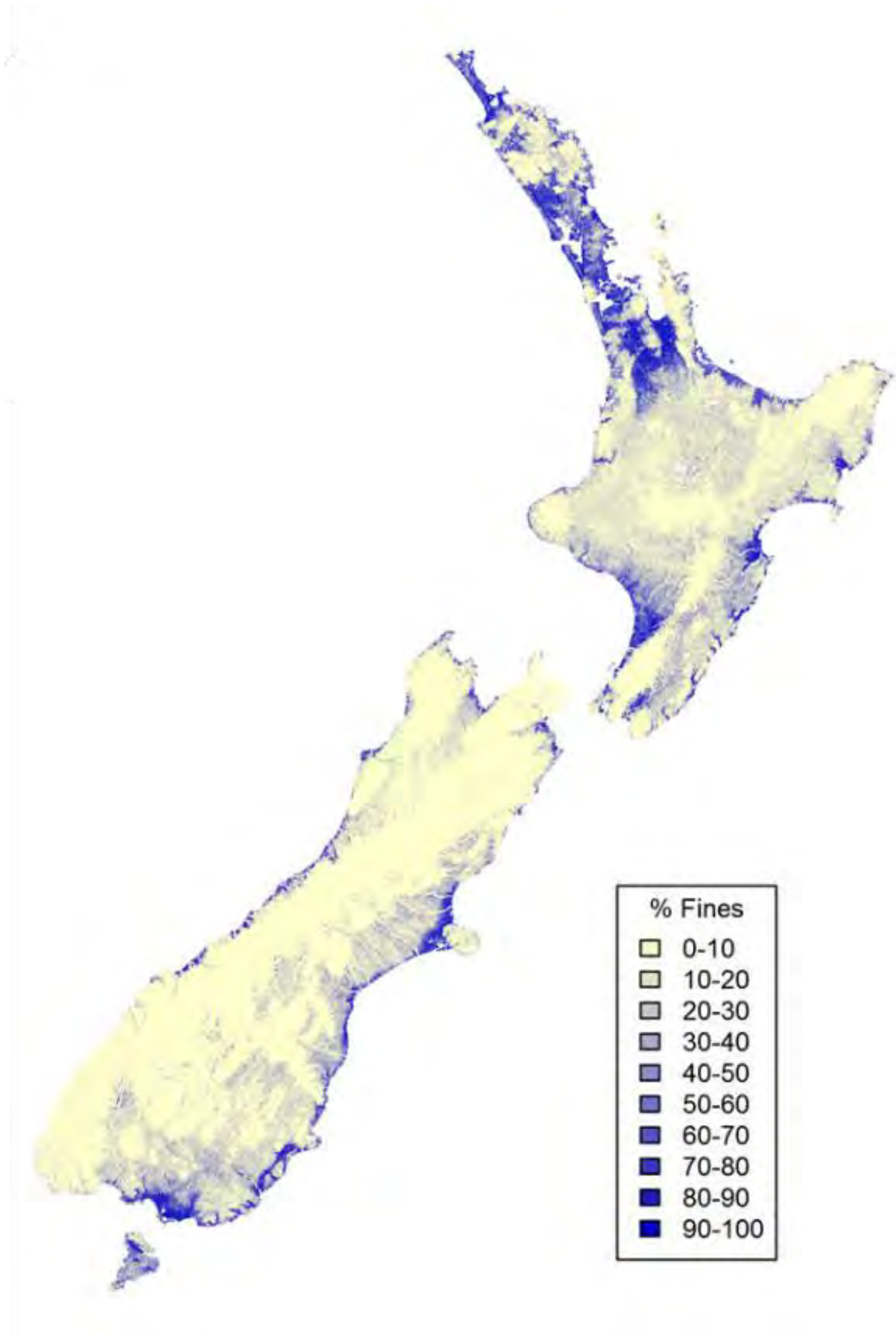


Figure 2-11: Map showing predicted reference sediment cover (% Fines) from BRT REF model (n = 2,022 sites from the % cover B dataset).

Predictive BRT model built using all data (BRT ALL)

We fitted the second BRT model using all sites data in the *% cover B* dataset (n=15,281), in two steps. First, we modelled sediment as a function of 5 land cover predictors (*Native vegetation*, *Exotic vegetation*, *Pastoral heavy*, *Pastoral light*, *Urban*) and *Surface water allocation*. This preferentially assigns the effect on sediment to the land cover variables, rather than to environmental variables where they may be collinear. Predictors were only retained in this step if they showed an expected relationship with sediment (necessary for setting values to zero), and if their inclusion significantly improved the percentage of total deviance explained (TDE). In order of importance, pastoral heavy, urban, exotic vegetation, and native vegetation were retained in the model (Figure 2-12), with a CV coefficient of 0.46 (fair-to-good), and explaining 23% (fair) of the TDE in the data. The land cover predictors of pastoral light and surface water allocation were excluded from subsequent analysis.

Then we modelled the residual variation in the land cover model using environmental predictors. All environmental predictor variables (n = 16, refer to Table 2-11) were retained in this step with 49% TDE (fair- good) and a CV of 0.50 (fair-good). Combined, these ‘two steps’ can be used to predict contemporary and reference sediment levels for all stream segments of the national river network.

To estimate reference state, we set the 4 retained land cover predictor variables to zero to estimate the average value of % sediment cover in the absence of anthropogenic pressures using the first step of the model. To estimate contemporary sediment cover, we added the fitted functions from the second step of the model (i.e., the 16 environmental predictors) to the average % sediment cover value for each site.

The model performance statistics suggest a fair-to-good predictive model: NSE = 0.46 and low bias suggested slight underestimation of predicted % sediment cover values (Figure 2-12). Comparison of natural state predicted from the BRT ALL model with observed sediment values at 2,022 reference sites defined by land cover rules (correlation coefficient of 0.5, data not shown) also suggests good model performance.

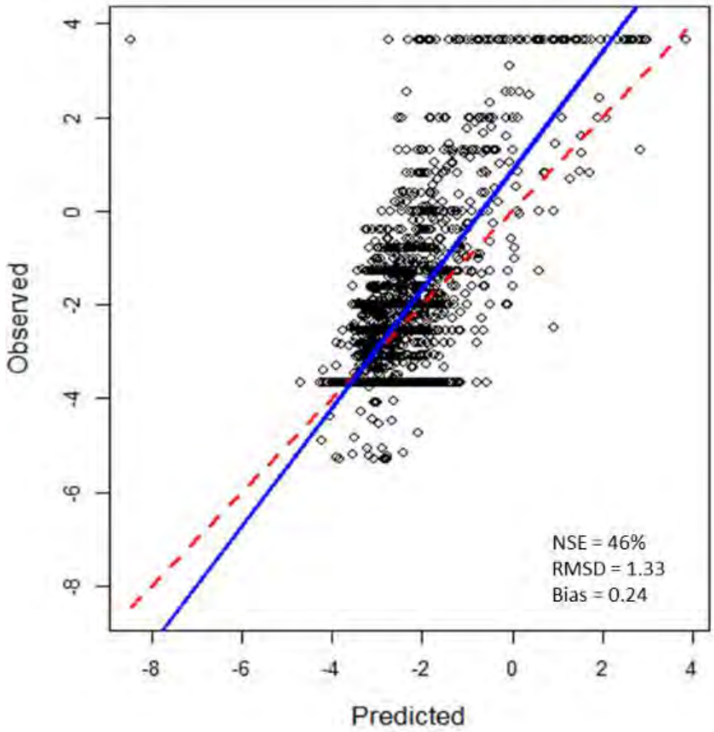


Figure 2-12: Scatterplot of the relationship between observed and predicted (logit transformed) sediment cover values from the BRT ALL model (n = 15,281). The dashed line is the 1:1 line and the blue line is the line of best fit. Model performance statistics are explained in text.

We predicted reference state for all stream segments in New Zealand using the BRT ALL model and plotted output on a map (Figure 2-13). Based on expert opinion, the distribution of sediment reference state at a national scale seemed reasonable, for example, higher sediment values in Stewart Island, Waikato, Bay of Plenty and Northland were predicted. High values in central Otago and the Clutha River, on the other hand, are not expected based on expert opinion.

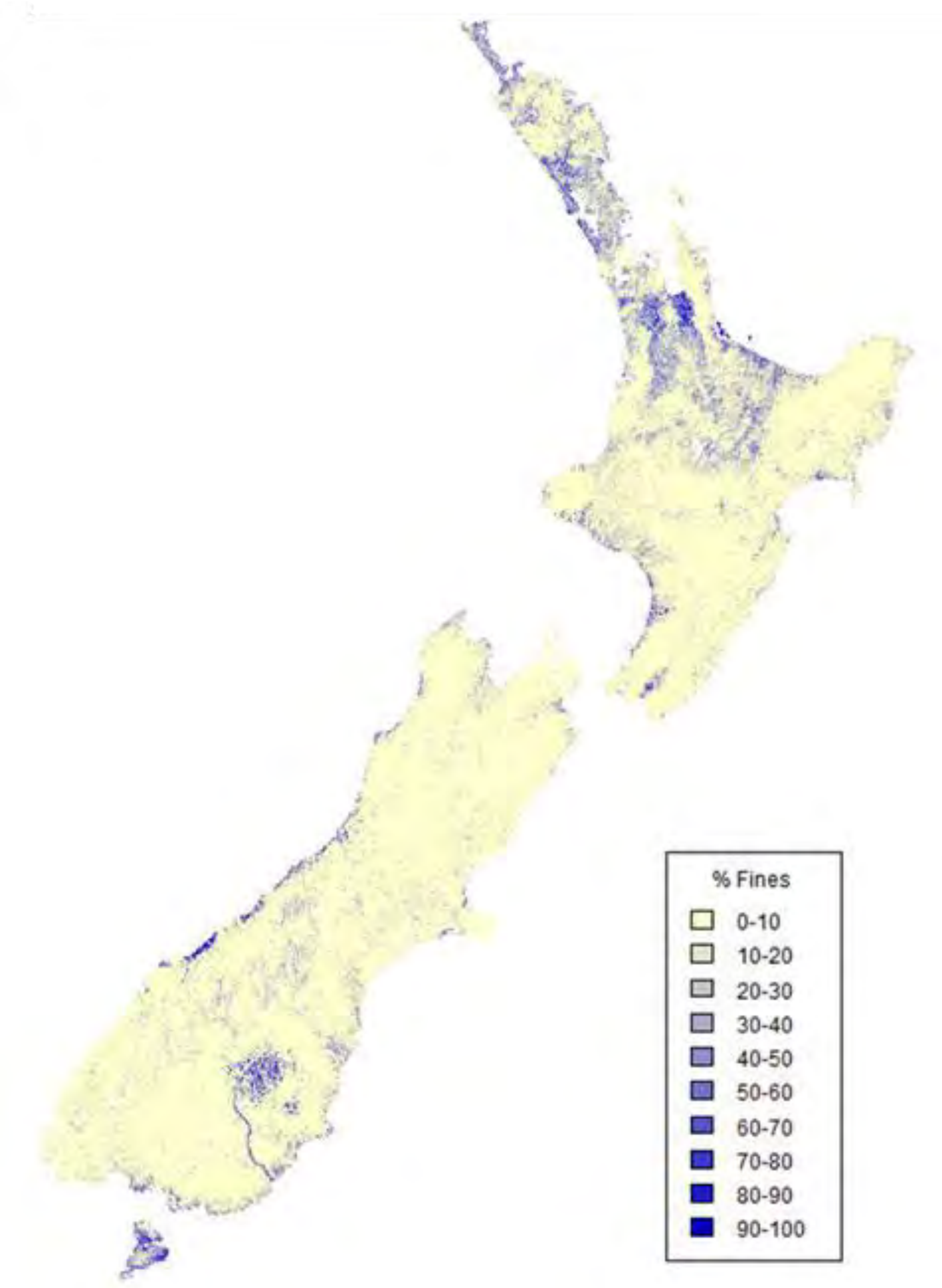


Figure 2-13: Map showing predicted reference (landcover predictor variables set to zero) *% sediment cover* from BRT ALL model using the '% cover B' dataset (n=15,281).

Evaluation of BRT modelling approach

We used two model approaches (BRT REF and BRT ALL) to predict reference state for the national stream network. Both models showed good predictive performance when validated internally on the training dataset using cross validation. Model performance was much improved compared to that observed in an earlier analysis (Clapcott & Goodwin 2017), and can probably be attributed to the increased size of the training dataset.

The BRT REF model based on variation in existing reference site data (n=2,022) resulted in meaningful spatial predictions at the national scale. However, in areas where training data was relatively scarce (e.g., low slopes, few rain days greater than 25 mm, and low flows), predicted reference state was often higher than what was expected using expert opinion. Model performance statistics suggested ‘good’ model performance and did not detect any model bias (NSE = 0.53, RMSD = 1.23, Bias = 0), but there were no sites within these environmental gradients to test predictions.

By contrast, the BRT ALL model (developed using all available deposited sediment cover data, n=15,281) assumes that land cover is a major predictor of deposited sediment in streams and assigned almost half the total deviance explained in the sediment data to land cover predictors. This means that any covariance between natural environmental gradients and land cover was assigned to land cover, which may have resulted in unnaturally low sediment levels in these stream types, e.g., lowland, low slope streams. Also, the BRT ALL model statistics also suggested ‘fair- good’ model performance (NSE = 0.46, RMSD = 1.33, Bias = 0.24), which was not as good as the BRT REF model.

The latter model also predicted some odd spatial patterns in the national river network compared to the BRT REF model output. Based on this, we chose to further explore the spatial variation in natural state predictions from the BRT REF model.

Reference state predictions for deposited sediment were also developed using a generalised linear model (GLM) as a component of developing sediment thresholds for fishes with an REC-based classification (see section 6). We have not compared the model output from the BRT and GLM models to test which model provides the highest predicative accuracy; however, a summary comparison of the methods is provided in Appendix C.

2.4.5 Grouping of reference state predictions from the BRT REF model

Classification and regression tree approach

We used a classification and regression tree (CART) method (De'ath & Fabricius 2000) to explore whether continuous reference site predictions could be meaningfully divided into a small number of categories, such as *high*, *medium* and *low* deposited sediment. Each NZReach segment was weighted equally. We did not restrict the number of splits or resulting categories that the model could produce. CART performance was assessed using the predicted residual error sum of squares or ‘PRESS’ statistic, calculated using hold-one-out validation during model building. The response variable was predicted natural sediment cover and our predictor variables were the same 16 environmental variables used in BRT model development (Table 2-11).

The PRESS statistic suggested very good CART model performance (R^2 = 0.78) for the nationwide reference state predictions from the BRT REF model. The initial split at the head of the tree was made based on *sqrt Elevation* (≥ 6.4). Below that, group splits were based on *USAvgSlope*, </≥ 15 and </≥ 14. *Geology*, *Stream order* and *Specific stream power* were other important variables in classifying sediment predictions (Figure 2-14).

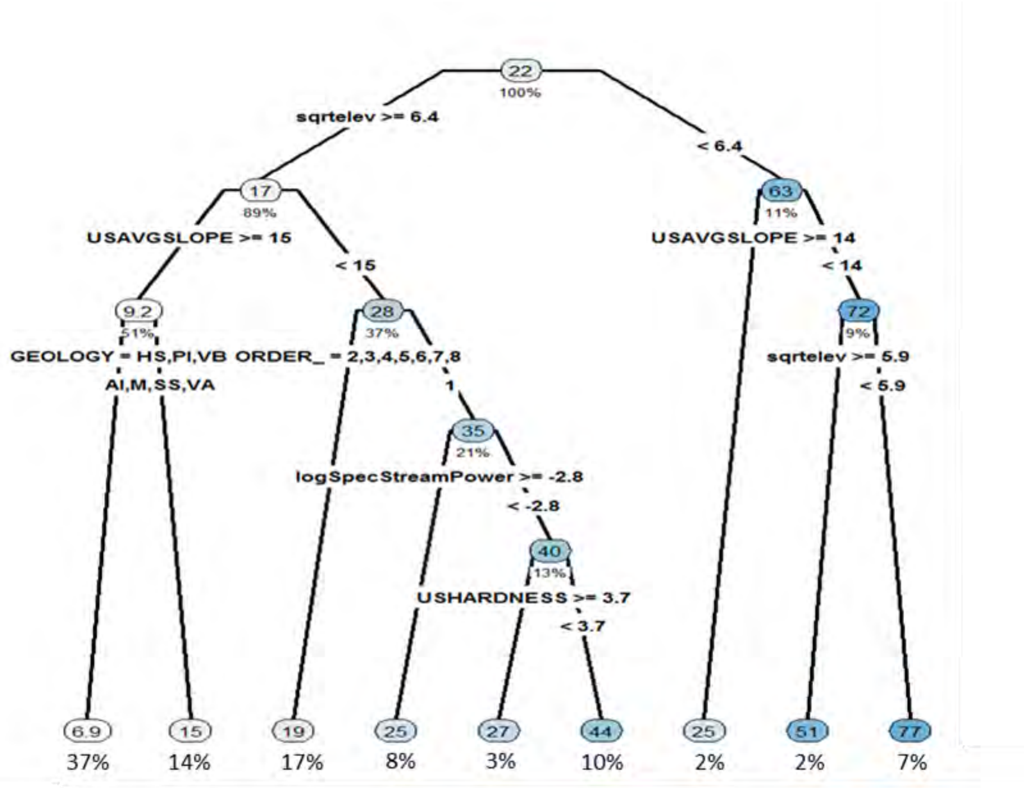


Figure 2-14: Decision tree structure for a CART model of reference sediment predictions from the reference site only data in the *% cover B* dataset. Each node is labelled by the average sediment value for the group resulting from this model and beneath each node is the percentage of the training data. At each intermediate node is the condition that determines whether the case goes to the left or right child node.

The CART resulted in 9 classes with overlapping distributions (Figure 2-14). The largest group (representing 37% of the stream network) had a mean sediment cover value of 7%. The second and third largest groups (representing 14% and 17% of the stream network) had medians of 15% and 19% sediment cover, respectively. Visual inspection of summary CART output (Figure 2-15) showed that streams could be pragmatically grouped into three broad classes based on predicted reference state – naturally hard-bottom streams with mean sediment values less than 30% sediment cover ('low-med'), and those with mean sediment values between 30% and 60% ('high'), and naturally soft bottom streams with greater than around 60% sediment cover.

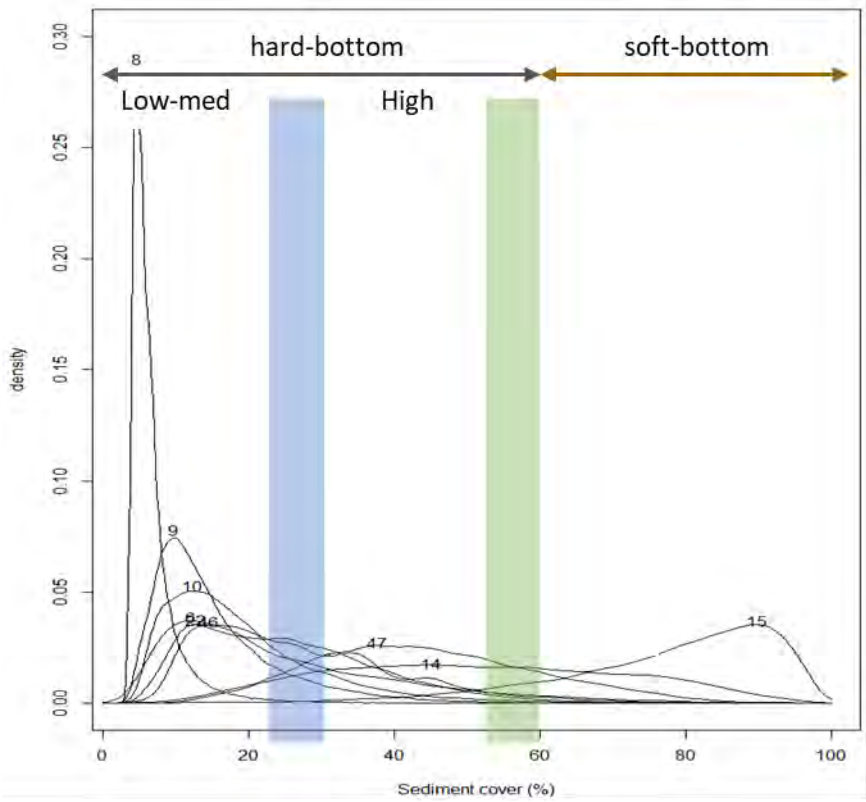


Figure 2-15: Distribution of % sediment cover predictions for all stream segments in the river network within each of nine sediment classes determined with a CART model. Blue and green bands represent transition regions (visually assessed) which pragmatically differentiate the continuum of predicted reference state condition of % sediment cover into three broad classes. Predictions were made using the BRT REF model, which was trained on data collected from reference sites (n = 2,022).

River Environment Classification approach

We tested the usefulness of the River Environment Classification (REC) for sediment classification by grouping the predictions from our preferred model (BRT REF) by the components of the REC classification separately: climate, source of flow, geology, and land cover (rather than testing REC groups that are hierarchically determined). Testing for difference between pairs within these categories suggests that most have distinct sediment levels (Figure 2-16), but the majority of REC 'classes' averaged less than 20% sediment cover and the median of classes was 13% (mean = 22%). Based on these results it was determined the REC was unsuitable for grouping segments into low, medium and high deposited sediment levels, i.e., all but wetland streams would be grouped together.

- For climate, all class medians were less than 12%, except for cold-dry climates (22%), warm-dry climates (WD = 63%), and warm-wet climates (26%).
- For source-of-flow, all class medians were less than 14%, except for lowland streams (48%).
- For geology, all class medians were less than 17%, except alluvium (34%), soft sedimentary (26%) and miscellaneous (34%) geologies found in areas of Auckland, Waikato, Manawatu and Canterbury.
- For land cover, all class medians were less than or equal to 18%, except wetlands (W = 89%) and miscellaneous (38%) land covers including riparian areas, mangroves and coastal dunes. Urban (73%) and pasture (28%) streams also had higher predicted reference values which is

likely to reflect the areas in which these land uses occur, for example, predominantly low slope, coastal areas, rather than the land cover.

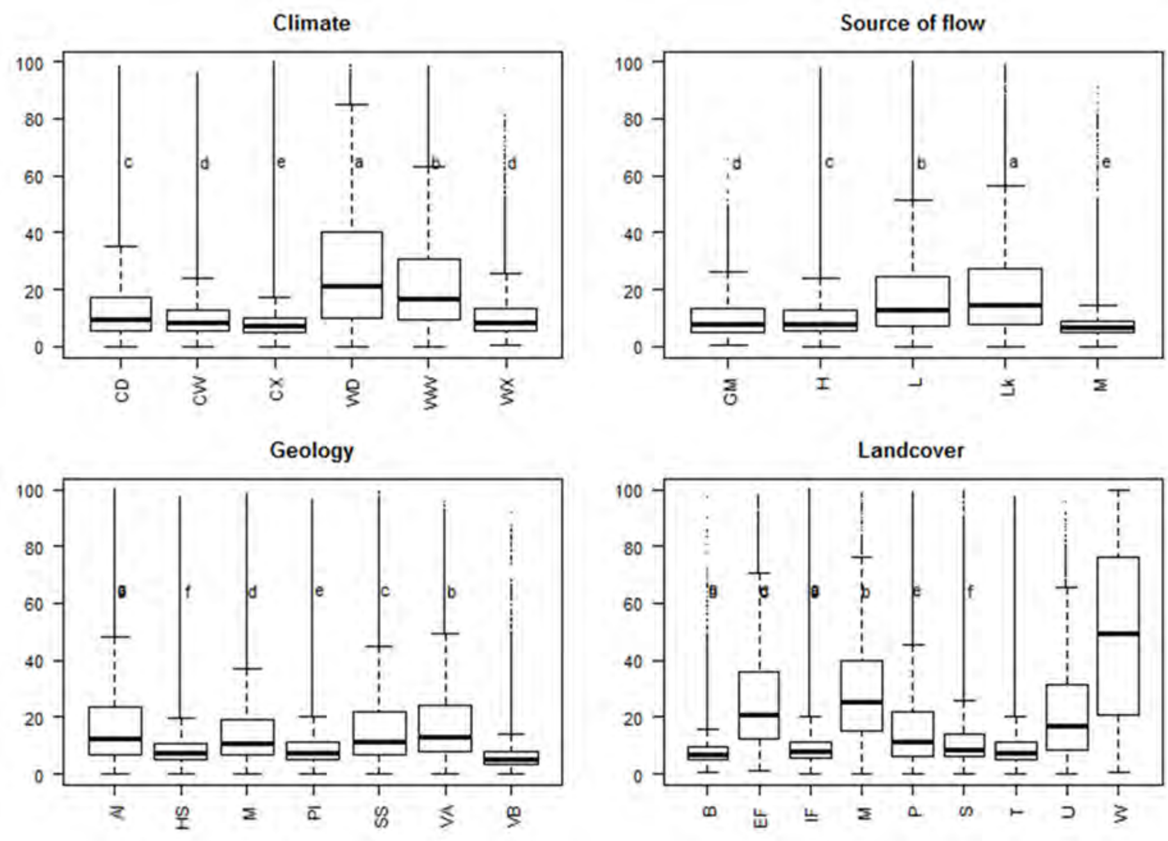


Figure 2-16: Box plots of deposited sediment reference state predicted from reference sites only (BRT REF model) in the *% cover B* dataset grouped by REC components Climate, Source of flow, Geology, and Land cover. Letters indicate a significant difference between groups (Tukey's $p < 0.05$).

2.4.6 Recommended deposited sediment classification based on reference state predictions

The CART procedure distinguished nine groupings in the continuous distribution from 0 to 100% fine sediment cover, driven by elevation, catchment slope, geology, and stream power (Figure 2-14). These groups further combine into three sediment classes that represent low (<30%), medium (30%-60%) and high (>60%) levels of reference state sediment (Figure 2-15). We recommend that these three classes are used for the application of attribute band thresholds: 1) <30% sediment cover ('hard-bottom streams with low-medium sediment levels'), 2) 30-60% sediment cover ('hard-bottom streams with high sediment levels') and 3) >60% sediment cover ('soft-bottom streams'). We further recommend that continuous (segment specific) predictions of natural sediment state from the BRT REF model are used to determine which class any given stream site belongs to.

2.5 Future work

Analysis of natural state variation using another (higher) percentile statistic, as some studies report that ecological effects are less well correlated to medians, than less frequent, events involving higher levels of suspended sediment (i.e., 75th or 80th percentiles). It was noted that the distributions of median values were considerably different when compared to higher percentile values.

If higher environmental quality band thresholds (i.e., A/B and B/C) are required for the management of suspended sediment then a comprehensive classification system will be required. The current GLMM method (McDowell et al. 2013; and Chapter 6) that predicts reference state conditions for different REC classification levels may be sufficient, however it is possible that for some of these classes, the amount of suitable data underpinning the interpolation is inadequate.

Use of the particle size distribution layer in the Fundamental Soil Layer which provides the proportion of catchment soils described as 'clayey'. As opposed to the +1 NTU 'offset' applied to 'warm' REC climate classes, it would be useful to the 'offset' (if required at all) related to a 'landscape setting' that directly influences suspended sediment levels in streams.

A robust comparison of BRT model approaches and the GLM model approach (used in fish threshold analyses – see Chapter 6) to predicting deposited sediment would provide further evidence of appropriate natural sediment levels for streams where there was greatest deviance among models, e.g. warm-wet and warm-dry streams.

We further note that while our identified thresholds were mainly informed by SAM2 (instream) measures, predictive models are mainly informed by SAM1 (bankside) measurements. These two metrics are strongly correlated[10] at values less than 30% cover but diverge at higher values so that a 60% instream cover is equivalent to 43% bankside cover. It may be precautionary to assign classes on underestimates of reference state rather than potential overestimates. We recommend a more robust determination of the offset between % cover bankside and % cover instream to support the most suitable classification of sites for the assignment of attribute classes.

[10] Based on data used in BRT model threshold analysis in section 4.3 N = 89, R^2 = 0.50, p<0.001

3 Development of macroinvertebrate metrics as ecological indicators of deposited sediment effects

3.1 Chapter summary

The chapter presents the multiple research tasks that contributed to development of new macroinvertebrate metrics as ecological indicators of deposited sediment effects. In the Introduction, a short review is provided on sediment-specific macroinvertebrate metrics developed overseas. A field study conducted as part of this project showed that a broad gradient in specific sediment yield translated into a broad gradient of deposited sediment. Increasing sediment measures of deposited sediment (namely, % sediment cover and SIS), in turn, negatively affected macroinvertebrate communities. This overall suggests that macroinvertebrate metrics can be linked to sediment loads and hence may become useful tools for setting limits as required by the NPS-FM.

A systematic literature review was undertaken using a formal causal criteria analysis and the Eco Evidence software to identify what ecological evidence exists to inform sediment-specific metric development and support development of management thresholds. Major findings were 1) that EPT metrics, which are commonly used to indicate overall stream health, may be useful for quantifying deposited sediment effects and 2) that strong negative or positive response of several macroinvertebrate taxa support proof of concept for development of a sediment-specific invertebrate metrics based on the sensitivities/tolerances of the different taxa.

Sediment-specific metrics were developed using a large national dataset (N = 306). The three main steps involved were:

1. identification of indicator taxa for deposited sediment and assignment of tolerance values to them, this was achieved using gradient forest (GF) analysis that builds a random forest model and species turnover function for each taxon,
2. using various ways of combining information on indicator taxa into a metric score resulting in several candidate metrics, and
3. validating the candidate metrics by testing whether the metrics respond to deposited sediment.

The sediment-specific metrics were developed so that they can be calculated for macroinvertebrate samples collected with a kicknet (or Surber) and the taxa identified to MCI-level resolution. The 'Sediment MCI' and 'No. of decreasers' metrics were found to be most strongly related to % sediment cover hence used in subsequent analyses alongside the 'MCI' and '%EPT taxa' for development of deposited sediment management thresholds presented in Chapter 4. The latter two metrics ('MCI' and '%EPT taxa') were also used in the derivation of suspended sediment thresholds for macroinvertebrates (Chapter 5).

3.2 Introduction

Various stressor-specific invertebrate indices have been developed overseas as tools for measuring the instream effects of increased deposited sediment, and for determining ecologically relevant site-specific targets for managing sediment in streams. For example, Zweig and Rabeni (2001) developed the Deposited Sediment Biotic Index (DSBI) using sediment tolerance values for 30 taxa occurring in Missouri, United States. The tolerance values (three categories) were assigned according to the deposited-sediment (< 2 mm diameter) level at which the taxon reached 50% cumulative abundance across the data set using quantitative invertebrate data. The fine sediment tolerance values did not correlate with tolerance values developed for organic enrichment, suggesting that the DSBI has potential to discriminate between multiple stressors (Zweig & Rabeni 2001).

Extence et al. (2013) developed the Proportion of Sediment-sensitive Invertebrates (PSI), which is based on fine sediment sensitivity ratings of British benthic invertebrate species and families. Taxa were assigned to one of four categories based on information available in the literature and an assessment of anatomical, physiological and behavioural traits. Calculation of the PSI score involves fine sediment ratings as well as abundance weighted scores (relative abundance data). The PSI was shown to be more strongly related to % sediment cover than a flow-specific invertebrate index or % EPT abundance, suggesting that the PSI can also potentially discriminate between multiple stressors (Glendell et al*.* 2014).

Also, Bryce et al. (2010) assigned deposited sediment tolerance values to taxa using the weighted-averaging technique based on relative abundances determined from a mountain stream dataset in the western United States. The weighted average, or optimum tolerance value of each taxon, identifies the sediment values (streambed cover of % fines (< 0.06 µm) or % fines and sand (< 2 mm)) at which the highest relative abundances of each taxa occurred. These tolerance values were used to calculate an index of biotic integrity score.

Sediment-specific macroinvertebrate metrics have not yet been developed for New Zealand streams and rivers. The aim of this project (in parallel with the MfE macroinvertebrate project) was to develop new sediment-specific invertebrate metrics. Several research tasks were undertaken to inform metric development including field studies (section 3.3), systematic review of the literature, use of expert opinion to develop tolerance scores (section 3.4) and finally an extensive analysis of a large national dataset, composed of research data, which pairs benthic macroinvertebrate data to measures of stressors including deposited sediment, nutrients and periphyton (section 3.5).

3.3 Deposited sediment field studies

3.3.1 The Waihi Dam study

In early November 2015 the gate of the Waihi Dam (Hawkes Bay) failed, releasing water and sediment into the downstream river network including Waihi Stream, Waiau River and Wairoa River (and into Hawkes Bay). The gate was repaired and closed in March 2016. Monthly monitoring of suspended and deposited sediment along the length of the river network began in February 2016. The dam failure event and subsequent sediment monitoring provided an opportunity to quantify the effect size and recovery time of the benthic community in response to a large sediment addition event. This information was useful for informing the sensitivity of ESVs and macroinvertebrate taxa (i.e., responsiveness to a large sediment event) and possible frequency criteria for attribute development.

During the first sampling rounds after dam failure, in February and March 2016, suspended sediment ESVs and benthic invertebrate metrics significantly differed between the sampling points upstream and downstream of the Waihi Dam confluence (Clapcott 2016). From April to July 2016, on the other hand, there was no discernible difference above and below the dam confluence. High background levels of deposited sediment in the Waiau River along with high within-site spatial variation were suggested as the likely reasons why increased sediment loads did not translate into increases in deposited sediment downstream of the Waihi confluence (Clapcott 2016).

Further analyses of benthic community composition suggested that genus-level identification was suitable to detect sediment effects and further identified several sediment-sensitive taxa (Clapcott 2016). Due to the high background levels of deposited sediment observed and large river size (the Waiau River being on the borderline of suitable for using deposited sediment methods), further sampling effort in the Waiau River was not recommended to aid the development of a sediment attribute. Instead, a national survey study design which would allow testing of the relationships between sediment load or suspended sediment and deposited sediment, and relationships between deposited sediment and benthic invertebrate communities was recommended.

3.3.2 The 2017 field study

Following the recommendations of the Waihi Dam study and the Sediment Attributes Stage 1 report (Hicks et al. 2016), a field study was designed to provide robust data to address the following aims:

1. Link deposited sediment measured using SAM 1 (bankside visual assessment of % sediment cover) and SAM4 (SIS) methods with newly developed sediment-specific invertebrate metrics and for comparison also with a commonly used 'stream health' metric (%EPT taxa).

2. Link deposited sediment ESVs to catchment sediment yields and precedent suspended sediment regime (Appendix C).

3. Test the accuracy of the SAM4 method (Appendix D).

Only the first aim is directly relevant to the development of macroinvertebrate metrics as ecological indicators of deposited sediment effects and reported here. Methods and results that address the other two aims can be found in Appendices.

Methods

A targeted spatial survey of 16 sites was conducted in February 2017 (Table G-1, Appendix G). Sites were chosen to represent a wide gradient of sediment yield and also stream power which has been shown to be a major determinant of deposited sediment in New Zealand streams (Hicks et al. 2016). The wide range represents an expected sediment load range for all New Zealand streams (Hicks et al. 2011). Sites were also chosen where long-term suspended sediment data were available.

At the study sites sites, deposited sediment data was collected using SAM1 and SAM4 methods (Clapcott *et al.* 2011b). All sites were visited after a prolonged period of low flow (> 30 days), except the Motupiko at Christies site which was sampled 11 days after a high flow event (Table G-2, Appendix G).

Macroinvertebrate samples were collected from run habitats using five Surber samplers (mesh size = 500µm, area = 0.1m^2) placed across the wetted stream width. All five samples were pooled and preserved in 70% ethanol alcohol. In the laboratory, macroinvertebrates were identified to the lowest taxonomic level possible using a dissecting microscope following a full count protocol (Protocol P3 from Stark et al. (2001)). We calculated two of the best-performing sediment-specific macroinvertebrate metrics (see section 3.5) and % EPT taxa, a commonly used 'stream health' metric, i.e., not a stressor-specific metric. The relationships of these three macroinvertebrate metrics with deposited sediment ESVs were tested using simple linear regression models. Field study data also contributed to the large research dataset collated as part of the parallel MfE-funded macroinvertebrate project and was used in the development of the sediment-specific invertebrate metrics (see section 3.5).

Results

Eighty seven taxa were recorded across the 16 study sites. Taxa abundance data along with deposited data (SIS and % cover bankside) was added to the dataset compiled for sediment-specific invertebrate metric development (section 3.5). Exploration of the relationships between % EPT taxa, two of the new sediment-specific macroinvertebrate metrics and the broad deposited sediment gradient measured at the 16 study sites confirmed a strong relationship between the 'Number of decreasers' and SIS (Figure 3-1). There was a lot of noise in the relationships especially for %EPT taxa.

Despite the relative small sample size of our field study (N = 16), we were able to show that a broad gradient in SSY translated into a broad gradient of deposited sediment ESVs. Both increasing % sediment cover and SIS, in turn, negatively affected macroinvertebrate metrics '%EPT taxa', 'Sediment MCI' and 'Number of decreasers'. This overall suggests that macroinvertebrate metrics can be linked to sediment loads and hence may become useful tools for setting limits as required by the NPS-FM.

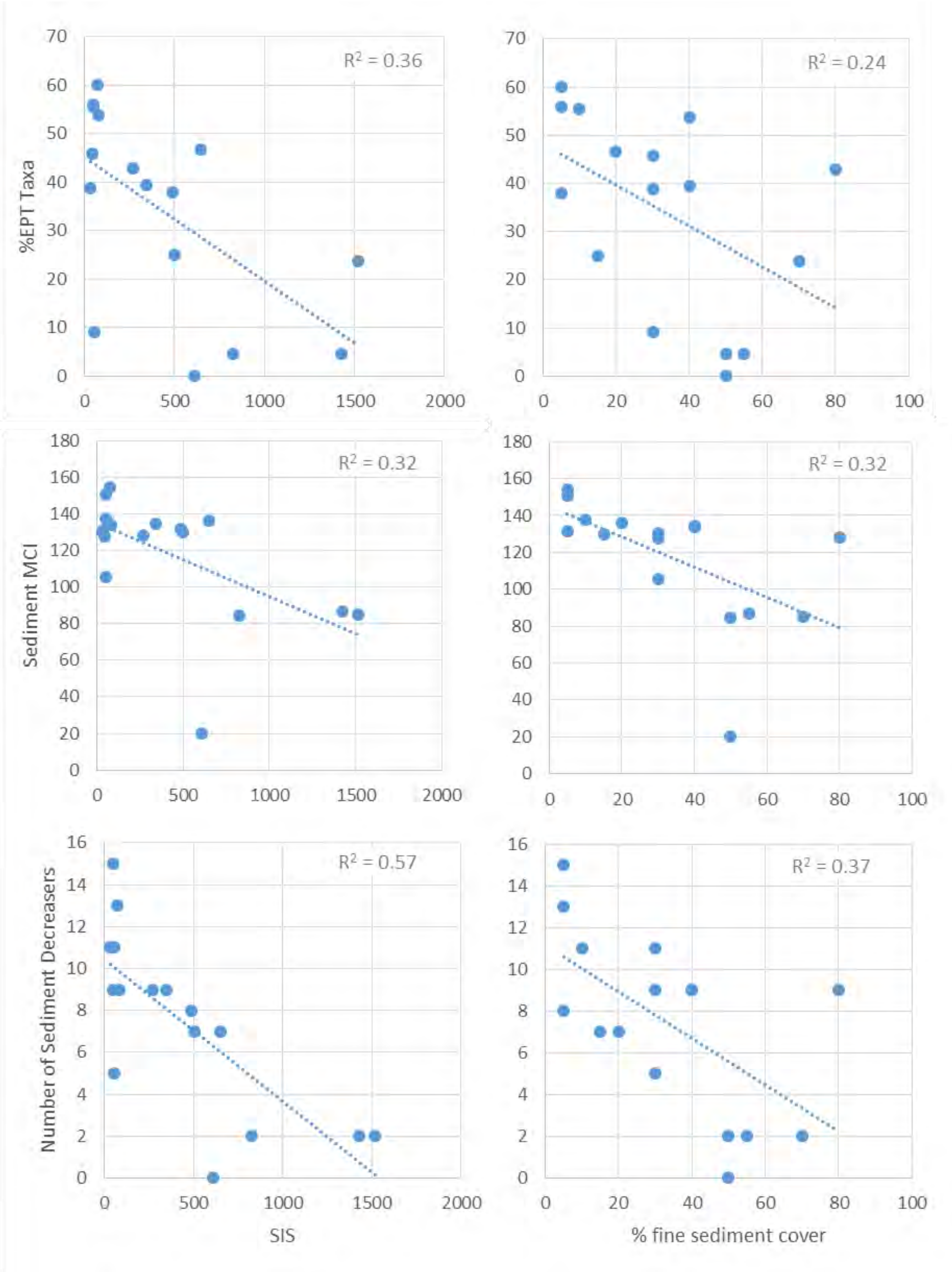


Figure 3-1: Response of % EPT and two of the new sediment-specific macroinvertebrate metrics to deposited sediment measures at 16 field study sites in February 2017.

3.4 Systematic review of the literature

3.4.1 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 1995a; Wood & Armitage 1997c; Clapcott *et al.* 2011b; Collins *et al.* 2011; Kemp *et al.* 2011a; 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.* 2015a). 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 (Appendix I) 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.

3.4.2 Eco Evidence

Eco Evidence is a form of systematic review that is based upon causal criteria analysis (Webb et al. 2015). In contrast to narrative reviews, systematic reviews treat relevant literature as data (Khan et al. 2003) and employ statistical analysis to succinctly analyse and summarise a large body of literature, thereby testing the level of support for hypotheses across numerous studies (Webb et al. 2015). Despite a call for improved defence and transparency of decision making in environmental management (e.g., for the setting of resource limits and freshwater targets/objectives), systematic syntheses such as Eco Evidence have not been conducted much to date (Webb et al. 2013).

Methodological details of the Eco Evidence systematic literature review are provided in Appendix E. Reference for the 65 studies interrogated using the Eco Evidence approach in this project are given in Appendix J. Here, we provide a brief outline of the results and key findings of the Eco Evidence approach.

Overall, 655 cause-effect hypotheses were tested, e.g. *Deleatidium* decrease as % deposited sediment cover increases. Most hypotheses had insufficient evidence (i.e not enough data) to test the cause-effect relationship, with only 111 of the 655 hypotheses returning sufficient support for a conclusion other than insufficient evidence (Table 3-1).

In response to a general increase in deposited fine sediment (all ESVs), 14 cause-effect hypotheses were supported by the analysis including a decrease in 8 taxa, 3 invertebrate 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 3-1).

There was little consistency among responses when comparing relationships with deposited sediment ESVs either assessed at the patch-scale or reach-scale, other than for decreases in EPT richness and abundance (Table 3-1). Only a single causal relationship for suspended sediment was supported by the literature, mainly due to a lack of published evidence; macroinvertebrate abundance decreased with increasing suspended sediment There was no evidence for macroinvertebrate abundance responding to deposited sediment.

In summary, while there were several limitations to the Eco Evidence framework (discussed in detail in Appendix I), we consider this approach to be potentially 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. Nevertheless, the results support the findings of our 2017 field study in that EPT metrics may be useful for investigating the effects of deposited sediment on benthic macroinvertebrates. The results also support proof of concept for development of sediment-specific invertebrate metrics based on the sensitivities/tolerances of the different taxa.

Table 3-1: 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 reported 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

3.5 Sediment-specific metrics

The development of stressor-specific invertebrate metrics, including metrics for deposited sediment and for nutrient enrichment, was explored in the parallel MfE-funded macroinvertebrate project (Contract No. 21630). The first stage of metric development is detailed in Clapcott et al. (2017). Here we provide a brief summary of stressor-specific metric development.

There were three main steps involved:

4. identification of indicator taxa (may be sensitive or tolerant) for deposited sediment or nutrient enrichment or both and assignment of tolerance values to them

5. using various ways of combining information on indicator taxa into a metric score resulting in several candidate metrics, and

6. validating the candidate metrics by testing whether the metrics responds to the stressor for which it was developed.

3.5.1 Identification of sensitive/tolerant taxa

At first, hypothetical tolerance values were assigned by the project team using expert opinion during and following a workshop on taxa sensitivity. However, most taxa were assigned the same or a similar tolerance value for both deposited sediment and nutrient enrichment, suggesting that the metrics calculated from those tolerance values would not be stressor-specific. It was hypothesised that experts may not be able to tease apart the individual effects of these two stressors which often act in concert. Hence, a second, data-based approach to metric development was used.

A large research dataset was compiled from 26 New Zealand studies (both field surveys and experiments), where deposited sediment or nutrient data had been collected alongside benthic macroinvertebrate data. In total, the dataset consisted of 1,850 paired samples (~900 sites or experimental units) out of which 306 samples were selected for which information on both stressors (% sediment cover and chlorophyll a) was available and which were collected from streams. Macroinvertebrate data was expressed as relative abundances and the level of taxonomic resolution was that of New Zealand's Macroinvertebrate Community Index (MCI, Stark 1985). A novel statistical approach to indicator taxa identification and tolerance value assignment, gradient forest (GF) analysis, was applied. This approach uses random forest (RF) models to relate taxon abundance to the focal stressors and to potentially other important environmental variables, taking into account potential interactive effects. These multi-predictor models were expected to be able to tease apart the effects of the individual stressors. Taxa that consistently increased or decreased in response to % sediment cover (referred to as 'increasers' and 'decreasers', respectively) were determined by visual inspection of the RF partial dependence plots. Expert opinion was then used to decide whether response patterns conform to ecological theory or to field observations and only those taxa made it into the final list of indicator taxa for each respective stressor.

We then used gradient forest (GF) output, the species turnover functions, to determine for each indicator taxon a threshold across the stressor gradient at which 25% of the change had occurred. Tolerance values (1-10) were assigned by grouping taxa by their threshold value into 5 bins for both sensitive and tolerant taxa. Tolerance values for *decreasers* were assigned to range from 10 to 6 (i.e., 10, 9, 8, 7 or 6) with values of 10 being assigned to the most sensitive taxa, i.e., those with the lowest thresholds. Tolerance values for *increasers* were assigned to range from 1 to 5 (i.e., 1, 2, 3, 4 or 5) with values of 1 being assigned to the most tolerant taxa. For deposited fine sediment, we identified 25 *decreasers* and 12 *increasers*, i.e., a total of 37 indicator taxa available for metric calculation..

3.5.2 Metric calculation

We constructed the following sediment-specific invertebrate metrics that are based on MCI-level taxonomic information determined from an invertebrate sample. All metrics can be calculated from a semi-quantitative sample collected with a kicknet (or a quantitative Surber sample). Some metrics require taxon counts while others only require presence-absence data.:

- 'No. of decreasers'
- 'proportion (%) of decreasers' ('No. of decreasers' / richness x 100)
- 'No. of increasers'
- 'proportion (%) of increasers' ('No. of increasers' / richness x 100)
- 'sediment MCI' score (average tolerance value / total number of scoring taxa *20)
- 'sediment QMCI' ((average tolerance value * abundance of scoring taxa) / total number of scoring taxa x20).

3.5.3 Metric validation

We explored the response of the sediment-specific invertebrate metrics to the % sediment cover in the research dataset. As expected, sensitive taxon metrics as well as the 'sediment-MCI' and 'sediment-QMCI' responded negatively to increasing *% cover instream*, while tolerant taxon metrics responded positively (Figure 3-2). The 'sediment-MCI' produced the best linear regression model (R^2 = 0.33).

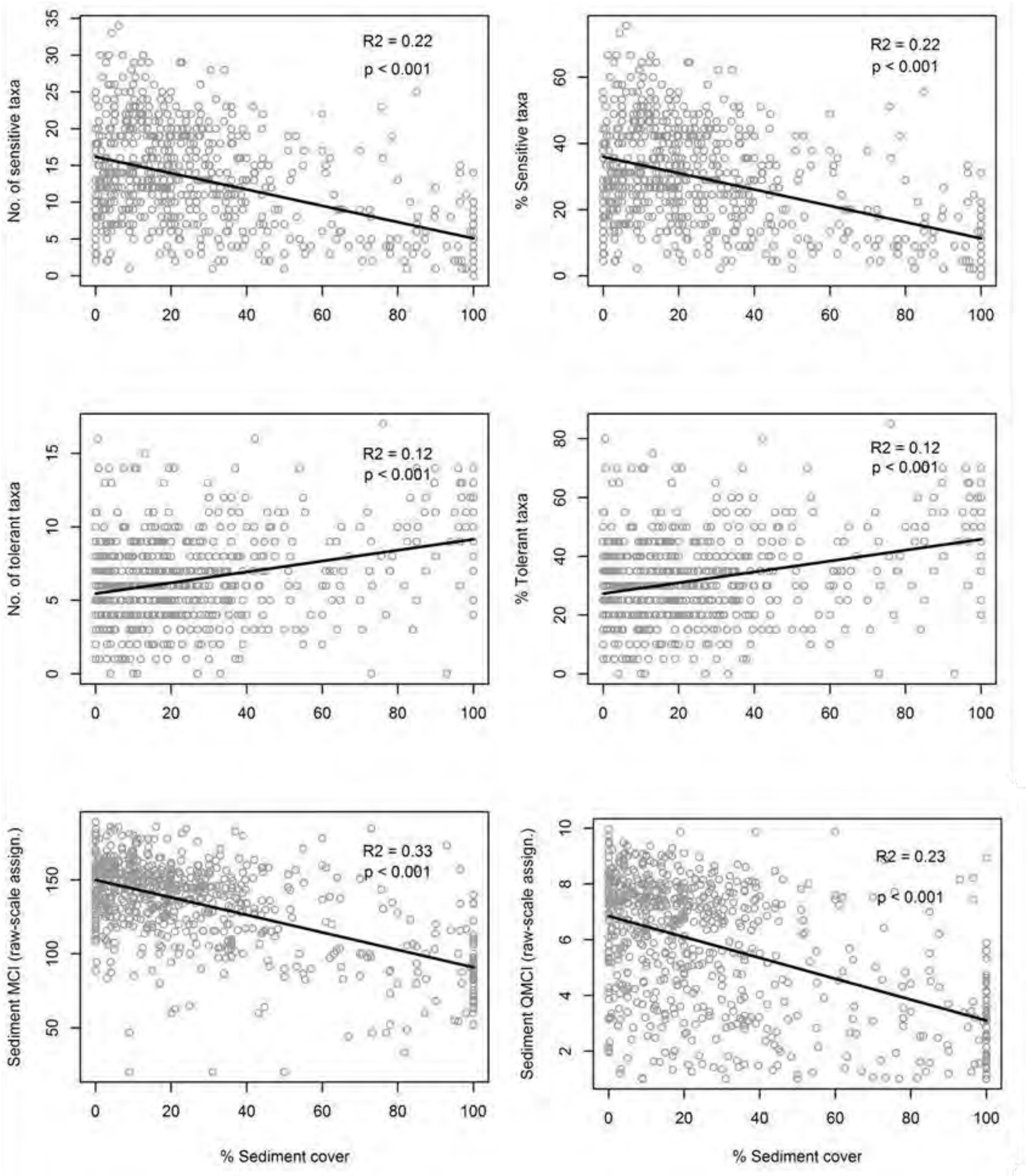


Figure 3-2: Sediment-specific invertebrate metric responses to % sediment cover.

Overall, the results suggest that stressor-specific metrics can be developed, and furthermore may have the potential to be able to discriminate between sediment and nutrient effects. The suggested *sediment MCI* and *no. of decreasers* metrics look particularly useful based on training dataset validation and are further explored in Chapter 4 (and were used in the analysis of the 2017 field study reported in this Chapter also). Further work related to development and validation of these stressor-specific invertebrate metrics had been suggested (outlined in Appendix K) and some of these ideas are currently explored during a follow-up project (due to be finalised in July 2018).

4 Derivation of deposited sediment thresholds based on macroinvertebrate responses

4.1 Chapter summary

This chapter reports on the statistical quantification of the relationship between benthic macroinvertebrate communities and deposited sediment in rivers and streams. This information is used to derive scientifically sound and justifiable thresholds for deposited sediment. The derived thresholds were defined to be consistent with NPS-FM ecosystem health 'bottom line' thresholds which define the transition from C band ('generally represents a minimum safe level before an ecological tipping point' MfE 2014) to an unacceptable D band state.

Threshold derivation was focussed on a level of effects consistent with NOF ecosystem health C/D band for the two hard-bottom stream types, but not for soft-bottom streams, defined in Chapter 2. Briefly summarised, the three sediment reference state classes were:

1. 'hard-bottom streams with low-medium sediment cover' with 0-30% sediment cover
2. 'hard-bottom streams with high sediment cover' with 30-60% sediment cover
3. 'soft-bottom stream*s*' with >60% sediment cover

Three different approaches were selected to model response curves for definition or investigation of sediment thresholds, these were: 1) simple linear quantile regression (QR) (related to the QR approach used for deriving suspended sediment thresholds for macroinvertebrate in Chapter 5); 2) boosted regression tree (BRT) analysis; and 3) gradient forest (GF) analysis. Threshold analyses were conducted on a national dataset containing many sites with information on macroinvertebrates and deposited sediment measures. The BRT and GF approaches, and in particular, the resulting response curves across % sediment cover (instream), were most informative for threshold development.

A distinct threshold in the response of four benthic macroinvertebrate metrics (MCI, 'EPT richness', 'Sediment MCI' and 'No. of decreasers') at 20-30% sediment cover was evident in the BRT output, and the GF analysis suggested that the most prominent change in community composition occurs from 30-55% sediment cover. Consequently, the proposed sediment thresholds (consistent with a NOF C/D band transition) for the two hard-bottom stream reference state classes were:

- 30% sediment cover for the 'hard-bottom streams with low-medium sediment cover' reference state class; and
- 60% sediment cover for the 'hard-bottom streams with high sediment cover' reference state class.

The thresholds contributed to multiple lines of evidence (including fish-based thresholds – Chapter 6, literature effects thresholds, guideline values and expert opinion) for the final proposed C/D band threshold for the deposited sediment attribute (Chapter 7). Further analyses of taxon-level responses are recommended to provide additional scientific support for the A/B and B/C band thresholds. In addition, refinement of the current classification system (Chapter 2), or alternatively, definition of sediment attribute thresholds as a percentage reduction from predicted reference state are recommended.

4.2 Introduction to analytical approaches used to derive thresholds

Broadly speaking, two analytical approaches to deriving sediment thresholds for river management have been commonly suggested.

The first is simply based on sediment data. Within this type of approach, sediment thresholds can, for example, be defined at some upper percentile of the sediment values observed at reference streams. For example, the 75th percentile of the distribution of percent fine deposited sediment values observed at 19 reference sites was used to inform sediment criteria in Clapcott et al. (2011) (Figure 4-1A1).

The second broad approach that has been mainly adopted in this report is related to combined use of sediment and ecological data to derive so-called 'effects-based' sediment thresholds, i.e., thresholds that are defined based on some ecological effect and in this chapter based on the effects on benthic macroinvertebrate communities. Within the effects-based approaches there are also several ways of how to derive sediment thresholds, and commonly stressor-response relationships between macroinvertebrate metrics and a sediment measure built the basis for threshold definition. These relationships, represented by some statistical model, are commonly derived from a spatial data set (many streams sampled across a broad spatial area) using a space-for-time approach to tell how macroinvertebrate communities would change in a stream that is increasingly affected by sediment deposition.

Within the effects-based approaches, there are two fundamentally different approaches. 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. Both, the linear regression and quantile regression models (note, Chapter 5 uses non-linear quantile regression models) assume a linear response shape between predictor and response (Figure 4-1B), although if predictor and/or response were transformed prior to the analysis, the relationship depicted for raw predictor and response values will show curvature.

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. 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 (Wagenhoff et al. 2017, Qian 2014). The assumption of an abrupt response across the full stressor gradient is particularly questionable given that macroinvertebrate metrics were developed to respond in a relative gradual fashion. The use of a flexible modelling approach, boosted regression tree (BRT) analysis, allows modelling of complex response shapes. For example, 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 (Figure 4-1C). This conceptual framework can be useful for definition of management thresholds.

Another important consideration in defining sediment thresholds from a spatial data set is the potential effect of other stressors and other environmental variables. Typically, spatial data sets span gradients of several natural environmental variables (e.g., geology, climate, flow etc) and gradients of multiple human-induced stressors. In particular, if stream sites where selected across a pastoral land-use gradient, one can expect that the data set spans a wide gradient in sediment as well as nutrient conditions. There are different ways of trying to single out the effects of sediment, or in other words to avoid that the relationship is confounded by other stressors or variables that lead to inaccurate estimates of sediment thresholds. One way is to account for the effects of other potentially confounding variables by entering them into a model as predictors. BRT models or random forest models, allow multiple predictors and also automatically take into account the effects of potential interactions. The weaknesses of these models include:

- relate to that there is no commonly accepted consensus around when a model is strong enough to infer an ecological effect;
- that the model can be influenced by unusual data points;
- that there is no measure of uncertainty around the sediment threshold; and
- that response shapes are investigated visually making threshold definition somewhat subjective.

Information on multiple variables measured at a stream site is often not available, restricting the size of suitable data sets for such multi-predictor modelling approaches. One commonly used approach of dealing with potentially confounding variables in simple linear models is to use a quantile version. Instead of estimating the mean (i.e., central tendency) of the macroinvertebrate metric with the values of the sediment measure as it is done in a simple linear regression model, the quantile regression model fits a conditional upper quantile (e.g., 95th) of the macroinvertebrate response. In the presence of multiple stressor gradients within the data set, the quantile model (a related approach was used for threshold analysis of suspended sediment on macroinvertebrates; Chapter **Error! Reference source not found.**) is likely to be more meaningful than a linear regression model as i t reduces the confounding effects of stressors other than sediment

Effects-based approaches are not restricted to using macroinvertebrate metrics, which are calculated from previously assigned indicator taxa, i.e., aggregate previous knowledge about the response of individual taxa to stressors into a single metric. It has been argued and shown evidence for that such aggregate metrics are relatively insensitive to synchronous threshold declines of several taxa (King and Baker 2010). In response, the same authors have developed a statistical approach, called Threshold Indicator Taxa ANalysis (TITAN), that investigates threshold responses to a stressor of individual taxa and aggregates the information to calculate ecological community thresholds (Baker and King 2010). The statistical robustness of TITAN has been questioned by some (Cuffney and Qian 2013, Cuffney et al. 2011) and defended by the authors (King and Baker 2011, Baker and King 2013) but the conceptual framework and TITAN seem to appeal to stream ecologists and have been widely applied by several authors (e.g., Sundermann et al. 2015, Taylor et al. 2014). Another statistical technique to identifying ecological community thresholds by modelling the responses of individual taxa first and secondly combining the information, called gradient forest, has been developed by Ellis et al. (2012). Gradient forest analysis builds random forest models for each of the occurring taxa in the data set and then aggregates the response patterns to investigate where there are sediment thresholds at which macroinvertebrate communities change more dramatically. Random forest models are, like BRT models, ensemble tree-based regression models and able to take multiple predictors, automatically account for interactive effects, and to describe complex response shapes.

For these reasons, gradient forest analysis appears to be more appealing than TITAN and has recently been applied to a New Zealand regional data set (Wagenhoff et al. 2017).

Finally, the conceptual framework of using responses of individual taxa and then calculating sediment thresholds to protect macroinvertebrate communities, as TITAN and gradient forest suggest, has also been adopted by Cormier et al. (2008) but the statistical approach again differs. Cormier et al. (2008) used 90th-quantile regressions for each of 21 macroinvertebrate taxa to estimate sediment thresholds at which a 20% reduction in taxon abundance from reference conditions occurs. The resulting taxon-specific sediment thresholds were then ordered and a curve fitted to produce a species sensitivity distribution (SSD) plot from which the sediment threshold was determined (at about 8% deposited fine sediment) at which no more than 5% of the species were reduced. Such an approach allows some flexibility around the accepted reduction in abundance of individual taxa (here 20%) to increasing sediment as well as the desired level of species protection (here 95%). This approach was used in the derivation of suspended sediment thresholds for macroinvertebrates in Chapter 5 – it was also investigated as a potential method for sediment thresholds, based on fish presence absence data, but it was found to unsuitable (Chapter 6).

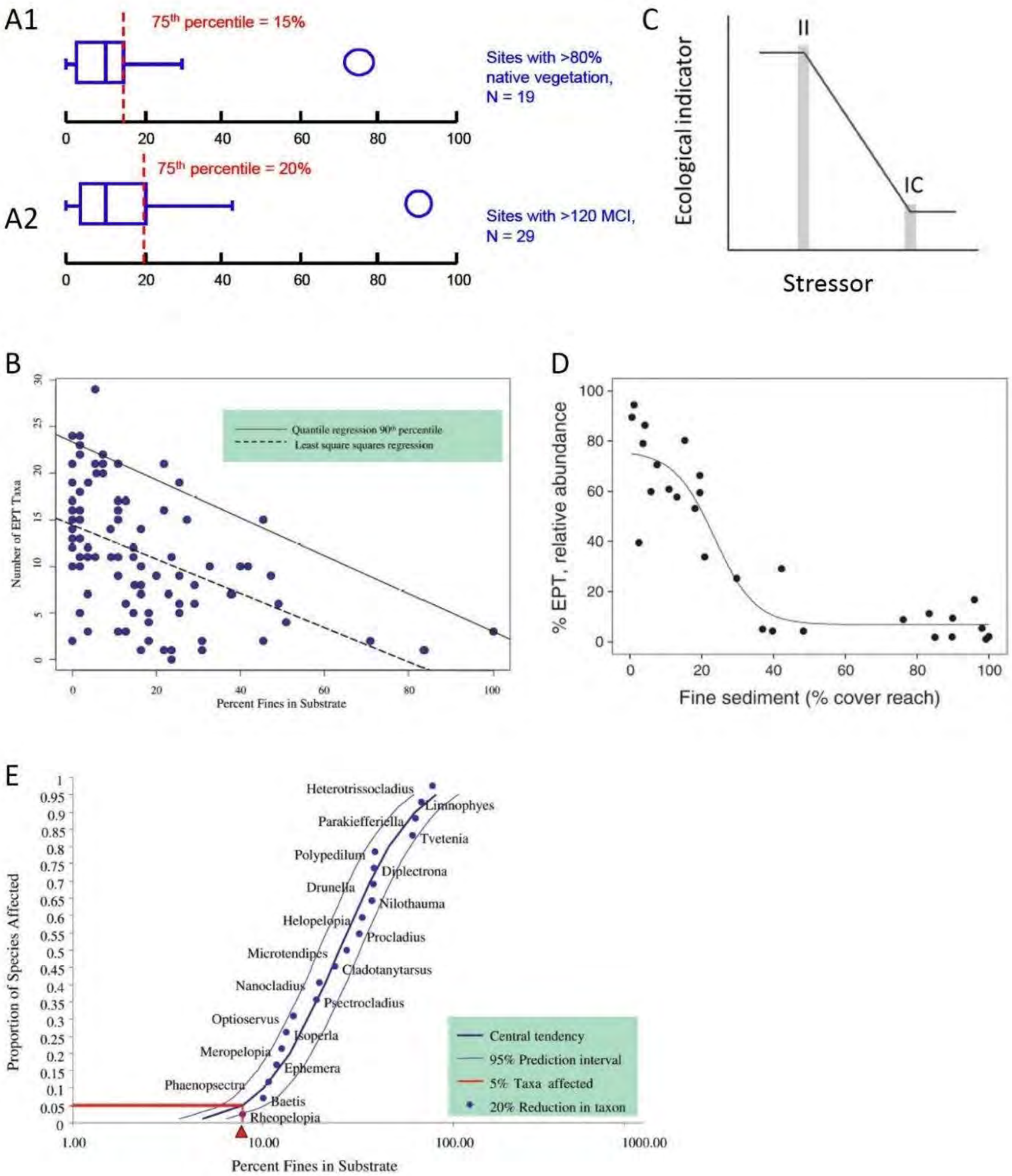


Figure 4-1: Examples of statistical approaches to identifying deposited fine sediment thresholds for resource management. A. The 75th percentile of the distribution of percent deposited fine sediment values observed at reference sites was used to inform sediment criteria in Clapcott et al. (2011b); reference 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 in Cormier et al. (2008).C. Conceptual diagram illustrating impact initiation (II) and impact cessation (IC) thresholds presented in Wagenhoff et al. (2017b). D. Example of a four-parameter sigmoidal regression model to quantify the relationship between %EPT and percent deposited fine sediment cover presented in 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 in Cormier et al. (2008).

4.3 Analysis of macroinvertebrate responses to deposited fine sediment to determine relevant threshold values

The response of benthic macroinvertebrate communities to measures of deposited fine sediment was investigated using a selection of the above approaches (see section 4.2) to inform suitable management thresholds for inclusion in a deposited fine sediment attribute in the National Objectives Framework (NOF). Analyses were performed on a national dataset linking macroinvertebrate data with deposited sediment and other stressor data (details provided in 4.3.1). For approaches 1 and 2, we used a range of macroinvertebrate metrics commonly used to indicate stream health and also recently developed sediment-specific macroinvertebrate metrics (see section 3.5). With all three statistical approaches, models were built separately for three different measures of deposited sediment, namely:

1. the percentage of sediment cover on the streambed assessed by standing in the stream (% sediment cover instream);
2. the percentage of sediment cover on the streambed assessed from standing on the stream bank (% sediment cover bankside); and,
3. suspendable inorganic sediment (SIS) assessed using the Quorer method.

The technical details of the three statistical approaches are provided in section 4.3.2. All three approaches come with their different strengths and weaknesses and were used to provide weight-of-evidence of the ecological response of macroinvertebrates to deposited sediment from which to derive management thresholds for a deposited fine sediment attribute presented in section 4.5.

4.3.1 Dataset compilation

The macroinvertebrate-stressor dataset was specifically created from a large national macroinvertebrate dataset containing SoE data provided by regional and unitary councils as well data collected by NIWA from National River Water Quality Network (NRWQN) sites[11], typically collected on an annual basis (see details in Appendix L). To boost sample size and the spread of sites around New Zealand, we added research data that had been compiled for a companion MfE-funded macroinvertebrate project (Clapcott et al. 2017).

Stressor datasets

Deposited sediment and other stressor data were retrieved from three separate datasets.

- The Sediment Stage 2 dataset consisted of deposited fine sediment data that had been compiled during the Sediment Stage 1 project and updated as part of this project as described in section 2.4.2.
- Water quality data collected at SoE monitoring sites (typically monthly), was retrieved from the LAWA (Land, Air and Water Aotearoa) website https://www.lawa.org.nz/explore-data/river-quality/ (downloaded 5 May 2017).

[11] Compiled by Martin Unwin, NIWA, Christchurch.

- Periphyton data, compiled as part of this project as described in Appendix L; mostly assessed at SoE or NRWQN sites monthly or annually.

Full details of dataset compilation, including matching with macroinvertebrate data are provided in Appendix M. Briefly, macroinvertebrate data were matched to sediment data and then water quality data from LAWA by using a fuzzy matching algorithm prioritised to match site name and date. Periphyton data was matched with macroinvertebrate data using site name and date, RCSID and NZReach ID. Of the 15,508 macroinvertebrate samples contained within the macroinvertebrate dataset, we matched 4,717 samples with at least one measure of deposited fine sediment (Table 4-1), consisting of '% sediment cover instream', '% sediment cover bankside' and/or SIS.

Table 4-1: Sample size of each of seven deposited fine sediment measures within the national SoE macroinvertebrate-stressor dataset. Includes total sample size and number of sites in the final combined SoE and research dataset for three selected deposited fine sediment measures (grey shading).

Deposited sediment measure	Sample size (SoE data only)	No. of sites (SoE data only)	Total sample size (SoE and research data)	Total no. of sites (SoE and research data)
% sediment cover instream	571	188	1,039	593
% sediment cover bankside	2,620	467	2,708	555
SIS	83	47	449	302
Wolman pebble count	1,403			
SBSV	33			
Shuffle test score	74			
Bankside visual assess. (RHA protocol)	591			

In additional to SoE monitoring data, we added a subset of samples from a recently compiled research dataset for the MfE-funded macroinvertebrate project used to develop sediment-specific macroinvertebrate metrics. This research dataset predominately contains stressor data from a single observation taken on or close to the day of sampling macroinvertebrates. The final total sample size and number of sites for each the '% sediment cover instream', '% sediment cover bankside' and SIS can be found in Table 4-1. Figure 4-2 shows the spread of sites across the country for each of these three focal deposited sediment measures.

Out of the total number of 571 macroinvertebrate samples with matching '%sediment cover instream' data, about half of the samples were matched with a single sediment observation, and for about a third, two observations were available (Table 4-1). Out of the total number of 2,620 macroinvertebrate samples with matching '% sediment cover bankside' data, about half of the samples were matched with a single sediment observation and a third with two observations. All 83 macroinvertebrate samples were matched with a single observation of SIS (Table 4-1).

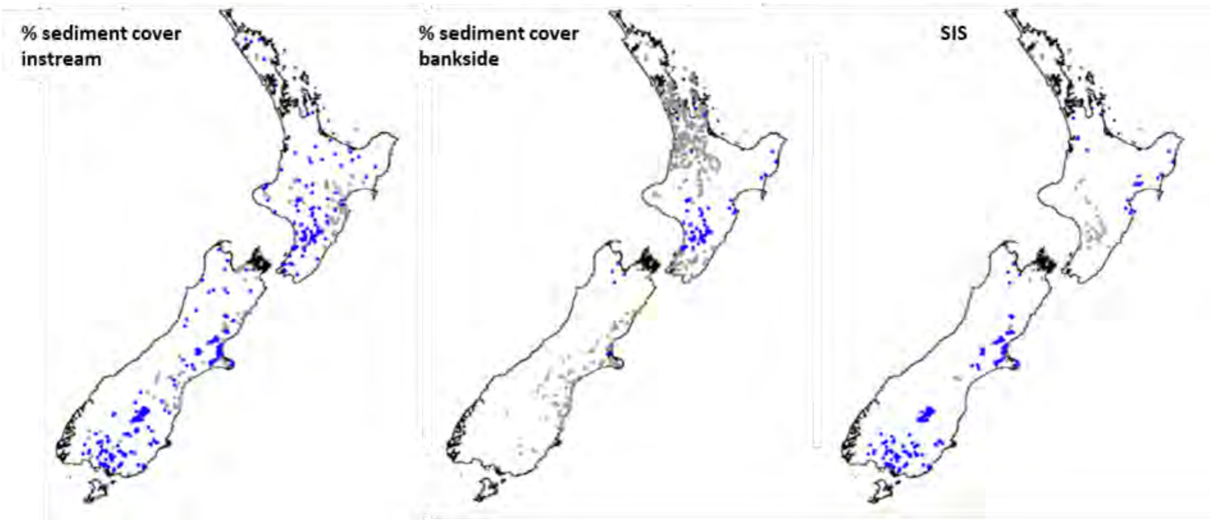


Figure 4-2: Spread of sample sites across New Zealand. Sites colour-coded as to whether data were retrieved from a national SoE dataset (grey), or from a research dataset (blue) for each '% sediment cover instream', '% sediment cover bankside', and SIS; see Table 4-1 for sample size and number of sites.

4.3.2 Methods

To provide multiple lines of evidence, three analytical approaches were adopted for threshold identification using the specifically compiled national macroinvertebrate-stressor dataset (refer to Appendix M). For each of these approaches, separate analyses were performed for each of the three deposited sediment measures ('%sediment cover instream' '% sediment cover bankside' and SIS).

The first approach is a single-stressor analysis while the other two approaches incorporate multiple predictors. The first two approaches use aggregate macroinvertebrate metrics (e.g., MCI, QMCI and EPT) as response variables whereas the third approach produces models for individual taxa and combines the information to determine macroinvertebrate assemblage thresholds. The three approaches employed were:

1. Simple linear quantile regression (referred to as QR method) to relate each macroinvertebrate metric to a deposited sediment measure as a single predictor to calculate a sediment threshold at a pre-defined reduction (i.e., effect level) in the macroinvertebrate metric
2. Boosted regression tree (BRT) analysis (referred to as BRT method) to relate each macroinvertebrate metric to a deposited sediment measure taking into account the effects of other stressors and environmental variables which enter the model as multiple predictors; the stressor-response shapes are visually investigated for impact initiation and cessation thresholds.
3. Gradient forest analysis (referred to as GF method) that uses random forest models to relate the abundance of individual macroinvertebrate taxa to a deposited sediment measure taking into account the effects of other variables, the sediment threshold is determined from plots that visualise where along the sediment gradient that the macroinvertebrate communities change most dramatically.

We considered a set of 16 macroinvertebrate metrics (Table N-1, Appendix N) that were expected to respond to deposited sediment based on previous research, including commonly-used metrics by regional councils and recently developed sediment-specific macroinvertebrate metrics (refer to section 3.5). All metrics were calculated from MCI-level taxonomic resolution.

Based on scatterplots of these metrics across the deposited sediment gradients, we selected the following four macroinvertebrate responses for threshold analyses (using method 1 and method 2):

- MCI
- ‘EPT taxon richness’
- ‘sediment MCI’
- ‘No. of decreasers’ (i.e., taxa that decline with increasing deposited sediment).

Summary statistics for the four-selected metrics are given in Table 4-2. All analyses were performed in statistical programme R (R Core Team 2016), with specialised functions from a range of R packages.

Table 4-2: Summary statistics of the 4 selected macroinvertebrate metrics used for the quantile regression (QR) and boosted regression tree (BRT) methods. A short description of the new sediment-specific macroinvertebrate metrics, indicated with *, can be found in section 3.5.

Macroinvertebrate metric	Min	Max	Mean	Median
MCI (hard-bottom)	27	173	101	102
‘EPT taxon richness’	0	26	7	7
‘sediment MCI’ (raw scale assignment of bins)*	20	200	126	133
‘No. of decreasers’*	0	22	6	6

Method 1: Quantile regression analysis (QR method)

We adopted a single-stressor analytical threshold approach were the sediment threshold is identified at a predetermined biological benchmark effect using a simple linear regression model (Cormier et al. 2008). The benchmark effect here is a reduction in a macroinvertebrate metric from reference condition to benchmarks of 5, 10, 15 and 20%, which are consistent with suggestions in the threshold literature (Cormier et al. 2008). Cormier et al. (2008) suggested that for criteria development in the United States, a 5% and 20% change from reference condition could be used to calculate candidate criteria for aquatic life use and marginal aquatic life use, respectively, and defined the biological reference condition at the y-intercept. We defined the biological reference condition at the macroinvertebrate metric value derived from the regression model at the minimum sediment value observed instead.

Regression analysis requires the response variables and residuals to approximate a normal data distribution. Scatterplots of macroinvertebrate metrics across the three measures of deposited sediment were mainly wedge-shaped. Quantile regressions were performed with R package *quantreg*. To compare model fit among the various quantile models, we calculated R^2 following the suggestions by Koenker and Machado (1999). Further methodological details are provided in Appendix N.

Method 2: Boosted regression tree analysis (BRT method)

Boosted regression tree (BRT) analysis is a flexible modelling approach that allows incorporation of multiple predictors. BRT analysis is well described in the statistical literature for ecology (De'ath 2007; Elith et al. 2008). BRT output provides a 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 can be used for visual threshold definition (Wagenhoff et al. 2017b). Inclusion of stressors in the model other than deposited sediment as well as of environmental predictors improves confidence in the fitted function depicting the response shape to sediment rather than the response to another predictor that is correlated with increasing sediment.

The BRT models contained 17 predictors including a single deposited sediment measure, chlorophyll *a* as a means of accounting for the effect of nutrients via periphyton biomass, turbidity as a measure of suspended fine sediment, and a range of predicted variables retrieved from large databases that hold information for each NZReach (Table N-2, Appendix N). Three separate BRT models were built for each macroinvertebrate metric, one for each focus deposited sediment measure.

Sample size and number of sites for the three datasets used for BRT analysis are provided in Table and maps of the distribution across New Zealand are given Figure 4-3. Model parameterisation was done following the suggestions by Elith et al. (2008) and the *gbm* R package and modified functions based on procedures published by Elith and Leathwick (2014).

Table 4-3: Sample size and number of sites of the subsets used for BRT analysis for each deposited sediment measure.

Deposited sediment measure	Sample size	No. of sites
% sediment cover instream	602	354
% sediment cover bankside	199	100
SIS	243	156

Figure 4-3: Spread of sample sites across New Zealand for the three subsets ('% sediment cover instream', '% sediment cover bankside' and SIS) used for BRT analysis. Refer to Table 4-3 for sample size and number of sites.

Method 3: Gradient forest analysis (GF method)

Gradient forest (GF) analysis was developed to identify community thresholds defined as a point(s) at which a small increase in a stressor will result in a disproportionally large change in community structure relative to other points across the stressor gradient (Ellis et al. 2012). This approach has recently been used to identify community thresholds of three different stream assemblages including macroinvertebrates (Wagenhoff et al. 2017a).

Detailed description of the GF approach can be found in Ellis et al. (2012). A detailed description of the GF method is provided in Appendix N.

Community thresholds were derived from aggregation of the information from RF models of the taxa into cumulative splits importance curves of the overall macroinvertebrate community. Thresholds can be visually identified from split density plots that take into account data density across the stressor gradient. The GF approach also produces cumulative splits importance curves of all taxa to investigate which taxa were contributing most to community thresholds.

The predictor variables used in the RF models were the same we used for BRT analysis (Table N-2, Appendix N), and the GF approach was implemented for the same three deposited sediment measures. Sample size and number of sites for the three datasets used for GF analysis can be found in Table 4-4 and maps of the spread across New Zealand in Figure 4-4.

Table 4-4: Sample size and number of sites of the subsets used for GF analysis for each deposited sediment measure.

Deposited sediment measure	Sample size	No. of sites
% sediment cover instream	380	211
% sediment cover bankside	184	97
SIS	66	31

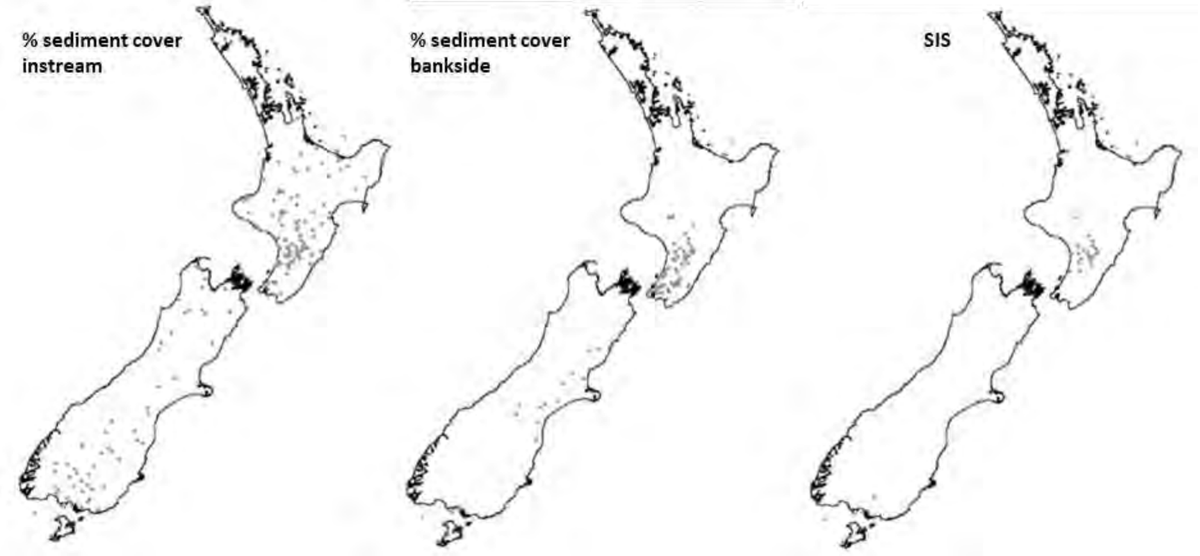


Figure 4-4: Spread of sample sites across New Zealand for the three subsets ('% sediment cover instream', '% sediment cover bankside' and SIS) used for GF analysis. Refer to Table for sample size and number of sites.

4.3.3 Results

Method 1: Quantile regression analysis (QR method)

Generally, R^2 values of the 85th quantile regression models were low, ranging from 0.001 to 0.075 (Table 4-6). For % cover instream (SAM2), the MCI model had the highest R^2 among the models for the other metrics. The same was true for SIS although the model for *number of decreasers* was very similar in model fit. The best model for *% cover bankside* was the *number of decreasers* (Table 4-5).

Table 4-5: Goodness-of-fit measure R^2 for the 85th quantile regression (QR) models.

Response variable	Predictor	R^2
MCI	('% sediment cover instream')$^{1/2}$	0.044
(EPT taxon richness)$^{1/2}$		0.001
'Sediment MCI'		0.039
('No. of decreasers')$^{1/2}$		0.024
MCI	('% sediment cover bankside')$^{1/2}$	0.013
(EPT taxon richness)$^{1/2}$		0.026
Sediment MCI		0.020
('No. of decreasers')$^{1/2}$		0.057
MCI	log(SIS)	0.075
(EPT taxon richness)$^{1/2}$		0.048
'Sediment MCI'		0.070
('No. of decreasers')$^{1/2}$		0.074

The sediment thresholds calculated from the 85th quantile regression lines based on a 5, 10, 15 and 20% decrease benchmark effects (relative to the maximum value) ranged widely across the focal macroinvertebrate metrics (Table 4-6). For example, a 5% benchmark effect on the MCI was predicted to be at 8% sediment cover assessed instream but at 73% sediment cover for EPT richness (Table 4-6, Figure 4-5). A 20% and 15% benchmark effect was often predicted to be beyond the observed sediment gradient (i.e., greater than 100% fine sediment cover). The inability of the QR method to define effect thresholds consistent with an ecosystem health bottom-line (i.e., 20% or more), indicates it is an unsuitable method (as applied to these data). The data in Figure 4-5 show the 'output' of QR method plots for '% sediment cover instream' – similar plots for '% sediment cover (bankside' and SIS are provided in Figure O-1 (Appendix O).

Table 4-6: Single-stressor sediment thresholds calculated from the 85th quantile regression model. Calculated for a benchmark effect of 5, 10, 15 and 20% relative to the *maximum metric value*. '-' indicates that the threshold was beyond the observed stressor gradient.

Metric	Sediment measure	Maximum metric value	5% Sed. Threshold[1]	10% Sed. Threshold[1]	15% Sed. Threshold[1]	20% Sed. Threshold[1]
MCI	'% sediment cover instream'	133	**8**	**32**	**72**	**-**
'EPT taxon richness'		12	**73**	**-**	**-**	**-**
'Sediment MCI'		151	**12**	**48**	**-**	**-**
'No. of decreasers'		13	**6**	**24**	**53**	**94**
MCI	'% sediment cover bankside'	132	**19**	**75**	**-**	**-**
'EPT taxon richness'		14	**9**	**33**	**75**	**-**
'Sediment MCI'		151	**16**	**65**	**-**	**-**
'No. of decreasers'		13	**4**	**15**	**33**	**57**
MCI	SIS (g/m^2)	140	**41**	**261**	**1,629**	**10,040**
'EPT taxon richness'		17	**26**	**117**	**501**	**2,084**
'Sediment MCI'		153	**67**	**720**	**7,449**	**-**
'No. of decreasers'		16	**19**	**69**	**230**	**753**

[1] indicative threshold based on the sediment metric value corresponding to a 5, 10, 15 or 20% decrease in the maximum value of the macroinvertebrate metric. A decrease of 20% is considered the minimum requirement to be consistent with a NOF bottom-line value. Macroinvertebrate response values corresponding to the 5, 10 and 20% reductions are provided in Table O-1 (Appendix O).

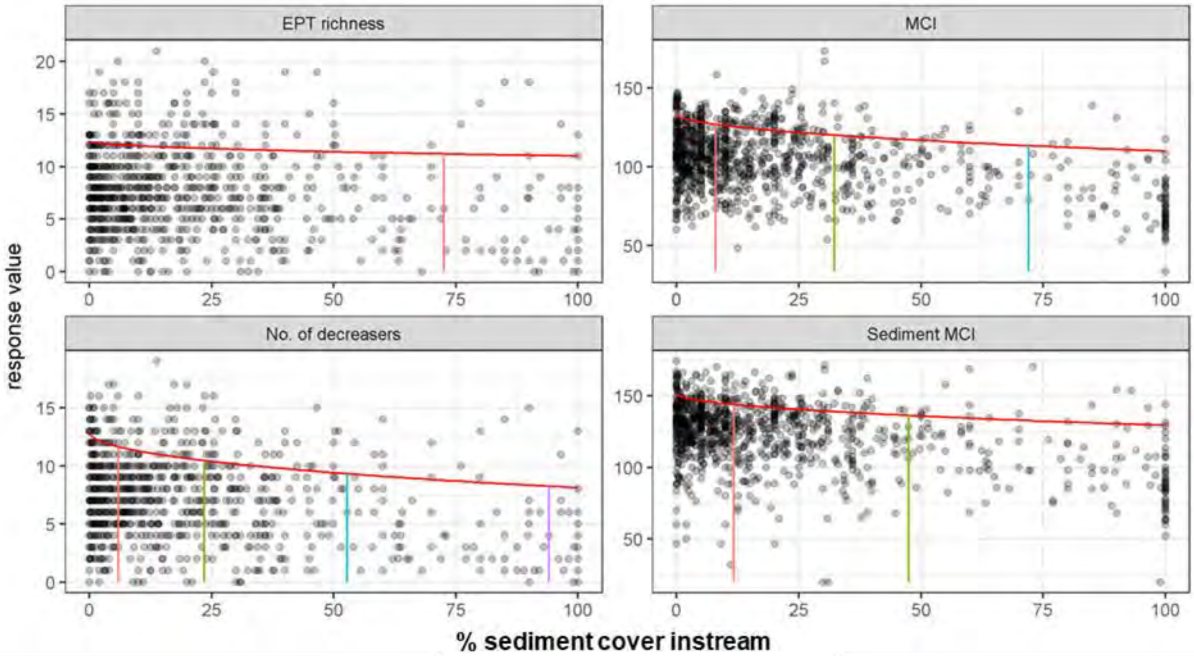


Figure 4-5: The 85th quantile regression models plotted for raw '% sediment cover instream' and metric values along with the sediment thresholds (vertical lines). Calculated at 5 (pink), 10 (green), 15 (blue) and 20% (purple) benchmark effect (i.e., reduction in maximum value of metric/response). Equivalent regression plots for all three deposited sediment measures (i.e., '% sediment cover instream', '% sediment cover bankside', and SIS, are provided in Figure O-1 (Appendix O).

Method 2: Boosted regression tree analysis (BRT method)

Overall the BRT model fit ranged from 43% to 69% TDE (total deviance explained), and the CV correlation coefficient ranged from 0.68 to 0.82, indicating good predictive performance (Table 4-7). The relative importance of '% sediment cover instream' as a predictor for macroinvertebrate response was indicated from the generally high ranking for all macroinvertebrate metrics (i.e., ranked 1st to 3rd of the 17 predictor variables). SIS also ranked highly, for the 'sediment MCI' and 'No. of decreasers' it was the 2nd most important predictor, while for EPT richness and MCI it was ranked 3rd and 4th, respectively. However, '% sediment cover bankside' generally ranked lowly, ranging from 11th to 16th out of a total of 17 predictors.

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

Macroinvertebrate metric	Deposited sediment predictor[1]	TDE (%)	CV correlation coefficient
MCI	'% sediment cover instream'	63	0.80
'EPT taxon richness'		52	0.73
'Sediment MCI'		60	0.77
'No. of decreasers'		61	0.78
MCI	'% sediment cover bankside'	58	0.78
'EPT taxon richness'		55	0.74
'Sediment MCI'		43	0.68
'No. of decreasers'		59	0.77
MCI	SIS	69	0.81
'EPT taxon richness'		61	0.76
'Sediment MCI'		59	0.76
'No. of decreasers'		66	0.82

The fitted functions for deposited sediment measures (i.e., '% sediment cover instream', '% sediment cover bankside' and SIS) presented in partial dependence plots were considered for threshold definition. Overall, the four metrics (i.e., MCI, 'EPT taxon richness', 'sediment MCI' and 'number of decreasers') showed similar response shapes for the 3 sediment measures (Figure 4-6). For '% sediment cover instream', no marked change in macroinvertebrate metrics could be observed until about 30% sediment cover after which metrics continued to decline up to 100% sediment cover (Figure 4-6A). The overall effect of '% sediment cover bankside' was small for EPT richness and 'No. of decreasers' and ceased around 20% sediment cover. In contrast, for MCI and 'sediment MCI' metrics there was not marked effect/response to increasing '% sediment cover bankside' (Figure 4-6B). Across the SIS gradient, little change in macroinvertebrate metrics could be observed until about 500 g/m^2, and metrics ceased to respond at about 2,000 g/m^2 of SIS (Figure 4-6C).

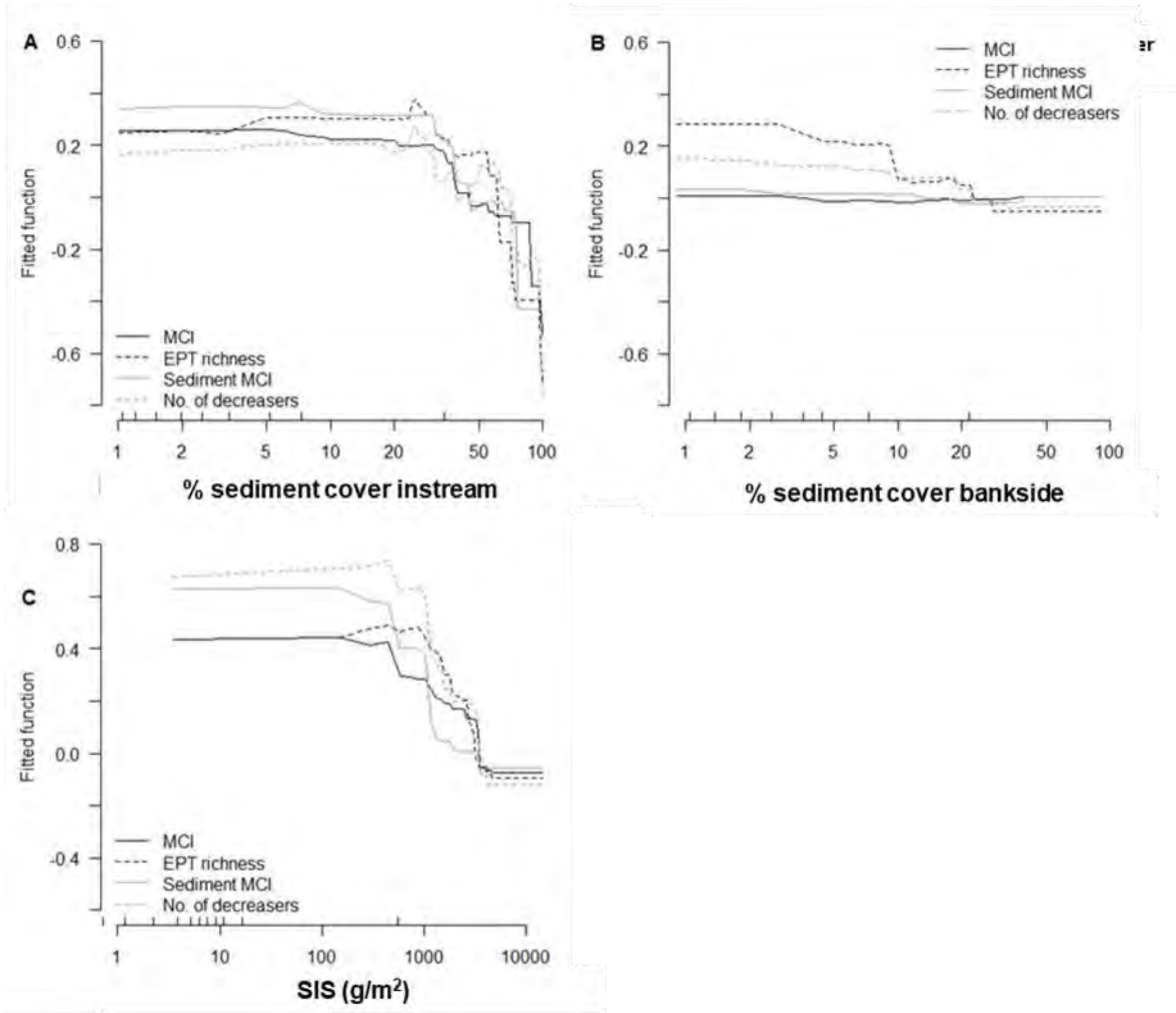


Figure 4-6: Partial dependence plots of the BRT fitted functions of four focal macroinvertebrate metrics applied across the gradient of '% sediment cover instream' (A), '% sediment cover bankside' (B), and SIS (C). Note that the y-axes in all three panels show change from mean response values in units of standard deviation and show the same range of values to make comparable the effect sizes among response variables and among panels. Also note all x-axes have been log-scaled to help with visual identification of thresholds; the x-axis for SIS has also been substantially shortened as there was no change in the fitted values beyond 10,000 g SIS/m^2.

Method 3: Gradient forest analysis (GF method)

In the three deposited datasets ('% sediment cover instream', '% sediment cover bankside' and SIS), there were over 100 taxa present. However only between about 40% and 60% of those exceeded 10 occurrences, which was the criterion for inclusion in the GF analysis. The final number of taxa that contributed to the GF threshold approach was relatively small, with 13, 37 and 17 taxa included for the analysis with '% sediment cover instream, '% sediment cover bankside' and SIS, respectively (refer to Table P-1, Appendix P).

The relative importance of the deposited sediment measures varied for the three GF analyses. Among all 17 predictors, ' % sediment cover instream', '% sediment cover bankside' and SIS, were ranked 10, 5 and 4, respectively (refer to Figure P-1, Appendix P).

Community thresholds were visually delineated from splits in density plots (top panels in Figure 4-7) at the zone(s) where the ratio of the split densities (blue line) is above the ratio of 1 (horizontal blue dashed line). For '% sediment cover instream', the most prominent community threshold was from about 30 to 55% fine sediment cover with a less important community threshold occurring between 0 and around 5% sediment cover (Figure 4-7a). The taxa most responsible for the large change in community structure were Orthocladiinae, *Potamopyrgus*, Chironomidae, *Aoteapsyche*, Tanytarsini and Hydraenidae (Figure 4-7d). The relative importance of the taxa reflected by the steeper response along the deposited sediment gradient.

For '% sediment cover bankside', the most important community threshold was from about 60 to 90% cover with two less important community thresholds from about 5 to 15% cover (Figure 4-7b). The taxa most responsible for the largest change in community structure were *Zephlebia* and Elmidae with much smaller contributions from other taxa (Figure 4-7e).

For the sediment measure SIS, the most important community threshold was from 0 to about 500 g/m^2 of SIS (Figure 4-7c). The taxa most responsible for the largest change in community structure were *Aoteapsyche*, *Paranephrops*, *Latia*, *Zephlebia* and *Orthopsyche* (Figure 4-7f).

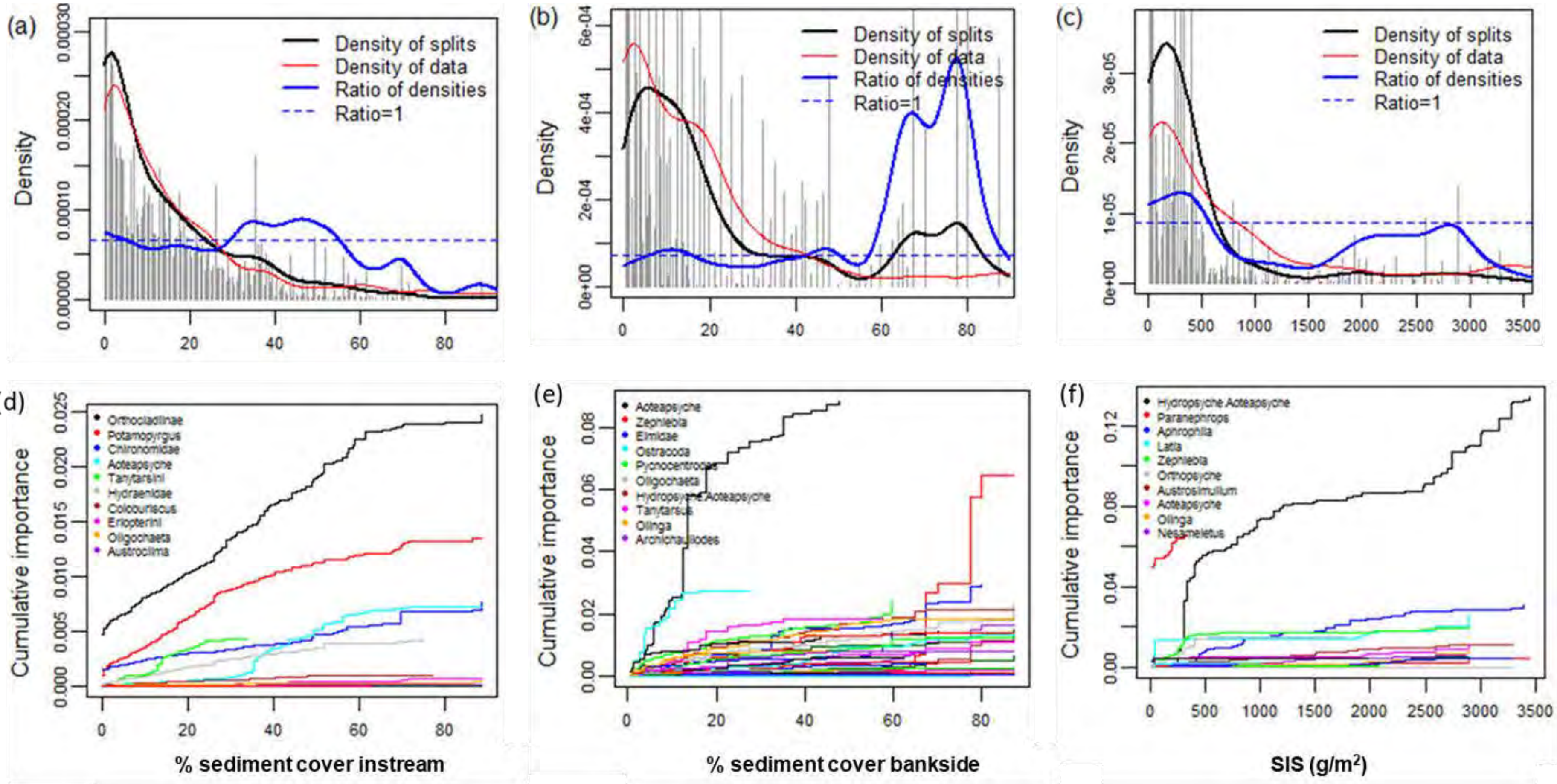


Figure 4-7: Graphical output of the GF analysis for '% sediment cover instream' (a,d), '% sediment cover bankside' (b,e) and SIS (c,f). The top panels are the split density plots showing the raw split importance density function computed from split point density weighted by importance (black line), the binned raw split importance density (grey bars), the density function of the observed predictor values (red line), and the estimated *importance function* computed as the ratio of split and data density (blue line). Standardization of the split density by data density accounts for bias toward values where sampling was more intense (Ellis et al. 2012). Ratios >1 of the *importance function* indicate that compositional rate of change is relatively larger than elsewhere across the stressor gradient. Thus, thresholds were visually identified at locations of the peaks of the blue line provided they exceeded the ratio of 1, and the zones around those thresholds where ratios still exceeded 1 were also delineated. The bottom panels are the cumulative importance curves of all taxa, scaled by their respective R^2, of which only those of the ten most important taxa are labelled. Note that directionality of the taxa responses is not shown by these plots.

4.3.4 Discussion

Method 1: Quantile regression analysis (QR method)

Quantile regression (QR) analyses were limited and appeared to be influenced by unusual data points. For example, the slope of the model for EPT richness and sediment cover assessed instream was very low, likely due to a few unusually high values in the response at high sediment levels. In addition, limited data were available in the sediment cover range from about 50% despite the overall large sample size of over 1000 data points. Consequently, this method did not provide meaningful thresholds. In particular, it did not allow for significant effects threshold scenarios, such as those required to estimate bottom-lines based on >20% effects level used for suspended sediment (see Chapter 5). However, it is noted that for suspended sediment ESV measures, a related QR approach was successful applied to macroinvertebrate data (at both the individual taxa and community metric level) presented in Chapter 5. This highlights the need for multiple approaches and that because of the differences in sediment ESVs, responses of different organism types, and ecological modes of action of sediment, there is unlikely to be a 'one size fits all' analytical approach.

Method 2: Boosted regression tree analysis (BRT method)

Boosted regression tree (BRT) models best demonstrate the macroinvertebrate community response to deposited sediment drivers. Overall, model fit and predictive performance were within a similar range to previous national models for macroinvertebrate metrics (Clapcott et al. 2012; Booker et al. 2014). The largest sample size and best spread of sample sites across the country was available for '% sediment cover instream' data. The metrics did not start to respond up to about 20-30% 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.

By contrast, macroinvertebrate metrics did not show a strong response to '% sediment cover bankside', suggesting that these data are of limited use for threshold identification for deposited sediment, despite the EPT richness metric indicating that a negative effect occurs between 10 to 20% 'sediment cover bankside'. The overall small effect size observed across the sediment cover gradient, compared to that observed for the corresponding 'instream' measurements may be due to assessment error. Bankside assessments are more prone to error as the assessor makes a single overall estimate of '% sediment cover' compared to multiple patch-scale estimates to calculate an average for the instream assessment. This suggests that a sediment attribute based on '% sediment cover' should be based on data collected via the instream sediment assessment method, SAM2.

Macroinvertebrate metrics also showed a negative response to SIS from about 500 g/m^2. The lack of response below 500 g/m^2 is not an indication of resistance of the macroinvertebrate community because SIS values within that range are typically observed at reference sites (Clapcott *et al.* 2011b). The response shape suggests that negative effects on macroinvertebrate community structure can potentially be observed when SIS exceeds natural levels. In contrast to the response to '% sediment cover instream', macroinvertebrate metrics ceased to respond at about 3,000 to 4,000 g/m^2 SIS, indicating that beyond this point there is no further significant change in community structure based on these metrics.

The strengths of BRT models include the flexibility in modelling complex, nonlinear response shapes, accounting for the effects of multiple predictors and their interactions. Limitations of the BRT approach, however, are that the models are fitted to the data regardless of whether response shapes are ecologically meaningful or not, and no measure of statistical significance or confidence around the visually determined threshold values. Hence, the response shapes are potentially influenced by unusual data points, although BRT models are known to be relatively unaffected by single outliers (Elith et al. 2008).

Despite these potential caveats, the inclusion of environmental predictors relevant to macroinvertebrate community structure in BRT models, and the relatively large dataset used, provides a good level of confidence in the identified deposited threshold values.

Method 3: Gradient forest analysis (GF method)

Gradient forest (GF) analysis of '% sediment cover instream' data identified that the largest change in community structure occurred between about 30 to 55%, which was consistent with '% sediment cover instream' threshold of 20 to 30% identified from the BRT model output (i.e., Method 2).

Across the gradient of '% sediment cover bankside' data, the results of the GF analysis suggested that the most substantial change in community structure occurred from about 60 to 90% sediment cover, a threshold that was not apparent from BRT model output (Method 2). Two further possible community thresholds, but of lesser importance (i.e., ecologically), were observed from about 5 to 15% sediment cover. However, as discussed earlier, we suggest definition of a deposited sediment attribute for sediment cover assessed instream.

GF analysis identified the largest change in community structure occurring between the lowest SIS value observed and about 500 g SIS/m^2. This value also was the impact initiation threshold identified using the BRT model approach which appears contradictory to the GF approach that indicated major changes in community structure would have already occurred at 500 g SIS/m^2. The discrepancy in thresholds likely stems from using taxon-level information in the GF approach as opposed to using macroinvertebrate metrics as the input variable for the BRT approach. The GF community threshold for SIS was mainly attributed to changes in 2 to 5 taxa, and these taxa appeared to not have influenced metric values significantly enough to detect a response change (via the BRT method).

In general, it has been argued that use of aggregate macroinvertebrate metrics for threshold definition may fail to protect a range of sensitive taxa that have their thresholds at very low levels of a stressor gradient (King & Baker 2010). Giving more weight to community thresholds identified with the GF approach at potentially lower sediment levels compared to thresholds identified from metric responses would be warranted if protection of the most sensitive taxa driving the community threshold or biodiversity generally was a management goal. A limitation of the GF analysis was that only a limited number of taxa met the occurrence criteria for inclusion in the analysis and that even further taxa get excluded because their RF models did not have an $R^2>0$, which may also be simply due to data limitation. For those reasons, rare taxa are less likely to contribute to the GF analysis although they may be among the most sensitive taxa (Wagenhoff et al. 2017a).

Our GF analysis (Method 3) was based on fewer data points relative to the BRT analysis (Method 2). Smaller sample size makes it less likely for rare taxa to be included in community threshold analysis, and only 13 and 17 taxa contributed to community thresholds across the sediment cover (assessed instream) and the SIS gradient, respectively.

Despite this limitation, we suggest that the GF analysis was useful in that it supported the indicative threshold of 20-30% 'sediment cover instream' identified via the BRT method. The results of the GF analysis suggested that individual macroinvertebrate taxa are affected by % sediment cover at levels below 20-30% and at SIS levels below 500 g/m^2. These taxa would probably not be protected by a sediment attribute defined using the shapes of metric responses. Using a smaller, but regional dataset, Wagenhoff et al. (2017a) found that while the major changes in community structure identified from taxon-level responses across a stressor gradient (nitrogen) were generally congruent with thresholds identified using a BRT approach (Wagenhoff et al. 2017b), the largest change in community structure was at the very low end of stressor gradient. The authors (Wagenhoff et al. 2017a) suggested that loss of biodiversity may not be as gradual as metric response suggest. This may have implications for assigning higher levels of protection for sensitive species (i.e., A/B and possibly B/C band thresholds).

4.3.5 Analysis of temporal variation to inform attribute frequency/duration criteria

We explored temporal variation in SIS and '% sediment cover' metrics to determine the minimum/recommended number of sampling events required to assess compliance against proposed deposited sediment attribute thresholds.

Full methodological details of the analyses used to inform the monitoring requirements for assessing against a proposed threshold for a deposited sediment attribute are provide in Appendix Q. The data, methods and summary are briefly discussed in this section.

For SIS, the only suitable data available was from the Whatawhata integrated catchment management research project (NIWA, unpublished data). Eight sites were sampled quarterly or biannually between 1995 and 2013, providing between 22 and 53 replicate samples per site. Five of the 8 sites were treatment sites and subject to a change in catchment vegetation, and the remaining 3 sites were controls (with greater than 69% native vegetation), as described in Quinn et al. (2009).

For the % cover sediment dataset, we identified sites from the collated database where the '% sediment cover' of fine sediment had been repeatedly measured, over time, using either bankside (SAM 1) or instream (SAM2) visual assessment methods. For both SAM1 and SAM2 methods the variance for all sites with ≥4 temporal samples were calculated. The relationship between the mean and standard deviation was then used to estimate the number of samples required to determine the mean within an absolute 10% fine sediment cover margin of error (i.e., +/- 5%).

Based on the analyses undertaken with '% sediment cover' and SIS datasets, from the temporal variation in deposited sediment data, we concluded the following:

- % sediment cover instream: approximately 24 monthly samples are required to accurately estimate the, considering the likely methodological error of visual estimates due to observer bias.
- SIS: requires up to 6 years of quarterly measurements (i.e., 24 samples) to accurately estimate mean values.

4.4 Macroinvertebrate-based deposited sediment thresholds consistent with NPS-FM ecosystem health 'bottom-line'

4.4.1 Deposited sediment threshold values for macroinvertebrates

We defined effects thresholds for macroinvertebrates that we considered were consistent with a transition from a 'C' band state. The 2014 draft implementation guide defined an ecosystem health 'C' band as generally representing a minimum safe level before an ecological tipping point. The two thresholds derived for deposited sediment that we consider are consistent with that definition relate to *impact initiation* thresholds indicated by marked changes in macroinvertebrate community metrics/structure.

The BRT model approach identified a distinct threshold in the response of commonly used stream health metrics (MCI and EPT taxon richness) as well as two newly developed sediment-specific metrics (Sediment MCI, Number of decreasers) at 20-30% cover, at which point a further increase in sediment cover led to a continuing decrease in these metrics. Also, GF analysis suggested that from 30% up 55% sediment cover the macroinvertebrate community changes most significantly. Furthermore, the 20-30% sediment cover threshold we found in our analyses is similar to thresholds defined by previous New Zealand studies. For example, 20% sediment cover was the threshold at which %EPT started to significantly decline using a sigmoidal model on a relatively small dataset consisting of 30 Canterbury stream sites (Burdon et al. 2013), and also streams supporting macroinvertebrate communities indicative of very good stream health (MCI > 120) were typically associated with sediment cover of 20% or less using a small national data set (Clapcott et al. 2011).

Together, these results support a proposed C/D band threshold of 30% sediment cover (instream, SAM2) to protect ecological health of hard-bottom streams (or rivers) with naturally low-to-medium sediment levels (sediment reference state of 0-30% sediment cover). Our BRT analysis did not provide indication of a threshold beyond 30% sediment cover but GF analysis suggested that 55% could be a point beyond which macroinvertebrate communities do not further change significantly. As this threshold was close to the boundary defined for the hard-bottom stream class with naturally high sediment levels (sediment reference state of 30-60% sediment cover), we suggest that 60% sediment cover is a pragmatic bottom line threshold (based on macroinvertebrates) in this stream class.

Table 4-8: Macroinvertebrate-based thresholds for deposited sediment defined by marked changes (i.e., impact initiation) in macroinvertebrate metrics and community structure.

Predicted reference class (range of natural-state '% sediment cover')	name (descriptor) for reference class	Recommended threshold comparable to NPS-FM bottom-line (% sediment cover instream)[2]
0-30%	low-to-medium	30%
30-60%	high	60%
>60%	soft-bottom	na[1]

[1] streams classed as naturally soft-bottom are exempt from deposited sediment thresholds. [2]Assessed as annual mean, based on a monthly monitoring regime. The minimum record length for grading a site based on an instream visual assessment of % fine sediment cover (SAM2) is 2 years.

We do not recommend deposited sediment thresholds for soft-bottom streams, which are defined as those with naturally >60% sediment cover. The proposed thresholds for deposited sediment, based on macroinvertebrate effects are summarised in Table 4-8. These thresholds are considered, along with other lines of evidence, in the synthesis presented in Chapter 7, which proposes C/D band thresholds for the NOF deposited sediment attribute table.

4.4.2 Limitation of our proposed C/D band threshold to protect biodiversity values

Biodiversity is an important characteristic of ecological integrity, and the current NPS-FM (MfE 2014b) also defines 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'. One caveat of using macroinvertebrate metrics as indicators for ecological health, such as in our BRT analysis, is that they may not be good indicators for threshold definition that adequately protects macroinvertebrate biodiversity values (see concerns raised by King & Baker 2010; Wagenhoff et al. 2017a). In other words, we do not know how many sensitive macroinvertebrate species our proposed C/D threshold will protect and how many it would not protect as macroinvertebrate metrics were used to investigate community change. Hence, while our proposed macroinvertebrate=derived thresholds are likely to protect streams from major changes in the macroinvertebrate community, it is unlikely to protect all sensitive macroinvertebrate species that can be found in streams with no human-induced elevated sediment levels.

The GF method modelled taxon responses then aggregated that information hence community change was used to support definition of the C/D band threshold at 30% sediment cover. The GF analysis suggests that there were changes in individual taxa at levels lower than 30% sediment cover, more specifically at levels as low as between 1 and 5%. This suggests that there are species sensitive to very low levels of deposited sediment and that an A/B (and potentially B/C) threshold would likely have to be defined at a very low % cover threshold to provide a high species protection level. One caveat of the GF method is, as previously discussed, that only a small number of taxa make it into the analysis and these taxa are often the more common taxa and less likely rare species. Hence, the GF approach is less likely to be useful for definition of those more stringent attribute band thresholds that may want to protect biodiversity values at a high level.

Overall, while our BRT and GF analyses have limitations in that we do not know at what level our proposed thresholds (based on marked changes in metrics/community structure) protect biodiversity, we believe that they are a useful first step for deriving a national bottom line that does not necessarily aim protect the most sensitive species.

4.4.3 Limitations of stream classification according to sediment reference state

A large proportion of New Zealand streams have natural amounts of fine sediment in the substrate and such streams are typically referred to as hard-bottom streams. Hence, excessive fine sediment that enters streams through human land use is a major stressor. However, some streams, the so-called soft-bottom streams, are naturally characterised by a high percentage of fines in the substrate as a result of certain catchment geologies, and little is known about how additional human-induced fine sediment affects the health of these streams. Obviously, sediment thresholds applicable to hard-bottom streams are not applicable to soft-bottom streams. In fact, classification of hard-bottom vs soft-bottom streams (as sometimes done for SoE reporting) is very coarse compared to what naturally is rather a continuum (i.e., different reference streams can have any natural level of fines in their substrate) and not useful for setting meaningful thresholds that are stringent enough to adequately protect hard-bottom streams but at the same time do not exceed natural levels. From a practical perspective, it is desirable to define a small number of different stream types but with respect to sediment levels, and sediment thresholds associated with them, that is able to balance these requirements.

Our stream classification work (section 2) distinguished the following three reference stream types with respect to their natural sediment levels; reference streams with 0-30% sediment cover (we called these hard-bottom reference streams with 'low-to-medium' sediment levels), 30-60% sediment cover (we called these hard-bottom reference streams with 'high' sediment levels), and >60% sediment cover (we called these soft-bottom streams). Our threshold analyses were done on a dataset that incorporated streams of the hard-bottom and soft-bottom types as stratifying the analyses would have required classifying streams a priori (which again is subjective and maybe even quite difficult) and probably would have led to data limited models, especially with the BRT and GF methods.

Due to the relatively coarse classification into these two sediment reference state classes for which we proposed thresholds, our thresholds will be necessarily most stringent on streams that naturally have fine sediment levels around the higher end of the class boundary (i.e., 30% or 60% sediment cover, for those two classes respectively) and less stringent on streams that naturally have fine sediment levels around the lower end of the class boundaries. For example, the 30% threshold applicable to a stream that naturally has 25% sediment cover is relatively restrictive with respect to the amount of fine sediment that is allowed to enter the stream due to human impacts compared to a stream that naturally only has 5% sediment cover.

4.4.4 Frequency of sediment monitoring

The proposed frequency of measurement of 2 years for sediment cover (Table 4-9) was determined by an analysis of temporal variation in % cover data at the national scale (details provided in Appendix Q). Variation was related to the mean value and while as few as 12 monthly samples were sufficient to estimate mean values with an accuracy of +/- 6% of the mean at low-sediment sites, approximately 2-times as many (i.e., 24 monthly) samples are needed to estimate mean values with the same precision at high-sediment sites.

Temporal variability of suspendable inorganic sediment (SIS) was also analysed and suggested that up to 6 years of quarterly measurements are needed to assess mean values with a similar accuracy as that determined for the sediment cover data. This analysis however was data limited and possibly overestimated the number of samples needed to provide sufficient accuracy.

4.4.5 Tentative C/D band thresholds for suspendable inorganic sediment (SIS)

We think it is unlikely that regional council staff will implement the SIS method because of its limited application (e.g., cannot be applied in deeper streams, in very fast flowing water and in streams with cobble or boulder substrates), and significant cost related to effort required in the field as well as in the lab (e.g., Clapcott 2016; Clapcott 2017). Nevertheless, our threshold analyses provided some interesting results. As for sediment cover, where we proposed a C/D band threshold based on the impact initiation threshold determined in the BRT analysis, we propose a tentative C/D threshold for SIS at 500 g/m^2 to protect ecosystem health for the streams that classify as hard-bottom streams with low-medium sediment levels (<30% cover). This level of 500 g/m^2, however, can be considered close to reference condition for the average reference site (Clapcott et al. 2011) allowing little additional increase in SIS from human impact. However, as for the C/D threshold for sediment cover, our GF analysis across the SIS gradient also suggests that there are changes in individual taxa at levels lower than 500 g SIS/m^2. Hence, the same note of caution and the same suggestion for additional analysis are associated with this threshold.

In our dataset, a strong relationship was apparent between SIS and specific sediment yield (SIS could be linked to land management options), and between SIS and % cover instream (SIS could be linked to visual assessments of deposited sediment) that could be used to derive attribute band thresholds. Hence, SIS thresholds can be derived simply by translating the 30% and 60% sediment cover thresholds into SIS units using a simple linear regression model. We used the same dataset which resulted in a sample size of 343 where data was available for both sediment cover (% cover instream) and SIS. Data transformation was performed as previously (square root for sediment cover, natural log for SIS). Given the noise in environmental data, the regression model had a relatively good fit (R^2 of 0.37, P<0.001). The 30% and 60% cover thresholds translated into 460 and 1,070 g SIS/m^2, respectively. The former SIS threshold for the 0-30% sediment reference state class is very close to the tentative ecologically defined threshold of 500 g SIS/m^2 from the BRT model. Our analyses did not suggest a specific SIS threshold that could be adopted for the 30-60% sediment reference state class, but the SIS threshold translated from the 60% sediment cover threshold can be adopted. This equated to about 1000 g SIS/m^2, which is well below the point at which macroinvertebrate metrics ceased to respond to SIS according to the BRT analysis. Any bottom-line thresholds defined for ecological health protection should be below levels that suggest no further change in a stream health metric (Wagenhoff et al*.* 2017b). We note though that sediment cover and SIS measure different aspects of sediment deposition. Sediment cover is a measure of surficial sediment, whereas SIS is a measure of surficial and subsurface sediment. Hence, simple translation of the sediment cover thresholds may not accurately reflect the effects of SIS on macroinvertebrates. Hence, we propose a tentative C/D band threshold of 500 g SIS/m^2 for the 0-30% sediment reference state class (hard-bottom streams with low-medium sediment levels) and of 1000 g SIS/m^2 for the 30-60% sediment reference state class (hard-bottom streams with high sediment level) (Table 4-9).

4.5 Future work

We identified two important limitations of our analyses.

1. The limitation of not being able to provide information on to what level our proposed thresholds (that we consider consistent with ecosystem health bottom-line values) protect biodiversity. This type of information will be particularly important for definition of the more stringent attribute bands (A/B and B/C band thresholds). For example, it may be desirable to define attribute band thresholds based on predetermined species protection levels similar to how the ammonia NOF attribute has been defined (A/B = 99%, B/C = 95% and C/D = 80% species protection level). Hence, we suggest future threshold analyses on taxon-level responses (other than GF analysis) to determine a sediment threshold based on a desired species protection level for all attribute band thresholds to identify sensitive taxa and develop sediment attributes that adequately protect biodiversity values, highly valued macroinvertebrate species and ultimately ecosystem functioning. These analyses likely require collection of more data.

2. We discussed the limitations of implementing attributes based on our relatively coarse stream classification (0-30% and 30-60% sediment cover). An alternative way of managing the negative effects of excessive sediment is by defining attribute band thresholds as a deviance from reference state. Our current reference state predictions for deposited sediment are based on a model of moderate accuracy and precision (TDE = 53%, see section 2), which is comparable to overseas attempts to predict broad-scale spatial patterns in deposited sediment. Future work could focus on improving the model by incorporating data with improved spatial and temporal representation of reference state. However, model predictions is only the first step, secondly it needs to be determined what an ecologically meaningful deviance would be for definition of the attribute band thresholds. For example, threshold analyses as the ones we conducted could be done separately for a set of different sediment reference state classes. These analyses also likely to require more representative data.

5 Derivation of suspended sediment thresholds based on macroinvertebrate responses

5.1 Summary

This chapter reports on the response of benthic macroinvertebrates to increasing suspended fine sediment (SS) in rivers and streams to derive threshold values that are broadly consistent with a C/D band management threshold. These values contribute to a *multiple lines of evidence approach* (along with fish response thresholds, literature threshold values, regulatory guidelines and expert opinion) to develop a proposed C/D band threshold for suspended sediment in New Zealand streams and rivers (see Section 7). The section begins with a brief literature review of selected macroinvertebrate-SS effects, and regulatory guideline values of SS used around the world.

While suspended sediments typically are found to have detrimental effects on macroinvertebrate communities, there was also evidence of a potential subsidy/stress relationship. A non-linear quantile regression was chosen to reduce the potential confounding effects of other stressors or natural environmental variables (as discussed in Chapter 4).

The quantile regression approach used for macroinvertebrate suspended sediment measures differ from that used for 'macroinvertebrate vs. deposited sediment' and 'fishes vs. sediment' as these sediment effects do not result in subsidy/stress response relationships, and were based on different databases for their derivation. As mentioned above, fish data was limited to presence/absence data, whereas macroinvertebrate monitoring comprises abundance data. The quantile regression approach was found to be insensitive when applied to deposited sediment gradients.

Seven biotic community diversity and/or abundance metrics were selected as response variables to establish SS thresholds, using turbidity and visual clarity as measures of SS from 67 sites across the National River Water Quality Network (NRWQN). Models were fit for each macroinvertebrate indicator (either a metric or the abundance of selected taxa). The potential confounding effects of a range of other potential stressors were apparent but did not have strong relationships with macroinvertebrate responses.

For each stressor-response pair, turbidity and visual clarity thresholds were visually estimated from the respective quantile regression model at the point where the macroinvertebrate response variable was reduced by 30% as compared to a maximum value.[12] A predefined 30% reduction level for metrics and taxon abundance was selected to represent likely chronic effects on individual macroinvertebrate 'measures'. These thresholds subsequently were used to construct a plot visualising the cumulative 'species' sensitivity distribution (SSD).[13] The SSD is then used to provide management band thresholds according to differing levels of environmental protection. Consistent with other attributes (i.e., nitrate and ammonia toxicity), a species protection level of 80% was used to estimate threshold values of turbidity and visual clarity from the SSD curves. These threshold values were consistent with NOF C/D band thresholds (i.e., transition to significant adverse effects).

The 80% protection level thresholds (used to inform the proposed NOF C/D threshold) were 4.3 NTU and 0.9-1.0 m for turbidity and visual clarity, respectively. These values contributed to a multiple

[12] For the non-linear models, the maximum response value was the peak of the subsidy/stress relationships while for the linear models, the maximum value was at the lowest end of the stressor gradient
[13] Note that individual taxa as well as macroinvertebrate metrics were included in the SSD plot

lines of evidence approach to define the proposed thresholds for suspended sediment, which is described in Chapter 7 (Synthesis of results).

5.2 Literature Review: suspended sediment effects thresholds and regulatory guideline values

5.2.1 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 1991a; Wood & Armitage 1997b; 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 & 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.

5.2.2 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 & Hickey 1993). These discharges contain both particulate organic SS and nutrients which result in stimulation of benthic periphyton growths (see Figure 5-1). 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.

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 & Quinn 2011a).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 hr). 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.

5.2.3 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 5-1). 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 7-10 NTU.

Table 5-1: Summary of study results on effects of suspended sediments on invertebrates. (adapted from (Bilotta & 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 & feeding unaffected	Australia	(Kefford *et al.* 2010)
Macro-invertebrates	8 (±2) NTU	Winter dataset	Moderate impairment of riffle macroinvertebrate scores	USA	(State of Oregon Department of Environmental Quality 2010)
	9 (±2.2) NTU	Not stated	20% decrease in PREDATOR[a] score		

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

5.2.4 Numeric standards

Numeric thresholds for management (e.g., standards, guidelines, criteria or trigger values) for suspended sediments can be expressed either as the mass concentration, measured as total suspended solids (TSS) or expressed in units of some surrogate measure (e.g., turbidity and visual clarity). For example, the amount of suspended sediment in water influences water clarity and turbidity. TSS is a measure of the mass of organic and inorganic particles suspended in a water sample, whereas turbidity is a relative measure of the amount of scattering of light by particles (Davies-Colley & Smith 2001).

Davies-Colley et al. (1992) reported close correlations between the log-transformed measures of turbidity and TSS (r = 0.94), turbidity and visual clarity (r = -0.96), and TSS and visual clarity (r = -0.94) in West Coast streams, so that any one of these three closely interrelated variables could be used to define the loading of suspended inorganic sediment (Davies-Colley et al. 1992);however, the correlation should be checked for each water body. Numeric standards are best determined by dose-response studies along natural or artificial gradients to find thresholds/tipping points and 'safe levels'. In the absence of such data, "trigger values" for non-toxic stressors, including TSS and turbidity, are usually determined by examining naturally occurring background values, or variability around these values, to establish critical reference thresholds within an eco-region (Culp et al. 2009). A common approach for defining trigger values above which ecological impacts may occur, whilst accounting for natural variability, is to select a particular percentile of naturally occurring values (see (ANZECC 2000); (US EPA 2006)). The ANZECC (2000) physico-chemical guidelines use the 80th (or 20th for inversely related visual clarity) percentile for distributions from "slightly to moderately disturbed" ecosystems was used as a guide. The statistical data were from the lowland rivers (three rivers) and upland rivers (18 rivers) from the National River Water Quality Network (NRWQN) (Davies-Colley 2000). These values do not represent any basis for the establishment of adverse ecosystem effects. As such, they should not be used in a regulatory context. Ultimately, the method used to define thresholds will depend upon the ecosystem type, the desired level of protection, and the availability of suitable reference systems and adequate data for these systems (ANZECC 2000; US EPA 2006).

A summary of existing numeric standards for suspended sediments from jurisdictions worldwide shows that many use turbidity as a surrogate measure (Table 5-2). Most of the international studies on suspended sediment effects on macroinvertebrates rely on short-term laboratory-derived data or did not provide a concentration-response relationship suitable for deriving effects thresholds. Field monitoring studies have been used to generate guidelines and criteria for 'impairment' – with macroinvertebrate impairment occurring in the turbidity level range 7-10 NTU (including New Zealand studies).

Table 5-2: Water quality guidelines and/or criteria for suspended sediments (SS) or surrogate measure (i.e., turbidity) to protect aquatic life by jurisdiction. (from State of Oregon Department of Environmental Quality 2010, unless stated).

Jurisdiction	Guideline or criteria
Australia & New Zealand	Trigger values for upland rivers >150 m to <1500 m and lowland rivers <150 m. South East Australia: Upland rivers 2–25 NTU, Lowland rivers 6–50 NTU. Tropical Australia: Upland & Lowland rivers 2–15 NTU. South West Australia: Upland & Lowland rivers 10–20 NTU. South Central Australia: Upland & Lowland rivers 1–50 NTU. **New Zealand**: Upland rivers 4.1 NTU, Lowland rivers 5.6 NTU. ANZECC (2000).
European Union	Apart from in exceptional circumstances such as storms or droughts, concentrations should not exceed 25 mg/L SS in waters suitable for both salmonid and cyprinid fish populations Freshwater Fisheries Directive (78/659/EEC) (2006/44/EC) (Environment Agency 2011).
Europe	Maximum of 25 mg/L SS measured as annual mean (for high level of protection). Alabaster and Lloyd (1982); (EIFAC 1964).
Austria	Maximum of 4,500 mg/L SS for flushing operations. (Schneider *et al.* 2006).
Canada	For maintenance of biotic values thresholds of 13 mg/L SS and 8 NTU. Seasonal or annual averaging is recommended. (Culp *et al.* 2009).
Italy	Maximum of 100,000 mg/L SS for 1 h and a maximum daily mean of 6,500 mg/L SS (for flushing operations). (Vitali *et al.* 1995).
Northeast Italy	Maximum of 30,000 mg/L for 2 h, and a mean of 6,000 mg/L for the entire duration of flushing. (Regional Government of Veneto 2006).
Northwest Italy	Maximum of 40,000 mg/L for 0.5 h and less than 5,000 mg/L for the entire duration of flushing (equal or less than 1 week). (Regional Government of Piemonte 2008).
Switzerland	Limits for flushing ranging from 5 to 10 ml/L SS. (Gerster & Rey 1994).
Alaska	May not exceed 25 NTU above natural conditions.
Arizona	Maximum concentrations. Cold water fishery: Not to exceed 10 NTU in rivers, streams, other flowing waters, lakes, reservoirs, tanks and ponds. Warm water fishery: Not to exceed 50 NTU in rivers, streams, and other flowing waters.
British Columbia	Change from background of 8 NTU at any one time for a duration of 24 hours in all waters during clear flows or in clear waters; change from background of 2 NTU at any one time for a duration of 30 days in all waters during clear flows or in clear waters; change from background of 5 NTU at any time when background is 8-50 NTU during high flows or in turbid waters; change from background of 10% when background is >50 NTU at any time during high flows or in turbid waters. Maximum increase of 25 mg/L from background level in 24 h, and a maximum mean increase of 5 mg/L from background in 30 days (when background is less than or equal to 25 mg/L British Columbia Ministry of Environment, Lands and Parks (1998).
California	North Coast Region as typical example: Turbidity shall not be increased more than 20% above naturally occurring background levels. Central Valley Region. Where natural turbidity is: <1 NTU controllable factors shall not cause d/s turbidity to exceed 2. Where 1-5 NTU, increases shall not exceed 1 NTU. Where 5-50 NTU, increases shall not exceed 20 percent. Where 50-100 NTU, increases shall not exceed 10 NTU. Where > 100 NTU, increases shall not exceed 10 percent.
Idaho	Cold Water Aquatic Life. Turbidity, below any applicable mixing zone, shall not exceed background turbidity by more than 50 NTU instantaneously or more than 25 NTU for more than 10 consecutive days.
Washington	Char Spawning and Rearing; Core Summer Salmonid Habitat; Salmonid Spawning, Rearing, and Migration; and Non-anadromous Interior and Redband Trout: 5 NTU over background turbidity when the background turbidity is 50 NTU or less, or a 10% increase in turbidity when the background turbidity is greater than 50 NTU. Salmonid Rearing and Migration *Only*; and Indigenous Warm Water Species: 10 NTU over background turbidity when the background turbidity is 50 NTU or less, or a 20% increase in turbidity when the background turbidity is greater than 50 NTU.

5.3 Introduction to conceptual framework and selection of macroinvertebrate indicators

Subsidy/stress relationships have been reported for benthic macroinvertebrate communities in New Zealand streams in agricultural environments along gradients of deposited fine sediment and nutrients (Quinn & Hickey 1990a; Wagenhoff *et al.* 2011) and for sites impacted by a dilution gradient of sewage oxidation ponds (Quinn & Hickey 1993). Such relationships are non-linear in nature with thresholds for adverse effects occurring at a SS concentration greater than that which results in a maximum taxonomic richness or species abundance. Species responding to SS will differ depending on both the chemical nature of SS and the concentration of SS – potentially confounding the establishment of sensitivity/response relationships based on correlative responses with surrogate measures of SS, such as turbidity of visual clarity. A marked subsidy/stress response was observed for analysis of the macroinvertebrate community responses to both turbidity and clarity in the NRWQN (this study).

A conceptual diagram for a subsidy/stress relationship illustrated for generic stressors and pastoral development is shown in Figure 5-1. Fine POM is expected to cause a subsidy/stress relationship – with adverse effects at high concentrations occurring because of clogging of gills and/or settling of POM to overly enrich the sediments, resulting in hypoxia (low dissolved oxygen) and build-up of decomposition products (e.g., toxic ammonia) within the streambed. In contrast, the 'toxic' response (Figure 5-1) would be expected to occur where the SS is dominated by inorganic fines that clog the bed without providing food resources and hence no stimulation in biodiversity or key species (Quinn 2000).

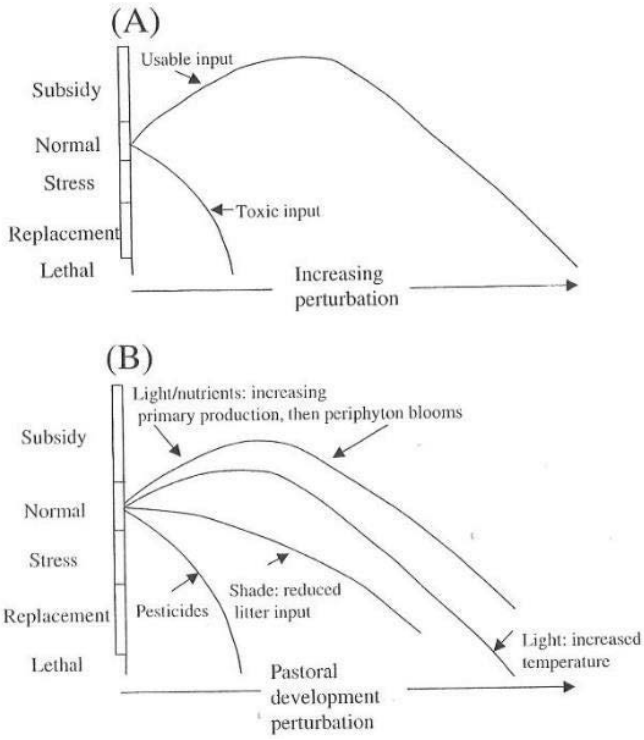


Figure 5-1: Conceptual models for subsidy/stress effects of environmental perturbations (A), and as adapted to summarise the key effects of pastoral agriculture on stream macroinvertebrates. (from (Quinn 2000))

A conceptual framework for the effects of suspended sediment on stream macroinvertebrate communities is shown in Figure 5-2 and the mechanisms of effect of different suspended solids components are summarised in Table 5-3. Figure 5-2 identifies potential modes of action of the change in concentration and composition of SS, together with the likely most sensitive biotic groups. Importantly, several indirect or 'proximate' stressors are identified with increasing SS concentration as pathways underlying potential adverse effects. These proximate stressors are associated with reduced light penetration in the water column and associated reduction in benthic biofilm production. The primary modes of action for inorganic SS with invertebrates are (Figure 5-2):

- changes in feeding efficiency – affecting filter-feeders, and
- gill damage – affecting taxa with exposed gills.

An understanding of key modes of action for macroinvertebrate changes in response to SS was informative in selecting the measures for detection of adverse effects in community biomonitoring datasets. However, the nature of the combined stressors of organic POM and inorganic SS means that specific sensitive species which are responsive to a specific mode of action are not expected. Thus, we have included key representative species which are likely to be sensitive based on the broad mode of action, including:

- mayfly (Ephemeroptera), *Deleatidium* spp. – exposed gills, potentially sensitive to a change in biofilm food quality
- filter-feeding caddis (Trichoptera), *Aoteapsyche colonica*; net-spinning species, potentially sensitive to net clogging
- a snail, *Potamopyrgus antipodarum* – potentially sensitive to a change in biofilm food quality
- caddis (Trichoptera), *Pycnocentria evecta* – widespread occurrence, collector-browser species potentially sensitive to a change in biofilm food quality
- caddis (Trichoptera), *Oxyethira albiceps* – feeds on filamentous algal cells, potentially affected by reduced food quality in biofilms
- stonefly (Plecoptera), *Zelandobius* sp. - algal and detrital browsers, potentially affected by reduced food quality in biofilms, and
- stonefly (Plecoptera), *Zelandoperla* sp. – algal and detrital browsers, potentially affected by reduced food quality in biofilms.

Additionally, a range of macroinvertebrate community metrics were evaluated, including:

- total taxa richness
- total density
- EPT richness[14]
- EPT abundance
- %EPT
- MCI, and
- QMCI.

[14] EPT = Ephemeroptera, Plecoptera and Trichoptera

Table 5-3: Macroinvertebrate suspended sediment stressors.

Stressor	Effect	Mechanism
Particulate Organic Material (POM)	Subsidy/Stress	Initial food increase followed by stressor effects (e.g., ammonia, sediment deoxygenation)
Inorganic Suspended Sediment (SS)	Stress	Gill clogging, filter net clogging, reduced biofilm quality
Visual clarity	Indirect effect on predation	Fish and visual predator reduced feeding efficiency
Light penetration	Stress: Plants or periphyton biofilm productivity reduction	Food quantity (and quality) reduction for macroinvertebrates
Light quality	Subsidy(?)	UV radiation penetration affects invertebrate behaviour

Note, that although bespoke sediment metrics ('sediment MCI' and 'number of decreasers') were developed as part of the deposited sediment work (refer to Section 3), this process was done in parallel with the suspended sediment effects workstream and so the suitability of these metrics could not be tested for suspended sediment gradients. The use of MCI and EPT metrics for determining effects thresholds is consistent with their use in deposit sediment (2 of the 4 macroinvertebrates metrics modelled in Section 4). The sensitivity of the biotic measures from macroinvertebrate monitoring data are related to the surrogate measures of SS - turbidity and visual clarity - to determine thresholds for adverse effects.

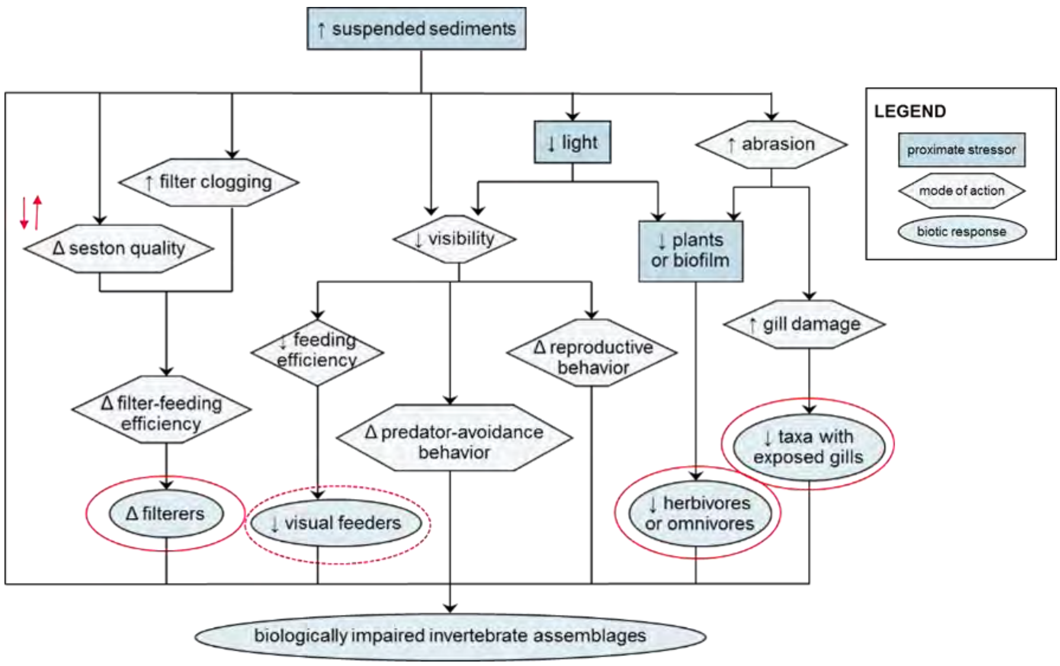


Figure 5-2: Conceptual framework for effects of suspended sediments on river invertebrate communities. Section from (US EPA 2014). Red circled indicate a range of biotic responses which may be expected from a change in suspended sediment.

5.4 Methods

5.4.1 Introduction: the quantile regression approach

The approach to establish thresholds for suspended sediments (in the form of turbidity and/or visual clarity) uses quantile regression methods, which are particularly suitable for identifying thresholds in both linear and non-linear distributions (Cade & Noon 2003).

As described by Cade and Noon (2003), "Quantile regression is a way to estimate the conditional quantiles of a response variable distribution in the linear model….". These authors express the intended benefit of quantile regression as follows:

> "*Typically, all the factors that affect ecological processes are not measured and included in the statistical models used to investigate relationships between variables associated with those processes. As a consequence, there may be a weak or no predictive relationship between the mean of the response variable (y) distribution and the measured predictive factors (X). Yet there may be stronger, useful predictive relationships with other parts (quantiles) of the response variable distribution*."

Some background information on the application of quantile regression to environmental data is provided in Appendix S.

5.4.2 Defining effects-based thresholds

The effect threshold encompasses two distinct components. The first component is the method of establishing the chronic effect value for the individual species, and community diversity and/or abundance metrics (e.g., EPT richness, EPT abundance, MCI, QMCI). The second component is combining the individual species/metric data to calculate a species sensitivity distribution (SSD). The SSD is then used to provide guideline values for differing levels of environmental protection. The key focus of the analysis was to derive a threshold for the C/D band, otherwise referred to as the "national bottom-line (threshold)".

The effect threshold for use in environmental guideline derivations is generally based on long-term toxicity data for laboratory-exposed organisms. The statistical measure used is generally either a no observed effect concentration (NOEC) value, or a minimal effect value determined from the concentration-response relationship (e.g., and EC_{10} or LC_{10} value). This approach is used as the basis for input to the SSD for the ANZECC guidelines (ANZECC 2000; Warne et al. 2015) and the updated procedure for those guidelines.

The SSD for the cumulative number of species is used to calculate a range of thresholds for differing levels of protection (i.e., 99%, 95%, 90%, 80%). A statistical calculation procedure is used to calculated these guideline values using a log-logistic procedure for datasets with five or less values or a BUR type III distribution for larger numbers of data (Warne et al. 2015).

Water quality guidelines have rarely been derived from field-based effects and SSD data – though this is the method used for all environmental sediment quality guideline derivations. The recent exception to this is the method used by the US EPA for the derivation of field-based conductivity criteria (US EPA 2011). The US EPA approach used multi-species field-based data for the extirpation threshold for each species. Use of a species extirpation value is particularly robust compared with trying to estimate statistical thresholds for lower effects compared to reference site values for each of the species. This data was then used in an SSD to calculate a 95% protection threshold for conductivity as a criteria value – noting that this represented extirpation of those species rather than sub-lethal effects on those sensitive species. This use of field-based concentration-response data is the basis of the approach used here to derive 'bottom line' values for suspended sediment thresholds for protecting ecosystem health.

The derivation of a 'bottom line' value for application as National Objectives Framework (NOF) standards in the NPS-FM seeks to identify a threshold level of protection which is protective of a 'tipping point' for environmental decline of key biotic characteristics (e.g., biodiversity, key functional or endangered species) of the environment. For example, the derivation of the NOF standards for the toxicant nitrate-nitrogen was based on laboratory-based toxicity data, with the 'bottom line' corresponding to an 80% protection threshold in the SSD. This threshold represented the transition between chronic sub-lethal effects (e.g., growth/reproduction) and the onset of significant effects on species survival (i.e., lethality) (Hickey 2013). Based on long-term laboratory data such effect thresholds can be rigorously statistically determined with effects attributable to the contaminant of concern. However, the derivation of field-based guidelines is generally less robust because of the uncertainties in linking the causative stressor with the measured biological response.

For these water column SS guideline derivations, we sought to determine *bottom-line* values (i.e., C/D band thresholds) for a range of representative species and for key community macroinvertebrate measures of biodiversity (e.g., taxa richness, EPT richness) and abundance using field-based data. The derivations are based on surrogate measures of SS, namely turbidity and visual clarity. Inspection of the data showed that extirpation of key species or excessive reductions in major indices, such as taxonomic richness, did not occur at very high median turbidity (or low median visual clarity) conditions, however, a marked decline in these measures did occur in a concentration-dependent manner. The lack of extirpation indicated that several tolerant species (or individuals for many species) remain at high concentrations. The quantile regression approach used is suitable for application to these macroinvertebrate datasets.

5.4.3 Data

The primary stressor/response analysis was undertaken on the NRWQN dataset for the period 1990 to 2013 as used for establishing benthic periphyton biofilm relationships with macroinvertebrate communities (Matheson et al. 2012; Matheson et al. 2015). This data was 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.

All macroinvertebrate data was analysed together because the natural state variation of suspended sediment (refer to Section **Error! Reference source not found.**) was significantly lower than literature effects thresholds. For example 80% of reference sites had turbidity values of between 0.4 and 2.2 NTU, compared to the lower end of turbidity-based effects thresholds in the literature of around 5 to 7 NTU (Table 7-6, Section 7), which were assumed to be representative of potential C/D band threshold values. This situation would not apply for lower environmental thresholds (i.e., A/B and B/C band thresholds) as these would most likely overlap with natural state values.

Pragmatically, the annual median of the monthly water quality monitoring data for each site provides a robust measure of the annual data. Conceptually, high seasonal variability in SS could result in the adverse effects on communities occurring during the high flow season which persist without sufficient recovery time occurring for the macroinvertebrate populations at the time of summer sampling. We examined the relationships between turbidity and clarity measurements for concurrent, annual median and annual 80th percentile measures for NRWQN and regional SoE monitoring datasets. We found reasonable agreement between a turbidity and visual clarity for the annual medians (r^2 = 0.81 for all NZ data; r^2 = 0.88 for NRWQN data for linear correlation of log transformed variables, Appendix T), but a marked increase in the scatter of the 80th percentile data (r^2 = 0.45 for all NZ data; r^2 = 0.45 for NRWQN data for linear correlation of log transformed variables, Appendix T).

Conceptually, at least, we would have favoured selecting a summary statistic away from the central tendency (median) of the data (such as the 80th percentile), for evaluating relationships with macroinvertebrates as these statistics shown greater variability. This is based on the hypothesis that the higher SS 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 (Section 2) used long-term medians and it was not known whether the assumption for medians would be valid if using 80th percentiles values of suspended sediment. Thus, for pragmatic reasons, we have selected our effects analysis to be based on the annual median for river sites. This is consistent with the classification work (based on long-term site medians, (Depree 2017)) so it was a logical step to focus our response analyses using the same summary statistic.

SoE regional data

A comparative analysis of NZ monitoring data with the NRWQN dataset for visual clarity and turbidity is provided in Appendix U, and the NRWQN datasets with stream order and REC classification classes is provided in Appendix V. This analysis indicated that the NRWQN dataset has a representative range in both turbidity and visual clarity values compared with the larger New Zealand SoE monitoring dataset (Appendix U). This comparative similarity extends to both the monthly turbidity/visual clarity relationships (log-transformed) and to the spread in these measures for the 80th percentile measurements. Accordingly, we decided to conduct our primary analysis for effects assessment on the NRWQN dataset, because of the consistency of the methodology, the quantitative macroinvertebrate monitoring and the availability of other data on potential confounding stressors (Scarsbrook *et al.* 2000; Davies-Colley *et al.* 2015b). Given the wide gradient in SS provided by the NRWQN sites, the inclusion of a greater number of data from regional SoE sites was not anticipated to change the magnitude of thresholds for a generic derivation; but it could reduce uncertainty in relation to river classification class relationships. However, without the same measurements and availability of data to inform potentially confounding stressors – we decided that the 'pros' outweigh the 'cons' for basing the generic derivation on the NRWQN dataset, rather than also including regional SOE data.

5.4.4 Stressor elimination

We identified a range of potential stressors at sites in the NRWQN dataset which may confound establishing a defining causative relationship with water column SS (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 other stressors confounding the quantile regression analysis was undertaken using data visualisation software (DataDesk, (Velleman 1989)). Relationships with each of the potential stressors was examined to determine:

i. if there was an apparent stress effect with increasing concentration (or content);

ii. whether the stressor leveraged the quantile regression region of the data cloud.

This analysis indicated that many of these stressors did result in apparent marked 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 (see Appendix W).

A second, alternative approach to reducing the potential impact of other stressors on the focal stressor-response relationship was investigated and involved removing all sites/occasions where other potential stressors exceeded a pre-defined threshold, which represented an adverse effect (see Appendix X for definition of those thresholds). A reduced dataset consisting of 401 sample points was then evaluated using quantile regressions. However, due to the considerably reduced sample size and the the general scatter in the data, this analysis was considered unsuitable for turbidity and visual clarity threshold definition. Hence, we decided to conduct a quantile regression analysis on the full NRWQN dataset consisting of 1275 samples collected at 67 sites.

5.4.5 Quantile regression analysis

Data exploration involved visual observation of log-transformed data was used to establish the nature of the biotic relationships with turbidity and clarity. Most of the biotic measures showed the presence of a subsidy/stress 'hump'.

Quantile regression and linear regression was performed using the 'quantreg' package (Koenker 2013) and the Ricker equation (Cade & Guo 2000; Grace *et al.* 2014) as the best form in fitting quantiles to these data. The Ricker equation derives a linear regression on $\log_e$+1 transformed data with the curve being consistent with the subsidy/stress pattern of the stressor-response relationship. 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. The model provides quantitative relationships for any of the quantiles examined (99%, 95%, 90%, 80% and 50%) and an ability to calculate levels of effect relative to specific stressor concentrations or as a change from the maximum measure.

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 estimate 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. We fitted the Ricker quantile relationship to the inverse of visual clarity so that the response shape would be consistent with that of turbidity. The response relationships are shown with visual clarity decreasing to the right on the x-axis, and turbidity increasing to the right on the x-axis.

5.4.6 Defining thresholds and guideline derivation

The macroinvertebrate-derived threshold values for suspended sediment that we considered to be the best approximation of a C/D band threshold (i.e., national bottom-line) uses a SSD model fitting procedure as applied to the ANZECC guidelines derivation procedure (Warne et al. 2015). The two-step process to deriving this 'community threshold' is illustrated in Figure 5-3. The turbidity and visual clarity values corresponding to a 30% reduction from peak are calculated from the 95th percentile quantile regression relationships (left plot, Figure 5-3). These effect values (e.g., turbidity indicated by red dashed arrow of species 'x' in Figure 5-3) from seven individual species and seven macroinvertebrate metrics were plotted as a SSD curve (right plot, Figure 5-3), and an 80th percentile protection level was used to approximate an effects level that was a NOF C/D band threshold for ecosystem health. As such, some sensitive measures may be below the protection concentration value, while other less sensitive measures will be above the protection concentration value (e.g., species 'y' and 'x', respectively – Figure 5-3).

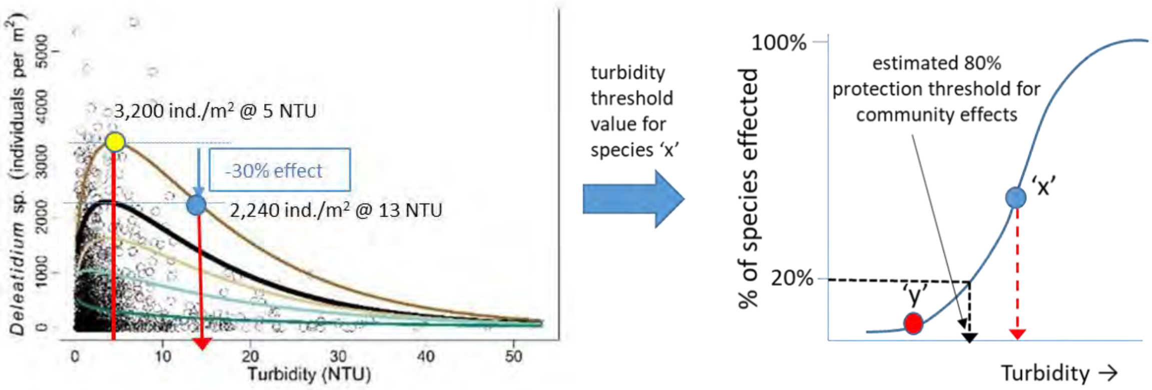


Figure 5-3: Illustration of the 2-step process used to derive community macroinvertebrate thresholds (turbidity and visual clarity) that were used to inform the proposed C/D band threshold for the suspended sediment attribute. The left plot shows how -30% effect levels, which in this example corresponds to a reduction in maximum density of 3,200 to 2,240 individuals per m^2 (left plot, blue and yellow dots, respectively). The 30% reduced density corresponds to a turbidity of around 13 NTU. The right plot shows how these individual thresholds are used to construct a species sensitivity distribution (SSD) curve. An 80 percent protection level was used to derive the macroinvertebrate community threshold. In this example, species 'x' would be protected from adverse effects, but not necessarily species 'y'.

Importantly, the indicative macroinvertebrate thresholds derived in this chapter are not proposed as C/D bands thresholds for suspended sediment. Rather the 80th percentile threshold values contribute to multiple lines of evidence used to develop the proposed C/D band threshold values for (see Chapter 7). The multiple lines of evidence include:

- macroinvertebrate effects threshold values;
- fish effects threshold values;
- literature effects threshold values;
- regulatory guideline values; and
- expert opinion.

5.5 Results

5.5.1 River classification

The macroinvertebrate datasets were visually examined for the distribution of key biotic and stressor measures in relation to stream order and to river environment classification (REC) categories for climate, geology and land use (Scarsbrook *et al.* 2000; Snelder & Biggs 2002b). Some marked differences were observed between factors such as stream order, and with some of the REC classifications (Appendix V). These categorical differences indicate that effect threshold for SS (or its surrogates) could be derived, particularly for specific climates/geologies. However, for this analysis we consider that the generic analysis of a large and inclusive dataset provides the most robust application of this approach for baseline guideline derivation. This was largely consistent with the suspended sediment classification analysis (Chapter 2), which showed that natural state levels of suspended sediment (typically <2.5 NTU) were significantly less than the lowest reported effects thresholds (e.g., 5-7 NTU).

5.5.2 Model fitting

Examples of the 95th percentile Ricker model fits to the species (taxa) richness and mayfly *Deleatidium* abundance relationships with annual median clarity and turbidity values from the NRWQN dataset are shown in Figure 5-4. These log-scale plots show apparent marked changes in biodiversity (i.e., taxa richness) and a key mayfly species (*Deleatidium*) in relation to the clarity and turbidity gradients in New Zealand rivers. The Ricker model shows good bounding of the data and fitting to the subsidy/stress relationship for both the visual clarity and turbidity datasets. Primary criteria for the model acceptability were that it provides both a good fit to bound the data cloud and to fit the peak location.

The quantile regressions relationships for a range of quantiles (99%, 95%, 90%, 80% and 50%) fitted to selected macroinvertebrate metrics and individual species in relation to visual clarity and turbidity are shown in **Error! Reference source not found.** Quantile regression for the 14 macroinvertebrate response measures are provided in Appendix Z. Most variables are fitted with a subsidy/stress Ricker model, except for MCI, QMCI and %EPT which were modelled with log-linear quantile regressions. We consider that the 95th percentile quantile provides a robust relationship for most of the metrics and selected species.

The Ricker and log-linear model coefficients and summary statistics are summarised in Appendix Y. The relationships indicate a high level of statistical significance for many of the biotic measures and provide sufficient confidence of use in the SSD guideline derivation. The Ricker equations for the 95th percentile were used for calculation of the effects based on a 30% reduction from the maximum or the reference condition. A reference condition of 0.5 NTU turbidity and 6 m visual clarity (the 5th and 95th percentile values, respectively, of the NRWQN dataset, Table 0-1) were used for Ricker equation relationships and for linear regression relationships which showed a linear decline and absence of a subsidy/stress peak in diversity or abundance.

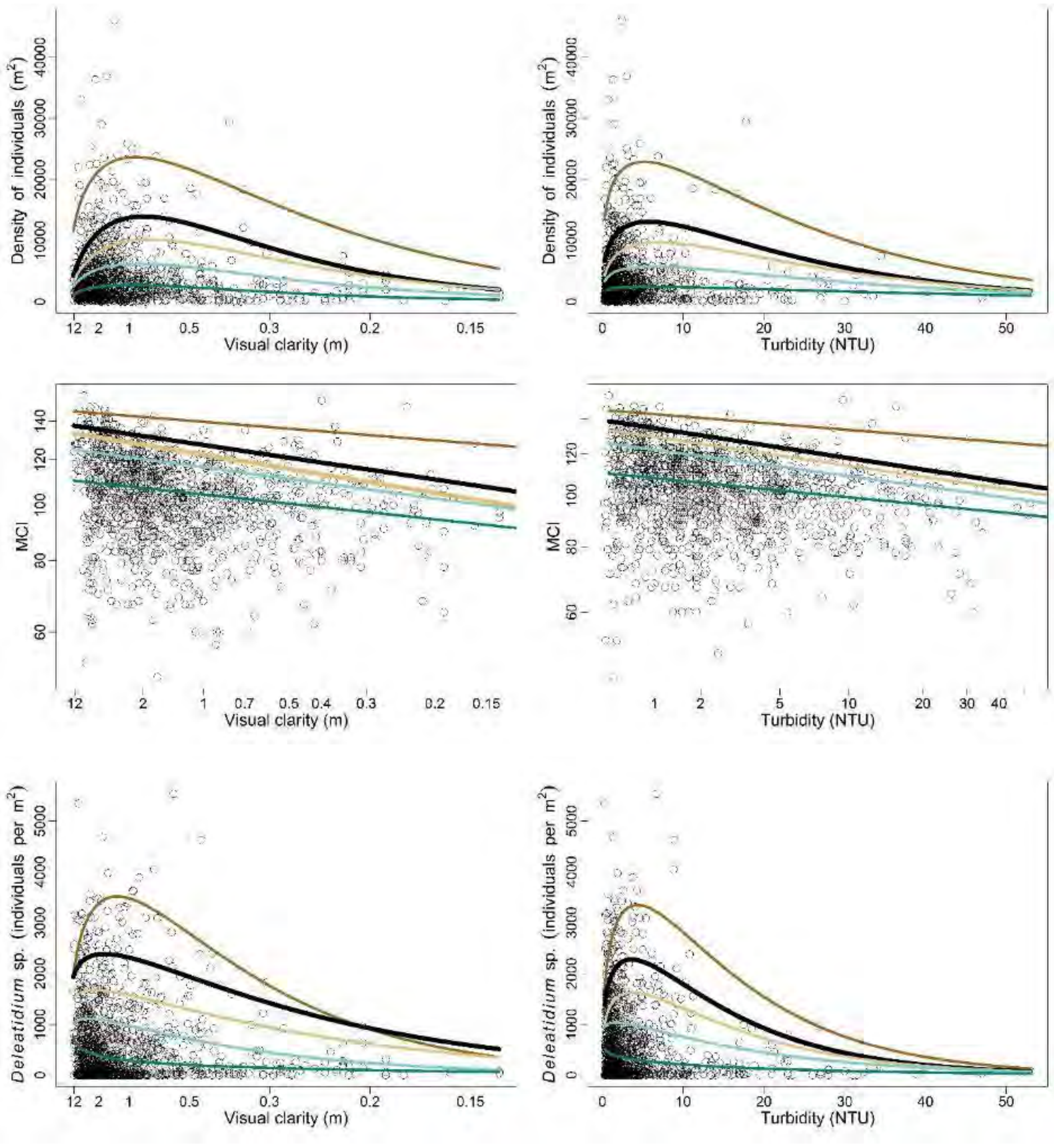


Figure 5-4: Selected examples of quantile relationships fitted to *density of individuals* (top), MCI (middle) and density of individual taxon *Deleatidium* (bottom). Ricker equation fit to quantiles (99% brown, 95% black, 90% light brown, 80% cyan and 50% green shown), except for MCI, which is fitted to log-linear quantile regression. Bold black 95th percentile curve was used for threshold effect calculations. Quantile response (QR) curves for density of individuals (top) and *Deleatidium* (bottom) show pronounced subsidy/stress relationship to sediment, whereas MCI is a relatively insensitive metric in response to increasing suspended sediment (measured as decreasing visual clarity and increasing turbidity). QR relationships for all 14 macroinvertebrate response metrics are provided in Appendix Z.

5.5.3 Indicative thresholds from quantile regression (QR) and species sensitivity distribution (SSD) curves

A summary of the model-derived maximum biota condition for the 95th percentile and the calculated 30% reductions from the maximum value of the metric (y-axis) are shown in Table 5-4 for visual clarity and turbidity, respectively. Further details (i.e., biotic response and associated SS measure for peak and 30% reductions) are provided in Table AA-1 (clarity) and Table AA-2 (turbidity) in Appendix AA.

Table 5-4: Summary of 30% effect thresholds for visual clarity (m) and turbidity (NTU) based on the 95th percentile quantile relationships. The blue highlighted variables are derived from log-linear regressions and a 30% reduction from a high quality biotic condition. ND indicates model fit not suitable for use in effects-based threshold determination. Macroinvertebrate metric response scores (maximum and at 30% reduction) are provided in Appendix AA.

Macroinvertebrate response variable	Visual clarity threshold for 30% reduction (m)	Turbidity threshold for 30% reduction (NTU)
Taxa richness	0.26	17.0
Density	0.33	19.0
MCI	<0.15	>50
QMCI	<0.15	>50
EPT taxa	0.33	8.2
EPT individuals	0.52	12.2
%EPT	<0.15	ND
Deleatidium	0.38	12
Aoteapsyche	0.39	15
Potamopyrgus	0.28	14.8
Pycnocentria	0.71	0.8
Zelandobius	1.2	8.2
Oxyethira	1.4	ND
Zelandoperla	5.2	0.6

The SSD plots for each visual clarity and turbidity (Figures 5-6 and 5-7, respectively) were constructed by bringing into order the 'chronic-effect thresholds' (points at which there was an estimated 30% reduction from the maximum response value of either a metric or a taxon abundance, Table 5-5) and plotting them on the x-axis while spreading the data points equally across 0-100% to visualise the cumulative distribution. Then a curve was fitted to these data points (representing the thresholds) from which a desired 'species' protection level can be selected and then the respective SS attribute threshold or management band threshold be determined. We defined an 80% protection level to best approximate thresholds that were consistent with NPS-FM ecosystem health bottom-lines (i.e., C/D band thresholds. The C band state generally represents a minimum safe level before an ecological tipping point (MfE 2014b).

Visual clarity: estimated threshold value corresponding to 80% protection level (from SSD curve)

For visual clarity (Figure 5-5) indicates that an 80% protection level would be achieved for median visual clarity values greater than around 0.9-1.0 m. For waters with visual clarity values above this (i.e., clearer water) biotic measures would be protected from a 30% effect, with the possible exception of densities of *Zelandoperla*, *Oxyethira* and *Zealandobius* taxa, which require greater clarity for their protection. This biodiversity measures of taxa richness and EPT taxa are impacted at thresholds of 0.26 m and 0.33 m, annual median visual clarity respectively. The metrics MCI, QMCI and % EPT individuals were relatively insensitive indicators (<0.15 m of visual clarity required to result in a 30% reduction, Table AA-1) and were not included in the SSD for threshold calculations.

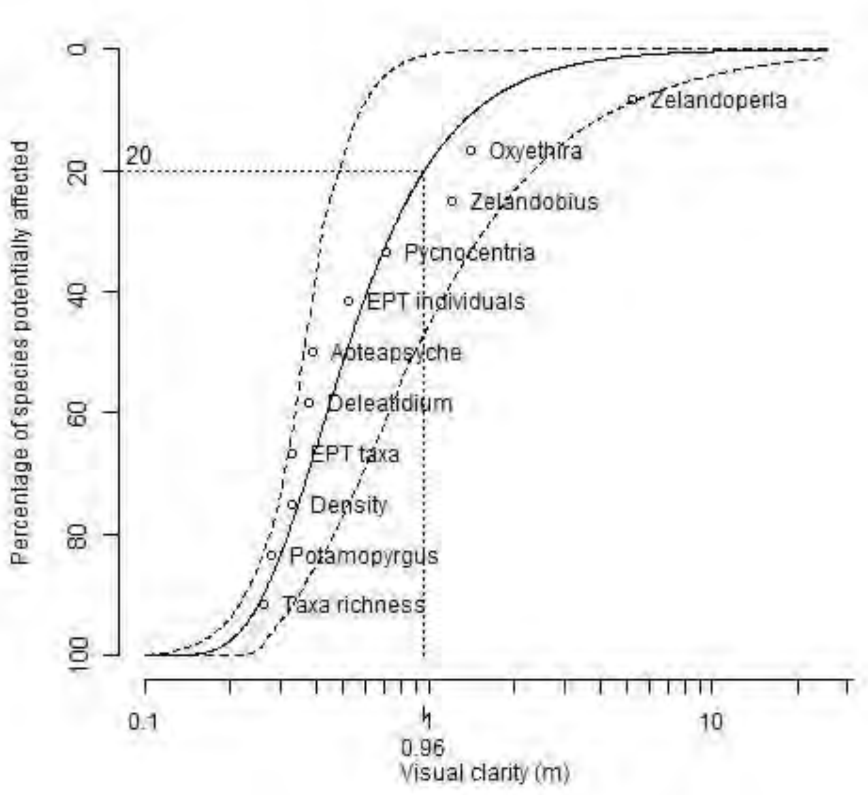


Figure 5-5: Species sensitivity distribution for clarity based on 30% effect on biotic measures using the 95th percentile quantile. Indicative threshold shown is for 80% 'species' level protection, which has been defined as the C/D threshold. BUR type III distribution fitted (Campbell et al. 2000; Warne et al. 2015). Note that the y-axis values have been inverted due to macroinvertebrate indicators increasing with increasing visual clarity.

Turbidity: estimated threshold value corresponding to 80% protection level (from SSD curve)

The turbidity threshold corresponding to 80% protection of test 'species' making up the SSD (based on a 30% effects level) was estimated as 4.3 NTU (Figure 5-6). This was protective for all biotic measures, except for densities of *Pynocentria* and *Zelandoperla* which appear highly sensitive to increases in SS (i.e., 30% effects thresholds of 0.8 and 0.6 NTU, respectively). The biodiversity measures of 'taxa richness' and 'EPT taxa' metrics are impacted at thresholds of 17 and 8 NTU (annual median turbidity), respectively. The SSD response relationships indicate that a wide range of biotic measures (individual taxa and metrics) are adversely affected when the median annual turbidity is in the range 15-20 NTU. The results (Table 5-4, Figure 5-6) show that the routinely used macroinvertebrate index MCI and the quantitative QMCI, together with %EPT individuals are relatively insensitive to elevated annual median turbidity. The results of the analysis indicate that annual median turbidity values of >50 NTU are required for a 30% reduction of MCI and QMCI metric scores Table AA-2).

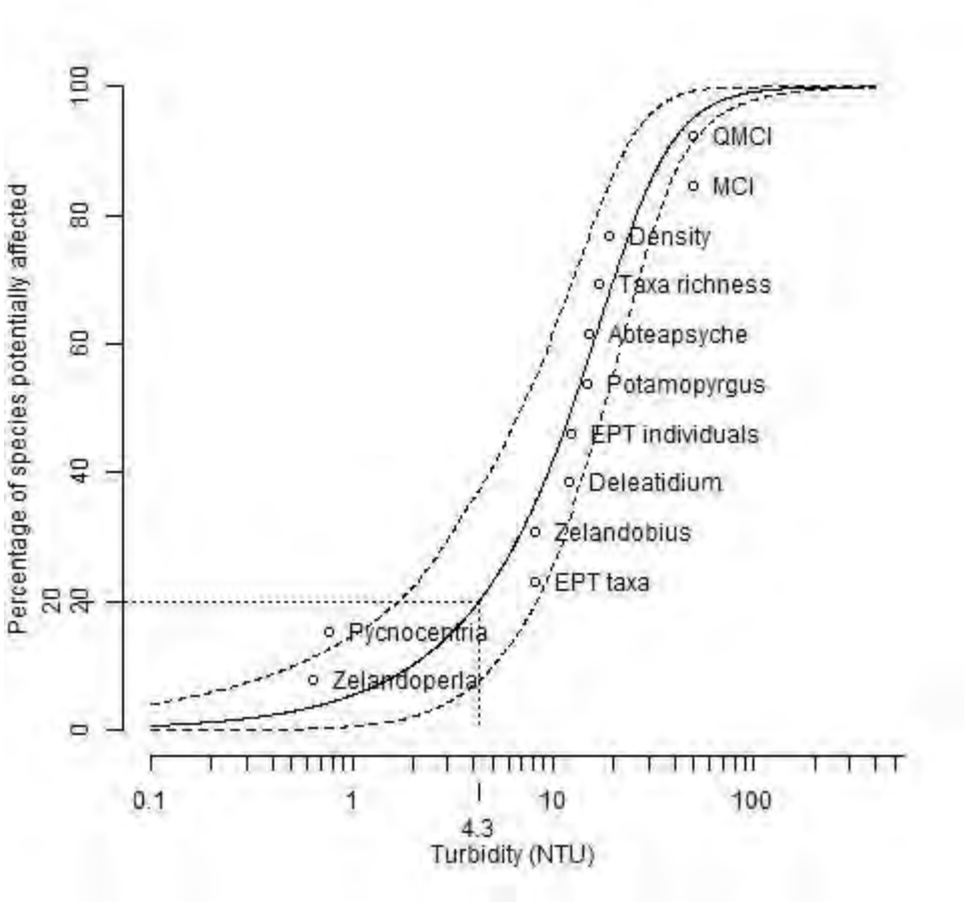


Figure 5-6: Species sensitivity distribution for turbidity based on 30% effect on biotic measures using the 95th percentile quantile. Indicative threshold shown is for 80% 'species' level protection, which has been defined as the C/D threshold. BUR type III distribution fitted (Campbell et al. 2000; Warne et al. 2015).

5.6 Discussion

5.6.1 Defining thresholds from quantile regression and SSD curves

The chosen analytical approach for the macroinvertebrate-suspended sediment (SS) ESV responses reflected both the availability of suitable national data which was 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 of other potential stressors which could confound establishing a causative relationship with water column SS (or its surrogates turbidity and black disk visual clarity). The nature of water column SS in rivers is a combination of particulate organic matter (POM) and inorganic SS – 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.

We emphasise that the quantile regression approach used for macroinvertebrate SS ESVs differs from that used for macroinvertebrate deposited sediment ESVs and fish ESVs – with these latter two ESVs being based on SS measures which do not result in subsidy/stress response relationships and were based on different databases for their derivation. The quantile regression approach was found to be insensitive when applied to deposited sediment gradients.

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. For this generic analysis, we consider that our inclusion of seven key biotic community measures and seven representative species based on their expected response to POM and/or inorganic SS provides a robust method for applying the quantile regression approach to derive water column SS guidelines in New Zealand streams and rivers. Our choice of a 30% reduction in the macroinvertebrate measures (e.g., individual abundances and metrics) for input to the species sensitivity distribution (SSD) 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 recommend use of the 30% effect measure as a pragmatic measure for this type of quantile analysis.

The numeric response thresholds for the seven macroinvertebrate metrics and individual species/taxa provided the basis of the SSD. Our selection of the 80% protection threshold from the SSD as being presentative of ecosystem health 'bottom line' (i.e., attribute band C/D transition) values is consistent with the approach used in the NPS-FM for other toxicants (e.g., ammoniacal-N, nitrate-N). This SSD-based approach may be used for derivation of higher protection thresholds; however, we consider that additional species should be included prior to the derivation of the other threshold bands to increase the confidence level around the SSD model – especially the 'tails' – providing greater information on species which might be susceptible to elevated SS conditions. An important caveat here, however, is that before defining thresholds for higher protection bands (i.e A/B and B/C thresholds) it will be necessary to revisit the suspended sediment classification system as natural state variation will overlaps with these band thresholds (certainly the A/B band).

5.6.2 Importance of understanding non-suspended sediment -related stressors

Several stressors were identified as having a potential effect on river macroinvertebrate populations based on a visual analysis of response relationships with the stressor concentration – or habitat composition change. These included: sand composition of the substrate and periphyton abundance (measured as weighted composite cover (WCC), (Matheson *et al.* 2015)) (plots for taxa richness, number of EPT taxa and *Deleatidium* abundance shown in Appendix W). The variables: sand; periphyton abundance; conductivity and water colour all showed a progressive decline with increasing concentration (or composition of sand in the sediments). A tolerance plateau followed by a decline was shown for pH; while temperature showed a subsidy/stress response, peaking at about 15°C. Average site water velocity showed minimal apparent response for taxa richness and EPT species, but a peaked subsidy/stress effect for *Deleatidium*. This analysis indicated that a range of biotic responses are likely to be influencing the biotic measures in river macroinvertebrate communities – and that many will be non-linear in nature and differ markedly in their mode of action.

The presence of other natural stressors may constitute exceptions to a generic application of a guideline for SS (i.e., turbidity and/or visual clarity). These would include water column stressors, such as highly coloured waters and glacial flour, and bed related stressors, such as sand fraction, periphyton abundance and iron flocs. The final implementation of a guidelines for SS could also include more specific classifications for specific environments where macroinvertebrate communities are naturally limited by other environmental constraints (e.g., regional temperature, flow variability).

5.6.3 Turbidity versus visual clarity and the statistical measure

The correlation between the annual median turbidity and visual clarity was high (r^2 = 0.88). Using this regression relationship gives a close predictive relationship between the guideline values (e.g., a turbidity value of 4.3 NTU corresponds to a predicted visual clarity of 0.9 m). Thus regression-derived relationships between turbidity and visual clarity and vice versa could be used for general comparisons between the stressor measures.

5.6.4 Summary of advantages and limitations of the quantile regression approach used

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 for input into SSD relationships for calculation of macroinvertebrate community protection 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.

The 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 in order 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 in order 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 20% community change level for the ‘bottom line’ (i.e., C/D attribute band transition) is arbitrary – although this is consistent with other derivations used in the NPS-FM for toxicants.

5.7 Macroinvertebrate-based suspended sediment thresholds consistent with NPS-FM ecosystem health 'bottom-line'

For ecosystem health, the draft guidance document (MfE 2014) defined the C-band as a state that 'generally represents a minimum safe level before an ecological tipping point'. Accordingly, our focus for deriving relevant macroinvertebrate thresholds for suspended sediment was to define 'effect levels' that were broadly consistent with the C-band narrative.

Our choice of a 30% reduction in the macroinvertebrate measures (e.g., individual abundances and metrics) for input to the species sensitivity distribution (SSD) is based on this being a substantial reduction. Our selection of the 80% protection threshold from the SSD as being presentative of ecosystem health 'bottom line' (i.e., attribute band C/D transition) values is consistent with the approach used in the NPS-FM for other toxicants, for example ammoniacal-N and nitrate-N.

Table 5-5: Suspended sediment thresholds for different levels of protection defined based on the species sensitivity distribution (SSD) for visual clarity (black disc) and turbidity expressed as annual medians. See text for more details.

Level of protection for macroinvertebrate communities	Visual clarity (m) (±95% CI)	Turbidity (NTU) (±95% CI)
99%	5.5 (0.9->12)	0.2 (<0.1-1.3)
95%	2.2 (0.7-8.2)	0.9 (0.14-8.0)
90%	1.5 (0.6-4.7)	2.0 (0.43-5.4)
80%	**1.0 (0.5-2.4)**	**4.3 (1.4-8.1)**

The macroinvertebrate-based thresholds are summarised in **Error! Reference source not found.**, i ncluding 95% confidence intervals (CI) estimates from the SSD analysis.[15] These threshold values are considered along with other lines of evidence (including fish-based threshold, literature, regulatory guidelines and expert opinion) in the synthesis present in Chapter 7, which presents proposed national bottom-line (C/D band) thresholds for the suspended sediment attribute. multiple lines of evidence approach discussed in.

While higher levels (i.e., A/B and B/C thresholds) of protection could, in theory, be based on the calculated higher protection thresholds derived from the SSD relationship as summarised in Table 5-5. However, further work would be required to enable feasible higher protection threshold to be derived, given that the 95th and 99th (and possibly even the 90th) threshold value overlap with natural (or minimally disturbed condition) sites. The certainty around these higher protection thresholds might be improved with the inclusion of a larger number of species in the SSD and validation compared with a larger dataset. This would require additional effort to amalgamate the larger dataset, refine the quantile regressions, expand the number of relevant macroinvertebrate species, refine/justify operationally-define limits applied to the regressions and optimise the SSD models to minimise uncertainties. For now, we only recommend the use of the 80th percentile protection value for consideration of threshold development for the suspended sediment attribute.

[15] The relatively wide CI intervals for the turbidity values represent the uncertainty in the modelled relationship for the two sensitive species (*Pynocentria* sp. and *Zelandoperla* sp.).

5.8 Future work

Possible future work to improve suspended sediment effects threshold derived for macroinvertebrate communities are provide below. It is emphasised that macroinvertebrate thresholds do not define the proposed C/D band thresholds, they contribute to a *multiple strands of evidence approach*, which considers derived fish thresholds, literature threshold values, regulatory guideline values and expert opinion:

- Derive additional species response data for inclusion in the SSD relationship for threshold determination. This would include species proposed as "*decreasers*" based on the deposited sediment analysis.
- Develop a robust method for determining higher protection attribute thresholds (i.e., A/B, B/C band) and stressor elimination based on the quantile regression and SSD approach. It is emphasised, however, that this would require a more comprehensive classification system to be developed as these thresholds would overlap with natural state levels of suspended sediment.
- Evaluate the 80th percentile turbidity (20th percentile clarity) as a response measure in streams (compared with the median values used in this analysis);
- Test the river environment classification classes relative to 80th percentile turbidity (20th percentile clarity) values for relationships with macroinvertebrate responses. This work is based on the observation that high SS events occur in many river classes and the hypothesis that macroinvertebrate communities may be limited by those events.
- Incorporate other major regional datasets with suitable data for quantile relationships.
- Undertake laboratory/field chronic testing for key species and mesocosm testing for macroinvertebrate community responses to definitively establish causative relationships. Testing should include POM and inorganic SS; suitable long duration for long-term effects (i.e., 30-40 d); have an experimental treatment scale that includes key low density functional groups (e.g., predators); and be designed to distinguish settled and suspended sediment-related effects.

6 Derivation of deposited and suspended sediment thresholds based on fish responses

6.1 Chapter summary

This chapter reports on the response of fish communities to increasing deposited and suspended fine sediment in rivers and streams from which to derive scientifically sound and justifiable thresholds.

The chosen analytical approach for characterising fish-sediment ESV responses reflected both the availability of suitable data, and the ecology and biology of NZ's freshwater fishes. In contrast to macroinvertebrates, freshwater fish data are not typically collected at standard SoE monitoring sites. Consequently, there were few fish observations that could be paired in both space and time with SoE data on sediment ESVs. Standardised abundance data are also rare for freshwater fish because of the wide variety of methods used to collect data. These factors meant that it was inappropriate to directly apply the analytical methods used for macroinvertebrates to fish.

Fish communities are also strongly influenced by landscape setting (e.g., distance inland). This is partly because many New Zealand native species spend some part of their life-cycle at sea, and exhibit different abilities to penetrate inland. Consequently, it was necessary to apply methods that could account for these influences to increase the chances of being able to detect the 'real' effects of elevated sediment on fish communities. Landscape setting was incorporated into the fish-sediment ESV model by including the first two levels of the REC as 'random effects', which encapsulated many aspects of hydrology, geomorphology and climate that may influence fish presence.

The statistical analyses were performed on data available through the New Zealand Freshwater Fish Database (NZFFD). Presence-absence data in the NZFFD was paired with data of deposited sediment, and predicted long-term median visual clarity and turbidity values. Eleven fish species were selected for analysis based on abundance and potential sensitivity to sediment gradients

Fish probability of capture (FPC) for each of the 11 fishes was modelled,[16] using a generalised linear mixed model (GLMM). The FPC model provides an estimate of the probability of capturing a species in a given landscape setting at a given sediment ESV value. Predicted FPC responses to sediment ESVs for individual fishes were translated into a metric (i.e., ΔC) that describes the overall expected change in fish community relative to the community that might be expected at reference ESV state. The greater the reduction in the ΔC metric from reference, the greater the risk to fish community integrity.

A 20% decline in fish community integrity (-20% ΔC) was defined to be consisted with ecosystem health bottom-line values (i.e., C/D band threshold). Across landscape settings, these predicted threshold values (-20% ΔC) fall into the ranges of: 30-60% for deposited sediment; 2-3 NTU for turbidity in cool climate classes; 5-7 NTU for turbidity in warm climate classes; and 0.8-2.0 m for visual clarity. These 20% community change thresholds are set such that they provide a level of protection at an overall fish community level and may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations.

These threshold values (-20% ΔC) for fish community change were used we considered along with other lines of evidence for proposing NOF C/D band thresholds for sediment (see Chapter 7).

[16] based on presence-absence data from the NZFFD

6.2 Introduction

6.2.1 Rationale

Elevated fine sediment inputs are widely acknowledged to impact negatively on freshwater fish and a range of causal mechanisms are known to underpin the exhibited responses (Ryan 1991b; Kemp *et al.* 2011b). Fish communities are a fundamental and highly valued components of healthy ecosystems in New Zealand and worldwide. Evaluating the impact of elevated fine sediment inputs on the state of freshwater fish communities is, therefore, essential for determining appropriate protection levels for managing ecosystem health.

6.2.2 Scope

This section of the report evaluates and proposes possible protection levels for both deposited and suspended sediment environmental state variables (ESVs) for freshwater fish in New Zealand. There are two main components of this evaluation:

1. a review of literature from both New Zealand and overseas with the purpose of identifying documented biological effects levels relevant to managing the impact of sediment ESVs on freshwater fish species, and

2. analyses of fish versus sediment ESV relationships based on existing datasets of fish and sediment from New Zealand.

The information gained from these two components is subsequently used to inform recommendations on the development of sediment NOF attribute bands suitable for protecting freshwater fish communities.

6.3 Literature review

6.3.1 Introduction

Sediment plays a pivotal role in determining the biological integrity of fish communities (Ryan 1991b; Bilotta & Brazier 2008; Kemp *et al.* 2011b). 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 & Brazier 2008; Collins *et al.* 2011).

Several comprehensive reviews have been published within New Zealand detailing the effects of sediment in aquatic systems (Ryan 1991b; Crowe & Hay 2004; Reid & Quinn 2011b; Cavanagh *et al.* 2014; Davies-Colley *et al.* 2015c). 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, kōura (*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 & Brazier 2008; Clapcott *et al.* 2011a; 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.

6.3.2 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 1991b; Kemp *et al.* 2011b). 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 and Boustead 2001). These impacts have been well documented overseas, but have received limited attention in New Zealand (Newcombe & Macdonald 1991; Wood & Armitage 1997a; Bilotta & Brazier 2008; Kemp *et al.* 2011b). A summary of the key findings relevant to New Zealand is provided below. For more details of the individual studies, see Table BB-1 in Appendix BB.

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 & Boustead 2001), redfin bully (*Gobiomorphus huttoni*) (McEwan & Joy 2014a) 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 & Joy 2014b), shortjaw kōkopu (*Galaxias postvectis*) (McEwan & Joy 2014b), dwarf galaxias (*Galaxias divergens*) (Jowett et al. 1996) and kōura (Usio & 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 & Persson 1988; Wood & Armitage 1997a; Kemp *et al.* 2011b; Louhi *et al.* 2011) and may lead to premature hatching (Olsson and Persson 1998). 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 and Persson (1998) 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. (2005) 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.* 2011a) 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.* 2011a). 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 inānga (*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 1991b; Kemp *et al.* 2011b). 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 & 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).

6.3.3 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 CC-1 in Appendix CC.

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 1995b), and even mortality due to suffocation or stress (Ryan 1991b; Wood & Armitage 1997a). 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 & 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 LC50) 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 inānga were around 100% up to a turbidity of 1000 NTU, but declined with increasing turbidity above this level. LC50 thresholds were around 3,050 NTU for smelt and 20,000 NTU for inānga, with 100% mortality at around 15,000 NTU and 30,000 NTU for smelt and inānga 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 1,000 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 & Meyer 2007; Cumming & Herbert 2016) and physiological stress (Herbert & 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.* 2011b).

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 & Dean 1998; Harvey & 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 & 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 & 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 0-160 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 & 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 & Armitage 1997a; Kemp *et al.* 2011b). 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 1991b). 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, 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, 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.

6.3.4 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 & Brazier 2008; Collins *et al.* 2011; Kemp *et al.* 2011b; 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 6-1. Sediment impacts occur primarily at a sub-lethal level for most life stages through changes in behaviour, food availability and habitat quality and quantity.

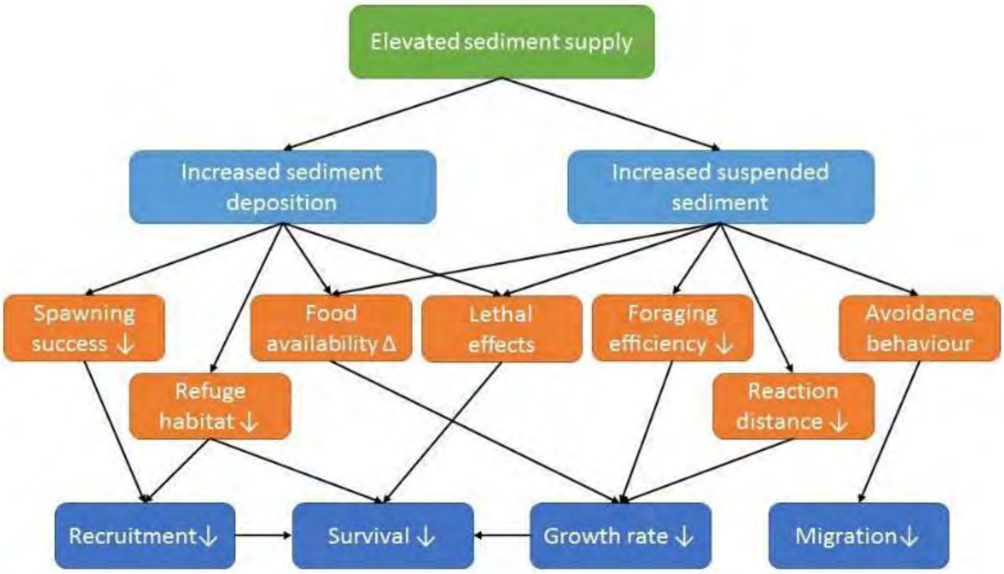


Figure 6-1: Summary of key mechanisms governing impacts of elevated sediments on freshwater fish.

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 6-1). 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 6-1: 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	

6.4 Analyses of fish responses to deposited and suspended sediment s

6.4.1 Introduction

The objective of this component of the project was to test for, and characterise, relationships between fish and sediment (deposited and suspended) that could be used to inform the development of suitable sediment thresholds. Existing datasets on fish and sediment were utilised for these analyses. The main text of this report focuses on derivation of effects thresholds that are consistent with NOF bottom-line threshold values for ecosystem health. For attribute relating to ecosystem health the C-band state generally represents a minimum safe level before an ecological tipping point (MfE 2014). However, by defining a smaller effects level, values relevant to setting to setting A/B and B/C thresholds were also determined. This information is presented in the technical appendices (Appendix DD and Appendix EE) along with details of the methods used to derive the proposed C/D thresholds. It is noted, however, that band threshold value of sediment ESVs will be closer to (or overlap) with the natural variation at reference sites (measured or predicted). As such, to be meaningful, lower thresholds corresponding to A/B and B/C NOF bands will require more comprehensive sediment classification systems or greater certainty regarding predicted reference state (i.e., if threshold change relative to reference state were advocated). .

6.4.2 Explanation of analytical approach

The chosen analytical approach for characterising fish-sediment ESV responses reflected both the availability of suitable data, and the specific ecology and biology of New Zealand's freshwater fishes. In contrast to macroinvertebrates, freshwater fish data are not typically collected at standard state of the environment (SOE) monitoring sites. Consequently, there were very few fish observations that could be paired in both space and time with SOE data on sediment ESVs. Standardised abundance data are also rare for freshwater fish because of the wide variety of methods used to collect data required to meet different survey objectives. To use abundance data for fish would rely on standardisation for fishing effort. This was not possible due to a lack of consistent data on fishing effort, and data being derived from various fishing methods (electric fishing, netting, etc). As a result, the fish analyses had to be based on presence/absence data, rather than abundance data. These two factors meant that it was inappropriate to directly apply the analytical methods used for macroinvertebrates to fish.

Fish communities are also strongly influenced by landscape setting (e.g., distance inland). This is partly because many New Zealand native species spend some part of their life-cycle at sea, and exhibit different abilities to penetrate inland. Consequently, it was necessary to apply methods that could account for these influences. This was for two reasons. First, there are many locations (e.g., more inland) were we would not expect to find certain species regardless of sediment conditions. Second, by accounting for the influence of landscape setting we expect to increase the chances of being able to detect the 'real' effects of elevated sediment on fish communities.

To maintain consistency with the model structure for the predicted sediment ESV reference state, landscape setting was incorporated into the fish-sediment ESV model by including the first two levels of the REC as random effects. It was assumed that these random-effects encapsulated many aspects of hydrology, geomorphology and climate that may influence fish presence through effects from more proximate variables such as temperature, hydraulic conditions, food supply and natural barriers to migration such as steep slopes in hill and mountain settings. This contrasts with including these parameters as multiple individual explanatory variables in the model.

Co-variance in explanatory variables may be a concern, particularly with multiple explanatory variables. This is because some variation in ecological state that should be attributed to the ESV could be attributed to another landscape variable because of a correlation between that landscape variable and the ESV. No evidence was found to support differences in slopes of fish-sediment ESV relationships between landscape classes. A universal relationship between sediment ESV and fish probability of capture (FPC) was therefore modelled, whilst simultaneously accounting for the influence of climate and topography (which we took to represent differences in temperature, physical habitat, etc) on the overall level of FPC. This produced a fish-sediment ESV relationship at the national-level, climate-level, and topography-within-climate level. Thus, the fish-sediment ESV model could be amalgamated with the sediment ESV reference model.

6.4.3 Methods

The main steps involved in this stage of the project for fish were to:

1. Determine the availability of suitable datasets;
2. Derive reference state for the sediment ESVs as a function of landcover;
3. Model fish probability of capture as a function of sediment ESVs within landscape settings;
4. Evaluate fish community change in response to deviation of ESV state from reference conditions; and
5. Derive potential sediment ESV thresholds.

The key elements of each of these steps are summarised below. Full technical details of each step are provided in Appendix DD.

Data availability

Fish

The New Zealand Freshwater Fish Database (NZFFD) contains 42,154 unique observations of fish from across the country. The methods applied by (Crow *et al.* 2016) and (Crow *et al.* 2014) were applied 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 national river network version 1), 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 (visual, netting, trapping, combinations of methods and electric fishing). A total of 34,364 NZFFD records remained for 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 observed conditions. Analyses were, therefore, carried out using presence-absence data.

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. These data were used to investigate the relationships between fish presence-absence and deposited fine sediment.

Suspended sediment data

For the purposes of this project we utilised the long-term site median suspended sediment ESV values derived from the dataset compiled by (Unwin & Larned 2013). This consisted of data assembled from regional council State of the Environment monitoring and NIWA's National River Water Quality Network programme. Site medians for visual clarity were available for 722 sites and site medians for turbidity were available for 833 sites. All sites had at least one full year of data (monthly spot samples). Note, that turbidity is not technically a suspended sediment ESV, however, for simplicity, both visual clarity (an ESV) and turbidity (a suspended sediment measure) are often referred to as suspended sediment ESVs (or simply ESVs).

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 2002a).

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 section 0).

Matching fish data with observed suspended and deposited sediment data

To evaluate sediment ESV – fish responses it was necessary to try to pair sediment ESV observations with fish observations by matching up where and when they were taken. In the case of deposited sediment, it was decided to use the deposited sediment data (% cover of fine sediment) associated with the NZFFD records. These paired observations were ideal for this purpose. For suspended sediment data, NZ reach numbers for observed ESVs and NZFFD records were compared to identify any spatial matches (i.e., places where both sediment and fish data were available from the same place). Several matches between independent ESV observations and NZFFD records on the same reach, but on different dates, were found (Turbidity = 133, TSS = 143, Clarity = 158). Very few of these cases also matched in time and, in most cases, the fish observations and independent ESV observations taken from the same site were taken more than 5 years apart. Following data checks, it was concluded that it was not appropriate to pair the fish and deposited sediment ESV observations from the same site when they were so far apart in time (e.g., >5 years).

To advance the analyses, modelled median visual clarity and turbidity derived by (Unwin & Larned 2013) 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.

Deriving reference state for suspended and deposited sediment as a function of landcover

Replicating the method of (McDowell *et al.* 2013), the response of each sediment ESV was statistically modelled by applying regression models to relate each ESV to the proportion of the upstream catchment covered by artificial landcover. (McDowell *et al.* 2013) used heavy pasture, but we also added exotic vegetation and urban landcovers to the analysis. The regression models were applied to also take account of different landscape settings (REC climate and topography classes). For visual clarity and turbidity, the observed median visual clarity and turbidity data from (Unwin & Larned 2013) were used to fit the statistical models. In the case of % sediment cover, the observations from the NZFFD records were used to fit the models.

Median visual clarity and median turbidity were each modelled by applying a generalised additive mixed model. Percent sediment cover was modelled using a generalised linear mixed model with a binomial error distribution as is appropriate for proportion data. Crossed random-effects were applied in all cases.

The following models were selected as the most appropriate for predicting sediment ESV reference state. Heavy pasture and exotic vegetation are modelled with smoothers (s) in the visual clarity and turbidity models:

$$\text{Log}_{10}(\text{Clarity}) \sim s(\text{heavy pasture}) + s(\text{exotic veg}) + \text{urban} + \text{heavy pasture}|\text{climate} + 1|\text{climate/topography} \quad (1)$$

$$\mathrm{Log_{10}(Turbidity)} \sim \mathrm{s(heavy\ pasture) + s(exotic\ veg) + urban + heavy\ pasture|climate + 1|climate/topography} \quad (2)$$

$$\mathrm{\%\ fine\ sediment\ cover} \sim \mathrm{heavy\ pasture + exotic\ veg + urban + network\ position + heavy\ pasture|climate + 1|climate/topography} \quad (3)$$

Reference states for suspended and deposited sediment were defined by setting the values of heavy pasture, exotic vegetation and urban landcover in each model to zero. Using the same approach as (McDowell *et al.* 2013), the intercept on the y-axis of each statistical model under these conditions was used to estimate the average expected sediment ESV value under natural landcover within each landscape setting (Figure 6-2).

Confidence in the ability of the models to distinguish differences in sediment ESV conditions between landscape setting (topography and climate classes) was quantified using empirical Bayes estimates of the standard errors on these group-level terms.

For full details of the model derivation and selection process, see Appendix DD.

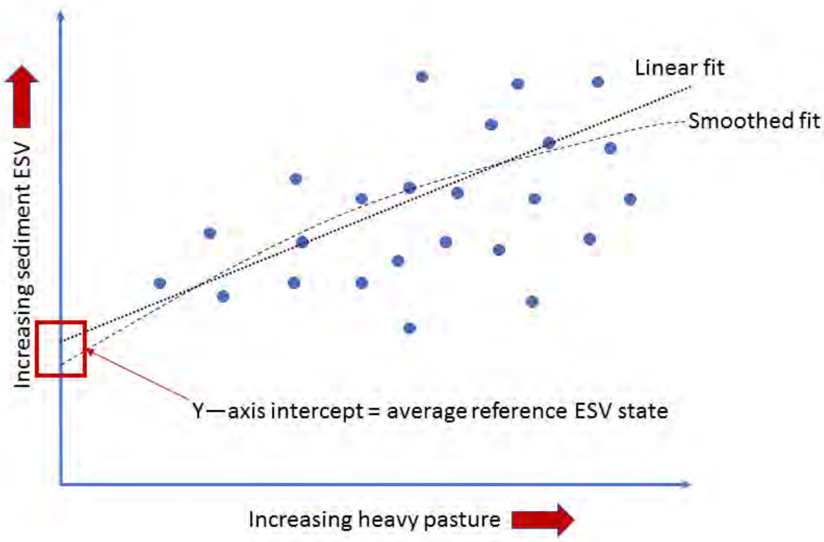


Figure 6-2: Simplified example of how average reference sediment ESV state is derived from the regression model for a given landscape setting. The location where the line fitted through the data crosses with the y-axis equates to the average predicted value of the sediment ESV under conditions of zero heavy pasture (i.e., reference state).

Fish probability of capture as a function of suspended and deposited sediment within landscape settings

Eleven species were selected for this analysis. These species were included in the analysis because each was found across New Zealand and was present in a reasonable proportion of samples in the NZFFD (at least 7%). Ten species were natives. Despite not being a native species, brown trout was also included in the analysis due to the strong likelihood of this species showing a response to the ESVs and because of its high recreation value. 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 6-3).

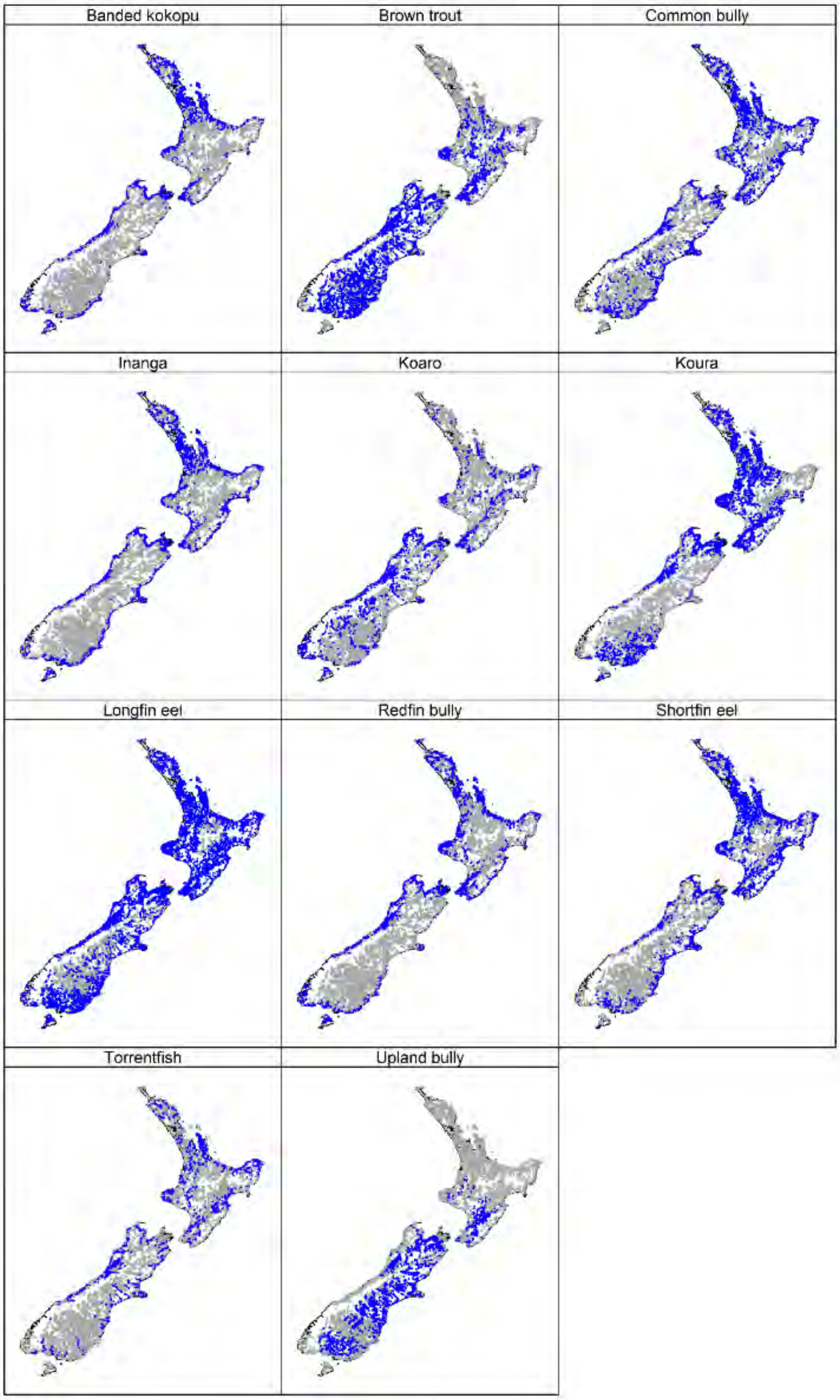


Figure 6-3: Maps of presence (blue) and absence (grey) in the NZFFD records for the eleven species used in these analyses.

Fish probability of capture (FPC) for each of the 11 species was statistically modelled using a regression model as a function of each sediment ESV (Figure 6-4). All species were modelled individually, for each separate sediment ESV, within different landscape settings (e.g., climate, topography, network position, distance inland). These landscape setting incorporated variables considered important for describing expected fish distributions due to influences on factors such as habitat types, flow regime, and migration ability.

The following model was selected as the most appropriate for describing the response of fish probability of capture to changes in sediment ESVs:

$$FPC \sim ESV + \text{fishing method} + \text{distance to sea} + \text{network position} + 1|\text{Climate/topography} \quad (4)$$

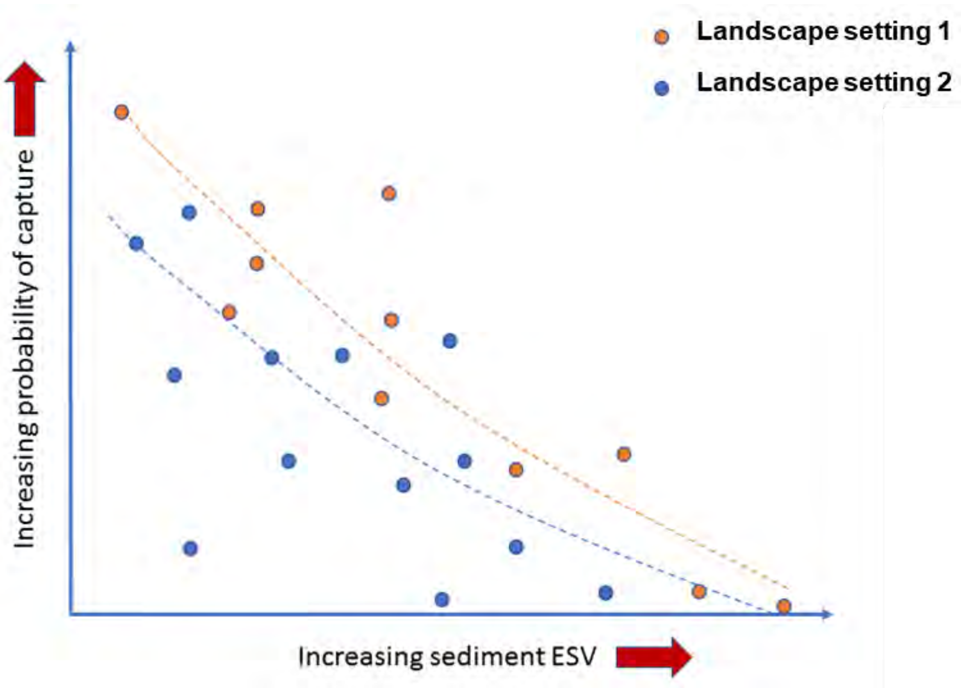


Figure 6-4: Simplified example of how variations in fish FPC with increasing sediment ESV are modelled across different landscape settings. The fitted curves derived from this step of the analysis (Equation (4)) are subsequently used to develop the metrics of community change. Settings refer to landscape settings (e.g., Warm-wet hill).

The FPC model (Equation (4)) provides an estimate of the probability of capturing a species in a particular fish setting (climate/topography/network position/distance inland) at a given sediment ESV value. These probabilities can be translated to expected 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.

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. The FPC threshold at which Cohen's kappa was maximised (maxKappa) was calculated for each species (Figure 6-5). In effect, if FPC > maxKappa the species is more likely present than absent, and if FPC < maxKappa the species is more likely absent than present.

For full details of the model derivation and selection process, see Appendix DD.

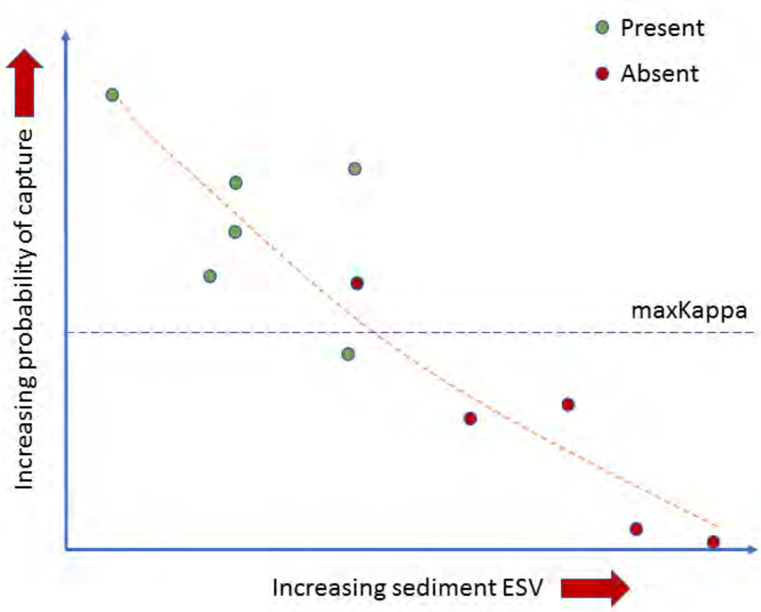


Figure 6-5: Illustration of how maxKappa is derived relative to the observed fish data (presence-absence) and the FPC for a species. In effect, when FPC > maxKappa (above the purple dashed line) a species is most likely to be predicted as present. When FPC < maxKappa (below the purple dashed line) a 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 fish community change resulting from suspended and deposited sediment state

Several steps were required to translate the predicted FPC sediment ESV responses for individual species into a metric of expected fish community change at different suspended and deposited sediment states (Figure 6-6). In simple terms this first involved determining the FPC at reference sediment ESV state and an array of different sediment ESV states for each individual species in each fish setting. These values were then combined into a metric (ΔC) describing the overall expected change in fish community relative to the community that might be expected at the reference state condition for suspended and deposited sediment.

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

The methodology applied required that each landscape setting applied in the fish regression models should have a predicted reference state for each sediment ESV. It was beneficial to apply the same modelling structure to predicting ESV reference states to that for modelling fish presence/absence for this reason.

A more detailed explanation of how ΔC is derived is provided in Appendix DD.

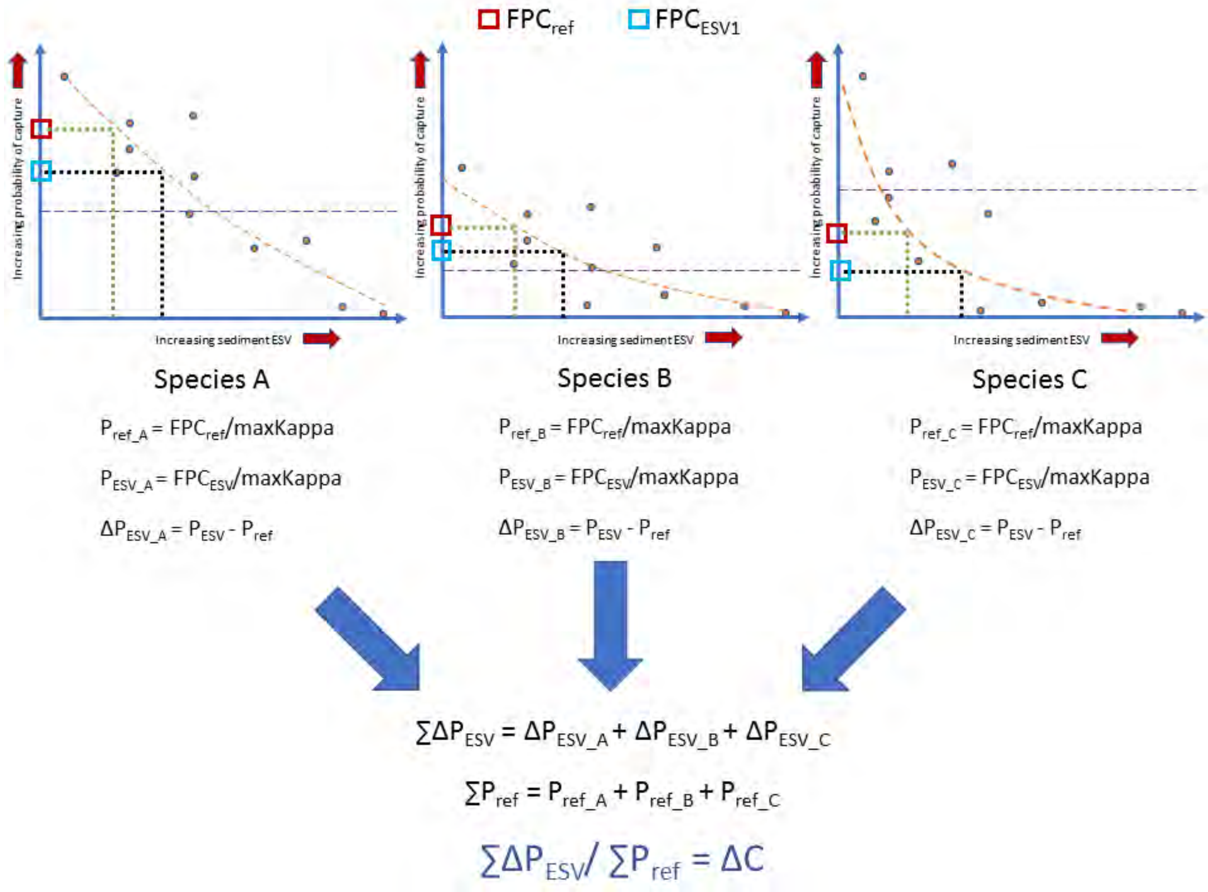


Figure 6-6: Illustration of how ΔC is derived from the FPCs for each species for different sediment ESV states. The FPC at reference ESV state (green dashed line links the predicted reference ESV, for zero heavy pasture, with the corresponding predicted reference FPC) is first derived for each species. FPC 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).

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 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 fish community integrity can be used to derive potential A/B and B/C band thresholds. In the main report, only information on the C/D bottom-line values are reported, but potential A/B and B/C bands are enumerated in the technical appendices (see Appendix EE).

6.4.4 Results

Predicted sediment ESV reference state

The average reference state condition of each ESV was determined by setting the % cover of heavy pasture in the catchment to zero in each reference model (Equations (1), (2) and (3) for visual clarity, turbidity and % sediment cover, respectively). It was established that % cover of heavy pasture had a much greater influence on sediment ESV state than exotic forest and thus it was valid to use heavy pasture as the pressure controlling the sediment ESVs (see Figure EE-1, Figure EE-2 and Figure EE-3 in Appendix EE). Predictions of reference state for each sediment ESV were determined for each REC climate/topography class. For % sediment cover, predictions of reference state were also made for different groups of aggregated stream orders (Low-order = stream order 1-2; Medium-order = 3-4; High-order = 5-8) within each climate/topography class. The predicted reference states for each of the sediment ESVs are shown in Figure 6-7, Figure 6-8 and Figure 6-9 for total fines, visual clarity and turbidity, respectively. For more detailed results of the reference state modelling see Appendix EE.

Inspection of AIC values and Empirical Bayes estimates of the standard errors on these group-level terms both demonstrated that inclusion of the second level of the REC (topography within climate classes) with the reference state models was justified. See Appendix EE for further details.

Predicted average reference state for deposited sediment (% sediment cover) varies from around 5% cover in high-order warm-wet hill landscape settings to around 60% cover in low-order cool-wet lake-fed landscapes settings. In general, lake-fed and lowland settings are predicted to have higher sediment cover at reference state than hill and mountain settings and low-order streams are predicted to have higher sediment cover at reference state than high-order streams. For most landscape settings, predicted average reference state is <30% sediment cover.

Predicted average reference state for visual clarity ranges from around 1.4 m in warm-wet lake-fed settings to 3.9 m in cool-extremely wet hill landscape settings. For turbidity, predicted average reference state varies from 0.6 NTU in the cool-extremely wet hill setting to around 3 NTU in the warm-wet lake-fed setting. Reference state turbidity levels are generally lower for cool climate settings compared to warm climate settings.

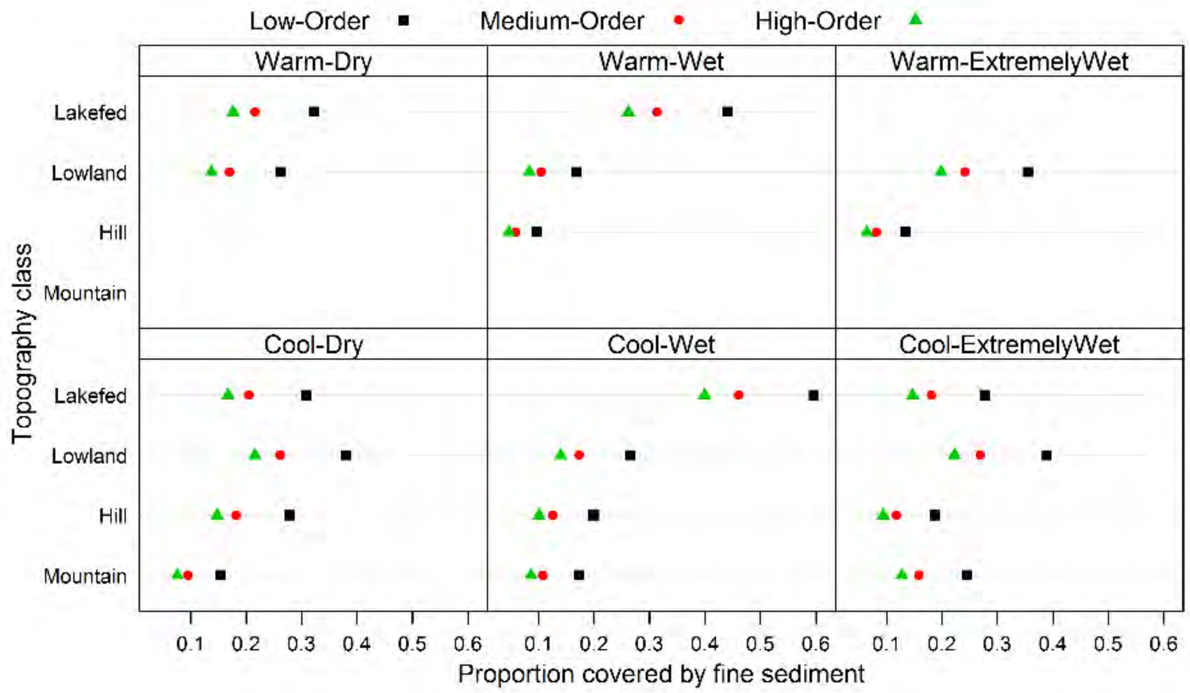


Figure 6-7: Predicted average reference state for the proportion of the stream bed covered by fine sediment. Reference state is predicted for all REC climate/topography settings that occur in more than one NZ reach of the river network for each of three stream size classes (Low-Order = Strahler stream order 1-2; Medium-Order = 3-4; High-Order = 5-8).

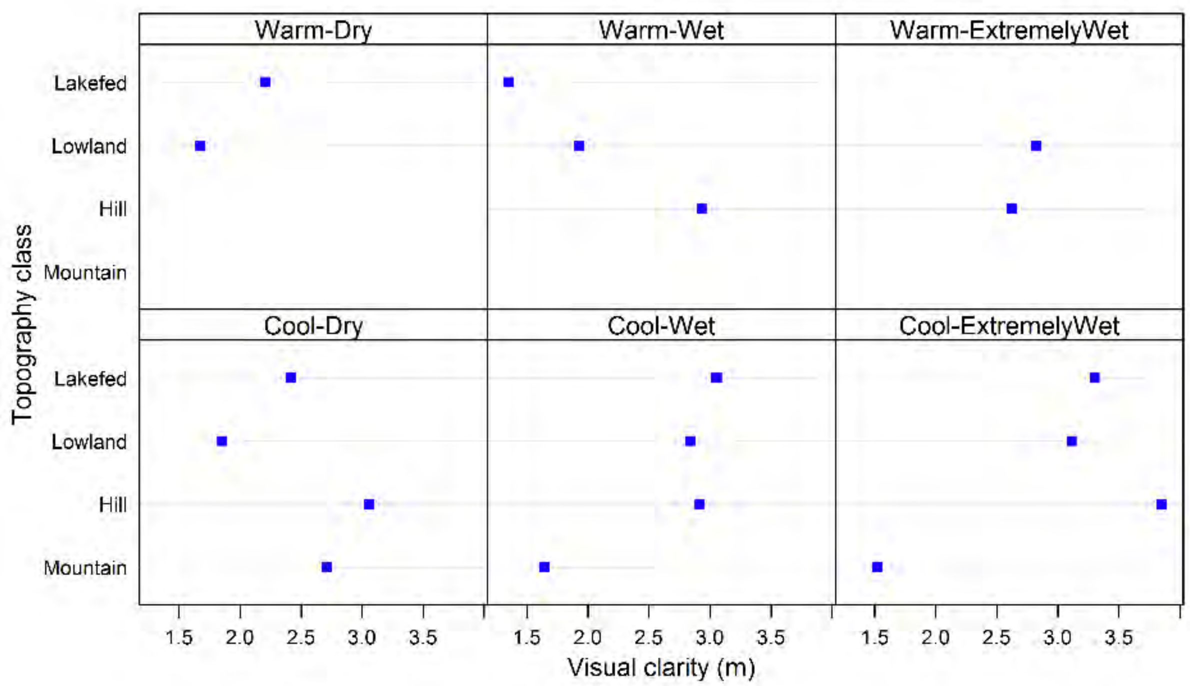


Figure 6-8: Predicted average reference state for visual clarity. Reference state is predicted for all REC climate/topography settings that occur in more than one NZ reach of the river network.

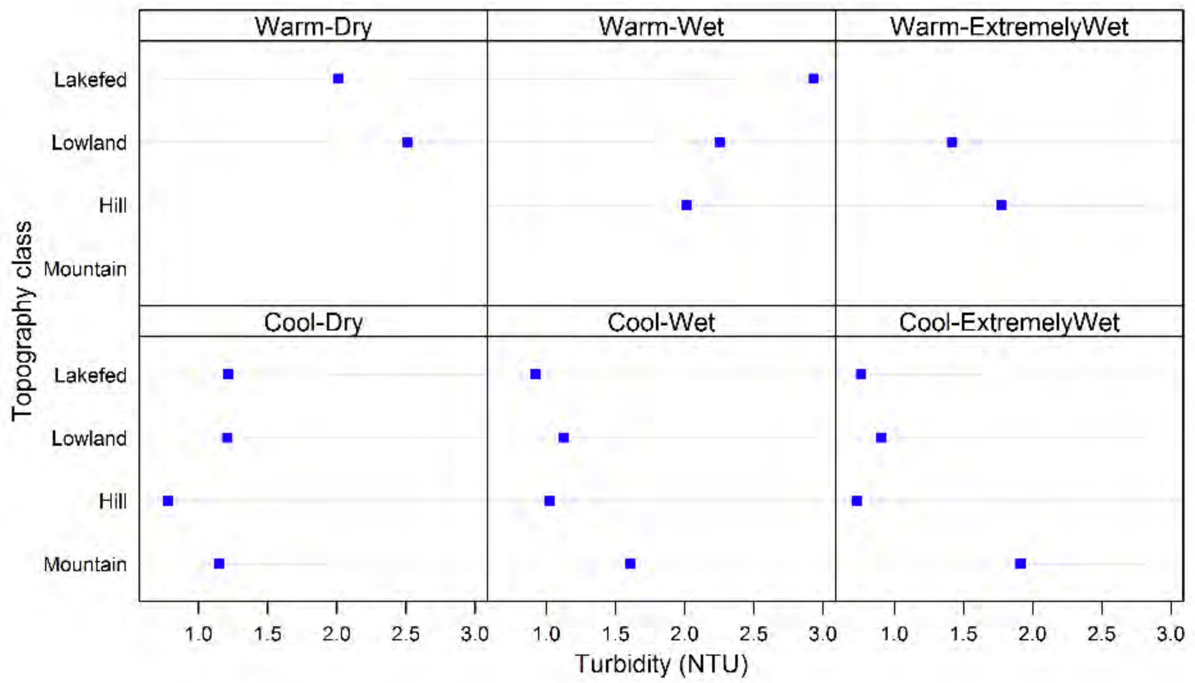


Figure 6-9: Predicted average reference state for turbidity. Reference state is predicted for all REC climate/topography settings that occur in more than one NZ reach of the river network.

Predicted fish probability of capture as a function of sediment ESVs

The predicted changes in fish probability of capture (FPC) with changes in the sediment ESVs varied between species and across fish settings (topography within climate plus network position, distance inland) as expected (e.g., Figure 6-10 to Figure 6-12). For some species, e.g., shortfin eel and ināŋga, FPC was correlated positively with increasing sediment ESV. That is, it was expected that the chance of them being present would increase with increasing sediment ESV. Other species, e.g., kōaro and redfin bully, had a FPC that was negatively correlated with elevated sediment ESV values, and thus it was expected they would become less likely to be present as sediment ESVs increased. In general, each species responded in the same way to all sediment ESVs, i.e., FPC either decreased with all sediment ESV metrics, or FPC increased with all sediment ESV metrics. For some species, e.g., banded kōkopu, the variation in FPC across fish settings was much greater than the variation in FPC associated with changes in the sediment ESV, but for others, e.g., kōaro, the opposite was true.

The FPC response curves for all species and sediment ESVs are provided in Appendix EE.

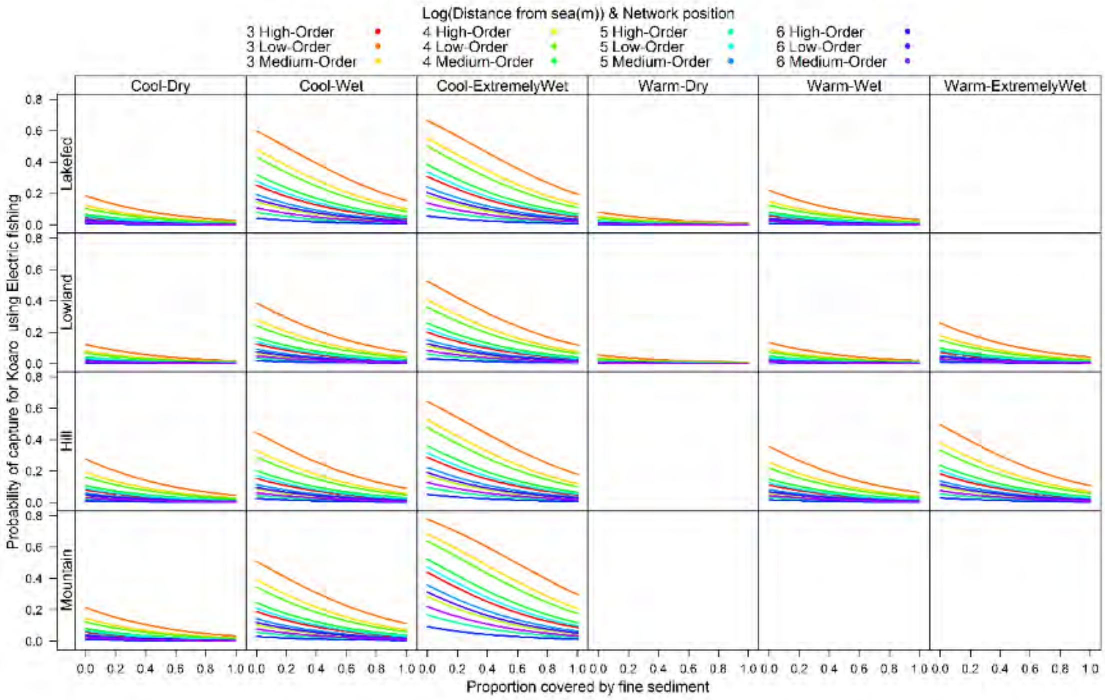


Figure 6-10: Predicted fish probability of capture (FPC) by electric fishing for kōaro relative to proportional cover of deposited fines. Predictions are made for each fish setting (climate/topography/ network setting (stream order)/distance to sea). Distance to sea (Log_{10} transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

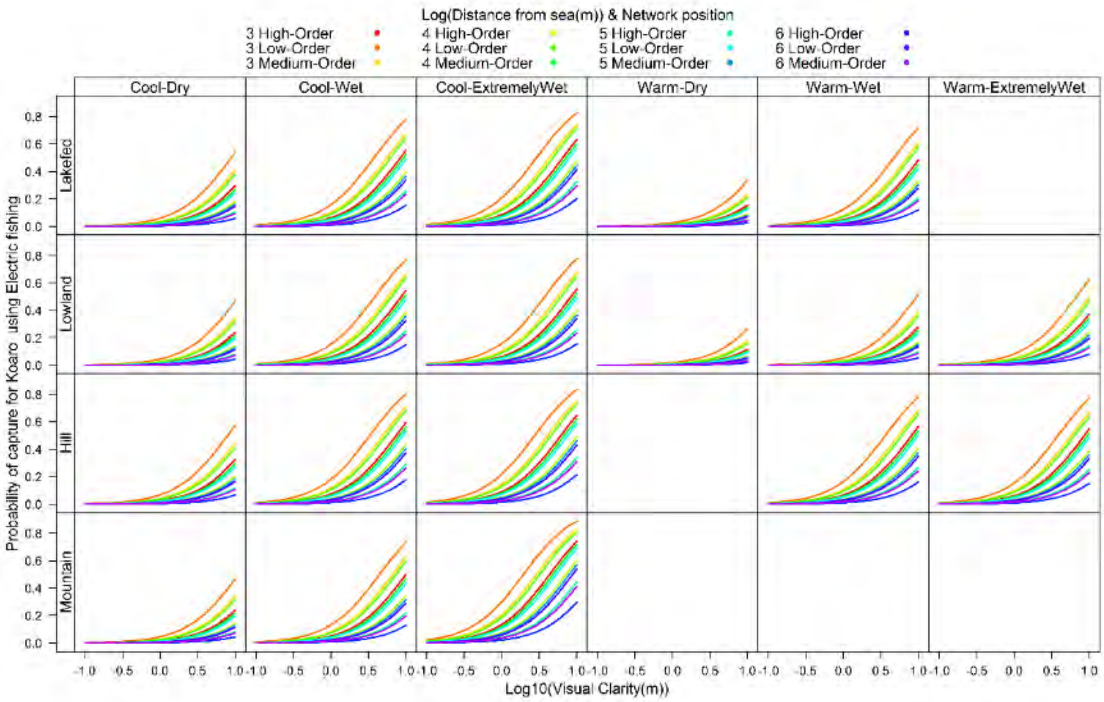


Figure 6-11: Predicted fish probability of capture (FPC) by electric fishing for kōaro relative to visual clarity (Log_{10} transformed). Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log_{10} transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

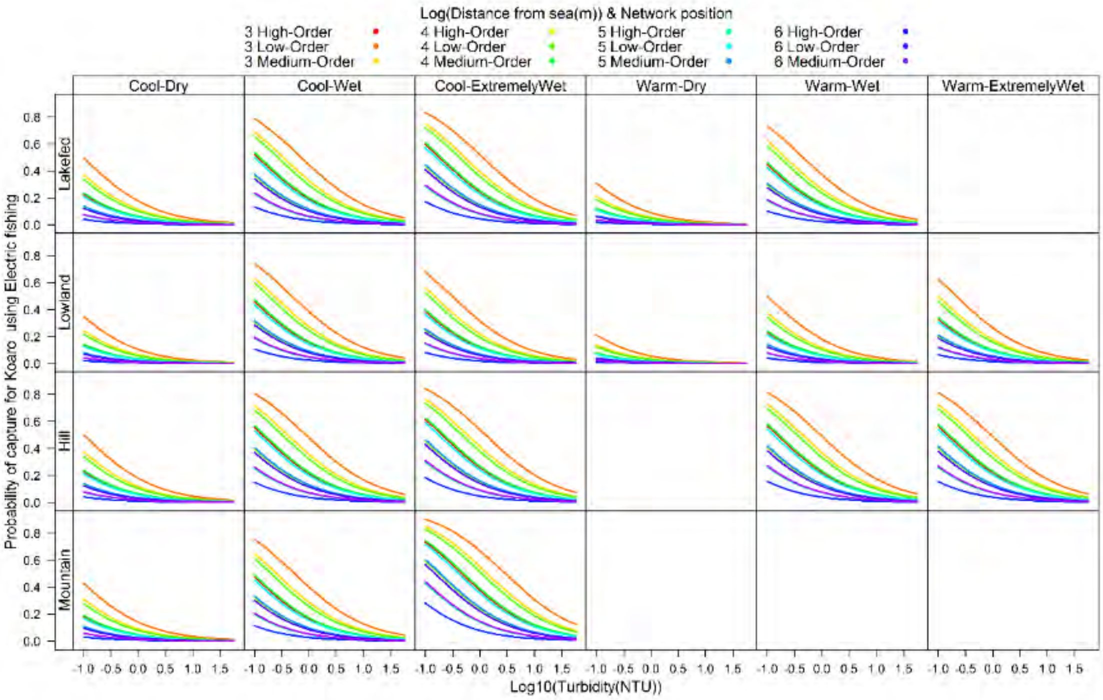


Figure 6-12: Predicted fish probability of capture (FPC) by electric fishing for kōaro relative to turbidity (Log_{10} transformed). Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log_{10} transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

Fish community change as a function of suspended and deposited sediment

Changes in fish community integrity relative to expected reference communities are represented by the term ΔC (i.e., 'delta' C). Only species with a negative correlation with the sediment ESV were used in the calculation of ΔC. The values of ΔC were calculated across a gradient of each ESV and are presented in Figure 6-13, Figure 6-14 and Figure 6-15, for % sediment cover, visual clarity and turbidity, respectively. The steeper the gradient of the response curve, the greater the potential impact on fish communities with increasing deviation of the sediment ESV from reference state. Cool climate classes are generally subject to a more rapid change in community integrity to changes in the sediment ESV than the warm climate classes. Furthermore, community integrity in lowland topography classes is generally less sensitive to changes in the sediment ESVs than for other topography classes. By contrast, community integrity in mountain topography classes was typically more sensitive to changes in the sediment ESVs compared to the other topography classes.

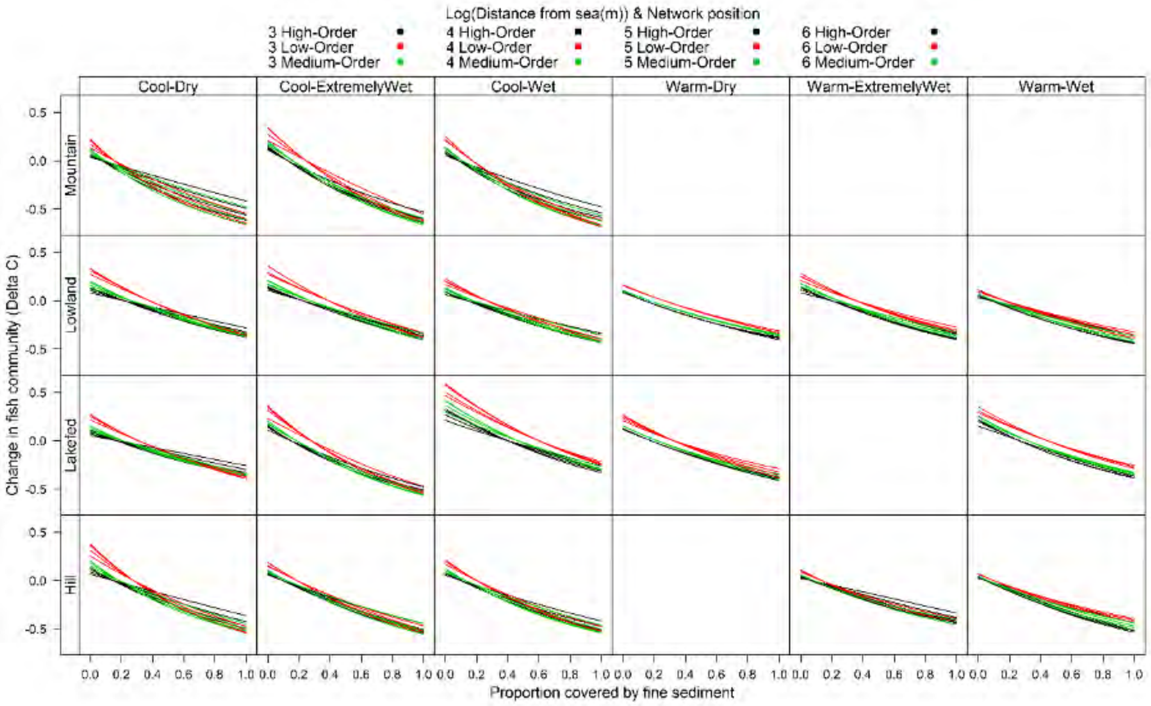


Figure 6-13: Response of the fish community integrity index, ΔC, to increasing proportion of the substrate covered by fine sediment. The individual lines in each colour represent the combinations of stream order and distance inland. Reference condition occurs when ΔC = 0. Negative values represent a decline in community integrity from reference condition.

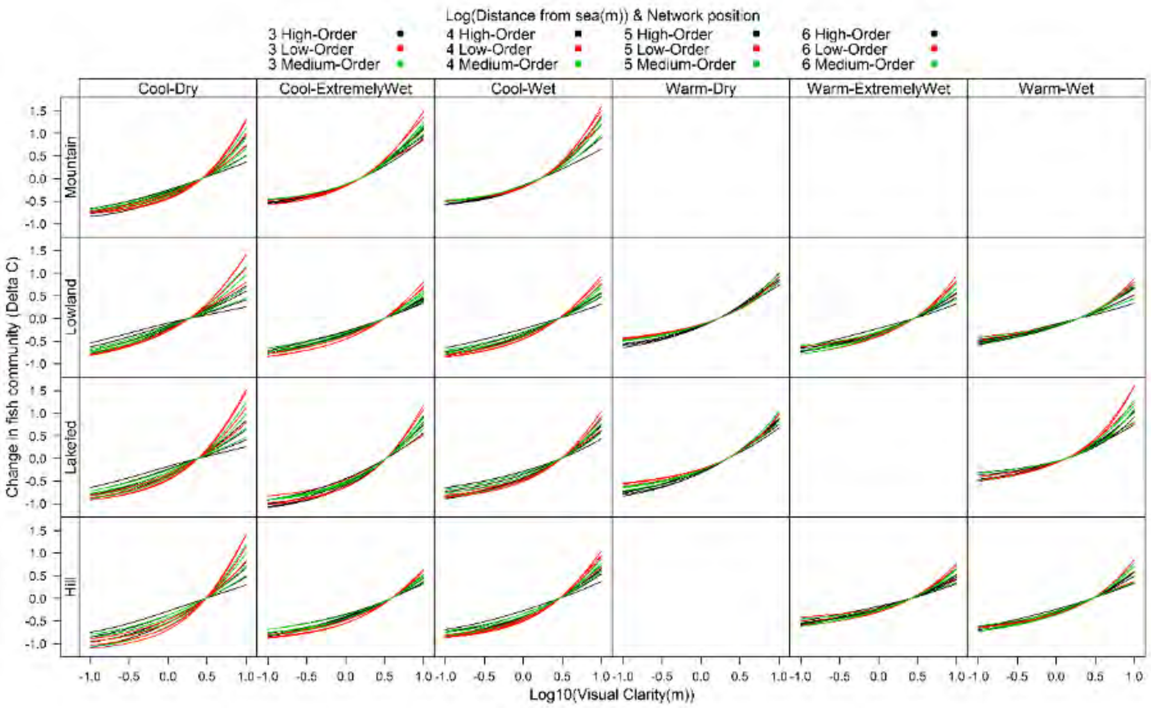


Figure 6-14: Response of the fish community integrity index, ΔC, to increasing visual clarity (Log_{10} transformed). The individual lines in each colour represent the combinations of stream order and distance inland. Reference condition occurs when ΔC = 0. Negative values represent a decline in community integrity from reference condition.

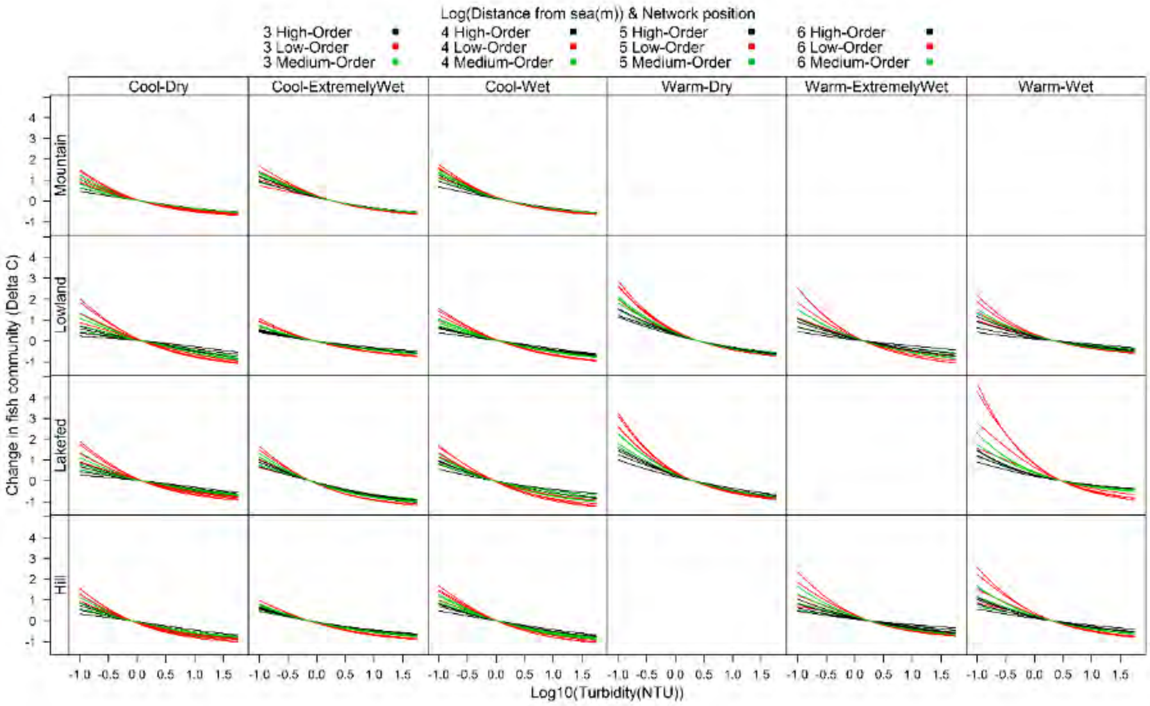


Figure 6-15: Response of the fish community integrity index, ΔC, to increasing turbidity (Log_{10} transformed). The individual lines in each colour represent the combinations of stream order and distance inland. Reference condition occurs when ΔC = 0. Negative values represent a decline in community integrity from reference condition.

Deposited and suspended thresholds based on a 20% decline in the fish community integrity metric (ΔC)

For ecosystem health, the NPS-FM defines the C-band state as one that 'generally represents a minimum safe level before an ecological tipping point'. We consider that a 20% decline in the fish community integrity metric (ΔC, referred to as -20% ΔC), relative to reference ESV state, is largely consistent with the NPS-FM definition of the C/D band. Accordingly the -20% ΔC thresholds presented inFigure 6-16 are part of the multiple lines of evidence considered in proposing the C/D band threshold values for for suspended and deposited sediment attribute tables (Chapter 7).

The bottom-line values for deposited fine sediment generally fall in the range of 30-60% cover of total fines across the different landscape settings, which correspond to the threshold values determined for macroinvertebrates (Chapter 6). The turbidity thresholds for the cool climate classes are around 2-3 NTU and in the warm climate classes generally in the range of 5-7 NTU. The bottom-line values for visual clarity across the landscape settings generally sit in the range from 0.8-2.0 m.

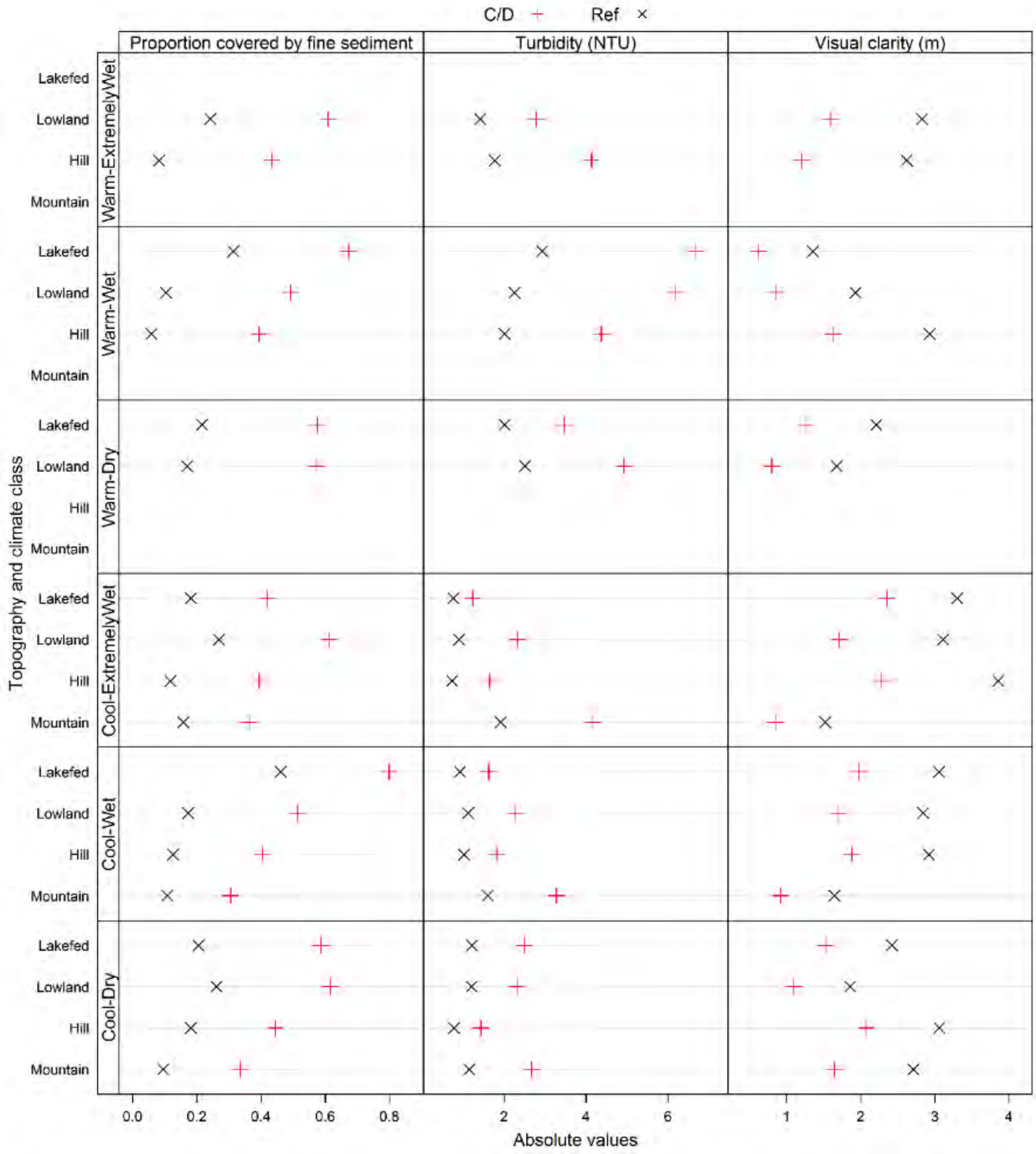


Figure 6-16: Deposited and suspended sediment thresholds thresholds at the second level (source of flow) of the REC classification based on a 20% decline in fish community integrity (i.e., -20% ΔC, red crosses) from sediment ESV reference state. We propose that a 20% decline in the fish community metric is consistent with NOF bottom-line thresholds (i.e., C band defined as minimum acceptable state prior to tipping point). The black crosses indicate the predicted average reference state for each source of flow class for each ESV.

6.5 Discussion

The literature review undertaken for this chapter has shown clear and consistent evidence for that elevated deposited and suspended sediment can have negative impacts on fish communities. There are a variety of mechanisms in operation including alterations to habitat quality and quantity, impacts on feeding and growth rates, elicitation of avoidance behaviour, direct lethal effects, and reduced reproductive success. Different fish species are impacted by elevated sediment through each of these pathways in differing ways depending on their habitat preferences, life-stage, feeding behaviours and movement capabilities.

There are relatively few examples in the literature of studies that quantify the effects of elevated deposited sediment on native fish in New Zealand. However, it is widely documented that many native fish species demonstrate a preference for coarser substrate sizes (Jowett & Richardson 2008). Fine sediment addition studies by (Jowett & Boustead 2001) and (Ramezani *et al.* 2014) provide quantitative support for a negative relationship between increased fine sediment loads and fish population size for a range of native fish species, but were not designed to identify specific thresholds.

Internationally, there are a wide range of studies that have documented the negative impacts of elevated deposited sediment on salmonid spawning success. While trout are not native to New Zealand, they are widely considered by the public as an indicator of ecosystem health and support an important recreational fishery. Spawning success in salmonids is highly sensitive to deposited sediment, with thresholds of between 10% and 20% fine sediment cover/volume frequently cited as being a threshold for protection of spawning habitats (Olsson and Persson 1998). It seems reasonable that these thresholds would translate to the protection of New Zealand native fish species, by preserving habitat for species associated with cobble substrate and by minimising egg smothering for benthic spawning fish.

A key factor to emerge from the literature review was that it may not be % cover of fine sediments per se that is the main functional control on fish responses, but rather the % embeddedness of the substrate that results from excessive sediment deposition (Collins et al. 2011). It is the infilling of the interstitial spaces in coarser substrates that is more problematic for many benthic species. The extent to which % sediment cover correlates with embeddedness was not evaluated as part of this study as there are no regular measures of substrate embeddedness collected in New Zealand.

A wider range of studies were available in the literature on the impacts of suspended sediment ESVs on native fish species. However, in almost all cases these were focused on short-term (<24 hr) effects and there is little information available on long-term, chronic impacts of elevated suspended sediment concentrations. Data on direct and indirect impacts are available for the native kōaro, inānga, banded kōkopu, smelt, redfin bully and eels, as well as introduced brown trout and rainbow trout. Much of this work has rightly focused on juvenile life-stages, which are expected to be more susceptible to sediment impacts. Of the species studied, the banded kōkopu has been identified as being most susceptible to the indirect, sub-lethal impacts of increased suspended sediment. Significant negative effects on banded kōkopu were observed relating to avoidance response, feeding rate, in-stream occurrence, and upstream migration as low as 10 NTU after 20 minutes in both laboratory and instream studies. Due to this sensitivity and their widespread distribution, banded kōkopu have been suggested as a useful benchmark species for the protection of fish in turbid waters in New Zealand (Rowe et al. 2002b) and have been used as an indicator species in a decision support system for managing suspended sediment concentrations for fish (Appendix FF;

accessed from https://www.niwa.co.nz/freshwater/management-tools/sediment-tools/setting-maximum-turbidity-levels-for-riverine-fish-dss). However, it should be noted that in nearly all cases documented in the New Zealand literature the 10-15 NTU threshold was the first sediment concentration tested above the 0 NTU control condition. Given that significant impacts on fish were observed over short durations at these levels, it is likely that significant effects could occur at lower turbidity levels (i.e., in the range 0-10 NTU) if tested, particularly over longer exposure durations.

Based on the currently available New Zealand literature on the sensitivity of fish to suspended sediment, it is reasonable to suggest that, where other factors are not limiting across the landscape as a whole, median turbidity levels of 0-5 NTU will likely have limited adverse effects on fish. However, as median turbidity increases from 5-15 NTU significant impacts on feeding effectiveness and growth would be expected in sensitive species resulting in reduced fitness and potential impacts on the long-term sustainability of fish communities. At median turbidity levels >15 NTU impacts on feeding and growth likely become more extensive and avoidance behaviour may begin to be expected. However, where other factors are limiting, e.g., food supply, the thresholds at which the ecological effects of elevated suspended sediment may become apparent may be lower.

It can be concluded from the literature review that both elevated deposited and suspended sediment can have significant impacts on the integrity of fish communities. Typically, as the level of impact increases, sensitive fish species are replaced with those more tolerant of higher sediment and poorer habitat quality, including undesirable exotic species that are generally more tolerant of these conditions.

Analyses of the broad-scale relationships between fish occurrence and sediment ESVs identified clear correlations between elevated sediment and changes in expected fish communities. Species such as shortfin eel and inānga demonstrated a positive correlation with increasing deposited and suspended sediment, indicating that they are likely tolerant of moderate increases in sediment. However, some species (e.g., kōaro) demonstrated significant negative responses to elevated sediment. Interestingly, the probability of capture for banded kōkopu, which have been identified as particularly sensitive to elevated turbidity in laboratory experiments, was found to be relatively insensitive to the effects of turbidity (or the other sediment ESVs) at a landscape scale. The reason for this apparent discrepancy is unclear, but may relate to the effects of multiple stressors and a failure to adequately take account of landscape scale influences to explain observations of the species in the field (e.g., distance inland and habitat types) that appear to have a greater effect on their probability of capture. These differences raise some questions over the use of this species as an indicator for sediment impacts on fish communities.

The analysis of the broad-scale relationships allowed the determination of community level response metrics that were subsequently used as the basis of deriving proposed attribute band thresholds. Visual inspection of these bands showed them to be relatively consistent with many of the thresholds identified in the literature, although in some landscape settings the thresholds for suspended sediment appear to be somewhat conservative. However, the studies in the literature tend to be based on short-term laboratory experimental studies where the effects of elevated sediments are tested in isolation from other potential influences on fish populations. The slightly different response patterns observed at the landscape scale may, therefore, reflect that they are based on long-term medians, the influence of multiple drivers of fish community integrity, and subsequently differing levels of sensitivity to elevated sediments across the landscape.

A key challenge for establishing legitimate relationships between the sediment ESVs and fish community responses was accounting for the landscape drivers that play an important role in influencing New Zealand's freshwater fish communities. Using REC classes as surrogates for key environmental characteristics known to influence fish communities was considered valid. In the process of undertaking the modelling for this project, nesting order of the REC (climate then topography then geology) was assumed to be a legitimate order in which to apply all random-effects. The physical basis and reasoning for the nested REC classification is justified and discussed by (Snelder *et al.* 2005). Nested REC classes were applied as legitimate random-effects. This is the methodology applied by (McDowell *et al.* 2013). It was assumed that application of crossed random-effects was a legitimate treatment. This means that the influence of a particular topography class within a particular climate class is related to the influence of that topography class within another climate class.

We chose species to include in the analyses that were widely distributed across New Zealand. This had the advantage of ensuring the same method was applied consistently across landscape settings. The disadvantage of this method was that it did not take account of species with restricted ranges. Many of these species with restricted ranges that were not included in this analysis are of high conservation value (Goodman et al. 2014) and may be highly susceptible to elevated sediment deposition due to their preference for coarse substrates and use of interstitial spaces as refuge habitat. However, these species are not suitable for undertaking landscape scale analyses of responses. We also modelled each species separately. This was the same method applied for fish species distribution modelling by previous studies (Leathwick et al. 2008; Crow et al. 2013). This method could not account for competition between species and how this may influence fish response to sediment stressors. Presence-absence data were used as the basis of the modelling work, rather than information on fish abundance. This is due to a lack of suitable information on fish abundance that can be paired with sediment ESV data and that span an appropriate gradient of landscape settings and sediment ESV pressures. It is likely that responses in fish communities, due to elevated sediments, would in the first instance be evident through changes in fish condition and abundance, prior to a fish becoming extirpated at a site. This justifies taking a precautionary approach to interpreting the landscape scale models that have been derived in this project.

Each sediment ESV was also modelled separately. This pragmatic approach allowed us to identify relationships between fish probability of capture and each sediment ESV, but did not allow for the possibility of confounding of our probability of capture due to co-variance between sediment ESVs. It is possible that a strong correlation between clarity and cover of deposited fines would result in a relationship between one of these sediment ESVs and fish probability of capture being caused by the other sediment ESV. This may not be an important issue in a practical sense because, in such a situation, managing for one sediment ESV would likely also manage for the other.

The proposed attribute band thresholds that have been derived in this study are based on expected community level responses of fish to elevated sediment. Due to the significant influence that landscape setting plays on determining fish community composition, it was considered necessary to account for this in deriving scientifically defensible and justifiable thresholds relevant to fish. Consequently, the proposed thresholds have been derived as absolute deviations in the sediment ESV from reference condition across a range of landscape settings. The full detail of these limits is provided in Appendix EE to allow more transparent derivation of final sediment ESV attribute tables. It would be possible to collapse these limits across landscape settings to simplify determination of attribute bands, but by reducing spatial specificity it increases the risk of adverse effects occurring due to the varying nature of expected fish communities across landscape settings.

Advantages and limitations of the approach used to derive sediment thresholds based on fish responses

Some of the key advantages of the generalised linear mixed model (GLMM) approach (i.e., to model fish probability of capture (FPC) include:

- The method explicitly accounted for the absence of a fish species because of distance inland, regardless of sediment ESV conditions.
- There is only one answer per landscape setting, not one per reach or one per observation site.
- Fish probability of capture (FPC) can only either continuously increase or decrease with sediment ESV for each species. The method is therefore conducive to finding A, B, C, and D bands.
- ESV states and fish communities were defined relative to reference conditions; therefore, there are no arbitrary decisions around the presence, or otherwise, of thresholds. This makes the derivation of limits more transparent.
- It is consistent with the McDowell et al. (2013) method (for estimating reference ESV values).
- Uncertainties in these models could be assessed by inspecting standard error of the modelled coefficients.

The main limitations to using the GLMM approach are:

- It is harder to explain than some of the methods used for the macroinvertebrate analyses.
- Changes in FPC will always be modelled as gradual declines; no sharp tipping points can, therefore, be identified.
- The selection of the 20% community change level as the threshold for the C/D boundary is arbitrary, although this was consistent with the approach used for deriving suspended sediment thresholds for macroinvertebrates.

The approach of translating the FPC-sediment ESV responses to species sensitivity distribution curves (SSD) was tested (as used for macroinvertebrate-suspended sediment threshold derivation – Chapter **Error! Reference source not found.**), but the low number of fish taxa meant that the SSD approach d id not work effectively. Consequently, the approach of calculating predicting fish community change was developed as an alternative for evaluating the expected consequences of elevated sediment ESVs.

6.6 Fish-based thresholds for sediment ESVs consistent with NPS-FM ecosystem health 'bottom-line'

Table 6-2 sets out the predicted reference and threshold values correspond to a 20% change in fish community (relative to reference) for each sediment ESV derived from the analysis of fish community responses. We propose that a 20% decline in fish community integrity is consistent with an ecosystem health bottom-line (i.e., C/D band threshold), and hence these thresholds were used along with other lines of evidence (including expert opinion) do propose C/D band threshold values for both deposited and suspended sediment attribute tables (Chapter 7). Results are set out at the second level (source-of-flow) of the REC and reflect the spatial variability in fish community responses. It should be recognised that these values are indicative of potential protection levels for overall fish community integrity and are intended to contribute towards deriving overall thresholds for ecosystem health alongside other ecosystem metrics (e.g., macroinvertebrates), rather than as definitive protection levels.

Table 6-2: Predicted reference state and threshold values for suspended and deposited sediment ESVs, based on a 20% decline in the fish community change metric (-20% ΔC); results are grouped by REC class (at the second 'Source-of-Flow' level). These differing thresholds reflect the spatial variability in fish communities captured by the REC categories.

	Visual clarity (m)		Turbidity (NTU)		% sediment cover	
REC source-of-flow class[1]	Predicted reference	Threshold (-20% ΔC)[2]	Predicted reference	Threshold (-20% ΔC)[2]	Predicted reference	Threshold (-20% ΔC)[2]
Cool-Dry.Hill	3.1	2.1	0.8	1.4	18.2	44.4
Cool-Dry.Lakefed	2.4	1.5	1.2	2.5	20.5	58.6
Cool-Dry.Lowland	1.9	1.1	1.2	2.3	26.1	61.6
Cool-Dry.Mountain	2.7	1.6	1.2	2.7	9.5	33.5
Cool-ExtremelyWet.Hill	3.9	2.3	0.7	1.6	11.8	39.4
Cool-ExtremelyWet.Lakefed	3.3	2.4	0.8	1.2	18.1	42.0
Cool-ExtremelyWet.Lowland	3.1	1.7	0.9	2.3	26.9	61.1
Cool-ExtremelyWet.Mountain	1.5	0.8	1.9	4.2	15.8	36.3
Cool-Wet.Hill	2.9	1.9	1.0	1.8	12.6	40.4
Cool-Wet.Lakefed	3.1	2.0	0.9	1.6	46.1	79.9
Cool-Wet.Lowland	2.8	1.7	1.1	2.3	17.3	51.4
Cool-Wet.Mountain	1.6	0.9	1.6	3.3	10.8	30.5
Warm-Dry.Lakefed	2.2	1.3	2.0	3.5	21.6	57.5
Warm-Dry.Lowland	1.7	0.8	2.5	4.9	17.0	57.4
Warm-ExtremelyWet.Hill	2.6	1.2	1.8	4.1	8.2	43.3
Warm-ExtremelyWet.Lowland	2.8	1.6	1.4	2.8	24.1	60.9
Warm-Wet.Hill	2.9	1.6	2.0	4.4	5.9	39.4
Warm-Wet.Lakefed	1.4	0.6	2.9	6.7	31.4	67.2
Warm-Wet.Lowland	1.9	0.9	2.3	6.2	10.5	49.2

[1] 2nd level REC (topography nested in climate). [2] '-20% ΔC' = thresholds based on 20% decline in the fish community integrity metric (ΔC) relative to reference state condition – intended to be consistent with NPS-FM bottom-line threshold (C/D band)

Threshold values (based on 20% decline in the ΔC metric) for clarity and turbidity were derived from the data of Unwin and Larned (2013). Threshold values for these suspended sediment measures should, therefore, be expressed as medians measured over the long-term. This would be a minimum of monthly samples over a two-year duration (24 samples).

Thresholds for bands for % sediment cover were derived from observations recorded in the NZFFD. The band thresholds should, therefore, be expressed as means of areal cover of deposited fine sediment derived from instream observations over a representative river reach.

It is also important to consider that these levels are set such that they provide a level of protection at an overall fish community level and may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations. For example, salmonid spawning habitats have a requirement for low levels of deposited sediment cover (<10%) to provide optimal habitat. However, such requirements are specific to protection of that species and will only be applicable in certain locations.

It should also be noted that, although there was strong evidence of predictive power and ability to distinguish landscape-scale patterns, there are statistical uncertainties within both predictions of the reference state conditions for both suspended and deposited sediment ESVs (see discussion by (McDowell *et al.* 2013) and details given in (Appendix DD) and predictions of fish presence/absence (see Appendix EE).

6.7 Future Work

We provide the following recommendations for future work:

- Test metrics other than the median suspended sediment ESV as explanatory variables.
- Evaluate the effect of the different methods for evaluating reference sediment ESV state.
- Compare between % sediment cover measurement methods SAM1 and SAM2 (bankside and instream visual assessments, respectively) based on run measurements vs. NZFFD whole reach sediment cover estimates.
- Collate and analyse standardised fish abundance data across a gradient of sediment states, while accounting for differences in landscape setting.
- Evaluate how sensitive the macroinvertebrate-sediment ESV responses in the BRT and RF models are to choose a different suite of environmental predictors.
- Investigate the sensitivity of the recommended thresholds to uncertainty in the sediment ESV reference state model.
- Consider appropriate methods for collapsing ΔC-based thresholds across landscape/REC settings.

7 Synthesis and final proposed management thresholds for deposited and suspended sediment based on multiple lines of evidence

7.1 Introduction

The diagram showing the workflow and connections between chapters was provided in Chapter 1, and for clarity is reproduced in below (Figure 7-1).

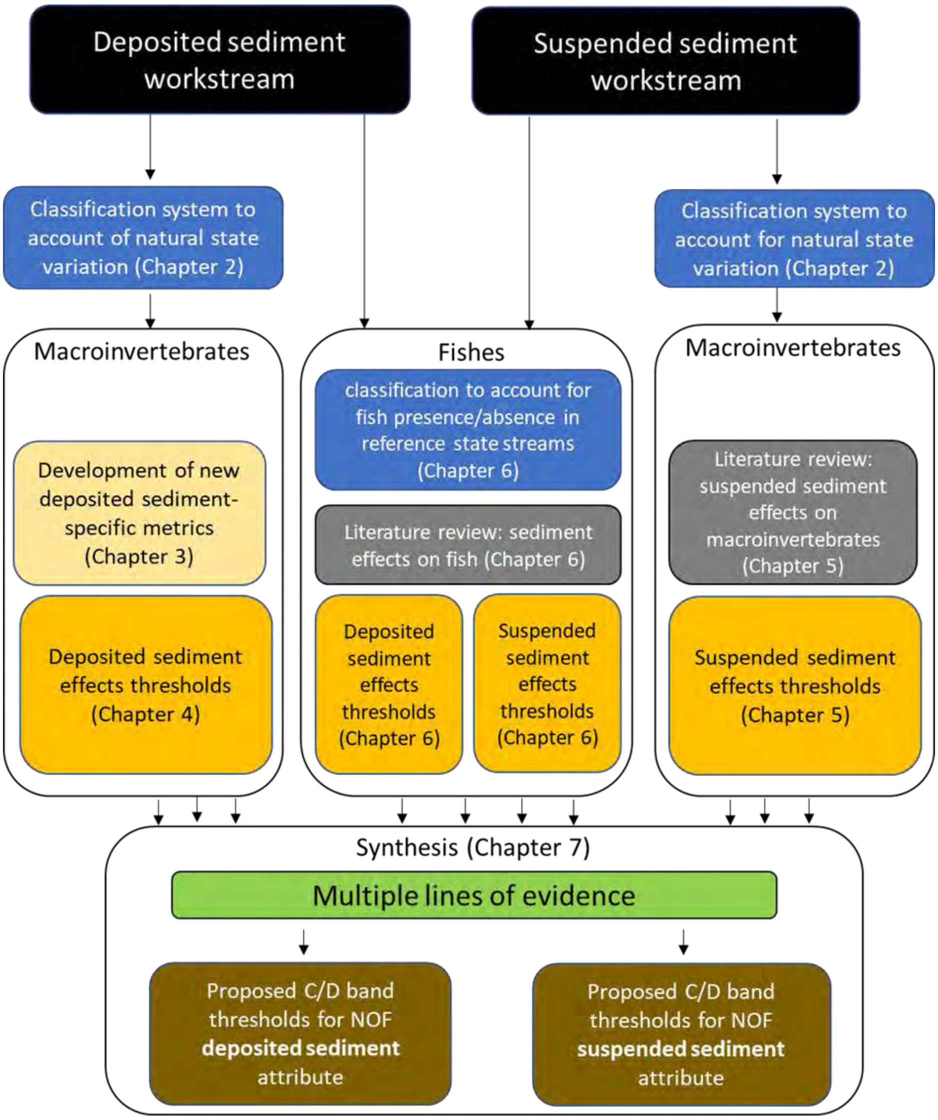


Figure 7-1: Summary of the workflow of the Stage 2 sediment threshold project, illustrating the major components of the chapters, and how these contributed to the final proposed C/D band thresholds for suspended and deposited sediment attributes. (refer to Figure 1-1 for full explanation).

Chapters 4-6 derived sediment threshold values based on the responses of fishes and macroinvertebrates. The effects level or 'magnitude of community change' was chosen to yield thresholds that were consistent with ecosystem health bottom-lines (i.e., C/D band thresholds). The thresholds derived from Chapters 4-6 contribute to multiple lines of evidence which also include:

- Literature effects thresholds for suspended and (to a lesser extent) deposited sediment;
- Existing regulatory guidelines for managing the effects of sediment (mainly suspended sediment); and
- Expert opinion.

The purpose of this Chapter is to briefly summarise the work from the 3 thresholds chapters and synthesise the multiple lines of evidence and propose C/D band thresholds for the deposited and suspended sediment attribute tables.

7.1.1 The use of different methods for deriving sediment thresholds

'Deposited sediment vs macroinvertebrates'; 'suspended sediment vs macroinvertebrates' and 'sediment (deposited and suspended) vs fishes' thresholds were derived by different analytical approaches. Where applicable, similar methodologies were trialled, however, methods that worked for one sediment-organism combination did not necessarily work for another. This is not surprising, giving the differences in the distribution, data and mode of action of suspended sediment vs deposited sediment, and the differences between data (i.e., presence absence vs abundance) and natural distribution of fishes compared with macroinvertebrates. Finally, methodological differences also reflect the experience and familiarity of the key investigators with different analytical, statistical and modelling approaches.

Justification for fish method and why it was different to the 2 macroinvertebrate approaches

The chosen analytical approach for characterising fish-sediment ESV responses reflected both the availability of suitable data, and the specific ecology and biology of New Zealand's freshwater fishes. In contrast to macroinvertebrates, freshwater fish data are not typically collected at standard state of the environment (SOE) monitoring sites. Consequently, there were very few fish observations that could be paired in both space and time with SOE data on sediment ESVs. Standardised abundance data are also rare for freshwater fish because of the wide variety of methods used to collect data required to meet different survey objectives. As a result, the fish analyses had to be based on presence/absence data, rather than abundance data. These two factors meant that it was inappropriate to directly apply the analytical methods used for macroinvertebrates to fish.

In addition, unlike macroinvertebrates, fish communities are strongly influenced by landscape setting (e.g., distance inland). This is partly because many New Zealand native species spend some part of their life-cycle at sea, and exhibit different abilities to penetrate inland. Consequently, it was necessary to apply methods that could account for these influences to increase the chances of being able to detect the 'real' effects of elevated sediment on fish communities.

Justification for different approach for suspended vs deposited sediment threshold based on macroinvertebrate response

The chosen analytical approach for the 'macroinvertebrate-suspended sediment' thresholds reflected both the availability of suitable national data (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 of other potential stressors, which could confound establishing a causative relationship with suspended sediment measures. The nature of suspended sediment in rivers is a combination of particulate organic matter and inorganic particulates – 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.

The quantile regression approach used for ‘macroinvertebrate vs suspended sediment’ differ from that used for ‘macroinvertebrate vs. deposited sediment’ and ‘fishes vs. sediment’ as these sediment effects do not result in subsidy-stress response relationships, and were based on different databases for their derivation (e.g., fish data was limited to presence/absence data, whereas macroinvertebrate monitoring comprises abundance data). Quantile regression approaches were investigated as an approach for analysing macroinvertebrate responses to deposited sediment gradients (refer to Chapter 4, Method 1), but were found to be relatively insensitive (reflecting the absence of a subsidy-stress relationship with a different sediment ESV). The construction of species sensitivity distribution curves (SSD) for macroinvertebrate-suspended sediment threshold derivation was trialled unsuccessfully in the derivation of fish-based sediment thresholds.

7.2 Deposited sediment ESV

7.2.1 Comparison of fish and macroinvertebrate thresholds

The ecological effects of deposited sediment on macroinvertebrates (Chapter 4) and fish (Chapter 6) were explored using different approaches. For macroinvertebrates, sediment effects were explored across a full environmental gradient, whereas for fish sediment effects were explored within REC source of flow categories. Both approaches relied on models to predict reference state of deposited fine sediment to apply (as was the case for macroinvertebrates) or develop (as was the case for fish) recommended band thresholds.

Using a combination of BRT and gradient forest (GF) analysis of macroinvertebrate data, a bottom-line threshold of 30% fine sediment cover for sites with naturally low-to-medium levels of deposited sediment (i.e., predicted reference state of <30% fine sediment cover) was identified. For sites identified as having naturally high levels of fine sediment cover (i.e., predicted reference state of 30-60% fine sediment cover) a bottom-line threshold of 60% fine sediment cover was proposed to protect ecosystem health. Sites with predicted reference state % fine sediment cover values of >60% were classed as soft-bottom sites, and were considered exempt from the deposited sediment attribute. The 20-30% sediment cover threshold we found in our analyses is similar to thresholds defined by previous New Zealand studies. For example, 20% sediment cover was the threshold at which %EPT started to significantly decline using a sigmoidal model on a relatively small dataset consisting of 30 Canterbury stream sites (Burdon et al. 2013), and also streams supporting macroinvertebrate communities indicative of very good stream health (MCI > 120) were typically associated with sediment cover of 20% or less using a small national data set (Clapcott et al. 2011).

Spatial variability was taken account of in the analysis of fish communities by using the second level of the REC (source-of-flow). This yielded 19 ‘environments’ (11 cool and 7 warm), where the bottom-line is defined as a 20% reduction in the fish community integrity index, ΔC (relative to reference state). The final thresholds can be presented as deviations from reference state, or as absolute values (i.e., already added to reference). Accordingly, an added flexibility of this approach is that the method can be applied using any reference state model (not just the GLMM approach described in this report).

The fish- and macroinvertebrate-based thresholds were derived using two different reference state prediction models and datasets (GLMM and BRTREF), although a comparison of model output shows consistency in predictions across many REC classes (Appendix E).

'Side-by-side' comparison of macroinvertebrate- and fish-based thresholds

A 'side-by-side' comparison to reconcile the two sets of thresholds (fish and macroinvertebrates) is provided in Table 7-1. In a nutshell, we want to know if the macroinvertebrate-based thresholds and and proposed 'national' classification for deposited sediment would be protective of fish.

Table 7-1: Side-by-side comparison of predicted reference state and thresholds derived independently for macroinvertebrates (Chapter 4) and fishes (Chapter 6) for deposited fine sediment (i.e., <2 mm particle size). The proposed national classification are the two stream types under macroinvertebrate, namely 'low-medium' and 'high'. The multiple REC-based classifications for fish reflect the strong influence of landscape setting on fish community composition. Methodological details for macroinvertebrates and fishes are provided in Chapters 4 and 6, respectively. Blue shading depicts sites predicted to have naturally *high* levels of deposited fine sediment (30-60%). * ↓ or ↑ symbolises predicted C/D band for fish is lower or higher than the relevant macroinvertebrate-derived C/D band threshold, respectively.

Macroinvertebrate		Fish (Chapter 6)			Fish ↓ or ↑ than macro-invertebrate threshold*
Reference state range (% sediment cover) Chapter 2	Equiv. C/D threshold (% sediment cover) Chapter 4	REC Source-of- flow	Reference state (% sediment cover)	equiv. C/D threshold (% sediment cover)	
		WWH	6	39	↑
		WXH	8	43	↑
		CDM	10	34	↑
		WWL	11	49	↑
		CWM	11	31	↑
		CXH	12	39	↑
		CWH	13	40	↑
'Low-to-medium'		CXM	16	36	↑
0-30%	30	WDL	17	57	↑
		CWL	17	51	↑
		CXLk	18	42	↑
		CDH	18	44	↑
		CDLk	21	59	↑
		WDLk	22	58	↑
		WXL	24	61	↑
		CDL	26	62	↑
		CXL	27	61	↑
'high'		WWLk	31	67	↑
30-60%	60	CWLk	46	80	↑

To compare the two sets of thresholds, the REC 'source-of-flow' reference classes are 'mapped' to the corresponding proposed 'national classification system' for deposited sediment (Chapter 2).[17] Except for two lake-fed classes (WWLk and CWLk), the fish-based reference states (which incorporate

[17] The 'low median' (0-30%), 'high' (30-60%) and 'soft-bottom' (>60% deposited sediment) described in Chapter 2 are proposed classification classes for managing deposited sediment in NZ rivers and streams. The REC-based (i.e., source of flow) was a requirement for developing fish-based threshold (using a 'fish community integrity index'), which required incorporation of landscape setting.

landscape setting – Chapter 6) map to the single ‘low-medium’ reference state (0-30%) which has a proposed C/D band threshold of 30% (grey-shaded rows in Table 7-2). The macroinvertebrate-based threshold was based on marked changes in multiple community metrics (including two developed for sediment response; Chapter 3) along a deposited sediment gradient (Refer to Method 2, BRT analysis in Chapter 4). Predicted reference states for fish REC classes (that incorporate ‘landscape setting’) spanned the full range of the ‘low-to-med’ class, ranging from 6 to 27% sediment cover. The important ‘take home message’ is that for all 17 fish-based classes in the grey-shaded region of Table 7-2, the proposed C/D threshold for fish (i.e., 20% decline in ΔC; 2nd column from right, Table 7-2) are all greater than the 30% bottom-line threshold proposed for the ‘low-medium’ deposited sediment class. Fish-based thresholds ranged from 31-62% sediment cover (median 44%). For the two lake-fed fish classes that mapped to the ‘high’ (30-60% sediment cover) deposited sediment class (unshaded rows in Table 7-2), the fish-based ‘C/D band-equivalent’ threshold values (i.e., 20% decline in ΔC) were both greater (67 and 80%) than the macroinvertebrate C/D-equivalent band proposed for this class of 60% sediment cover (refer to Method 3, RF method in Chapter 4).

The macroinvertebrate-based thresholds, and proposed dual classification system for hard-bottom streams (i.e., ‘low-med’ and ‘high’) appear to be adequately protective of all fish-based thresholds (based on 20% declines in ΔC). Accordingly, we recommend that these form the basis of the C/D band threshold for a deposited sediment attribute (Table 7-1, section 7.2.2).

7.2.2 Proposed NOF attribute table for management of deposited sediment in NZ streams

The proposed form of the deposited sediment attribute is presented in Table 7-3.

Table 7-3: Proposed NOF attribute table for assessing deposited fine sediment in wadeable rivers.

Value	Ecosystem Health		
Freshwater Body Type	Rivers (wadeable only)		
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	**Numeric Attribute State ‘Low-to-medium’ level (<30%)[1] of natural sediment**	**Numeric Attribute State ‘high’ level (30-60%)[1] of natural sediment**	**Narrative Attribute State**
	Annual mean[2]	Annual mean[2]	
A	NA	NA	
B	NA	NA	
C	<30%	<60%	Low to moderate cover relative to reference state providing excellent to fair habitat for biota. Risk of sensitive macroinvertebrate species being lost and change in community composition.
National Bottom Line	**30%[3]**	**60%[3]**	
D	>30%	>60%	High likelihood of sediment cover exceeding reference state providing poor habitat for biota. High probability of loss of sensitive macroinvertebrate species.

4) Classes are streams and rivers defined according to predicted reference state for deposited sediment, currently this is based on predicted reference state from the BRT REF model. Streams with greater than 60% fine sediment cover are classified as naturally soft-bottomed streams and are exempt. Based on a monthly monitoring regime.
5) The minimum record length for grading a site based on an instream visual assessment of % fine sediment cover (SAM2) is 2 years.
6) 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 integrity index*), however they may not always be sufficient for the protection of specific life-stages or habitat requirements in specific locations (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.

7.2.3 Indicative compliance with proposed national bottom-line for deposited sediment

Compliance with proposed bottom-lines was determined for 522 sites from a combined SoE-research database where deposited fine sediment was measured using the ‘% sediment cover’ (SAM2) method. For stream sites classified as ‘low-to-medium’ levels of fine sediment (predicted reference state <30% sediment cover), the proportion of ‘D-band’ sediment sites (i.e., >30% fine sediment cover) was 15% (51 out of 347 sites). For stream sites classified as ‘high’ levels of fine sediment (predicted reference state 30-60% fines), the proportion of ‘D-band’ sites (i.e., >60% fine sediment cover) was 3% (2 out of 69 sites). Approximately 20% of sites (106) were classified as naturally soft-bottom streams (reference state predictions of >60% sediment cover)– we recommend that the deposited sediment attribute is not applied to this stream class.

7.3 Suspended sediment

7.3.1 Comparison of fish and macroinvertebrate thresholds

The approach taken to derive fish-based suspended sediment thresholds for fish was analogous to the method for deposited fine sediment, except modelled predictions of long-term median turbidity and visual clarity values were paired with observations of fish from the NZFFD. Predicted reference states for turbidity and visual clarity were calculated according to the method of McDowell et al. (2013). Fish probability of capture (FPC) plots for 11 species were modelled against suspended sediment gradients (using the proxy measures of median visual clarity and median turbidity separately). Individual species plots where aggregated into single plots represented by a fish community integrity index. As for deposited sediment, a 20% decrease in the fish community integrity index (ΔC), relative to the reference state, was used to derive thresholds consistent with ecosystem health bottom-lines (refer to Chapter 6 for methods).

Macroinvertebrate-based thresholds for suspended sediment (both turbidity and visual clarity) were derived using a 30% effect level at taxon and metric level, and then ordering these 'individual' results to construct a species sensitivity distribution (SSD) curve. An 80% protection level was then chosen to derive a macroinvertebrate-based threshold consistent with an effect level represented by the ecosystem health bottom line.

'Side-by-side' comparison of macroinvertebrate- and fish-based thresholds

To facilitate the 'side-by-side' comparison between fish-based and 'individual' macroinvertebrate-based thresholds, which used different effect levels[18] to define a 'C/D band-equivalent' threshold, we recalculated macroinvertebrate and fish thresholds using both 20 and 30% effect levels. In addition to reduce the complexity of the comparison, we 'collapsed' the source-of-flow (topography nested in climate) classes to the 6 REC climate classes (cool dry, cool wet, cool extremely wet, warm dry, warm wet and warm extremely wet).

A 'side-by-side' comparison of the suspended sediment thresholds derived from fish and macroinvertebrates is provided for turbidity (Table 7-4) and visual clarity (Table 7-5).

[18] 'C/D band-equivalent' suspended sediment thresholds derived from fish and macroinvertebrates responses were originally based on 20% (chapter 6) and 30% (chapter 4) effect levels, respectively.

Table 7-4: Side-by-side comparison of 'C/D band -equivalent' suspended sediment thresholds (measured as turbidity, NTU) derived from the analysis of macroinvertebrate (Chapter 4) and fish (Chapter 6) datasets. The level of effects (20 or 30%) used for these thresholds were targeted towards C/D band transitions (i.e., bottom-line values). Bold text indicates macroinvertebrate diversity measures.

Macroinvertebrate: Turbidity (NTU) (Chapter 5)			Fish: Turbidity (NTU) (Chapter 6)			
Organism / metric	Turbidity (NTU) threshold (median)		REC climate class	Predicted reference turbidity[1]	Turbidity (NTU) threshold (median)	
	20% effect level	30% effect level			20% effect level	30% effect level
SSD (80% protection)[2]		**4.3**				
Taxa richness	**4.5**	**17**	CD	1.1	2.1	3.1
Density	12.4	19	CW	1.1	2.2	3.1
EPT taxa	**3.0**	**8.2**	CX	1.0	2.2	3.5
EPT individuals	9.4	12.2	WD	2.1	3.9	5.5
Deleatidium	9.7	12	WW	2.3	5.2	9.0
Aoteapsyche	12	15	WX	1.6	3.3	5.2

[1] determined via the method of McDowell et al. (2013). [2] 80% protection level taken from species sensitivity distribution (SSD) curve consistent of approximately 7 individual species and 7 community metrics based on annual median turbidity values.

Table 7-5: Side-by-side comparison of potential visual clarity (m) thresholds derived from the analysis of fish and macroinvertebrate datasets. The level of effects (20 or 30%) used for these thresholds were targeted towards C/D band transitions (i.e., bottom-line values). Bold indicates macroinvertebrate diversity measures.

Macroinvertebrates: Visual Clarity (m) (Chapter 5)			Fish: Visual Clarity (Chapter 6)			
Organism / metric	Visual clarity (m) threshold (median)		REC climate class	Predicted reference turbidity[1]	Visual clarity (m) threshold (median)	
	20% effect level	30% effect level			20% effect level	30% effect level
SSD (80% protection)[2]		0.95				
Taxa richness	0.48	0.26	CD	2.5	1.5	1.2
Density	0.39	0.33	CW	2.5	1.6	1.2
EPT taxa	0.61	0.33	CX	2.8	1.6	1.2
EPT individuals	0.64	0.52	WD	2.2	1.1	0.74
Deleatidium	0.51	0.40	WW	2.1	0.95	0.58
Aoteapsyche	0.56	0.39	WX	2.6	1.3	0.85

[1] determined via the method of McDowell et al. (2013). [2] 80% protection level taken from species sensitivity distribution (SSD) curve consistent of approximately 7 individual species and 7 community metrics based on annual median visual clarity values.

At a 20% effects level, thresholds for the 6 macroinvertebrate metrics (Table 7-4 and Table 7-5) ranged from approximately 3 to 12 NTU and 0.64 to 0.39 m for visual clarity. The 80% protection level (calculated from SSD, Chapter 5) was at the lower (i.e., cleaner) end of the range, with respective turbidity and clarity thresholds of 4.3 NTU and 0.95 m. The metrics most sensitive to suspended sediment (measured as turbidity) appeared to be taxa richness and EPT taxa, with a 20% effects threshold of 4.5 and 3 NTU respectively. At an increased effects level of 30%, the thresholds for EPT taxa increased from 3.0 to 8.2 NTU, and taxa richness from 4.5 to 17 NTU.

By comparison, fish-based thresholds using a 20% effect level (decline in ΔC) were more conservative, with values of 2 NTU for 'cool' REC climate classes and 3 to 5 NTU for warm REC classes. For visual clarity, the respective values for cool and warm REC climate classes were 1.5-1.6 m and 1-1.3 m. When we increased the effect level to a 30% decline in ΔC, the turbidity thresholds for 'cool' REC climate classes increased to around 3 NTU (from 2 NTU); and thresholds for 'warm' REC classes, increased to 5 to 9 NTU (from a range of 3-5 NTU). Visual clarity showed a similar pattern, with thresholds for a 30% effect level for cool climates decreasing (i.e., less clear water) to 1.2 m (from 1.6 m), and for warm classes, decreasing to 0.6-0.85 (from 1.0-1.3 m). With the exception of thresholds corresponding to a 30% effects level for 'warm' REC climate classes, the indicative thresholds for fish were generally lower than those based on macroinvertebrates.

As part of the synthesis, we needed to determine whether the experimental approach to deriving fish thresholds were overly cautious, or whether the results indicate that a macroinvertebrate-based threshold would not provide adequate 'bottom-line' protection for fishes in New Zealand streams and rivers. The fish-based thresholds relied on modelled values of suspended sediment (not measured) and a newly developed, and therefore largely untested, community fish integrity metric (ΔC) and what would be a likely effect level would produce threshold equivalent to a C/D band. As a first step, we looked at what the application of fish-based C/D thresholds would mean for compliance at the 832 turbidity sites across New Zealand. The conservative nature of the 'cool' climate threshold (particularly at a 20% effect level) was highlighted by the high number of D-band sites (Figure 7-2). For CD and CW climate classes, more than 50% of sites were classified as D-band using the 20% effect level as a C/D band threshold. Aggregated 'cool' climate class (i.e., CD, CW and CX) resulted in 44 and 48% of sites, being classified in the D-band based on the turbidity and visual clarity 20% effect level, respectively. This is not surprising given that for cool climate classes, the 20% effects threshold represents an increase of just 1 NTU above reference condition. In contrast, for 'warm' climates, the proportion of sites exceeding the 20% ΔC threshold was 23% and 34% for turbidity and visual clarity, respectively.

These finding suggest that for at least the 20% effect level, the fish-based thresholds derived for suspended sediment effects are probably not representative of the level of effects consistent with ecosystem health bottom line thresholds.

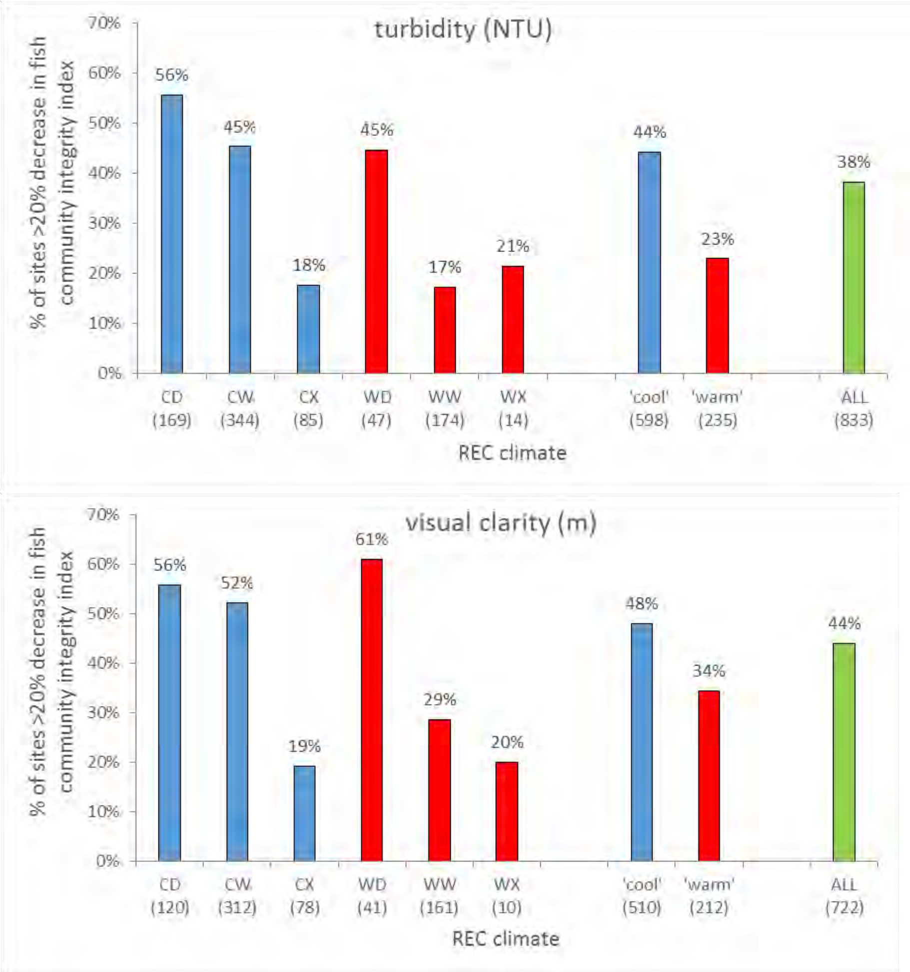


Figure 7-2: Compliance of sites (turbidity (top) n=833 and visual clarity (bottom n=722) with 20% effect level threshold using the fish community integrity index (i.e., 20% decline in ΔC metric). Sites are group by REC climate class, and aggregated 'cool' and 'warm' categories.

7.3.2 Comparison of fish- and macroinvertebrate-derived threshold values with literature/regulatory values

Reviews of relevant literature for fish and macroinvertebrates are provided in Section 5 and Section 6, however, selected literature at the cleaner end of the suspended sediment spectrum is briefly presented here (Table 7-6) to provide some context for the derived values in Table 7-4 and Table 7-5. Literature reviews of suspended sediment effects on macroinvertebrates and fishes are presented in Chapter 4 and Chapter 6. The literature, generally, seem to indicate that ecologically significant effects can occur between 5 and 10 NTU, and with marked impacts occurring at turbidity values in excess of around 15-20 NTU.

Although underlying severity of effect and reaction distance models of Newcombe's (2003) study have been criticised, the risk assessment framework is one of few that tackles the issue of duration of exposure – with the shorter the duration, the higher the level of suspended sediment for a given ecosystem effect (severity of effect values). The severity of effects values define ranges of *slight impairment* (1-3) and *significant impairment* (4-8) and *severely impacted* (9-14). Assuming an ecological tipping point would occur somewhere near *slight* to *significant impairment* (i.e., score of 4-6), this corresponds to visual clarity values of between 1.1 and 0.55 m (converting to turbidity values of between 3.5 and 7 NTU).

The range of turbidity effects in the literature (relevant to chronic exposures) are largely inconsistent with fish-based values derived via the newly developed approach around decline in fish community integrity. This approach relies of modelled reference state condition (via GLMM) and modelled median values for suspended sediment measures (turbidity and visual clarity), and uncertainty of the model approach was not quantified. Furthermore, the approach was developed to use biological data that is not generally suited to threshold derivation. Unlike macroinvertebrate data sets that include abundance data collected via standard methods, the fish database only has presence/absence data. For these reasons, we place more emphasis on macroinvertebrate-based estimates of C/D band thresholds.

Table 7-6: Selected literature thresholds targeting the lower end of the suspended sediment effects spectrum.

study	comment
Newcombe (2003)	reactive distance - EL50[a] = c. 7 NTU (brook, lake and rainbow trout)
Vogel Beauchamp (1999)	reactive distance LOEL[b] = 3 NTU (*Salvelinus namaycush*)
Quinn et al. (1992)	50% effects level (EL50) macroinvertebrates = 3.7 NTU
Lloyd (1987)	increase of 5 NTU (in cold, clear water stream) could reduce primary productivity by 3-13% high level of protection would be 5NTU above natural conditions for clear, cold water streams
Boubee et al. (1997)	avoidance response , estimated EL25[c] values of 6.7 and 6.5 NTU for banded kokopu and koaro, respectively
De Robertis et al. (2003)	5-10 NTU decreased rate at which sable fish pursue prey and the probability of capture
Cavanagh (2014)	21 day experiment tank trials, inānga, kōaro, eels and brown trout. Inānga showed a significant decrease in growth rates from 5 to 15 NTU.
Hay et al. (2006)	Predicted 50% reduction in the reactive distance of 520 mm brown trout at 10 NTU
Newcombe (2003)	Impact assessment model for fish – with duration exposures from 1 h to 11 months Severity score ranging 1-14; 1-3 = *slight impairment*; 4-8 *significant impairment* (feeding and other behaviour begin to change); 9-14 = *severely impacted*. 4 month duration: ‘3 to 4’ or ‘4 to 5’ transition is predicted to occur at 0.77 and 0.55 m (corresponding to c. 5 and 7 NTU), respectively. 11 month duration: ‘3 to 4’ or ‘4 to 5’or ‘5 to 6’ transition is predicted to occur at 1.1, 0.77 and 0.55 m (corresponding to c. 3-3.5, 5 and 7 NTU), respectively.

[a] EL50 = 50% effects limit. [b] LOEL lowest observed effects limit. [c]EL25 = 25% effects level

7.3.3 Proposed NOF attribute thresholds (ecosystem health) for suspended fine sediment in NZ streams and rivers

A proposed table in NPS-FM format for potential national objectives framework (NOF) implementation for visual clarity and/or turbidity in New Zealand streams and rivers is provided in Table 7-7. Our final recommendation for the proposed threshold was based on:

- macroinvertebrate-derived thresholds - 80% protection level from SSD curve;
- macroinvertebrate-derived thresholds from 20-30% effects levels of selected community metrics;
- relevant effects literature (Table 7-6)
- taxa richness, EPT taxa, EPT individuals;

- median reference state conditions indicate higher levels of suspended sediment in ‘warm’ as opposed to ‘cool’ climates – which although is not a driver of suspended sediment, to at least account for this empirical observation, we recommended introducing an ‘offset’ to the ‘general’ or ‘base’ C/D band threshold. This is described in more detail in Chapter 2, section 2.3.6). (e.g., respective median clarity of 2.0 and 1.0 NTU); and,
- expert opinion.

Based on these multiple lines of evidence with propose a suspended sediment C/D band threshold of 5 NTU, based on a median value from 2 years of monthly monitoring (24 samples). We also recommend a C/D band for visual clarity. This was based on converting the turbidity value via regression equations in Appendix D – this corresponded to a proposed C/D band of 0.85 m.

We emphasise that no single derived threshold was used to define the proposed turbidity value, this value was consistent with the 80% protection value derived from the macroinvertebrate SSD curve (4.3 NTU), but we also acknowledge that the 95 CI limits on this threshold value ranged between 1.4 and 8.1 NTU). Accordingly, it was imperative that the final proposed value was at least consistent with chronic effect threshold literature, which we belief the base value of 5 NTU is. The ‘C/D band state relates to the upper bound of a state (i.e., C-band) that generally represents a minimum safe level before an ecological tipping point (MfE 2014). Accordingly, we need to make sure that the level of the effect threshold applied to the data is consistent with this definition, and that the derived numbers are consistent with literature values and natural state variation.

Differentiating natural state variation

In the suspended sediment classification system (Chapter 2) both measured and modelled reference state levels of suspended sediment appear to be on average, higher (lower for visual clarity) for warm REC climate classes. We are hesitant to recommend ‘two classes’ based on climate (i.e., aggregated climate classes) because this is not a driver of suspended sediment. But we also recognise that the observation that ‘warm’ climates, on average, appear to have 0.5-1 NTU higher turbidity values at reference sites may have potential management that need to be flagged at this stage of attribute development. As a pragmatic step, we have recommended applying an ‘offset’ to the ‘general’ (or ‘base’) C/D threshold of 5 NTU for warm climate classes. Preliminary analyses and previous reference state determinations (McDowell et al. 2013, and Section 2.3, Chapter 2) indicate an ‘offset’ of around 0.5-1 NTU. We have recommended an offset of 1 NTU be applied.

For the proposed NOF attribute table for suspended sediment we have chosen to illustrate this offset as two separate ‘classes’; however, an alternative (and perhaps better) layout for the suspended sediment attribute table layout could have a single C/D threshold (for turbidity and visual clarity), with the ‘offset’ being applied to ‘warm’ climate classes via a table footnote (analogous to productive periphyton class). An advantage of this approach is that future refinements/amendments could be made to site criteria qualifying for an offset and the magnitude of the ‘offset’ as new knowledge/analyses are developed (multiple offsets/criteria could be accommodated if required). Such changes could be implemented without revisiting the classification system.

Based on an offset of 1 NTU for warm climate classes, resulting in a C/D threshold of 6 NTU (median based over 2 years of monthly data), the corresponding visual clarity C/D band threshold for warm climates would be 0.7 m (compared to 0.85 m for other sites). The regression used to convert turbidity C/D thresholds into visual clarity and provided in Appendix D.

Table 7-7: Proposed NOF attribute table for assessing suspended fine sediment in streams and rivers.

Value	Ecosystem Health				
Freshwater Body Type	Rivers				
Attribute	Suspended fine sediment quantity (Surrogate measures: visual clarity or turbidity)				
Attribute Unit	Visual clarity, m (metres); turbidity, NTU (Nephelometric Turbidity Units)				
Attribute State	**Numeric Attribute State: Visual clarity (m)** [1]		**Numeric Attribute State: Turbidity (NTU)** [1]		**Narrative Attribute State**
	Annual median[2,3]		Annual median[2,3]		
	'cool'[4]	'warm'[4]	'cool'[4]	'warm'[4]	
A					High level of protection corresponding to reference site conditions.
B					Protects biodiversity measures, such as species taxonomic richness and EPT richness, and sensitive species from >30% impact.
C	>0.85	>0.7	<5	<6	Protects biodiversity measures from >30% impact.
National Bottom Line[5]	0.85	0.7	5	6	
D	<0.85	<0.7	>5	>6	High likelihood of loss of sensitive species and marked reduction in biodiversity. High probability of extirpation of sensitive macroinvertebrate species.

6. Classes are for all wadeable streams and rivers with the following exclusions: (i) 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 inorganic sediment from the catchment).
7. Based on a monthly monitoring regime. The minimum number of samples is 20, this will generally mean that assessment again the thresholds will require 2 years of monthly data, or 5 years of quarterly data.
8. Interconversion of visual clarity and turbidity is acceptable as derivation based on database of annual median data for these parameters (i.e., not concurrent instantaneous measurements). The more sensitive of the visual clarity or turbidity measures will determine the site grading. Visual clarity will be a more sensitive measure of changes in river particulate organic material and inorganic SS in high quality (i.e., low turbidity) waters.
9. Aggregated REC climate classes: 'cool' consists of cool dry (CD), cool wet (CW) and cool extremely wet (CX); 'warm' consists of warm dry (WD), warm wet (WW) and warm extremely wet (WX). Suspended sediments summary statistics from minimally disturbed condition sites (Depree 2017) and predicted reference states (McDowell et al. 2013) indicate that 'cool' and 'warm' sites have respective median turbidity values of 1 and 2 NTU; 'cool' and 'warm' median visual clarity values for reference sites are around 3.5 and 2.5 m respectively. Insufficient data was available in the macroinvertebrate database to derive distinct thresholds for the cool and warm classes. Differentiation based on differences in reference conditions.
10. Visual clarity values based on average of 2 regression equations between visual clarity and turbidity. 1) all NRWQN data (n=76), 2) all SoE monitoring data (n=722) (refer to Stage 1B, Depree 2017).

Recommended criteria for assessing state against proposed thresholds

Estimates of medians and 95%iles become more precise as the number of samples is increased, assuming that the sampling programme is bias-free (McBride 2014). McBride (2014) showed that based on the improvements in confidence interval curves (Figure 7-3), that somewhere in the region of 20 to 40 samples, one reaches an area of rapidly diminishing returns, and was recommended as a suitable sample size for defining medians (McBride 2014).

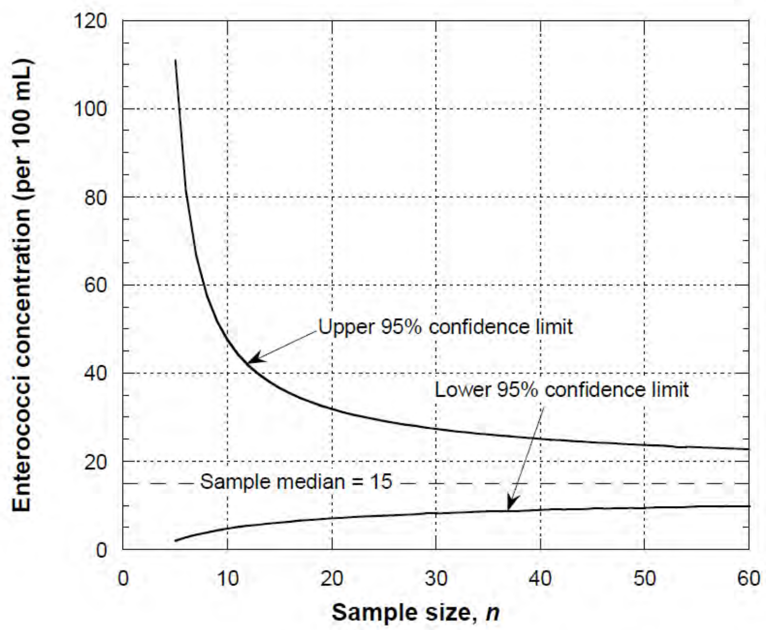


Figure 7-3: Effect of sample size on confidence limits for the median. Taken from McBride (2014); original source McBride 2005 (Figure 3.1)

Accordingly, for turbidity and visual clarity, we recommend sample sizes of at least 20, which for monthly and quarterly monitoring programmes, the minimum duration is 2 and 5 years, respectively. Monthly monitoring is recommended. Although data used in this report used long-term median from up to 10 years of data, such long periods are not recommended as presumably reporting by RC against threshold will require a duration more representative of current state. We consider a monitoring period of 2 or 3 years (n=24, 36) provides an acceptable balance between data requirements and a duration that is representative of current state. If using a 3-year period, then for each year the assessment is down, the median used would be calculated from the preceding 3-year period.

Indicative compliance with proposed C/D band thresholds for turbidity and visual clarity

The water quality data set comprised median visual clarity observed at 722 sites and median turbidity observed at 833 sites. Applying the proposed thresholds (Table 7-4), the overall estimated number of D-band sites was around 20%. If applying 'cool' thresholds uniformly across all environments, the proportion of D-band sites increased to around 25% (irrespective of the SS proxy measure).

For aggregated climate classes ('cool' and 'warm'), using turbidity the proportion of D-bands for cool (5 NTU) and warm (6 NTU) classes was 15 and 38%, respectively. Applying a single turbidity threshold of 5 NTU, increased the proportion of warm D-band sites from 38 to 51%. Using visual clarity, the proportion of cool and warm D-band sites was 15 and 30%, with the latter increasing to 41% if 0.85 m was uniformly across the network.

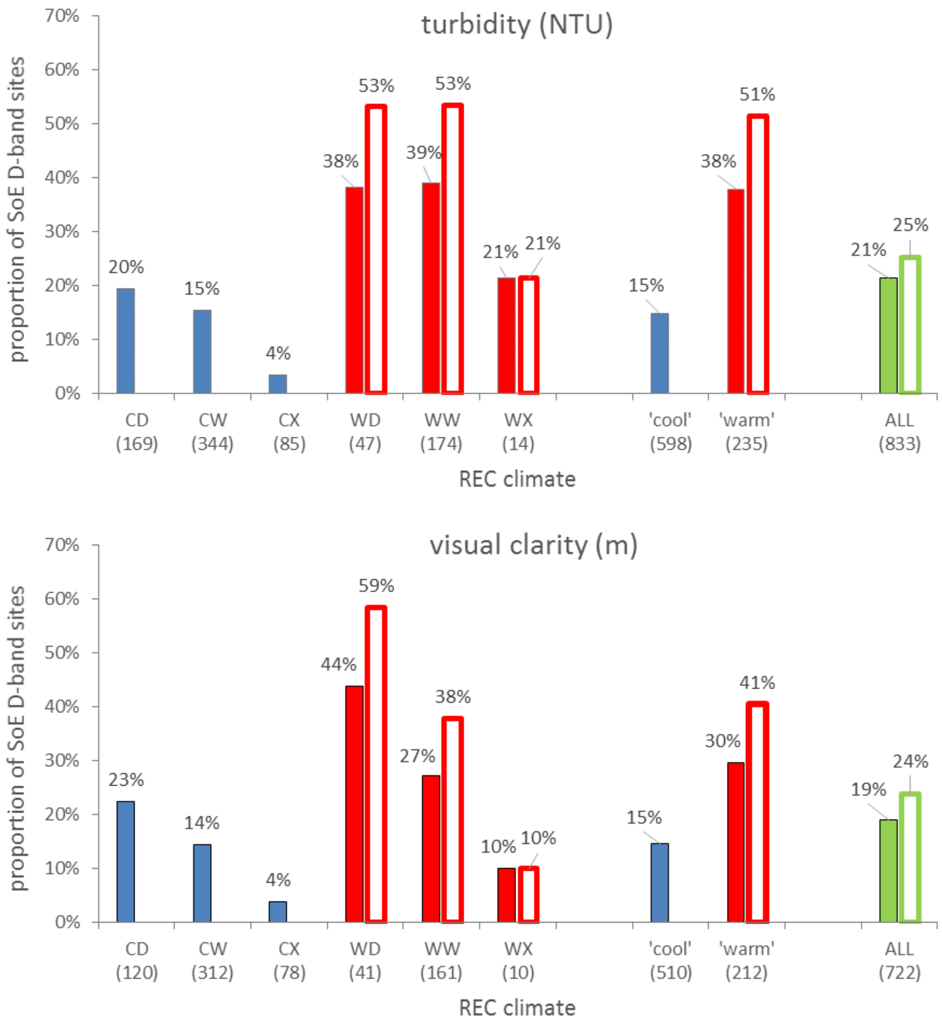


Figure 7-4: Compliance of sites (turbidity (top) n=833 and visual clarity (bottom n=722) with the proposed C/D band thresholds (national bottom lines) for the suspended sediment attribute. Sites are group by REC climate class, and aggregated 'cool' (blue) and 'warm' (red) categories. The proportion of D-band sites across all data is shown in green. The ‘hollow’ bars show the proportion of D-bands if the ‘cool’ threshold were applied across all climate class (i.e., 5 NTU and 0.85 m).

8 Acknowledgements

We gratefully acknowledge data provided by:

- Northland Regional Council, Auckland Regional Council, Waikato Regional Council, Bay of Plenty Regional Council, Gisborne District Council, Hawkes 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. Taranaki and Canterbury. With the 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 are thankful to Mark Newton, Jacob Thomson-Laing, Alasdair Macdonald, and Mark Hamer for assisting in the deposited sediment field study.

We are very grateful for the R-software input by Theo Mouton to the macroinvertebrate response to suspended sediments component of this project. We appreciate the assistance of Shan Crow in collating and processing the NZFFD data used in the study.

Finally, we thank John Quinn for his invaluable experience in reviewing this report, and Neale Hudson for his substantial input as project director (and a valued colleague to ‘bounce’ ideas off).

9 References

Akbaripasand A, Nichol EC, Lokman PM, Closs GP 2011. Microhabitat use of a native New Zealand galaxiid fish, *Galaxias fasciatus*. New Zealand Journal of Marine and Freshwater Research 45(1): 135-144.

Alabaster JS, Lloyd R 1982. Water quality criteria for freshwater fish (2nd edn). Boston, Butterworth Scientific.

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. October 2000 ed. Canberra, Australia, 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.

Ausseil, O. and Clark, M. (2007). Recommended Water Quality Standards for the Manawatu Wanganui Region: Technical Report to support policy development. Horizon's Regional Council, Report No. 2007/EXT/806.

Azrina MZ, Yap CK, Rahim Ismail A, Ismail A, Tan SG 2006. Anthropogenic impacts on the distribution and biodiversity of benthic macroinvertebrates and water quality of the Langat River, Peninsular Malaysia. Ecotoxicology and Environmental Safety 64(3): 337-347.

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

Baker CF 2003. Effect of adult pheromones on the avoidance of suspended sediment by migratory banded kokopu juveniles. Journal of Fish Biology 62(2): 386-394.

Baker ME, King RS 2010. A new method for detecting and interpreting biodiversity and ecological commun ity thresholds. Methods in Ecology and Evolution 1 (1): 25-37.

Baker ME, King RS 2013. Of TITAN and straw men: an appeal for greater understanding of community data. Freshwater Science 32 (2): 489-506.

Barrett JC, Grossman GD, Rosenfeld J 1992. Turbidity-Induced Changes in Reactive Distance of Rainbow Trout. Transactions of the American Fisheries Society 121(4): 437-443.

Barter PJ, Deas D 2003. Comparison of portable nephelometric turbidimeters on natural water and effluents. New Zealand Journal of Marine and Freshwater Research 37: 485-492.

Bilotta GS, Brazier RE 2008. Understanding the influence of suspended solids on water quality and aquatic biota. Water Research 42(12): 2849-2861.

Bilotta, G.S., Brazier, R.E. (2008) Understanding the influence of suspended solids on water quality and aquatic biota. *Water Research 42*: 2849–2861.

Booker DJ, Dunbar M, Ibbotson AT 2004. Predicting juvenile salmonid drift-feeding habitat quality using a three-dimensional hydraulic-bioenergetic model. Ecological Modelling 177: 157-177.

Booker, D.J. (2010) Predicting wetted width in any river at any discharge. *Earth Surface Processes and Landforms*, 35: 828–841.

Booker, D.J. (2013) Spatial and temporal patterns in the frequency of events exceeding three times the median flow (FRE3) across New Zealand. *Journal of Hydrology (New Zealand)*, 52(1): 15–39.

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(0): 227–239.

Booker, D.J., Snelder, T.H., Greenwood, M.J., Crow, S.K. (2014). Relationships between invertebrate communities and both hydrological regime and other environmental factors across New Zealand's rivers. *Ecohydrology,* 8: 13–32.

Booker DJ, Snelder TH, Greenwood MJ, Crow SK 2014. Relationships between invertebrate communities and both hydrological regime and other environmental factors across New Zealand's rivers. Ecohydrology 8: 13-32.

Boubee JAT, Dean TL, West DW, Barrier RFG 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.

Brain P, Cousens R 1989. An Equation to Describe Dose Responses Where There Is Stimulation of Growth at Low-Doses. Weed Research 29(2): 93-96.

Bryce SA, Lomnicky GA, Kaufmann PR 2010. Protecting sediment-sensitive aquatic species in mountain streams through the application of biologically based streambed sediment criteria. Journal of the North American Benthological Society 29(2): 657-672.

Burdon FJ, McIntosh AR, Harding JS 2013. Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. Ecological Applications 23(5): 1036-1047.

Cade BS, Guo QF 2000. Estimating effects of constraints on plant performance with regression quantiles. Oikos 91(2): 245-254.

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

Campbell E, Palmer MJ, Shao Q, Wilson D 2000. BurrliOZ: a computer program for calculating toxicant trigger values for the ANZECC and ARMCANZ water quality guidelines.

Cavanagh JE, Hogsden KL, Harding J 2014. Effects of suspended sediment on freshwater fishLC1986. 29 p.

Cedergreen N, Ritz C, Streibig JC 2005. Improved empirical models describing hormesis. Environmental Toxicology and Chemistry 24(12): 3166-3172.

Chapman JM, Proulx CL, Veilleux MAN, Levert C, Bliss S, André M-È, Lapointe NWR, Cooke SJ 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.

Chessman BC 2003. New sensitivity grades for Australian river macroinvertebrates. Marine and Freshwater Research 54: 95-103.

Clapcott J 2016. Effects of the Waihi Dam failure on the benthic invertebrates of the Waiau River. Prepared for Hawkes Bay Regional Council and Ministry for the Environment. 22 p.

Clapcott J, Young RG, Harding JS, Matthaei CD, Quinn JM, Death RG 2011a. Sediment Assessment Methods: Protocols and guidelines for assessing the effects of deposited fine-sediment on in-stream values. 108 p.

Clapcott J, Wagenhoff A, Neale M, Death R, Storey R, Smith B, Harding J, Matthaei C, Collier K, Quinn J, Young R 2017. Macroinvertebrate metrics for the National Policy Statement for Freshwater Management project: Report 1. Prepared for Ministry for the Environment. Cawthron Report No. 3012. 99 p.

Clapcott JE 2017. EMaR habitat and fine sediment survey. Prepared for Hawke's Bay Regional CouncilCawthron Report No. 3007. 16 p.

Clapcott JE, Goodwin EO 2017. Technical report on developing a deposited sediment classification for New Zealand streams. Prepared for Ministry for the EnvironmentCawthron Report No. 2994. 36 p.

Clapcott JE, Young RG, Harding JS, Matthaei CD, Quinn JM, Death RG 2011b. Sediment Assessment Methods: Protocols and guidelines for assessing the effects of deposited fine sediment on in-stream values. Nelson, New Zealand, Cawthron Institute. 94p. plus appendices

Clapcott JE, Collier KJ, Death RG, Goodwin EO, Harding JS, Kelly D, Leathwick JR, Young RG 2012. Quantifying relationships between land-use gradients and structural and functional indicators of stream ecological integrity. Freshwater Biology 57(1): 74-90.

Cohen J 1960. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20(1): 37-46.

Collins AL, Naden PS, Sear DA, Jones JI, Foster IDL, Morrow K 2011. Sediment targets for informing river catchment management: international experience and prospects. Hydrological Processes 25(13): 2112-2129.

Conover WJ 1971. Practical Nonparametric Statistics. New York, John Wiley & Sons.

Cormier SM, Paul JF, Spehar RL, Shaw-Allen P, Berry WJ, Suter GW 2008. Using field data and weight of evidence to develop water quality criteria. Integrated Environmental Assessment and Management 4(4): 490–504.

Crisp DT, Carling PA 1989. Observations on siting, dimensions and structure of salmonid redds. Journal of Fish Biology 34(1): 119-134.

Crosa G, Castelli E, Gentili G, Espa P 2010. Effects of suspended sediments from reservoir flushing on fish and macroinvertebrates in an alpine stream. Aquatic Sciences 72(1): 85.

Crow SK, Booker DJ, Snelder TH 2013. Contrasting influence of flow regime on freshwater fishes displaying diadromous and nondiadromous life histories. Ecology of Freshwater Fish 22(1): 82-94.

Crow SK, Booker D, Sykes J, Unwin M, Shankar U 2014. Predicting distributions on New Zealand freshwater fishesCHC2014-145. 100 p.

Crow SK, Snelder T, Jellyman PG, Greenwood MJ, Dunn A 2016. Temporal trends in the relative abundance of New Zealand freshwater fishes: Analysis of New Zealand Freshwater Fish Database recordsCHC2016-049.

Crowe A, Hay J 2004. Effects of fine sediment on river biota951. 40 p.

Cuffney T, Qian S, Brightbill R, May J, Waite I 2011. Response to King and Baker: limitations on threshold detection and characterization of community thresholds. Ecological Applications 21 (7): 2840-2845.

Cuffney TF, Qian SS 2013. A critique of the use of indicator-species scores for identifying thresholds in species responses. Freshwater Science: 471-488.

Culp JM, Benoy GA, Brua RB, Sutherland AB, Chambers PA 2009. Total suspended sediment, turbidity and deposited sediment standards to prevent excessive sediment effects in Canadian streams.

Cumming H, Herbert NA 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.

Davies-Colley RJ 2000. "Trigger" values for New Zealand rivers (http://www.mfe.govt.nz/publications/water/trigger-values-rivers-may00/trigger-values-rivers-may00.html). 4+App p.

Davies-Colley, R.J., Ballantine, D.J., Elliot, S.H., Swales, A., Hughes, A.O., Gall, M.P. (2014). Light attenuation – a more effective basis for the management of fine sediment than mass concentration? *Waster Science & Technology*. 69(9): 1867-1874.

Davies-Colley, R.J., Nagels, J.W. (2008). Predicting light penetration into river waters. *Journal of Geophysical Research-Biogeosciences 113*: G03028. doi10.1029/2008JG000722.

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

Davies-Colley RJ, Hickey CW, Quinn JM, Ryan PA 1992. Effects of clay discharges on streams, 1. Optical properties and epilithon. Hydrobiologia 248: 215-234.

Davies-Colley RJ, 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 reviewHAM2015-104. 105 p.

De'ath G 2007. Boosted trees for ecological modeling and prediction. Ecology 88(1): 243-251.

De'ath G, Fabricius KE 2000. Classification and regression trees: a powerful yet simple technique for ecologial data analysis. Ecology 81(11): 3178-3192.

Depree C 2017. Sediment Attribute Stage 1b. Proposed classification system. NIWA Client Report prepared for Ministry for the Environment. March 2017. 75 p.

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

Duarte, C.M (1991). Seagrass depth limits. *Aquatic Botany* 40: 363-377.

EIFAC 1964. Water quality criteria for European freshwater fish. Report on finely divided solids and inland fisheries. 21 p.

Elith J, Leathwick J 2014. Boosted Regression Trees for ecological modeling. Vignette for R dismo package version 1.0-5. http://cran.r-project.org/web/packages/dismo/.

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

Ellis N, Smith SJ, Pitcher CR 2012. Gradient forests: calculating importance gradients on physical predictors. Ecology 93(1): 156-168.

Environment Agency 2011. Freshwater Fisheries Directive (78/659/EEC) (2006/44/EC) (<http://www.environment-agency.gov.uk/business/regulation/31955.aspx) <http://www.environment-agency.gov.uk/business/regulation/31955.aspx>.

Extence CA, Chadd RP, England J, Dunbar MJ, Wood PJ, Taylor ED 2013. The assessment of fine sediment accumulation in rivers using macro-invertebrate community response. River Research and Applications 29(1): 17-55.

Freedman PE, Dilks DW, Holmberg HP, Moskus PE, McBride GB, Hickey CW, Smith DG, Striplin PL 2003. Narrative criteria in the TMDL process: Method development for addressing narrative criteria. 107 p.

Gerster S, Rey P 1994. Consequences ecologiques des curages des bassin de retenue.

Glendell M, Extence C, Chadd R, Brazier RE 2014. Testing the pressure-specific invertebrate index (PSI) as a tool for determining ecologically relevant targets for reducing sedimentation in streams. Freshwater Biology 59(2): 353-367.

Goodman JM, Dunn NR, Ravenscroft P, Allibone R, Boubée J, David B, Griffiths M, Ling N, Hitchmough R, Rolfe JR 2014. Conservation status of New Zealand freshwater fish, 20137. 16 p.

Grace JB, Adler PB, Harpole WS, Borer ET, Seabloom EW 2014. Causal networks clarify productivity-richness interrelations, bivariate plots do not. Functional Ecology 28(4): 787-798.

Gray LJ, Ward JV 1982. Effects of sediment releases from a reservoir on stream macroinvertebrates. Hydrobiologia 96(2): 177-184.

Greer MJC 2014. The effects of macrophyte control on freshwater fish communities and water quality in New Zealand streams. Unpublished thesis, University of Otago, Dunedin. 158 p.

Greer MJC, Crow SK, Hicks AS, Closs GP 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.

Greet J, Webb JA, Cousens RD 2011. The importance of seasonal flow timing for riparian vegetation dynamics: a systematic review using causal criteria analysis. Freshwater Biology 56: 1231-1247.

Gregory RS 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.

Gregory RS, Levings CD 1998. Turbidity Reduces Predation on Migrating Juvenile Pacific Salmon. Transactions of the American Fisheries Society 127(2): 275-285.

Grove MK, Bilotta GS, Woockman RR, Schwartz JS 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.

Harrison ET 2010. Fine Sediment in Rivers: Scale of Ecological Outcomes. Unpublished thesis, University of Canberra, Canberra, Australia.

Harvey BC, White JL 2008. Use of Benthic Prey by Salmonids under Turbid Conditions in a Laboratory Stream. Transactions of the American Fisheries Society 137(6): 1756-1763.

Hay J 2004. Survival of trout eggs in relation to sediment composition of redds in Waikakahi Stream. 965.

Hayes J, Stark JD, Shearer KA 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.

Hayes JW, Goodwin E, Shearer KA, 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.

Henley WF, Patterson MA, Neves RJ, Lemly AD 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 DW, Merkins JC 1961. The effect of suspended sediment mineral solids on the survival of trout. International Journal of Air and Water Pollution 5: 46-55.

Hickey CW 2013. Updating nitrate toxicity effects on freshwater aquatic species1207-ESRC255. 39 p.

Hickey CW, Golding LA 2002. Response of macroinvertebrates to copper and zinc in a stream mesocosm. Environmental Toxicology and Chemistry 21(9): 1854-1863.

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

Hicks DM, Clapcott J, Davies-Colley R, Dymond J, Greenwood M, Hughes A, Shankar U, Walter K 2016. Sediment Attributes Stage 1. Prepared for Ministry for the Environment. NIWA Client Report No: CHC2016-058. 189 p.

Holliday, C.P., Rasmussen, T.C. and Miller, W.P. (2003). In *Proceedings of the 2003 Georgia Water Resources Conference*, University of Georgia (April 23-24, 2003). Ed. Kathryn J. Hatcher. Institute of Ecology, University of Georgia, Athens, Georgia.

Jones JI, Murphy JF, Collins AL, Sear DA, Naden PS, Armitage PD 2012. The impact of fine sediment on macro-invertebrates. River Research and Applications 28(8): 1055-1071.

Jowett IG, 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.

Jowett IG, Richardson J 2008. Habitat use by New Zealand fish and habitat suitability models55. 147 p.

Jowett IG, Richardson J, McDowall RM 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.

Julian, J.P., Doyle, M.W., Stanley, E.H. (2008) Empirical modeling of light availability in rivers. *Journal of Geophysical Research-biogeosciences*, 113: G03022. doi:10.1029/2007JG000601

Kail J, Arle J, Jaehnig SC 2012. Limiting factors and thresholds for macroinvertebrate assemblages in European rivers: Empirical evidence from three datasets on water quality, catchment urbanization, and river restoration. Ecological Indicators 18: 63-72.

Kefford BJ, Zalizniak L, Dunlop JE, Nugegoda D, Choy SC 2010. How are macroinvertebrates of slow flowing lotic systems directly affected by suspended and deposited sediments? Environmental Pollution 158(2): 543-550.

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

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

Khan KS, 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.

King RS, Baker ME 2010. Considerations for analyzing ecological community thresholds in response to anthropogenic environmental gradients. Journal of the North American Benthological Society 29(3): 998-1008.

King RS, Baker ME 2011. An alternative view of ecological community thresholds and appropriate analyses for their detection: comment. Ecological Applications 21 (7): 2833-2839.

Kjelland ME, Woodley CM, Swannack TM, Smith DL 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.

Koenker R 2013. quantreg: Quantile Regression. R package version 5.05.

Koenker R, Machado JAF 1999. Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association 94(448): 1296-1310.

Koenker R, Hallock K 2001. Quantile regression: An introduction. Journal of Economic Perspectives 15(4): 43-56.

Kusabs IA, Quinn JM, Hamilton DP 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.

Lake RG, Hinch SG 1999. Acute effects of suspended sediment angularity on juvenile coho salmon (Oncorhynchus kisutch). Canadian Journal of Fisheries and Aquatic Sciences 56(5): 862-867.

Lange K, Townsend CR, Gabrielsson R, Chanut PCM, Matthaei CD 2014. Responses of stream fish populations to farming intensity and water abstraction in an agricultural catchment. Freshwater Biology 59(2): 286-299.

Leathwick JR, Julian K, Elith J, Rowe DK 2008. Predicting the distributions of freshwater fish species for all New Zealand's rivers and streamsHAM2008-005. 56 p.

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.

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.

Manel S, Williams HC, Ormerod SJ 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38(5): 921-931.

Matheson F, Quinn J, Hickey C 2012. Review of the New Zealand instream plant and nutrient guidelines and development of an extended decision making framework: Phases 1 and 2 final report (http://www.envirolink.govt.nz/Envirolink-tools/). 103 p.

Matheson F, Quinn J, Unwin M 2015. Instream plant and nutrient guidelines: Review and development of an extended decision-making framework Phase 3 (http://www.envirolink.govt.nz/Envirolink-tools/). 117 p.

McDowall RM 1990. New Zealand freshwater fishes: A natural history and guide. Auckland, Heinemann Reed. 553 p.

McDowell RW, Snelder TH, Cox N, Booker DJ, Wilcock RJ 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.

McEwan AJ 2009. Fine scale spatial behaviour of indigenous riverine fish in a small New Zealand stream. Unpublished thesis, Massey University, Palmerston North. 95 p.

McEwan AJ, Joy MK 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.

McEwan AJ, Joy MK 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.

MfE 2014. National Policy Statement for Freshwater Management 2014: Draft Implementation guide. Wellington: Ministry for the Environment.

MfE 2014b. National Policy Statement for Freshwater Management 2014. pp. 34. Ministry for the Environment, Wellington, New Zealand

Mitchell SB, West JR, Guymer I 1999. Dissolved-Oxygen/Suspended-Solids Concentration Relationships in the Upper Humber Estuary. Water and Environment Journal 13(5): 327-337.

Nash JE, Sutcliffe JV 1970. River flow forecasting through conceptual models part I - A discussion of principles. Journal of Hydrology (NZ) 10(3): 282-290.

NEMS 2016. Measurement, Processing and Archiving of Turbidity Data. Final Draft Version: 1.1 National Environmental Monitoring Standard Turbidity. Date of Issue: October 2016 (https://www.lawa.org.nz/media/3198184/nems-turbidity-final-draft-version-as-at-29-aug_sent-fior-lawa-prior-to-workshop-.pdf - downloaded 26 May 2018).

Newcombe CP 2003. Impact assessment model for clear water fishes exposed to excessively cloudy water. JAWRA Journal of the American Water Resources Association 39(3): 529-544.

Newcombe CP, Macdonald DD 1991. Effects of Suspended Sediments on Aquatic Ecosystems. North American Journal of Fisheries Management 11(1): 72-82.

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

Norris RH, Webb JA, Nichols SJ, Stewardson MJ, Harrison ET 2012. Analyzing cause and effect in environmental assessments: using weighted evidence from the literature. Freshwater Science 31: 5-21.

Nuttall PM, Bielby GH 1973. The effect of China-clay wastes on stream invertebrates. Environmental Pollution (1970) 5(2): 77-86.

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 PN, Batalla RJ, Collins AJ, Gomez B, Hicks DM, Horowitz AJ, Kondolf GM, Marden M, Page MJ, Peacock DH, Petticrew EL, Salomons W, Trustrum NA 2005. Fine-grained sediment in river systems: environmental significance and management issues. River Research and Applications 21(7): 693-717.

Piggott JJ, Townsend CR, Matthaei CD 2015. Climate warming and agricultural stressors interact to determine stream macroinvertebrate community dynamics. Global Change Biology 21(5): 1887-1906.

Piñeiro G, Perelman S, Guerschman J, Paruelo J 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological Modelling 216: 316-322.

Qian SS 2014. Ecological threshold and environmental management: A note on statistical methods for detecting thresholds. Ecological Indicators 38 (0): 192-197.

Quinn JM 2000. Effects of pastoral development. In: Collier KJ, Winterbourn MJ ed. New Zealand stream invertebrates : ecology and implications for management. Hamilton, NZ, New Zealand Limnological Society, NIWA. Pp. 208-229.

Quinn JM, Hickey CW 1990a. Characterisation and classification of benthic invertebrate communities in 88 New Zealand rivers in relation to environmental factors. New Zealand Journal of Marine and Freshwater Research 24(3): 387-409.

Quinn JM, Hickey CW 1990b. Magnitude of effects of substrate particle size, recent flooding, and catchment development on benthic invertebrates in 88 New Zealand rivers. New Zealand Journal of Marine and Freshwater Research 24(3): 411-427.

Quinn JM, Vickers ML 1992. Benthic invertebrates and related habitat factors in the Tongariro RiverNo 6025/2.

Quinn JM, Hickey CW 1993. Effects of sewage stabilization lagoon effluent on stream invertebrates. Journal of Aquatic Ecosystem Health 2: 205-219.

Quinn JM, Croker GF, Smith BJ, Bellingham MA 2009. Integrated catchment management effects on runoff, habitat, instream vegetation and macroinvertebrates in Waikato, New Zealand, hill-country streams. New Zealand Journal of Marine and Freshwater Research 43: 775-802.

Quinn JM, Davies-Colley RJ, Hickey CW, Vickers ML, Ryan PA 1992. Effects of clay discharges on streams 2. Benthic invertebrates. Hydrobiologia 248(3): 235-247.

R Core Team 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

Regional Government of Piemonte 2008. Reg. 29 January 2008, n. 1/R. In Italian.

Regional Government of Veneto 2006. DGR n.138, In Italian.

Reid D, Quinn J 2011a. Preliminary information for developing sediment guidelines for streams of the West Coast, New Zealand. 21 p.

Reid D, Quinn J 2011b. Preliminary information for developing sediment guidelines for streams of the West Coast, New ZealandHAM2011-012. 21 p.

Ritz C, Streibig JC 2005. Bioassay analysis using R. Journal of Statistical Software 12(5): 1-22.

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 DM, Wiens AP 1978. Effects of sedimentation on macrobenthic invertebrates on a northern Canadian river. Water Research 12: 753–763.

Rowe DK, Dean TL 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.

Rowe DK, 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.

Rowe DK, 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.

Rowe DK, 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.

Rowe DK, Suren AM, Martin M, Smith J, Smith B, Williams E 2002b. Lethal turbidity levels for common freshwater fish and invertebrates in Auckland streams337. 37 p.

Ryan PA 1991a. Environmental effects of sediment on New Zealand streams: a review. New Zealand Journal of Marine and Freshwater Research 25: 207-221.

Ryan PA 1991b. Environmental effects of sediment on New Zealand streams: a review. New Zealand journal of marine and freshwater research 25(2): 207-221.

Scarsbrook MR, Boothroyd IKG, Quinn JM 2000. New Zealand's National River Water Quality Network: long-term trends in macroinvertebrate communities. New Zealand Journal of Marine and Freshwater Research 34(2): 289-302.

Schneider J, Badura H, Knoblauch H, Frei R, Eberstaller J 2006. Flushing of the run-of river plant Bodendorf in Styrria/Austria in respect of technical and ecological impacts. The 7th international conference on hydroscience and engineering (ICHE-2006).

Shaw EA, Richardson JS 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.

Slaney PA, Halsey TG, Tautz AF 1977. Effects of forest harvesting practices on spawning habitat of stream salmonids in the Centennial Creek watershed, British Columbia. 35 p.

Snelder T, Biggs BJF 2002a. Multi-scale river environment classification for water resources management. Journal of the American Water Resources Association 38(5): 1225-1239.

Snelder T, Biggs BJF 2002b. Multi-Scale River Environment Classification for Water Resources Management. Journal of the American Water Resources Association 38: 1225-1239.

Snelder T., Biggs B. & Weatherhead M. (2004) New Zealand River Environment Classification User Guide. Ministry for the Environment. 145p.

Snelder TH, Biggs BJF, Woods RA 2005. Improved eco-hydrological classification of rivers. River Research and Applications 21(6): 609-628.

Stark J, Boothroyd I, Harding J, Maxted J, Scarsbrook M 2001. Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate Working Group Report No. 1. Prepared for the Ministry for the Environment, Sustainable Management Fund Project No. 5103. 57 p.

Stark JD 1985. A macroinvertebrate community index of water quality for stony streams Water and Soil Miscellaneous Publication 87: 1-52.

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 p.

Sundermann A, Leps M, Leisner S, Haase P 2015. Taxon-specific physico-chemical change points for stream benthic invertebrates. Ecological Indicators 57 (0): 314-323.

Suren AM, Martin ML, Smith BJ 2005. Short-Term Effects of High Suspended Sediments on Six Common New Zealand Stream Invertebrates. Hydrobiologia 548(1): 67-74.

Sutherland AB, Meyer JL 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.

Suttle KB, Power ME, Levine JM, McNeely C 2004. How fine sediment in riverbeds impairs growth and survival of juvenile salmonids. Ecological Applications 14(4): 969-974.

Sweka JA, Hartman KJ 2001. Influence of turbidity on brook trout reactive distance and foraging success. Transactions of the American Fisheries Society 130(1): 138-146.

Taylor JM, King RS, Pease AA, Winemiller KO 2014. Nonlinear response of stream ecosystem structure to low-level phosphorus enrichment. Freshwater Biology 59 (5): 969-984.

Townsend CR, Uhlmann SS, Matthaei CD 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45(6): 1810-1819.

Uncles RJ, Joint I, Stephens JA 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.

Unwin M, Larned ST 2013. Statistical models, indicators and trend analyses for reporting national-scale river water qualityCHC2013-033. 71 p.

US EPA 1976. Quality criteria for water (Red book).

US EPA 2000. Stressor identification guidance document. 228 p.

US EPA 2006. Framework for developing suspended and bedded sediments (SABS) water quality criteria. 168 p.

US EPA 2011. A Field-Based Aquatic Life Benchmark for Conductivity in Central Appalachian Streams (Final Report). 276 p.

US EPA 2014. Caddis - sediment stressor response (https://www3.epa.gov/caddis/ssr_sed4d.html). Retrieved January 2014, updated July 2010: http://www.epa.gov/caddis/ssr_urb_intro.html

US EPA 2017. CADDIS - stressor response analysis: Quantile regression (https://www3.epa.gov/caddis/da_basic_3.details.html). Retrieved June 2017. https://www3.epa.gov/caddis/da_basic_3.details.html

Usio N, Townsend CR 2001. The significance of the crayfish paranephrops zealandicus as shredders in a new zealand headwater stream. Journal of Crustacean Biology 21(2): 354-359.

Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE 1980. River Continuum Concept. Canadian Journal of Fisheries and Aquatic Sciences 37(1): 130-137.

Vant B 2011. Water quality of the Hauraki Rivers and Southern Firth of Thames2011/06. 40 p.

Vant B 2013. Water quality of the lower Piako River 2011-132013/15. 20 p.

Velleman PF 1989. The Data Desk handbook. Version 6.3. Ithica, New York, Data Description Inc.

Vitali R, G. B, E. I 1995. Enel experience in the environmental management of maintenance of hydroelectric reservoirs. Hydroe´col Appl 7: 51–74.

Wagener SM, LaPerriere JD 1985. Effects of Placer Mining on the Invertebrate Communities of Interior Alaska Streams. Freshwater Invertebrate Biology 4(4): 208-214.

Wagenhoff A, Townsend CR, Matthaei CD 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 CR, Phillips N, Matthaei CD 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.

Wagenhoff A, Clapcott J, Lau KEM, Lewis GD, Young RG 2017a. Identifying congruence in stream assemblage thresholds in response to nutrient and sediment gradients for limit setting. Ecological Applications 27(2): 469-484.

Wagenhoff A, Liess A, Pastor A, Clapcott JE, Goodwin EO, Young RG 2017b. Thresholds in ecosystem structural and functional responses to agricultural stressors can inform limit setting in streams. Freshwater Science 36(1): 178-194

Warne MS, Batley GE, van Dam RA, Chapman JC, Fox DR, Hickey CW, Stauber JL 2015. Revised Method for Deriving Australian and New Zealand Water Quality Guideline Values for Toxicants. 43 p.

Waters TF 1995a. Sediment in streams: sources, biological effects, and control. Bethesda, MA, American Fisheries Society. 251 p.

Waters TF 1995b. Sediment in streams: Sources, biological effects and control. Bethesda, MD, American Fisheries Society. 268 p.

Webb AJ, Miller KA, King EL, de Little SC, Stewardson MJ, Zimmerman JKH, 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.

Webb JA, Miller KA, Stewardson MJ, de Little SC, Nichols SJ, Wealands SR 2015. An online database and desktop assessment software to simplify systematic reviews in environmental science. Environmental Modelling & Software 64: 72-79.

Wilson KA, Westphal MI, Possingham HP, Elith J 2005. Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation 122(1): 99-112.

Winterbourn MJ, Rounick JS, Cowie B 1981. Are New Zealand stream ecosystems really different? New Zealand Journal of Marine and Freshwater Research 15: 321-328.

Wolman MG 1954. A method of sampling coarse river-bed material. Transactions of the American Geophysical Union 35(6): 951-956.

Wood P, Armitage PD 1997a. Biological Effects of Fine Sediment in the Lotic Environment. Environmental Management 21(2): 203-217.

Wood PJ, Armitage PD 1997b. Biological Effects of Fine Sediment in the Lotic Environment. Environ Manage 21(2): 203-17.

Wood PJ, Armitage PD 1997c. Biological effects of fine sediment in the lotic environment. Environmental Management 21(2): 203-217.

Wood S 2004. Stable and efficient multiple smoothing parameter estimation for generalised additive models. Journal of the American Statistical Association 99: 673-686.

Zweig LD, Rabeni CF 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.

10 Glossary of abbreviations and terms

Detection limit (DL)	The value below which a laboratory cannot confidently distinguish the analyte concentration from zero. In practice, an index of the concentration below which relative precision declines markedly. A particular problem with TSS data at baseflow is that filtrations are often done on insufficient sample volume so that considerable proportions of datasets for TSS are approaching or below the DL.
Disturbance plume	A plume of turbid water produced by disturbance of fine sediment (silt and clay) deposited in the interstices of (much coarser) bed sediment in rivers. Such plumes, created by wading in channels, must be avoided for measurement of visual water clarity and water sampling for indices of SPM.
ESV	Environment State Variable: a variable that captures an aspect of the state of the physical, chemical, or ecological environment.
LAWA	Land, Air, Water, Aotearoa: a website displaying information for more than 1100 freshwater monitoring sites throughout New Zealand.
LCDB	New Zealand's Landcover Database v3. Classifies land cover across New Zealand in 33 different categories.
NRWQN	National River Water Quality Network. A monitoring network of 77 river sites run by NIWA since 1989, with an aggregate catchment about 50% of NZ's land area (Davies-Colley et al. 2011).
NEMAR	National Environmental Monitoring and Reporting
NSE	Nash-Sutcliffe Efficiency: a measure of fit between observed values and model predictions. NSE ranges from $-\infty$ to 1, with 1 indicating a perfect match to predictions, 0 indicating that predictions are as accurate as the mean of the observed data, and negative values indicating that the observed mean is a better predictor than the model.
NZSegment	Individual river segment within REC2, with associated environmental information available. Segment boundaries occur at confluences.
REC	River Environment Classification
RF	Random Forest. A flexible regression technique in which final predictions are based on averages across an ensemble of regression trees.
RMSE	Root Mean Square Error. A measure of fit between observed values and model predictions. A lower RMSE indicates a better fit between observed and predicted values.
SAM1	Sediment Assessment Method 1: Bankside visual estimate of % sediment cover. Rapid qualitative assessment of the surface area of the streambed covered by sediment.
SAM2	Sediment Assessment Method 2: In-stream 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

SAM3	Sediment Assessment Method 3: Wolman pebble count. Semi-quantitative assessment of the particle size distribution, including fine sediment, on the streambed. At least 100 particle measurements are made within a single habitat
SAM4	Sediment Assessment Method 4: Resuspendible sediment (Quorer method). Quantitative measure of total suspendable solids deposited on the streambed. Six samples are collected from a single habitat. Samples are processed in the laboratory for Total Inorganic/Organic Sediment by areal mass and/or Suspendable Benthic Solids by Volume.
SAM5	Sediment Assessment Method 5: Resuspendible sediment (Shuffle index). Rapid qualitative assessment of the amount of total suspendable solids deposited on the streambed. A score from 1-5 is assigned, where 1 is little/no sediment and 5 is excessive sediment.
Sediment load	The mass flux of sediment delivered from a catchment (typically in t/yr).
Sediment yield	The sediment load per unit catchment area (typically in $t/km^2/yr$).
SOF	Source of Flow category from REC2 (derived for REC1).
SS	Suspended Sediment.
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.
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	measurement of the 'cloudiness' of water sample, measurement generally provided as NTU (nephelometric turbidity units)
Visual clarity	quantified by the black disc visibility (in the horizontal direction).
WRENZ	Water REsources of New Zealand: a GIS model

APPENDICES FOR SECTIONS 1 to 4

Appendix A Executive summaries from Stage 1B reports

Executive summary: Stage 1B report on suspended sediment (Depree 2017)

The purpose of Stage 1B was to develop a classification system that differentiates (i.e., classifies) New Zealand rivers according to "reference state" variation in environment state variable (ESV) characteristics.

Work was divided into suspended sediment (NIWA, this report) and deposited sediment (Cawthron), which is *reported separately* (Clapcott and Goodwin 2017).

To characterise reference state variation, we derived up-to-date land-cover information for the catchments of more than 800 sites where suspended sediment (SS) is measured. This was done using the latest version of the New Zealand Land-cover Database (v4.1). Using these data, a series of thresholds (or rules) were developed to define the upper bounds of minimally disturbed condition (reference) sites that were used to define the extent of 'reference state' variation. Having established the extent of reference state variation, determination of a classification system required to address natural state variation commenced. The approach followed made use of the six levels of information embedded in the River Environment Classification (REC) system.

The 'reference state variation' for suspended sediment (determined using reference site medians) was relatively small, with maximum values of 6.0 g/m^3, 1.1 m and 3.3 NTU for TSS, clarity and turbidity respectively. Limiting the natural state variation to the 95th percentile of reference sites, the respective maximum values for TSS, clarity and turbidity were 3.9 g/m^3, 1.5 m and 2.5 NTU; and median values were 2.0 g/m^3, 3.5 m and 0.8 NTU.

Warm REC climate classes had, on average, 2-times higher turbidity values. This finding was supported by the work of McDowell et al. (2013), who predicted reference site median values for 12 cool and 6 warm climate categories to be approximately 2.0 and 1.0 NTU, respectively. In the context of ecological relevance, however, this difference is arguably not meaningful – for example, >95% of reference sites had clarity that exceeded 1.4 m, with turbidity values less than 2.5 NTU. Other work has demonstrated these latter values to be conservative biological effect thresholds, based on reaction distances of trout during drift feeding.

On balance, I recommend to not classify (i.e., differentiate) flowing waters in New Zealand for the purpose of managing SS for ecosystem health. That is, we recommend proceeding without a classification system for suspended sediment.

With respect to redundancy of sediment ESVs (including turbidity)

Following extensive analysis of national data representing all regions, we concluded the following:

- Euphotic depth is largely irrelevant for the management of SS in flow waters;
- In general, TSS has a method detection limit that is too high to enable it to be used to determine reference state variation;
- Clarity and TSS are not routinely monitored by all councils;
- Turbidity is measured by all councils;

- Turbidity is more suited (relative to clarity) for managing in the 'dirtier end of the SS spectrum';
- Correlations between TSS and turbidity and TSS and clarity were relatively weak (using site medians), which reduces confidence when converting between these metrics – this would be a potential limitation if TSS were adopted as the SS attribute;
- At national scale, clarity and turbidity correlated well (R^2=0.8), and for dirtier water, the conversion from turbidity to clarity was more robust than was conversion from clarity to turbidity.

Accordingly, we recommend that one ESV will be sufficient for managing suspended sediment in flowing waters for ecosystem health (analogous for deposited sediment). If a single suspended sediment metric (or ESV) is adopted, benefits may be realised from using one based on turbidity.

Executive summary: Stage 1B report on deposited sediment (Clapcott and Goodwin 2017)

One step towards the development of a sediment attribute for inclusion in the National Objectives Framework is knowledge of where attribute management bands should be applicable. A stream with naturally high deposited sediment volumes should not unnecessarily be categorised as degraded. Degraded streams should be identified as those where human activity is responsible for increasing deposited sediment from expected natural levels, to a degree that impacts the stream's values. As such, knowledge of the natural or reference condition is needed for any given stream. Reference condition can be estimated from sites with minimal land use or predicted from the relationship between sediment and land use. Reference condition will vary across the country due to natural environmental gradients such as the source and nature of the sediment (e.g., geology, soil), the delivery of the sediment (e.g., erosion, rainfall, elevation) and the ability of the stream to retain sediment (e.g., slope, flow). Understanding and classifying this variation is needed to determine where sediment attribute bands should be applied.

We explored a large body of data which describes the state of fine sediment deposited on the streambed to develop a classification of New Zealand streams based on variation in reference condition. The data was compiled from regional council application of standardised methods at their monitoring networks, research datasets, as well as observations recorded in the New Zealand freshwater fisheries database (NZFFB).

Reference sites were generally unrepresentative of the full range of environmental variation in the river network so we developed predictive models using flexible spatial regression models to predict reference condition for all stream segments. We then used a classification and regression tree (CART) model approach to partition sites by their environmental similarity into a number of classes. These classes were combined based on similarity in their sediment values into a small number of groups. We plotted these groupings to ascertain where in New Zealand levels of low, medium, or high sediment levels can be expected to occur naturally. Results suggest that the majority of New Zealand streams (>85%) can be expected to have less than 20% fine sediment cover. Higher sediment cover is generally expected in areas of the stream network at low elevation and on distinct geologies such as volcanic-acidic and alluvium.

This information forms the basis of a sediment classification for New Zealand streams. However, we recommend that a regional verification is needed to refine model predictions. We also recommend that the classification be revisited once sediment attribute bands have been developed based on the relationship between sediment and ecological responses.

Appendix B Technical justification for excluding euphotic depth as a suspended sediment ESV for flowing waters

Rob Davies-Colley, NIWA, March 2017

Summary

Suspended sediment attenuates light, leading to effects on aquatic ecosystems via two different aspects of water clarity: (reduced) *visual clarity* (restricting visual range of aquatic animals) and (reduced) *light penetration* (constraining aquatic plant photosynthesis). Visual clarity is relatively simply (linearly) related to concentration of suspended sediment of a particular type by its light beam attenuation, although the overall correlation with suspended sediment is weakened by a fairly wide range (~20-fold) in light beam attenuation per unit mass due to variation in particle size, shape and composition. Light penetration is more complicatedly (and non-linearly) related to suspended sediment concentration, because diffuse (sun)light attenuation reflects interaction of light scattering by fine particles with light absorption. Simple calculations suggest that light penetration into NZ rivers is not often limiting of benthic plants, and it is *downstream* waters (lakes, estuaries) where protection of light fields may be necessary. But protection of visual clarity by prospective NOF-objectives for fine sediment suspended in waters may also serve to protect light penetration in all but very 'unusual' cases.

Optical basis of water clarity

Two aspects of visual clarity in waters are important to ecology and human values: light penetration and visual clarity (Davies-Colley et al. 2003).

Visual clarity controls the sighting distance of aquatic animals, including fish and aquatic birds, as well as strongly affecting human use of waters for recreation (e.g., Smith et al. 1995) Visual clarity is quantified by the beam attenuation coefficient, c (m^{-1}), *the proportional loss of light energy from a perfectly collimated light beam per unit (small[19]) length of light path* by two optical processes:

- *absorption* (conversion of light energy to another form, ultimately heat, symbol a), and
- *scattering* (change in direction but not energy of light photons, symbol b).

These optical properties are associated as follows:

$$c = a + b,$$

The beam attenuation, absorption and scattering coefficients are all *inherent* optical properties (IOPs) (e.g., Kirk 2011), dependent only on the properties of the water and not on incident light.

Penetration of sunlight (diffuse light or irradiance) into waters, combined with water depth, controls the light field of benthic plants that are 'competing' for light with phytoplankton circulating through the overlying water column. Sunlight penetration is quantified by the irradiance attenuation coefficient K(m^{-1}), usually measured to detect down-welling irradiance with a cosine response[20] –

[19] The path length has to be small so that the proportional change in incident light does not change much. See formal definitions in standard texts like Davies-Colley et al. (2003) and Kirk (2011).

[20] Response proportional to the cosine of the angle of incidence. Light detectors fitted with flat plate diffusers typically closely approach the ideal cosine response (Kirk 2011).

denoted by subscript-d (Kirk 2011). Analogous to beam attenuation, irradiance attenuation can be defined as the *proportional change in irradiance over a (small) depth interval in water*. Irradiance attenuation can be measured using detectors having different spectral responses. The most commonly used device, particularly with regard to aquatic plant light fields, is a photosynthetically available radiation (PAR) sensor – one that is equally sensitive to all photons in the PAR band of 400-700 nm wavelength. Down-welling irradiance attenuation in the PAR band is denoted K_d(PAR). Irradiance attenuation is an *apparent* optical property (AOP) that depends (albeit weakly) on incident sunlight (Kirk 2011). One important consequence of this 'apparent' character is that it is not strictly correct to add the contributions of different light-attenuating constituents in water (such as fine sediment and phytoplankton) in order to estimate total irradiance attenuation.

Effects of fine sediment on water clarity

Both light penetration and visual clarity of waters are strongly affected by fine sediment (e.g., Davies-Colley et al. 2015) – which typically dominates light scattering in water and sometimes also contributes strongly to light absorption. Absorption of light by fine sediment is usually due to organic material sorbed onto mineral surfaces, rather than to intrinsic mineral 'colour', but counter examples of practical importance, notably (yellow to red-coloured) ferric sesquioxides exist.

Visual clarity is rather simply related to the amount of sediment suspended in waters. Beam attenuation (controlling visual clarity) depends linearly on suspended fine sediment concentration. However, fine sediment varies greatly in physical size and surface properties, and thus the light beam attenuation per unit mass concentration

$$c^* = c/\text{TSS},$$

units m^2/g, hence often referred to as an 'optical cross-section') also varies appreciably. So sediment 'quality' (physical properties) are almost equally important as sediment 'quantity' (concentration) as regards light beam attenuation. 'Fine' sediment, may be defined for current purposes as particles in the 0.1 to 10 μm diameter range, and in that range the attenuation cross-section (for equidimensional particles) varies roughly 20-fold (peaking at an intermediate size of about 1.2 μm; Davies-Colley et al. 2003: fig 2.9). Particle shape is also important, with layer clay minerals having, as could be expected, appreciably higher optical cross-sections than equidimensional particles (e.g., Gibbs 1978). Hicks et al. (2016) reported an approximate 20-fold range in average optical cross-section of fine sediment in NZ rivers in the NRWQN – which is most strongly due to variation in particle size but also particle shape. Unfortunately, the predictability of optical cross-section in NZ rivers was found to be weak, with only %silt-clay in catchment soils having any useful predictive power (Hicks et al. 2016), so visual clarity will probably have to be related empirically (locally) to suspended sediment concentration by pairing visual clarity with TSS measurements.

The relationship between light penetration and suspended sediment is more complex than that between visual clarity and suspended sediment. A classic paper by Kirk (1985) elucidated the mechanism:

- light absorption is what actually extinguishes light photons moving down through the water column, such that irradiance attenuation is proportional to absorption, but
- light scattering contributing indirectly to irradiance attenuation by forcing light photons to take a tortuous path, thereby increasing the probability of absorption over a given depth interval.

In a series of papers, Kirk showed that light attenuation was proportional to the square root of light scattering (Kirk 2011). For example, Kirk (1981) used stochastic modelling to derive the following expression for irradiance attenuation at the mid-point of the euphotic zone (i.e., the 10% light level):

$$K_d(z_m) = (a^2 + 0.256ab)^{0.5}$$

So if fine sediment is light-scattering, but negligibly light-absorbing (for example glacial flour), we expect a square-root relationship between K_d and TSS. More often fine sediment is light-*absorbing* as well as light-scattering, and, if the light absorption is comparatively strong ($a \sim b$), K_d is nearly linearly dependent on TSS. More often, fine sediment is more strongly light-scattering than absorbing ($b > a$), and the relationship between these K_d and TSS is best described by a power law with an empirical exponent between 0.5 and 1.0. Some studies have found a near-linear dependence (e.g., Vant 1990, and Gall et al. 2017 in prep. for estuaries), but more often the dependence is better fitted empirically by a power law with an exponent between 0.5 and 1.0 (e.g., Davies-Colley and Nagels 2008 reported an exponent of 0.50 for NZ rivers).

Note that fine sediment interacts with dissolved light absorption, primarily by coloured dissolved organic matter (CDOM, humic matter), to increase overall irradiance attenuation by the same mechanism discussed by Kirk (1985) – i.e., light scattering by suspended particles increases effective path and the likelihood of absorption of photons. This insight was the basis for development of a simple semi-empirical model of irradiance attenuation in NZ rivers as a function of beam attenuation and CDOM by Davies-Colley and Nagels (2008).

Protecting NZ waters from optical and non-optical impacts of fine suspended sediment

There is little doubt that visual clarity is a valued attribute of NZ waters, including rivers, and important to the habitat quality of higher animals (fish and birds). Therefore, visual clarity should be protected by national standards (NOF-bands). Given development of national standards for visual clarity, two further questions arise:

- Are NOF bands (standards) also required to manage the *non*-optical effects of suspended sediment?
- Are NOF bands (standards) also required to manage the light penetration effects of suspended sediment?

Non-optical effects of fine suspended sediment

In my opinion, because light beam attenuation per unit mass (attenuation cross-section, m^2/g) of sediment in NZ rivers varies appreciably (about 20-fold), we cannot rely on visual clarity (or suspended sediment concentration) alone, despite the good overall correlation of these quantities (e.g., Davies-Colley et al. 2014). It may be inconvenient, but New Zealand would ideally have standards for *both* visual clarity and suspended mass concentration to adequately protect from both optical and non-optical effects of fine suspended matter. (This assumes, in the absence of specific knowledge to the contrary, that non-optical effects will scale with mass concentration. That assumption should, ideally, be tested.)

Different optical effects of fine suspended sediment

The two main optical effects of fine suspended sediments, as we have seen, are reduced visual clarity and reduced light penetration. However, it should be noted that fine sediment also affects water colour and thereby, human aesthetic response to waters and, potentially, spectral light fields in waters, which are likely to effect plant photosynthesis and aquatic animal vision. For the moment we will ignore water colour and spectral effects and concentrate solely on water clarity.

Can standards for visual clarity adequately protect light penetration?

Davies-Colley and Nagels (2008) found that suspended sediment was the main controller of light penetration into NZ rivers, but CDOM (dissolved humic matter) also contributed, apparently by interacting with the light scattering of suspended matter via the process discussed by Kirk (1985). In waters downstream, that is, estuaries and lakes, phytoplankton may be expected to become important as a light-attenuating constituent (Davies-Colley et al. 2003), as has been shown for lakes (e.g., Vant and Davies-Colley 1984). Vant (1990) and Gall et al. (2017 in prep.) found that phytoplankton chlorophyll *a* contributed negligibly to irradiance attenuation in northern North Island estuaries.

In NZ's mostly shallow, small rivers light 'shading' by the water column is probably seldom a major constraint on plant growth, unlike shading by riparian vegetation (see calculations below and summarised in Table B-1). Deep, highly light-attenuating rivers can be severely light-limiting (e.g., Julian et al. 2008), but in NZ rivers such conditions are probably mostly confined to episodic and transient flood flows, which are both deeper and more light-attenuating than baseflows.

Furthermore, it is difficult to imagine a situation where visual clarity would not be changed to the extent that recommended standards were exceeded, but light penetration would be, because both aspects of clarity depend on light-attenuation, including by sediment, albeit in different ways.

So it is reasonable to hypothesise that controlling light beam attenuation (to protect visual clarity) will also protect irradiance attenuation (and thus light penetration) in New Zealand rivers.

To check this hypothesis, a model of benthic lighting for different NZ rivers would ideally be constructed, based on the BLAM (benthic light availability model) framework of Julian et al. (2008), with irradiance at the bed given by

$$E_{bed} = 0.93sE_o\exp(-K_\mathrm{d}z),$$

where E_o is incident irradiance, z is water depth, and the factor 0.93 accounts for an average of about 7% loss of irradiance by reflection at the water surface. (As a 'worst case' we neglect bank and riparian shading, *s*.) The model framework would use:

- the simple statistical model of K_d(PAR) (as a function of c and CDOM) from Davies-Colley and Nagels (2008), together with
- statistical models (to be constructed from available water quality and morphological data) of both
- optical properties (c, CDOM; Smith et al. 1997) and
- depth distribution (*z*) as a function of flow
- to estimate benthic lighting as a function of flow and thus time (using the flow-duration curve).

As a (very) rough indicator of how this kind of modelling would 'work', we can illustrate benthic irradiance as a fraction of incident for some particular cases (Table B-1).

1. For an averagely light-attenuating NZ river (median visibility = 1.28 m; g_{340}, an index of CDOM, = 4.1/m; Smith et al. 1997), the semi-empirical statistical model of Davies-Colley and Nagels (2008) predicts K_d(PAR) = 1.03/m. In 1 m water depth (average), the ratio E_{bed}/E_o = 38% (Table B-1), which would not be light-limiting for most benthic plant communities.

2. In a very 'dirty' and coloured NZ river with 95percentile visibility = 0.36 m and 95percentile CDOM, E_{bed}/E_o at 1 m depth is 10% – which is starting to constrain growth of some (light-demanding) benthic plants.

3. Obviously if the water was even deeper the light limitation would be more severe. For example, for a dirty and coloured river water at 2 m depth, the bed is approximately at the euphotic depth (irradiance has fallen to ~1% of surface value; Table B-1), extinguishing most benthic plants.

These simple calculations suggest that light limitation by water shading is generally not an issue in NZ rivers, and that specific protection of light penetration in rivers should not be needed.

Light penetration will need to be considered in lake and estuary receiving waters. There is ample evidence that keystone benthic plants in lakes (macrophytes) and estuaries (seagrasses) have declined historically in NZ because of reduction in euphotic depth (depth of the 1% light level, a useful rough index of the maximum depth of light growth – e.g. Vant et al. 1986), caused by increased suspended sediment concentrations. However, in particular cases it is difficult to decide whether increased suspended sediment and light attenuation is a symptom or a cause of the decline in keystone plant population (often probably both – Schallenberg and Sorrell 2009). But even in lakes and estuaries, protecting visual clarity (light beam attenuation) – depending on how the standards are formulated numerically – may also serve to adequately protect benthic light fields (irradiance attenuation) in all but very 'unusual' cases.

Table B-1: Irradiance (PAR) at the bed of a NZ river as a fraction of incident irradiance. Calculations used Model 1b of Davies-Colley and Nagels (2008) to calculate irradiance attenuation (Kd(PAR)) from visual clarity and g340 (CDOM index) statistics from the NRWQN as summarised by Smith et al. (1997). The benthic light availability model framework (BLAM) of Julian et al. (2008) was used to calculate the irradiance at the bed as a fraction of incident irradiance, assuming a 1 m and 2 m average depth (ignoring riparian shade and allowing for 7% loss of light by water surface reflection).

NZ rivers	Visual clarity (m)	CDOM (g_{340}, 1/m)	K_d(PAR) (1/m)	Bed irradiance as a % of incident irradiance E_{bed}/E_o	
				1m depth	2m depth
Median values	1.28	4.10	1.03	38	15
Dirty (95%ile clarity)	0.36	4.10	1.95	16	2.7
Dirty & coloured (95%ile)	0.36	12.2	2.46	10	1.0

References

Davies-Colley, R. J.; Nagels, J. W. 2008. Predicting light penetration into river waters. *Journal of Geophysical Research-Biogeosciences 113*: G03028. doi10.1029/2008JG000722.

Davies-Colley, R.J., Vant, W.N., Smith, D.G. (2003) *Colour and clarity of natural waters. Science and management of optical water quality.* reprinted by Blackburn Press, Caldwell, New Jersey: 310.

Davies-Colley, R.J., Ballantine, D.J., Elliott, A.H., Swales, A., Hughes, A.O., Gall, M.P. (2014) Light attenuation – a more effective basis for the management of fine suspended sediment than mass concentration? *Water science and technology*, 69(9): 1867-1874. doi: 10.2166/wst.2014.096

Davies-Colley, R.J.; Hicks, D.M. Hughes A.O.; Clapcott, J.; Kelly, D.; and Wagenhoff, A. (2015). Fine sediment effects on freshwaters, and the relationship of environmental state to sediment load. NIWA Client Report (HAM2015-154) for the Ministry for the Environment, November 2015. 99p.

Gall, M. P.; Swales, A.; Davies-Colley, R. J.; Bremner, D. (2017 in prep.). Light penetration and optical water quality of northern North Island, New Zealand estuaries. *Estuarine, coastal and shelf science*.

Gibbs, R.J. (1978) Light scattering from particles of different shapes. *Journal of Geophysical Research*, 83C1: 501-502.

Hicks, D. M. Clapcott, J.; Davies-Colley, R. J.; Dymond, J.; Greenwood, M.; Hughes A. O.; Shankar, U. and Walter, K. (2016). Sediment attributes Stage I. NIWA Client Report (CHC2016-058) for the Ministry for the Environment, June 2016. 166p + Appendices.

Julian, J.P., Doyle, M.W., Stanley, E.H. (2008) Empirical modeling of light availability in rivers. *Journal of Geophysical Research-biogeosciences*, 113: G03022. doi:10.1029/2007JG000601

Kirk, J.T.O. (1981) Monte Carlo study of the nature of the underwater light field in, and the relationships between optical properties of, turbid yellow waters. *Australian journal of marine and Freshwater Research*, 32: 517-532.

Kirk, J.T.O. (1985) Effects of suspensoids (turbidity) on penetration of solar radiation in aquatic ecosystems. *Hydrobiologia*, 125: 195-208.

Kirk, J.T.O. (2011) *Light and photosynthesis in aquatic ecosystems*. Cambridge University Press, New York, NY: 649.

Schallenberg, M., Sorrell, B. (2009) Regime shifts between clear and turbid water in New Zealand lakes: Environmental correlates and implications for management and restoration. *New Zealand Journal of Marine and Freshwater Research*, 43(3): 701-712. DOI: 10.1080/00288330909510035

Smith, D.G., Croker, G.F., McFarlane, K. (1995) Human perception of water appearance 1. Clarity and colour for bathing and aesthetics. *New Zealand Journal of Marine and Freshwater Research*, 29: 29-43.

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

Vant, W.N. (1990) Causes of light attenuation in nine New Zealand estuaries. *Estuarine, Coastal and Shelf Science*, 31: 125-137.

Vant, W.N., Davies-Colley, R.J. (1984) Factors affecting clarity of New Zealand lakes. *New Zealand Journal of Marine and Freshwater Research*, 18: 367-377.

Vant, W.N., Davies-Colley, R.J., Clayton, J.S., Coffey, B.T. (1986) Macrophyte depth limits in North Island (New Zealand) lakes of differing clarity. *Hydrobiologia*, 137: 55-60.

Appendix C Assignment of LCDB v4.1 land-cover classes to upstream catchments of 832 water quality sites

Sanjay Wadhwa (NIWA, Hamilton)

Revision of the definition of 'reference state' involved setting land-cover thresholds to bound an acceptable level of catchment 'disturbance'. This made use of recent land-cover data derived from the latest version of the New Zealand Land Cover Database (LCDB v4.1).

Each monitoring site has a corresponding unique reach identifier ('NZReachID') that links it to the River Environment Classification (REC) system. The REC system links individual reaches to form a drainage network. For each monitoring site:

- all upstream reaches were selected using 'from' and 'to' node information available in the REC stream network dataset
- watershed polygons defined for each associated reach in the REC database were then combined to create a single polygon that represented the catchment upstream of each monitoring site.
- Using a Geographic Information System, this catchment polygon was then intersected with the LCDB v4.1 dataset (layer), thereby
- generating the area of each land-cover class of the upstream catchment for each of 832 water quality sites.

This process was done using an automated script-driven analytical method within GIS. Given the importance of the intended use of this information (these data were used to define 'reference state' according to land-cover thresholds), several individual monitoring sites were selected to manually repeat the process and check the results.

The results obtained for three randomly selected sites (selected from the total of 833) were checked to determine if the areas calculated using the procedure outlined above were correct, and whether the area of the catchment polygons derived from the REC reaches matched those derived from the LDCB areas. Key data included:

1) The total area calculated for the sum of the REC polygons upstream of the monitoring site.
2) Total calculated area values derived from assigned LCDB v4.1 land-coverage.
3) The accuracy of the intersection of the polygons derived from the LCDB v4.1 data with catchment areas derived from GIS using the REC sub-catchment polygons initially.
4) Whether the differences in catchment areas estimated from these two processes were tolerable.

The three randomly-selected sites are listed in Table C-1**Error! Reference source not found.**, along with key information.

Table C-1: Details of sites randomly selected for method validation.

Site no.	NEMaR_ID	LAWA_ID	RC_ID	Sitename	NZReach
8	ARC-07811	ARC-00017	AC	Oteha Stream at Days Bridge	2004535
323	EW-1293-009	EW-00111	EW	Whangamarino River at Jefferies Rd Br	3008516
698	NRC-109098	NRC-00029	NRC	Waimamaku River @ SH12	1014099

Validation of results for site #8 (ARC-00017)

Referring to Figure C-1:

- the sub-catchments defined by the thin yellow lines are defined by the REC reach assignments
- the red line identifies the catchment boundary derived from GIS incorporating all the REC sub-catchments.
- Visual inspection indicates
  - the sub-catchments for all upstream reaches for site #8 appeared to be captured, and are inside the catchment polygon (red bold line)
  - the stream lines do not cross the catchment boundary.

The results of an intersection between the catchment polygon and the LCDB4 landcover overlay (indicated by the black line in Figure C-1, right) were exported to a .txt file that was opened in Excel for more detailed evaluation. The total areas (m^2) for the different classes for site #8 are summarised in Table C-2. These results indicate:

- total area of the polygon following overlay with LCDB4 was 11,975,407 m^2
- total area of the polygon derived from GIS was 11,978,100 m^2
- the difference in areas calculated manually was 0.02%.

The difference in area estimated from these two sources is negligible. Although no difference might be expected, the processes used to assign land cover class may lead to minor overlaps or gaps between areas. The difference observed here is within the error expected for the land cover assignment.

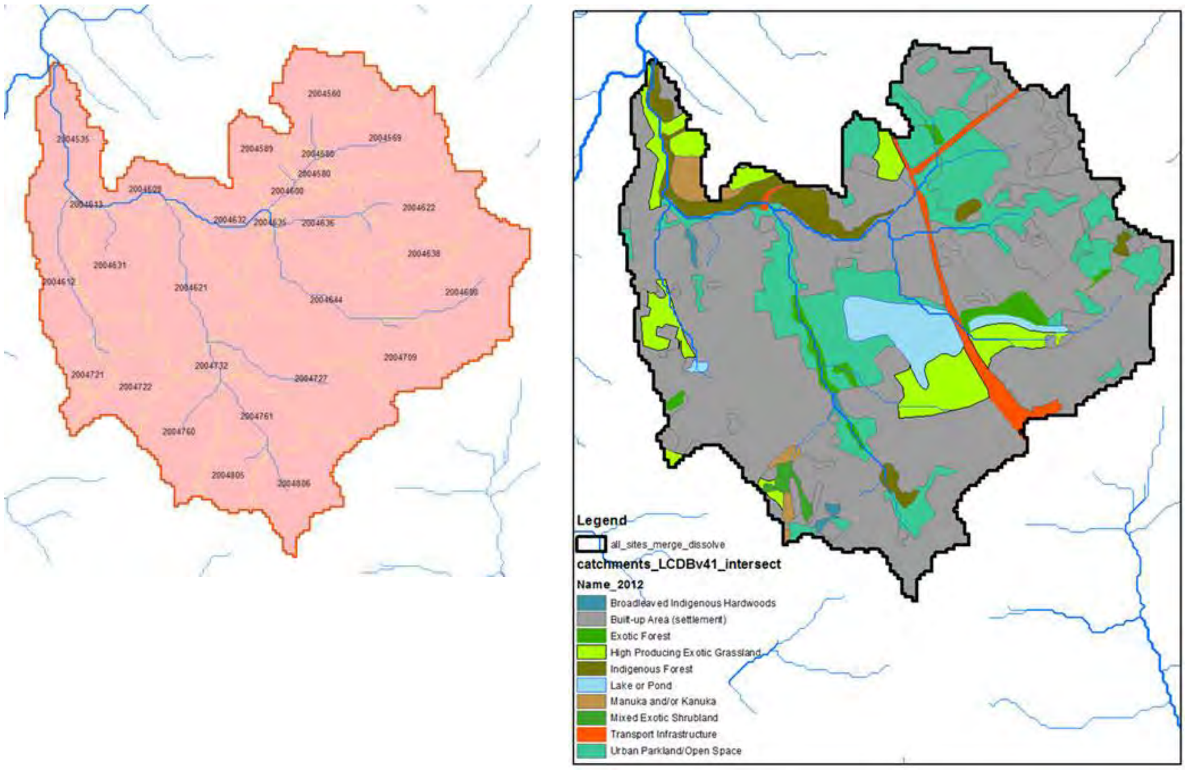


Figure C-1: Polygon map of the catchment upstream of site #8 (left), with LCDB v4.1 land-cover overlay (right). In the left-hand figure, the yellow lines identify catchments defined by the REC reaches, and the red line indicates the catchment boundary defined by GIS. The right-hand figure indicates the land cover classes across the entire monitoring site catchment area (denoted by the black outline).

Table C-2: Summary of LCDB v4.1 land-cover output for site #8 (ARC-00017).

LCDB v4.1 landcover (description)	LCDB v4.1 code	Area (m²)
Built-up Area (settlement)	1	7,520,159
Urban Parkland/Open Space	2	2,061,352
Transport Infrastructure	5	259,267
Lake or Pond	20	422,150
High Producing Exotic Grassland	40	802,904
Manuka and/or Kanuka	52	147,966
Broadleaved Indigenous Hardwoods	54	44,106
Mixed Exotic Shrubland	56	55,851
Indigenous Forest	69	456,182
Exotic Forest	71	205,471
TOTAL		**11,975,407**

Validation of results for Site #323 (EW-00111)

Water quality monitoring site #323 corresponds to NZReach 3008516. The catchment upstream of this site is more complex, comprising 236 individual watershed polygons defined by the REC.

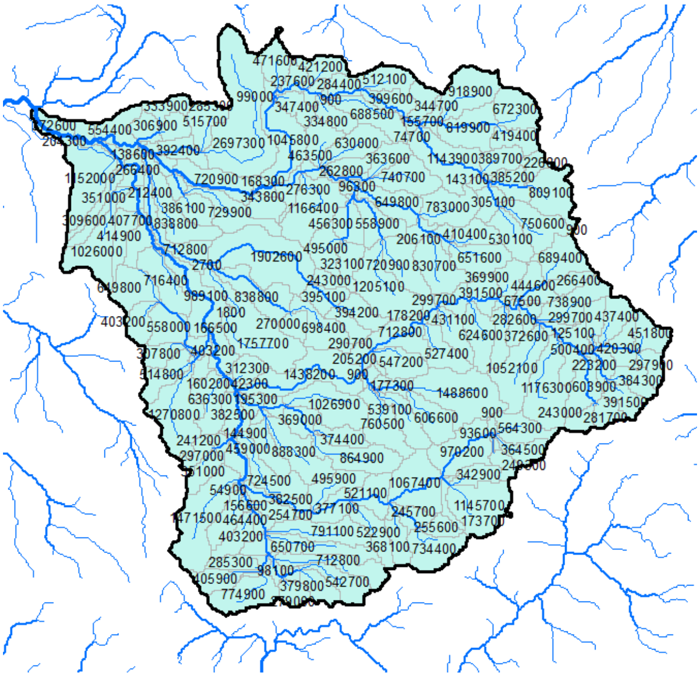


Figure C-2: Upstream reaches and catchment polygon for site #323 (EW-00111).

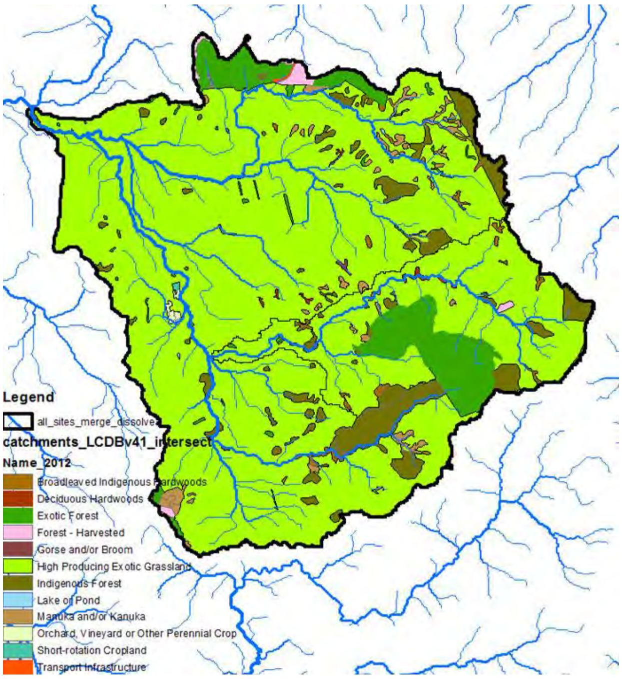


Figure C-3: LCDB v4.1 overlay on upstream catchment polygon of site #323.

Visual inspection of Figure C-2 shows that all upstream sub-reaches appear to be accurately captured within the overall catchment polygon (black line). The catchment polygon (in black) also appears to represent the catchment accurately.

The results of an intersection between the land cover classes and catchment boundary (Figure C-2 and Figure C-3) were assessed in Excel. The areas derived from this process are summarized in Table C-3.

Table C-3: Summary of LCDB v4.1 land-cover output for site #323 (EW-00111).

LCDB4 landcover (description)	LCDB4 code	Area (m^2)
Transport Infrastructure	5	36,383
Lake or Pond	20	14,656
Short-rotation Cropland	30	28,148
Orchard, Vineyard or Other Perennial Crop	33	129,456
High Producing Exotic Grassland	40	80,237,404
Gorse and/or Broom	51	41,277
Manuka and/or Kanuka	52	1,490,799
Broadleaved Indigenous Hardwoods	54	548,539
Forest - Harvested	64	505,425
Deciduous Hardwoods	68	220,149
Indigenous Forest	69	6,647,475
Exotic Forest	71	6,607,176
TOTAL		**96,506,887**

The total area estimated from LCDB4 was 96,506,887 m^2, 43,800 m^2 larger than the catchment polygon area derived from GIS. This represents a difference less than 0.05%

Validation of results for Site #698 (NRC-00029)

Water quality monitoring site #698 corresponds to NZReach 1014099. The catchment upstream of this site is the most complex of the three tested, comprising 332 individual watershed polygons defined by the REC.

Visual inspection of Figure C-4 shows that all upstream sub-reaches appear to be accurately captured within the overall catchment polygon (black line). The catchment polygon (in black) also appears to represent the catchment accurately.

The results of an intersection between the land cover classes and catchment boundary (Figure C-4 and Figure C-5) were assessed in Excel. The areas derived from this process are summarized in Table C-4.

The total area estimated from LCDB4 was 102,430,081 m^2, 115,920 m^2 smaller (0.1%) than the catchment polygon area derived from GIS (102,546,000 m^2).

In all cases the difference in area estimated by two independent techniques was less than or equal to 0.1%. In view of the processes used to generate catchment areas and areas of discrete land use (satellite and aerial photography, automated land use assignment with expert input etc., each of which has measurable error), we are confident that the method used to define ‘reference state’ in terms of land-cover thresholds to bound an acceptable level of catchment ‘disturbance’ is defensible, repeatable and introduces negligible error when defining reference states.

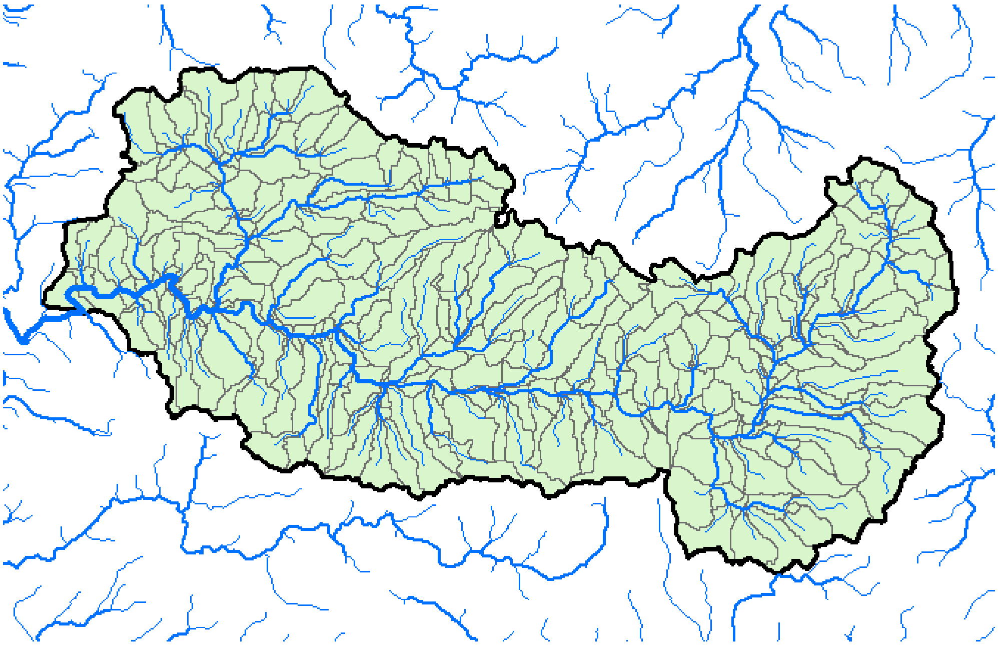


Figure C-4: Upstream reaches and catchment polygon for site #698 (EW-00029).

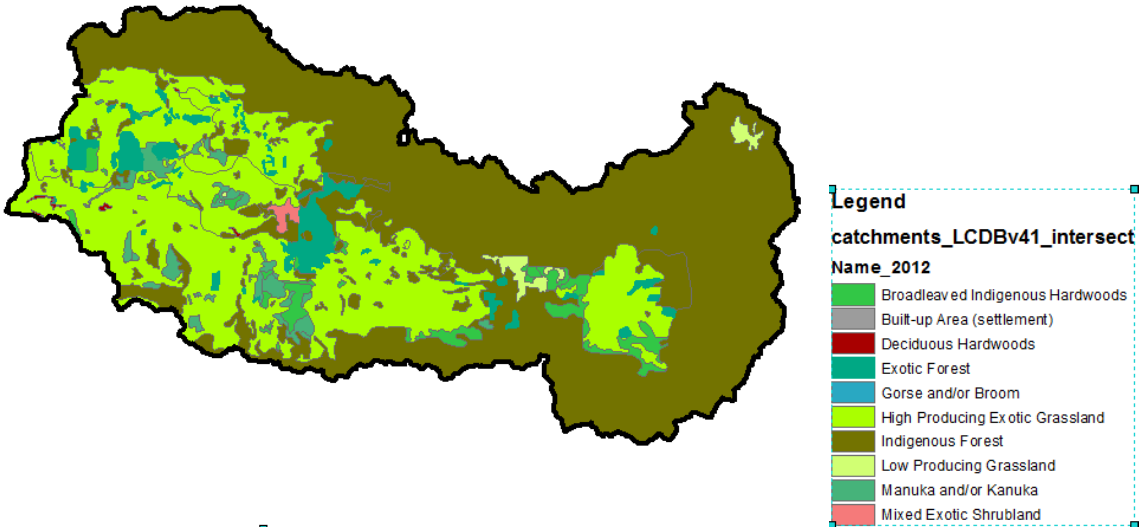


Figure C-5: LCDB v4.1 overlay on upstream catchment polygon of site #698.

Table C-4: Summary of LCDB v4.1 land-cover output for site #698 (EW-00029).

LCDB4 landcover (description)	LCDB4 code	Area (m^2)
Built-up Area (settlement)	1	12,460
High Producing Exotic Grassland	40	32,549,876
Low Producing Grassland	41	866,079
Gorse and/or Broom	51	834
Manuka and/or Kanuka	52	2,301,317
Broadleaved Indigenous Hardwoods	54	29,094,867
Mixed Exotic Shrubland	56	280,963
Deciduous Hardwoods	68	113,233
Indigenous Forest	69	59,696,786
Exotic Forest	71	3,699,045
TOTAL		**102,430,081**

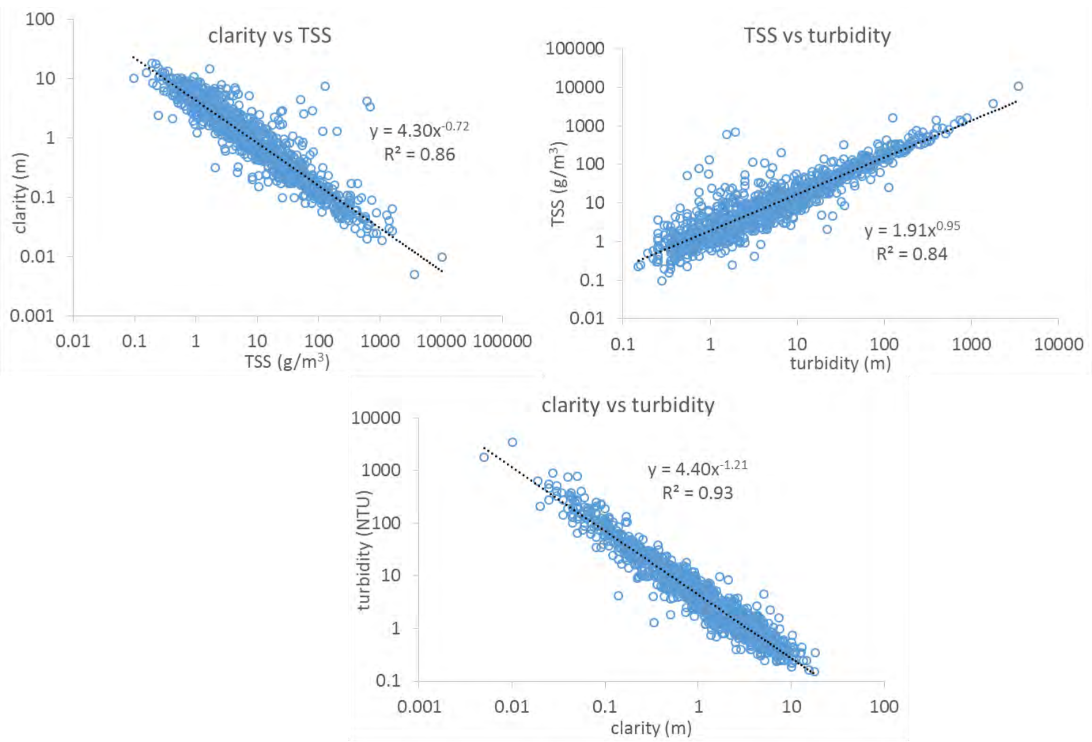
Appendix D Regressions for TSS, clarity and turbidity

TSS, clarity and turbidity are generally correlated with each other, however 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 & 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.

If there are robust regressions, then if, for example, the suspended sediment attribute only enumerated a C/D band threshold for visual clarity (m), a regional or unitary authority not monitoring this variable could either convert their data (to visual clarity), or convert the threshold value into either TSS or turbidity. If, on the other hand, there is considerable uncertainty in the regressions, then using the example of visual clarity, those RC's that do not measure this variable will be unable (with their currently monitoring data) to assess the suspended sediment state of waterways in their region. As such, a pragmatic step is to enumerate threshold bands (C/D band, in particular) in units that all RC's measure – i.e., both visual clarity and turbidity.[21]

NRWQN regressions between TSS, turbidity and clarity

NRWQN data (2011-2015) indicate strong correlations (R^2 of 0.84 to 0.93) between SS metrics, indicating that interconversion between three SS metrics may be undertaken with reasonable confidence (Figure D-1, from Hicks et al. 2016). Note that these figures are based on discrete data (approximately 60 data points per site), as opposed to site medians (1 data point per site). These data include stormflows, which is why the axes span a large range of SS values.



TSS is unlikely to be a convenient suspended sediment measure for characterising suspended sediment levels that correspond to chronic (long-term) exposure due to added cost and inadequate detection limits.

Figure D-1: Regression of TSS, clarity and turbidity using NRWQN data (2011-2015). Note data shown is all monthly data, as opposed to site medians.

Regional monitoring datasets (using site medians)

Using the regional SoE data set (collated as part of Larned et al. 2015) the regression of TSS vs turbidity and clarity vs TSS (Figure D-2) has correlation coefficients (R^2) of 0.69 and 0.52, respectively. It is unlikely that these regression equations would provide sufficient confidence to convert between TSS and turbidity, or TSS and clarity. This uncertainty would be a limitation when implementing a TSS attribute.

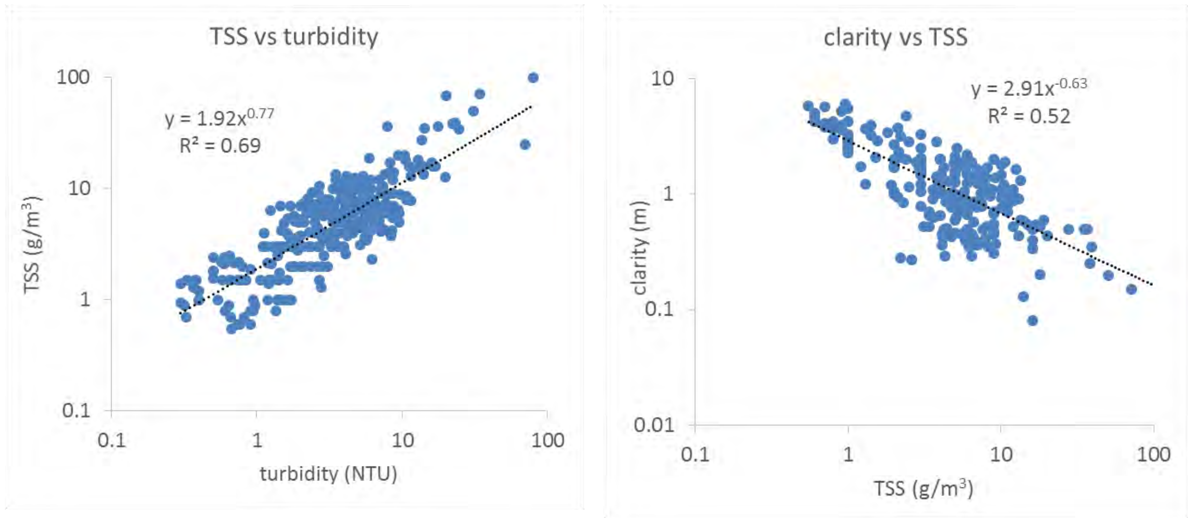


Figure D-2: Regression of TSS with visual clarity and turbidity, using site medians for all available monitoring data. Data points represent long-term site medians – usually derived from 7-10 years of monthly monitoring data.

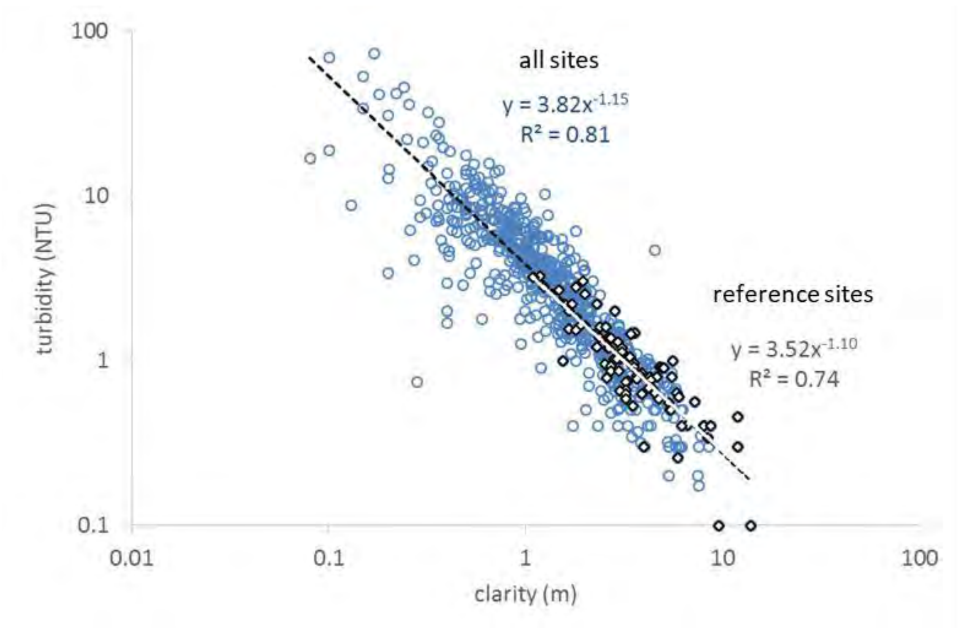


Figure D-3: Regression of turbidity and visual clarity using long-term site medians for all sites and reference sites. For all sites, n=722 (blue circles) and for reference sites, n=83 (black diamonds).

In contrast, using 722 paired turbidity and clarity data (derived from regional and NRWQN monitoring), the regression yielded a correlation coefficient of 0.81 (compared to 0.93 for NRWQN sites, Figure D-1). Limiting the number of sites to the subset of reference sites (n=83), the regression equation was similar and with an only slightly weaker R^2 value of 0.74. Using the three regression equations (for NRWQN, all sites and reference sites respectively), a range of clarity and turbidity values may be interconverted (Table D-1). These data show that because of the inverse relationship between turbidity and clarity, the compression of clarity at the 'dirtier' end of the spectrum results in greater absolute uncertainty when low clarity values (e.g., <0.5 m) are converted to turbidity. For example, 0.5 m clarity converted to turbidity values that ranged between 7.5 and 10.2 NTU; and 0.3 m clarity converted to turbidity values that ranged between 13.2 to 18.9 NTU, dependent on the regression equation used.

In contrast, when using the same regression equation to convert turbidity to clarity, the large range of turbidity is converted into the compressed range of clarity, resulting in less variability (or sensitivity) to the regression equation selected. For example, a turbidity of 5 NTU converted to clarity values that ranged between 0.9 and 1.1 m; and a turbidity of 20 NTU converted to clarity values that ranged between 0.2 and 0.3 m.

Table D-1: Conversion of clarity into turbidity (left side) and turbidity into clarity (right side) using three regression equations. The grey shaded area represents where management of suspended sediment is likely to focus, and hence where greater confidence is required when inter-converting values derived from different metrics. Dataset from Larned et al. (2015) collation.

Clarity (m) value to convert	Turbidity (NTU) via regression eqn.			Turbidity (NTU) value to convert	Clarity (m) via regression eqn.		
	All sites	Reference sites	NRWQN sites		All sites	Reference sites	NRWQN sites
0.3	15.3	13.2	18.9	0.5	5.9	5.9	6.0
0.5	8.5	7.5	10.2	1.0	3.2	3.1	3.4
0.6	6.9	6.2	8.2	1.5	2.3	2.2	2.4
0.8	4.9	4.5	5.8	2.0	1.8	1.7	1.9
1.0	3.8	3.5	4.4	2.5	1.4	1.4	1.6
1.4	2.6	2.4	2.9	3.0	1.2	1.2	1.4
1.6	2.2	2.1	2.5	4.0	1.0	0.9	1.1
2.0	1.7	1.6	1.9	5.0	0.8	0.7	0.9
3.0	1.1	1.1	1.2	10	0.4	0.4	0.5
5.0	0.6	0.6	0.6	20	0.2	0.2	0.3

Appendix E Comparison of predicted reference state condition (deposited sediment) from BRT and GLMM models

Reference state predictions for deposited sediment were also developed using a generalised linear mixed model (GLMM) as a component of developing sediment thresholds for fishes with an REC-based classification (see section 6). We have not compared the model output from the BRT and GLMM models to test which model provides the highest predicative accuracy. However, grouping site-specific predictions from the BRT REF and GLMM models post hoc by the REC classes illustrates where the models do perform similarly or differ (Figure E-1). All three models predict the majority of mountain and hill fed streams to have less than 30% fine sediment cover in deposited fine sediment. Likewise, all three models predict that most lowland and lake fed streams are likely to have <30% fine sediment cover. Lowland and lake-fed streams also showed the greatest discrepancy between model predictions especially for cool-dry, warm-dry, and warm-wet climates. These are also areas where there are fewer streams remaining in reference state, so model predictions are harder to validate. Regardless, the BRT REF model predictions based on environmental variability at reference sites are highest for these classes suggesting the assignment of attribute classes (i.e., <30%, 30-60%) based on these predictions may be overly permissive compared to the BRT ALL and GLM model predictions. Alternatively, the BRT REF predictions could be considered accurate in terms of where wetlands may have been once more dominate. Or, the BRT REF predictions could be considered to best reflect best attainable condition.

The SoE and research observations of '% sediment cover' were taken from runs whereas NZFFD observations were from all habitat types. The coverage of deposited fine sediment in runs is assumed to represent average reach cover, but may be less than that for whole reaches averaged over habitat types. This may partly explain why derived thresholds for macroinvertebrates were slightly more conservative than those derived for fish.

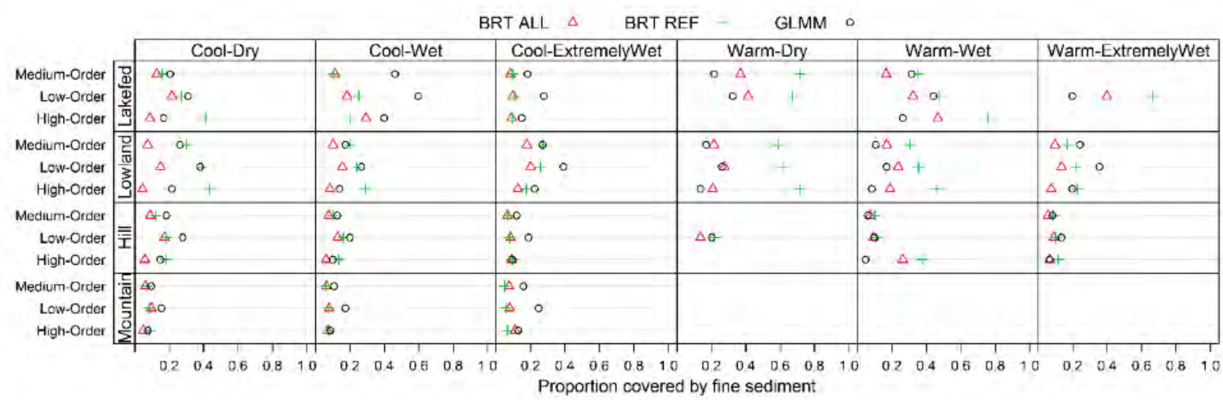


Figure E-1: Mean % sediment cover at reference state predicted from three different models used in this study. Black circles are predictions from a GLMM (McDowell et al. 2013; discussed in section 6), green crosses are predictions from a BRT REF model based on environmental variation at reference sites, and red triangles are from a BRT ALL model based on environmental variation at all sites after accounting for human land cover effects.

Appendix F Sediment Assessment Methods

Table F-1: Sediment Assessment Methods (SAMs) used for assessing the fine sediment content of stream channel beds. Summarised from Clapcott et al. (2011) excluding sediment depth (SAM6).

Name	Description	Variable measured	Applicability
Sediment Assessment Method 1 (SAM1)	A rapid visual estimation from the stream bank of the proportion of the channel bed covered by fine sediment (<2 mm)	% fine sediment cover	All streams
Sediment Assessment Method 2 (SAM2)	Semi-quantitative, in-stream visual assessment of the surface area of the streambed covered by fine sediment (< 2 mm), made by observing at least 20 locations within a single habitat	% fine sediment cover	Hard-bottomed streams
Sediment Assessment Method 3 (SAM3) (Wolman pebble count)	Semi-quantitative assessment of the particle size distribution, including fine sediment, on the streambed surface using a graduated template of "gravelometer". At least 100 particle measurements are made within a single habitat.	% by count of clast b-axis dimension into size fractions typically varying by a factor of 2; sediment too fine to measure (typically < 2 mm) labelled as "fines".	Hard-bottomed streams
Sediment Assessment Method 4 –(SAM4) (Quorer method)	Quantitative measure of total re-suspendible solids deposited on and within the streambed. A cylindrical tube is screwed into bed, the bed inside is stirred to suspend fine sediment, and the slurry is sampled and measured for suspended sediment concentration. Six samples are collected from a single habitat.	Samples are processed in the laboratory for total Inorganic/Organic sediment by area (SIS and SOS, respectively, in g/m^2) or Suspendible Benthic Solids by Volume (SBSV, g/m^3).	Hard-bottomed streams
Sediment Assessment Method 5 =(SAM5) (Shuffle method)	Rapid qualitative assessment of the amount of total re-suspendible solids deposited on a streambed. Made by observing turbidity created by disturbing the streambed by moving feet vigorously for five seconds.	A score from 1-5 is assigned (1 = little/no sediment; 5 = excessive sediment).	Hard-bottomed streams

Appendix G Linking the deposited sediment ESVs to catchment sediment yields and precedent suspended sediment regime

A targeted spatial survey of 16 sites was conducted in February 2017 (Table G-1). Sites were chosen to represent a wide gradient of sediment yield and also stream power which has been shown to be a major determinant of deposited sediment in New Zealand streams (Hicks et al. 2016). The wide range represents an expected sediment load range for all New Zealand streams (Hicks et al. 2011). Sites were also chosen where long-term suspended sediment data were available.

Table G-1: Sediment yields, catchment details and stream power of the 16 targeted sediment/macroinvertebrate monitoring sites (Feb 2017).

Region	Site name	Specific sediment yield ($t/y/m^2$)	Segment slope (°)	Upstream catchment slope (°)	Proportion of catchment in native vegetation	Segment mean annual flow (cumecs)	Stream power (W/m)
Waikato	Tapu at Tapu Coroglen Rd	29	0.57	18.26	0.94	1.19	6678
Waikato	Wharekawa at Adams Farm	36	0.17	14.53	0.51	2.22	3739
Waikato	Mangatutu at Walker Rd	38	0.01	10.74	0.42	4.56	255
Waikato	Whakapipi Stream at SH22	43	0.15	3.91	0.06	1.03	1560
Tasman	Lee at Meads	45	0.16	28.66	0.67	3.55	5570
Tasman	Wangapeka at Walters Peak	105	0.88	26.78	0.73	14.87	128219
Tasman	Motupiko at Christies	140	0.44	13.82	0.39	3.25	14003
Tasman	Motueka at Woodmans Bend	179	0.00	23.16	0.54	63.71	6243
Hawke’s Bay	Tukituki at Black Bridge	424	0.04	12.24	0.12	43.97	17236
Hawke’s Bay	Ngaruroro at Fernhill	658	0.74	17.97	0.56	46.31	335866
Hawke’s Bay	Maraetotara at Haumoana	789*	0.20	8.86	0.02	1.40	2745
Hawke’s Bay	Esk at Waipunga	1379	0.00	13.08	0.11	5.98	587
Gisborne	Te Arai at Pykes Weir	4392	1.72	16.71	0.24	2.89	48644
Gisborne	Waipaoa at Matawhero	7216	0.00	15.17	0.12	34.88	3418
Gisborne	Hikuwai at Willowflat	14200	0.24	16.71	0.29	8.09	19021
Gisborne	Mata at Pouturu	24739	0.00	13.38	0.17	11.68	1145

* Estimated from 90% confidence interval of incomplete rating curve. HB = Hawke’s Bay.

At the study sites sites, deposited sediment data was collected using SAM1 and SAM4 methods (Clapcott *et al.* 2011b). All sites were visited after a prolonged period of low flow (> 30 days), except the Motupiko at Christies site which was sampled 11 days after a high flow event (Table G-2).

Table G-2: Deposited sediment and related variables summarised for 2017 field study sites.

Region	Site name	SIS (g/m^2)	% cover bankside	Annual median turbidity (NTU)	Annual median flow (m^3)	Annual maximum flow (m^3)	Number of days since 3*median flow
Waikato	Tapu at Tapu Coroglen Rd	32	30	1.105	0.59	15.93	92
Waikato	Wharekawa at Adams Farm	46	30	3.95	0.96	72.03	127
Waikato	Mangatutu at Walker Rd	81	40	1.645	3.33	21.44	77
Waikato	Whakapipi Stream at SH22	53	30	4.4	0.50	9.52	78
Tasman	Lee at Meads	52	10	0.8	9.666†	347.08†	32†
Tasman	Wangapeka at Walters Peak	74	5	0.6	18.936	285.182	37
Tasman	Motupiko at Christies	48	5	0.8	1.617	21.288	11
Tasman	Motueka at Woodmans Bend	268	80	1.25*	53.758	732.385	31
HB	Tukituki at Black Bridge	1426	55	1.09	27.375	451.853	122
HB	Ngaruroro at Fernhill	487	5	2.01	28.705	482.400	126
HB	Maraetotara at Haumoana	501	15	0.915	0.614	13.650	134
HB	Esk at Waipunga	343	40	2.15	5.073	176.855	114
Gisborne	Te Arai at Pykes Weir	1518	70	1.95	0.183	48.473	90
Gisborne	Waipaoa at Matawhero	610	50	82.5	16.488	535.31	86
Gisborne	Hikuwai at Willowflat	823	50	2.7	1.193	278.602	71
Gisborne	Mata at Pouturu	648	20	41.3	4.626	105.304	84

† Flow summary from downstream Wairoa at Irvines site; * Turbidity summary from upstream Motueka at Woodstock site.

The relationships between deposited sediment and sediment yields, suspended sediment and additional driver variables, such as average stream flow and land cover, were analysed using generalised linear models (GLM).

There was a strong power relationship between SIS and suspended sediment yield, where SIS increased logarithmically as suspended sediment yield increased (R^2 = 0.45; **Error! Reference source not found.**). There was no correlation between sediment yield and % cover bankside at the 12 sites fro which reliable sediment yield data were available (R^2 = 0.005).

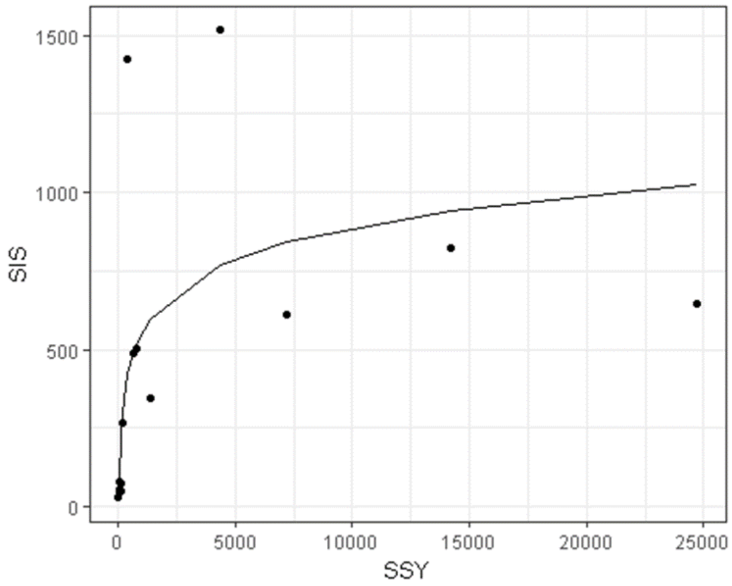


Figure 10-4: Relationship between specific sediment yield (SSY; t/km^2/y) and suspendable inorganic sediment (SIS; g/m^2) in the 2017 survey modelled by a negative binomial GLM with a log link.

SIS was best described by number of days since a significant fresh, as well as annual median flow, annual median turbidity, stream power, proportion of catchment in native vegetation and specific sediment yield.

The targeted field survey demonstrated a predictive relationship between specific sediment yield and the concentration of suspendable inorganic sediment (SIS) observed in the stream bed surface of a run habitat. It was not a 1:1 relationship; specific sediment yield ranged three orders of magnitude (25-24,800 t/km^2/y) and in response SIS ranged two orders of magnitude (32-1,518 g/m^2). The fact that a relationship between specific sediment yield (SSY) and SIS was demonstrated differs from the results of an earlier study which found no unidirectional relationship between these two variables at a national scale (Hicks et al. 2016). It may be that this relationship is highly variable, across a broader environmental gradient, dependent on where and when the sample is collected. The GLM showed that despite a strong relationship between SIS and SSY, the spatial variance in SIS is better predicted when catchment land use, flow and stream power are also taken into account. This confirms earlier attempts to model SIS nationally, where flow stability and stream power, as well as elevation and slope were important predictors of SIS (Hicks et al. 2016). It appears that flow in particular, and possibly the time elapsed since a last 'flushing' flow, may be important variables to consider for when SIS measurements should be taken that are to get a representative site estimate while reducing noise due to temporal within-site variation. Furthermore, most of the 2017 study sites had < 1° channel slope - if slope is an important predictor of SIS, then collection of data at sites with greater slopes could add further noise to the SIS to *specific sediment yield* relationship.

Appendix H Testing the accuracy of the SAM4 method

Methods

The original idea to use freeze-coring methods to collect the full sediment sample was quickly abandoned due to the inability to collect a ~25cm wide core. Most freeze-corers allow for the collection of < 10cm in diameter and there is a further difficulty of freezing loosely adhered sediments at the streambed surface (Dean Olsen, pers. comm.). As such, we designed a new technique to collect a full benthic sediment samples to test the accuracy of the SAM4 method. We filled buckets cut to 10cm in height with stream sediment and buried them at the stream surface (Figure H-1). Six replicates were placed in two streams with differing deposited fine sediment volumes. After 10 days we returned to the streams, inserted the rest of the bucket inside the bucket trays thus creating a watertight seal. Sediment was then stirred to a depth of 10cm for 30 seconds and a 125ml sample of resuspendable fine sediment was collected, following the SAM4 method. Samples were sent to Hill Laboratories (Christchurch) for measurement of total suspended sediment (TSS) and volatile suspended sediment (VSS). The remaining sediment and water slurry inside the bucket was then transferred to a larger container (Figure H-1), and transported to the laboratory for measurement of substrate size composition and TSS and VSS of the fine sediment portions (<0.63 mm and 0.63-2 mm).

Figure H-1: Sediment trays (a) buried at the stream surface were (b) retrieved along with the stream water following SAM4 sample collection.

Results

A significant difference in the benthic suspendable inorganic sediment concentration (SIS; g/m^2) of two study streams was detected when results were calculated from samples collected using standard methods (t-test, $p = 0.05$) or from processing the <63 µm portion of the entire benthic sediment sample ($p = 0.03$) (Figure H-2). Addition of the sediment portion between <63 µm and 2 mm resulted in no difference between the benthic fine sediment concentration of study sites. The same pattern was observed when data were expressed volumetrically (SIS; g/m^3) because samples were consistently 10-cm in depth. The standard SIS (Quinn et al. 1997) samples densities were double the Total < 63 um SIS by double the density at both sites (Figure C-2). Standard SIS was a less than half of the Total <2 mm density in the Maitai but only 20% lower in the Wakapuaka (Figure C-2).

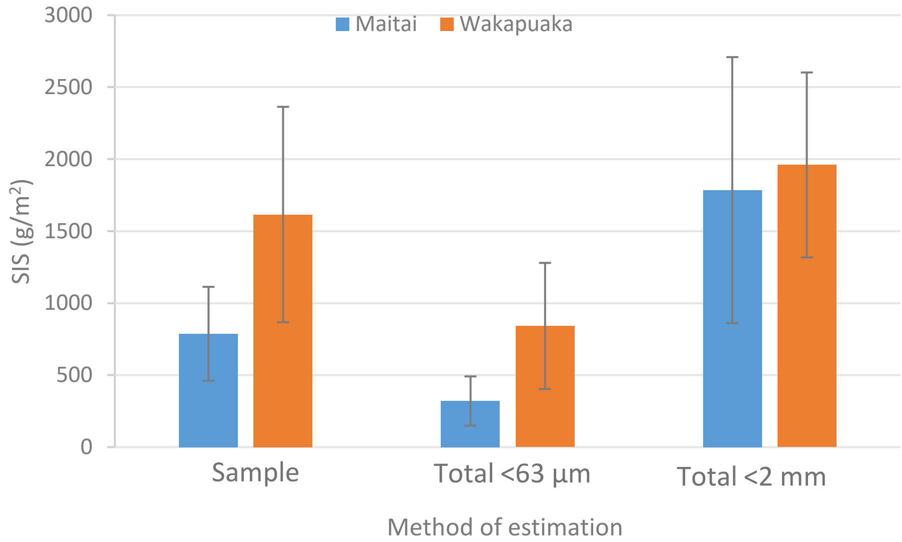


Figure H-2: Comparison of the re-suspendable inorganic sediment concentration (SIS, g/m^2) at two streams estimated from samples collected using standard method or by processing the entire benthic fine sediment sample.

The test of the SAM4 method (Appendix C) confirmed that SIS sampling provides a representative measure of SIS (mainly < 63 µm) entrained in the streambed. When stirred to an equal depth, the sampling results were able to discriminate between sites with varying SIS densities. However, the SIS appears to include both the <63 µm and variable amounts of the 63-2000 µm fraction since SIS was about double the <63 µm amount, but less than the total <2 mm amount. We did not test whether stirring to various depths influences SIS estimates, but expect that it would. Hence, we suggest that the method is only useful when the depth stirred can be standardised, as was done here.

Appendix I Eco Evidence: Systematic review of sediment effects literature

Introduction

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 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.

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

5. *Conceptual model*. Figure I-1.

6. *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.

7. *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).

8. *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.

9. *Catalogue and weight the evidence*. A total of 65 studies (Appendix J) 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 I-2), with greater weighting assigned to research having study design that controlled confounding influences and had greater replication of both controls and treatments.

10. *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 I-1) (Webb et al. 2013).

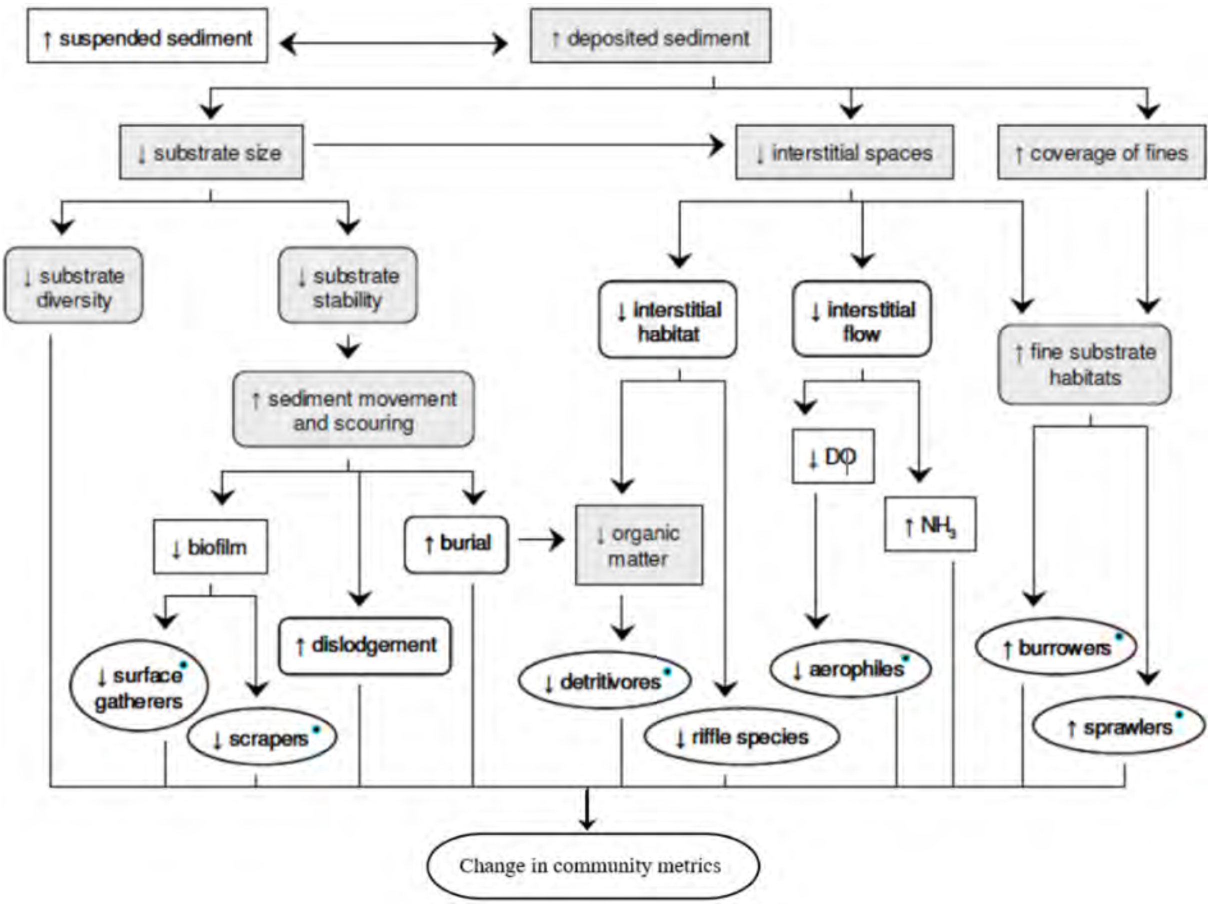


Figure I-1: Conceptual model for the effect of fine sediment on macroinvertebrates. Rectangular boxes are used for stressors, rounded rectangular boxes show an additional step in the causal pathway, and ovals are used for responses. Responses with blue-black dots indicate individual species are included 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 I-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 I-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 I-2).

Table I-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

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 I-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 I-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. In a second example, the hypothesis that an increase in the percent coverage of fine sediment at the patch scale causes an increase in macroinvertebrate density had a response of 19, but a consistency of association score of 20, thus supporting for an alternate hypothesis. In this case there was a 5% difference in the amount of evidence for each outcome, in contrast to the 250% difference seen in the first example.

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.

Appendix J Journal articles consulted as part of the Eco Evidence systematic 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 K Additional work to further to improve performance of sediment-specific metrics for macroinvertebrates

Further work is recommended to strengthen and validate tolerance value assignment and metric calculation to improve the performance of stressor-specific metrics. As outlined in Clapcott et al. (2017), this work could investigate:

- taking into account the strength of the relationships between the taxa and the stressor to inform tolerance values
- the inclusion of taxa classified as unclear in metric calculation
- the use of presence-absence macroinvertebrate data or density data instead of relative abundance data
- the use of expert opinion to assign tolerance values to taxa that were not present in the dataset or did not make it into the analysis as they had less than 10 occurrences
- the inclusion of taxa that showed subsidy-stress responses and were categorised as unclear in this analysis
- investigating other approaches to tolerance value assignment such as using individual, measured stressor gradients rather than an *a priori* overall pollution gradient based on expert opinion or the approach used for the MCIsb (i.e., an iterative rank correlation procedure following the method by Chessman (2003)).

Appendix L National macroinvertebrate and periphyton database collation

Macroinvertebrates

Data on benthic macroinvertebrates were obtained from 16 regional or unitary authorities, and the NRWQN. Collectively these represented a total of 2,005 sampling locations, not allowing for possible duplication between councils and the NRWQN. The largest contributions in terms of sampling locations were from the Waikato Regional Council (621 sites; 31.0% of the total) and Environment Canterbury (215 sites; 10.7% of the total). Four more councils (Bay of Plenty, Horizons, Environment Southland, Auckland) accounted for between 110 and 141 sites per region, collectively representing a further 516 sites (25.7% of the total).

Sampling dates ranged from 1990 to 2016, although the great majority (88.9%) were collected between 2000 and 2016. Excluding the NRWQN, which has operated continuously since 1990, data for the five years up to 2015 (the most recent year for which complete data were potentially available) were obtained from 12 of the 16 councils in the pooled data set.

Periphyton

Periphyton data were obtained from 10 regional or unitary authorities as well as the NRWQN, representing a combined total of 1,041 sampling locations. Of these, 1.009 could be unambiguously matched to a known macroinvertebrate sampling location. The geographical distribution of these locations was broadly similar to that of macroinvertebrate sites, with 429 (41.2% of the total) from the Waikato region, and a further 37.6% from Canterbury, Auckland, Southland, and Horizons. Sampling dates ranged from 2000 to 2016.

A distinguishing feature of the periphyton data set after pooling records across all contributing regions was the broad range of descriptors used to characterise periphyton abundance. These ranged from a single metric (Chlorophyll-a, total % cover) to more finely graduated descriptive indices of periphyton cover including (but not limited to) % epiphytic periphyton; % of long filaments; % of thick mats; % of thick green/light brown mats, % of thick dark brown/black mat, % of short green filaments, and % of long green filaments. Many of these descriptors, which numbered 122 in total, were specific to a single region.

To consolidate these variables into a more tractable subset of descriptors, and hence allow comparisons among regions, we created a parallel set of descriptors to relabel variables which we judged to be more or less equivalent across data sets. In some instances, this required synthesising new variables (e.g., % total cover) for data sets which included components (such as % mats and % filaments) but did not specifically report total cover. In particular, we identified four metrics which were sufficiently widely available to establish a sound basis for subsequent analyses: Chlorophyll-a (mg/m2); % cover of long filaments; % cover of thick mats; and % total cover (equivalent to the sum of long filament and thick mats). At least one of these variables was available for 5,810 samples from 2000 to 2016, representing a total of 1,031 sampling locations.

Appendix M Macroinvertebrate responses to deposited fine sediment: dataset compilation

Analysis of macroinvertebrate community responses to deposited fine sediment was performed on a national dataset linking macroinvertebrate data with deposited sediment and other stressor data. We chose three different analytical approaches for identifying ecological thresholds to provide multiple lines of evidence for determination of deposited sediment attribute thresholds.

Dataset compilation

The macroinvertebrate-stressor dataset was specifically created from a large national macroinvertebrate dataset containing SoE data provided by regional and unitary councils as well data collected by NIWA from National River Water Quality Network (NRWQN) sites[22], typically collected on an annual basis. To boost sample size and the spread of sites around New Zealand, we added research data that had been compiled for a companion MfE-funded macroinvertebrate project (Clapcott et al. 2017).

Stressor datasets

Deposited sediment and other stressor data were retrieved from three separate datasets.

1) The Sediment Stage 2 dataset consisted of deposited fine sediment data that had been compiled during the Sediment Stage 1 project and updated as part of this project as described in section 2.4.2.

- Deposited sediment measures were:
  - bankside visual assessment of sediment cover within the Rapid Habitat Assessment (RHA) protocol
  - bankside visual assessment of sediment cover (% cover bankside, SAM1)
  - instream visual assessment of sediment cover (% cover instream, SAM2)
  - Wolman pebble count (% fines, SAM3)
  - suspendable inorganic sediment (SIS, SAM4)
  - suspendable benthic sediment volume (SBSV, SAM4)
  - shuffle test score (SAM5).
- Suspended sediment measures included:
  - total suspended solids (TSS)
  - turbidity

[22] Compiled by Martin Unwin, NIWA, Christchurch.

– visual clarity.

The frequency of deposited sediment assessments at a single site generally varies largely. In some instances, assessments were made monthly, but in most cases deposited sediment assessments were made annually, or had been done only once for a site.

2) Water quality data collected at SoE monitoring sites (typically monthly), was retrieved from the LAWA (Land, Air and Water Aotearoa) website (downloaded 5 May 2017). The following water quality measures were of interest:

- turbidity
- visual clarity (black disk)
- ammonium-nitrogen (NH_4N)
- total oxidised nitrogen (NO_xN)
- total nitrogen (TN)
- dissolved reactive phosphorus (DRP)
- total phosphorus (TP).

3) Periphyton data, compiled as part of this project as described in Appendix L; mostly assessed at SoE or NRWQN sites monthly or annually. The following periphyton measures were of interest:

- benthic chlorophyll *a* (chl *a*)
- visual assessment of total periphyton cover.

Matching macroinvertebrate and stressor data

Threshold analyses required matching a single macroinvertebrate sample with a single deposited sediment value. Due to potentially significant annual variation of deposited sediment at a single site, we aimed at calculating a median value from all available data collected within the same month and the 12 months prior to macroinvertebrate sampling. Similarly, we aimed at calculating median values for periphyton, nutrient and suspended fine sediment data to match with a single macroinvertebrate sample for threshold analyses that accounted for variation of other stressors.

The macroinvertebrate and stressor datasets contained various identifiers with which samples could be matched. Matching deposited and suspended sediment data from the Sediment Stage 2 dataset with macroinvertebrate data was done first. The matching process first used site name (e.g., "Makotuku at Raetihi") and sampling date, accounting for possible multiple sediment sampling dates as described above. Inconsistencies in site names e.g., due to use of 'at' or '@', 'Road' or 'Rd', as well as spelling mistakes of rivers, road and place names, may have caused potential matches to be missed. Accordingly, uniform site names were created and "fuzzy matching" accommodated minor inconsistencies in site names. Exact matches were accepted without further checking whereas fuzzy matches were manually checked using their NZReach ID's. This process increased real matches.

For cases where no site name match was found, matching was then performed using regional council site ID (RCSID). Lastly, further matches between macroinvertebrate and deposited sediment data were found via NZReach ID, limited for those NZReach IDs where a single macroinvertebrate site existed. There were multiple occasions where sample sites were in close proximity, typically upstream and downstream of sewage treatment plants, resulting in very similar uniform site names and both having the same NZReach ID. These cases were matched manually before matching by NZReach.

After matching of macroinvertebrate with sediment data, macroinvertebrate data were matched with LAWA data, first by matching site name and date using the same fuzzy matching approach. Secondly, further matches were found by comparing regional council site ID (RCSID) and LAWA ID. Thirdly, further matches were found using NZREACH ID for those cases where only a single site was sampled within a NZReach ID. Finally, periphyton data was matched with macroinvertebrate data using site name and date, RCSID and NZReach ID as above.

Of the 15,508 macroinvertebrate samples contained within the macroinvertebrate dataset, we were able to match 4,717 samples with at least one measure of deposited fine sediment. Sample size for each of the seven deposited sediment measures is presented in Table M-1. For threshold analyses we decided to use three measures: 1) instream visual assessment of sediment cover *(% cover instream*), because it is a common measure used by councils and in research, 2) bankside visual assessment of sediment cover *(% cover bankside*), because it is a common measure used by councils and had a large sample size in this dataset, and 3) SIS, because it is a relatively good estimate of the amount of surficial and interstitial fine sediment.

Additional research data

In addition to SoE monitoring data, data was also included samples from a recently compiled research dataset for the MfE-funded macroinvertebrate project used to develop sediment-specific macroinvertebrate metrics. This research dataset predominately contains stressor data from a single observation taken on or close to the day of sampling macroinvertebrates. The final total sample size and number of sites for each the % cover instream, % cover bankside and SIS can be found in Table M-1. Figure M-1 shows the spread of sites across the country for each of these three focal deposited sediment measures.

Frequency of deposited sediment assessments

Frequencies of deposited sediment assessments within the calendar month or 12 months prior to macroinvertebrate sampling used to calculate median values varied largely among the macroinvertebrate samples for instream and bankside visual assessments of sediment cover. Out of the total number of 571 macroinvertebrate samples (Table M-1) with matching % cover instream data, about half of the samples were matched with a single sediment observation, and for about a third, two observations were available. The remaining 20% of the samples were matched with medians calculated from between three and 12 observations. Out of the total number of 2,620 macroinvertebrate samples (Table M-1) with matching % cover bankside data, about half of the samples were matched with a single sediment observation and a third with two observations. The remaining 20% of the samples were matched with medians calculated from between three and 23 observations. Finally, all 83 macroinvertebrate samples (Table M-1) were matched with a single observation of SIS.

Table M-1: Sample size of each of seven deposited fine sediment measures within the national SoE macroinvertebrate-stressor dataset. Includes total sample size and number of sites in the final combined SoE and research dataset for three selected deposited fine sediment measures (grey shading).

Deposited sediment measure	Sample size (SoE data only)	No. of sites (SoE data only)	Total sample size (SoE & research data)	Total no. of sites (SoE & research data)
% cover instream	571	188	1,039	593
% cover bankside	2,620	467	2,708	555
SIS	83	47	449	302
Wolman pebble count	1,403			
SBSV	33			
Shuffle test score	74			
Bankside visual assess. (RHA protocol)	591			

Data checking

For the majority of sample sites NZReach ID was known, and predictions of environmental variables were retrieved from various existing databases. For example, the percentage of intensive pastoral land use in the catchment (T2PastoralHeavy) and the percentage of native vegetation cover (T1NativeVeg) calculated from the Land Use Cover Data Base 3 (LCDB3) were retrieved and bivariate scatterplots produced for various macroinvertebrate metrics. These scatterplots were investigated for unusual values among the stressor attributes. We also checked data distributions of a set of macroinvertebrate metrics and compared summary statistics of these metrics between the SoE and research datasets. We also visually investigated if summary statistics of the macroinvertebrate metrics differed according to the sampling method (quantitative vs. semi-quantitative). Overall, the data looked fine.

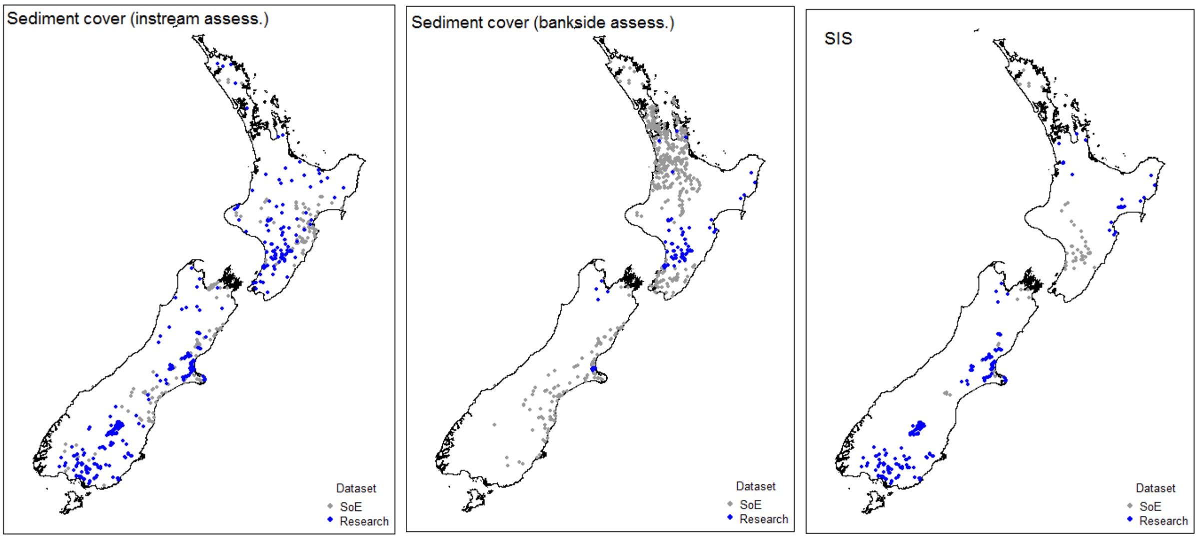


Figure M-1: Spread of sample sites across New Zealand. Sites colour-coded by as to whether data was were retrieved from a national SoE dataset (grey), or from a research dataset (blue) for each % cover instream, % cover bankside, and SIS; see Table M-1 for sample size and number of sites.

Appendix N Quantile regression (QR), Boosted regression tree (BRT) and gradient forest (GF) methods used to determine deposited sediment thresholds using macroinvertebrate data

Methods

To provide multiple line of evidence, three analytical approaches were adopted for threshold identification using the specifically compiled national macroinvertebrate-stressor dataset (Appendix L). For each of these approaches, separate analyses were performed for each of the three deposited sediment measures:

- *% cover instream*
- *% cover bankside*
- *suspendable inorganic sediment* (SIS).

The first approach is a single-stressor analysis while the other two approaches incorporate multiple predictors. The first two approaches use aggregate macroinvertebrate metrics (e.g., MCI, EPT) as response variables whereas method 3 produces models for individual taxa and combines the information to determine assemblage thresholds. The three methods employed were:

- Method 1: quantile regression model (referred to as QR method) – single stressor analysis using macroinvertebrate metrics as response variables.
- Method 2: boosted regression tree model (referred to as BRT method) – multiple predictor using aggregate macroinvertebrate metrics as response variables.
- Method 3: random forest model (referred to as RF method) – multiple predictors, uses individual taxa as response variables.

We considered a set of 16 macroinvertebrate metrics that were expected to respond to deposited sediment based on previous research including commonly-used metrics by regional councils and recently developed sediment-specific macroinvertebrate metrics (refer to Table 2-11). All metrics were calculated from MCI-level taxonomic resolution (Table N-1).

Table N-1: Summary statistics of the 16 candidate macroinvertebrate metrics. A short description of the new sediment-specific macroinvertebrate metrics, indicated with *, can be found in section 4.4.

Macroinvertebrate metric	Min	Max	Mean	Median
MCI (hard-bottom)	27	173	101	102
QMCI (hard bottom)	1	9	5	5
EPT taxon richness	0	26	7	7
EPT taxon richness (excl. Hydroptilidae)	0	26	7	7
% EPT abundance	0	99	39	37
% EPT abundance (excl. Hydroptilidae)	0	99	37	35
% EPT richness	0	91	40	43
% EPT richness (excl. Hydroptilidae)	0	91	37	40
Sediment MCI (raw scale assignment of bins)*	20	200	126	133
Sediment MCI (log-scale assignment of bins)*	20	200	117	123
Sediment QMCI (raw scale assignment of bins) *	1	10	5	6
Sediment QMCI (log-scale assignment of bins) *	1	10	5	5
No. of sensitive taxa*	0	22	6	6
No. of tolerant taxa*	0	9	3	3
% sensitive taxa*	0	100	35	37
% tolerant taxa*	0	75	19	18

Based on scatterplots of these metrics across the deposited sediment gradients, we selected the following four macroinvertebrate responses for threshold analyses (using method 1 and method 2):

- MCI.
- EPT taxon richness.
- Sediment MCI.
- Number of sensitive taxa.

All analyses were performed in statistical programme R (R Core Team 2016), with specialised functions from a range of R packages.

Method 1: Quantile regression analysis (QR method)

We adopted a single-stressor analytical threshold approach were the sediment threshold is identified at a predetermined biological benchmark effect using a simple linear regression model (Cormier et al. 2008). The benchmark effect here is a reduction in a macroinvertebrate metric from reference condition to benchmarks of 5, 10, 15 and 20%, which are consistent with suggestions in the threshold literature (Cormier et al. 2008). Cormier et al. (2008) suggested that for criteria development in the United States, a 5% and 20% change from reference condition could be used to calculate candidate criteria for aquatic life use and marginal aquatic life use, respectively, and defined the biological reference condition at the y-intercept. We defined the biological reference condition at the macroinvertebrate metric value derived from the regression model at the minimum sediment value observed instead.

Regression analysis requires the response variables and residuals to approximate a normal data distribution. Scatterplots of macroinvertebrate metrics across the three measures of deposited sediment were mainly wedge-shaped. This indicates that our dataset encompasses multiple stressor gradients and that an upper percentile rather than the mean value of the macroinvertebrate metric would be better suited to model the stressor-response relationship. Hence, we decided to use a linear quantile regression (QR) model. We calculated and plotted regression models for four upper percentiles, 80^{th}, 85^{th}, 90^{th} and 95^{th} percentile). Selection of the quantile to use for analysis was done following the criteria and plots suggested by Kail et al. (2012). Hence, the 85^{th} percentile was selected because it was an upper quantile with generally the narrowest confidence interval for the quantile regression line (Figure N-1). Quantile regressions were performed with R package *quantreg*. In order to compare model fit among the various quantile models, we calculated R^2 following the suggestions by Koenker and Machado (1999).

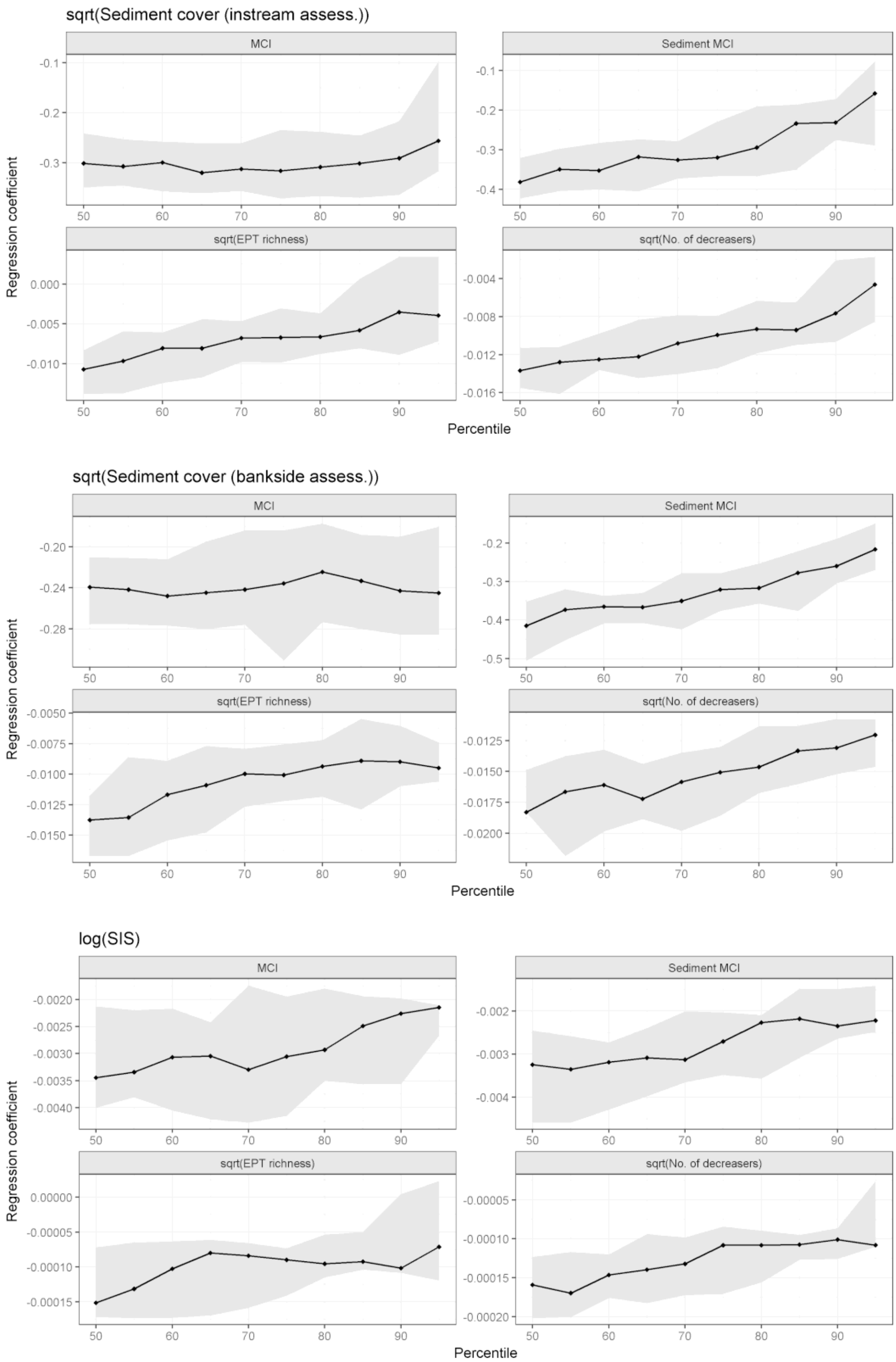


Figure N-1: Coefficients of quantile regression lines for percentiles ranging from 50th-95th along with their confidence interval. Plots were used to help select the percentile for threshold analysis according to Kail et al. (2012).

Method 2: Boosted regression tree analysis (BRT method)

Boosted regression tree (BRT) analysis is a flexible modelling approach that allows incorporation of multiple predictors. Contrary to multiple linear regression analysis, correlations between predictors are handled well. Interactions between predictors are automatically handled. BRT analysis is well described in the statistical literature for ecology (De'ath 2007; Elith et al*.* 2008). BRT output provides a 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 determined by a 10-fold CV procedure. 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 can be used for visual threshold definition (Wagenhoff et al*.* 2017b). Inclusion of stressors in the model other than deposited sediment as well as of environmental predictors improves confidence in the fitted function depicting the response shape to sediment rather than the response to another predictor that is correlated with increasing sediment.

Each BRT model was built for a single response variable which has to be approximately normally distributed, hence the same transformations were used as for the quantile regression analysis. Furthermore, the transformed response variables were standardised by dividing by the standard deviation in order to make the effects of deposited sediment comparable among the macroinvertebrate metrics. Predictors do not need to be transformed. The BRT models contained 17 predictors including a single deposited sediment measure, chlorophyll *a* as a means of accounting for the effect of nutrients via periphyton biomass, turbidity as a measure of suspended fine sediment, and a range of predicted variables retrieved from large databases that hold information for each NZReach (Table N-2). Three separate BRT models were built for each macroinvertebrate metric, one for each focus deposited sediment measure.

Table N-2: Set of 17 predictor variables used in BRT models along with their data source and description. Three flow statistics (Booker 2013; Booker & 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 & Biggs 2002), FENZ = Freshwater Ecosystems New Zealand database (Leathwick et al. 2011).

Predictor	Source	Description
instreamVis OR visualBank OR SIS	measured	% cover instream OR % cover bankside OR suspendable inorganic sediment (g/m^2)
CHLA	measured	Benthic chlorophyll *a* from rock scrapings
Turbidity	measured	Turbidity
ORDER	REC	Stream order
ELEVATION	REC	Altitude of the stream segment
SegJanAirT	FENZ	Summer air temperature for a segment
SegMinTNor	FENZ	Seasonal air temperature range for a segment
SegRipShade	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
USSlope	FENZ	Average slope in the catchment
DSDist	FENZ	Distance to coast
SegFlowStability	FENZ	Ratio of mean annual low flow/ mean annual mean flow
SpecMeanF	MfE website	Specific mean flow (= mean flow / catchment area)
SpecMALF	MfE website	Specific mean annual low flow (= mean flow / catchment area)
FRE3	MfE website	Annual frequency of flood events > 3x median annual flow

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 deposited sediment measure requiring non-missing values for chlorophyll *a*, which is an important stressor variable. Missing values were allowed for all other predictors. Sample size and number of sites for the three datasets used for BRT analysis are provided in Table N-3 and maps of the distribution across New Zealand are given in Figure N-2. 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 and Leathwick (2014).

Table N-3: Sample size and number of sites of the subsets used for BRT analysis for each deposited sediment measure.

Deposited sediment measure	Sample size	Number of sites
% cover instream	602	354
% cover bankside	199	100
SIS	243	156

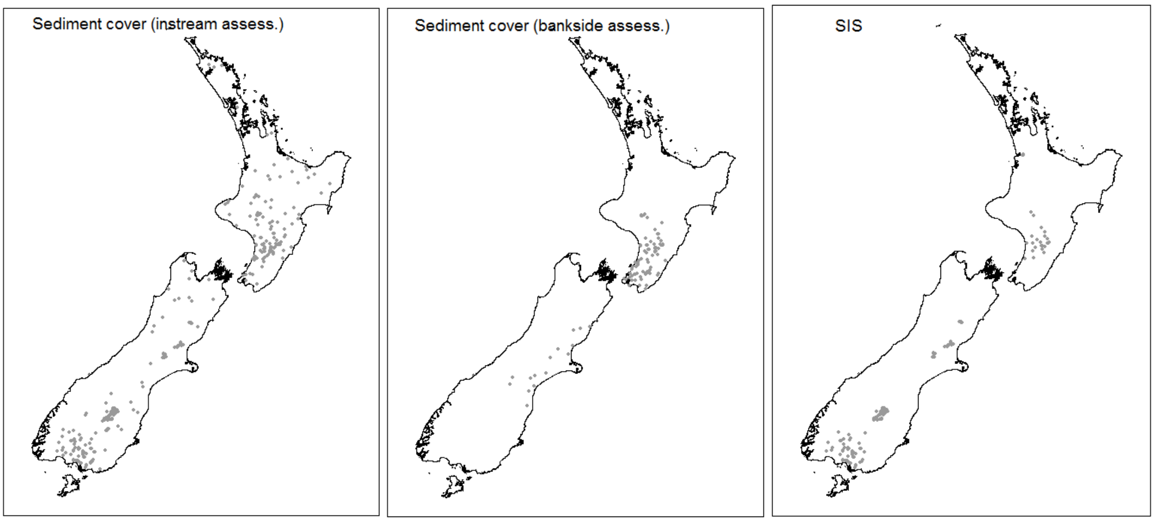


Figure N-2: Spread of sample sites across New Zealand for the three subsets (% cover instream, % cover bankside and SIS) used for BRT analysis. Refer to Table N-3 for sample size and number of sites.

Method 3: Gradient forest analysis (GF method)

Gradient forest (GF) analysis was developed to identify community thresholds defined as a point(s) at which a small increase in a stressor will result in a disproportionally large change in community structure relative to other points across the stressor gradient (Ellis et al. 2012). This approach has recently been used to identify community thresholds of three different stream assemblages including macroinvertebrates (Wagenhoff et al. 2017a).

Detailed description of the GF approach can be found in Ellis et al. (2012). Briefly, random forest (RF) models are built for each taxon that meets a cut-off for occurrence across the samples in the dataset. We set the cut-off at ≥10 occurrences. Random forest models, like BRT models, are flexible regression-tree models. Model assumptions require residuals to be approximately normally distributed and we chose to log-transform the macroinvertebrate data using the formula:

$$\ln(x + \min(x>0))$$

where x is the relative abundance, expressed as a proportion. As for the univariate threshold analyses using metric responses, taxon data was brought to MCI-level taxonomic resolution. Similar model outputs as for BRT models are produced by the RF procedure such as the relative importance of the predictors, partial dependence plots and a goodness-of-fit measure R^2, which is the proportion of the variance explained by the RF model and derived through cross-validation. Only taxa with models of an $R^2>0$ were taken into account for community threshold identification. Two R packages, *extendedForest* (based on package *randomForest*) and *gradientForest*, were used to implement the GF analysis.

Community thresholds were derived from aggregation of the information from RF models of the taxa into cumulative splits importance curves of the overall macroinvertebrate community. Thresholds can be visually identified from split density plots that take into account data density across the stressor gradient. The GF approach also produces cumulative splits importance curves of all taxa to investigate which taxa were mainly contributing to community thresholds. Only the fitted functions of the RF models, however, tell whether a taxon increased or decreased across the stressor gradient. GF output also provides the importance of each predictor for overall compositional turnover which is calculated by taking a weighted average of the taxon–specific predictor importances using the R^2 values of the RF models (Ellis et al. 2012).

The predictor variables used in the RF models were the same we used for BRT analysis (i.e., Table N-2), hence the GF approach was implemented for each 1) *% cover instream*, 2) *% cover bankside*, and 3) SIS. Model parametrisation was chosen as suggested by Ellis et al. (2012). As RF models do not allow missing values of the predictor variables, the dataset used for community threshold analysis was further reduced from that used for BRT analysis. Sample size and number of sites for the three datasets used for GF analysis can be found in

Table N-4 and maps of the spread across New Zealand in Figure N-3.

Table N-4: Sample size and number of sites of the subsets used for GF analysis for each deposited sediment measure.

Deposited sediment measure	Sample size	Number of sites
% cover instream	380	211
% cover bankside	184	97
SIS	66	31

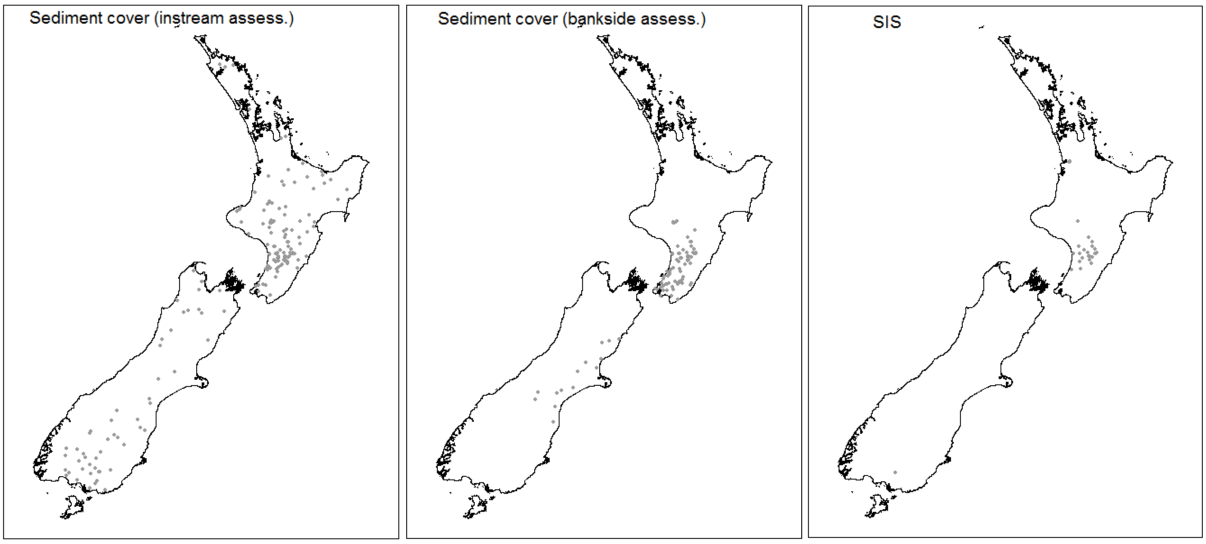


Figure N-3: Spread of sample sites across New Zealand for the three subsets (% cover instream, % cover bankside and SIS) used for GF analysis. Refer to Table J-4 for sample size and number of sites.

Appendix O Additional results from method 1: quantile regression (QR) method

Table O-1: Single-stressor sediment thresholds calculated from the 85th quantile regression model (QR method). Calculated for a benchmark effect of 5, 10, 15 and 20% along with the fitted metric value at the threshold as well as the maximum fitted metric value (i.e., reference); '-' indicates that the threshold was beyond the observed stressor gradient.

Macroinverte response Metric	Sediment measure/ predictor	Max.value of metric response	5% reduction in max.		10% reduction in max.		15% reduction in max.		20% reduction in max.[a]	
			Sed. thresh.	Fit at thresh.	10% Sed. thresh.	Fit at thresh.	Sed. thresh.	Fit at thresh.	Sed. thresh.	20% Fit at thresh.
MCI	% sediment cover (instream assess.)	133	**8**	127	**32**	120	**72**	113	-	-
EPT taxon richness		12	**73**	11	-	-	-	-	-	-
Sediment MCI		151	**12**	144	**48**	136	-	-	-	-
No. of decreasers		13	**6**	12	**24**	10	**53**	9	**94**	8
MCI	% sediment cover (bankside assess.)	132	**19**	125	**75**	118	-	-	-	-
EPT taxon richness		14	**9**	13	**33**	11	**75**	10	-	-
Sediment MCI		151	**16**	143	**65**	136	-	-	-	-
No. of decreasers		13	**4**	12	**15**	11	**33**	9	**57**	8
MCI	SIS (g/m^2)	140	**41**	133	**261**	126	**1629**	119	**10040**	112
EPT taxon richness		17	**26**	15	**117**	14	**501**	12	**2084**	11
Sediment MCI		153	**67**	145	**720**	138	**7449**	130	-	-
No. of decreasers		16	**19**	14	**69**	13	**230**	11	**753**	10

[a] approximate magnitude of effects (i.e., reduction in maximum or peak value of response) most likely to correspond to C/D band threshold (i.e., national *bottom-line*)

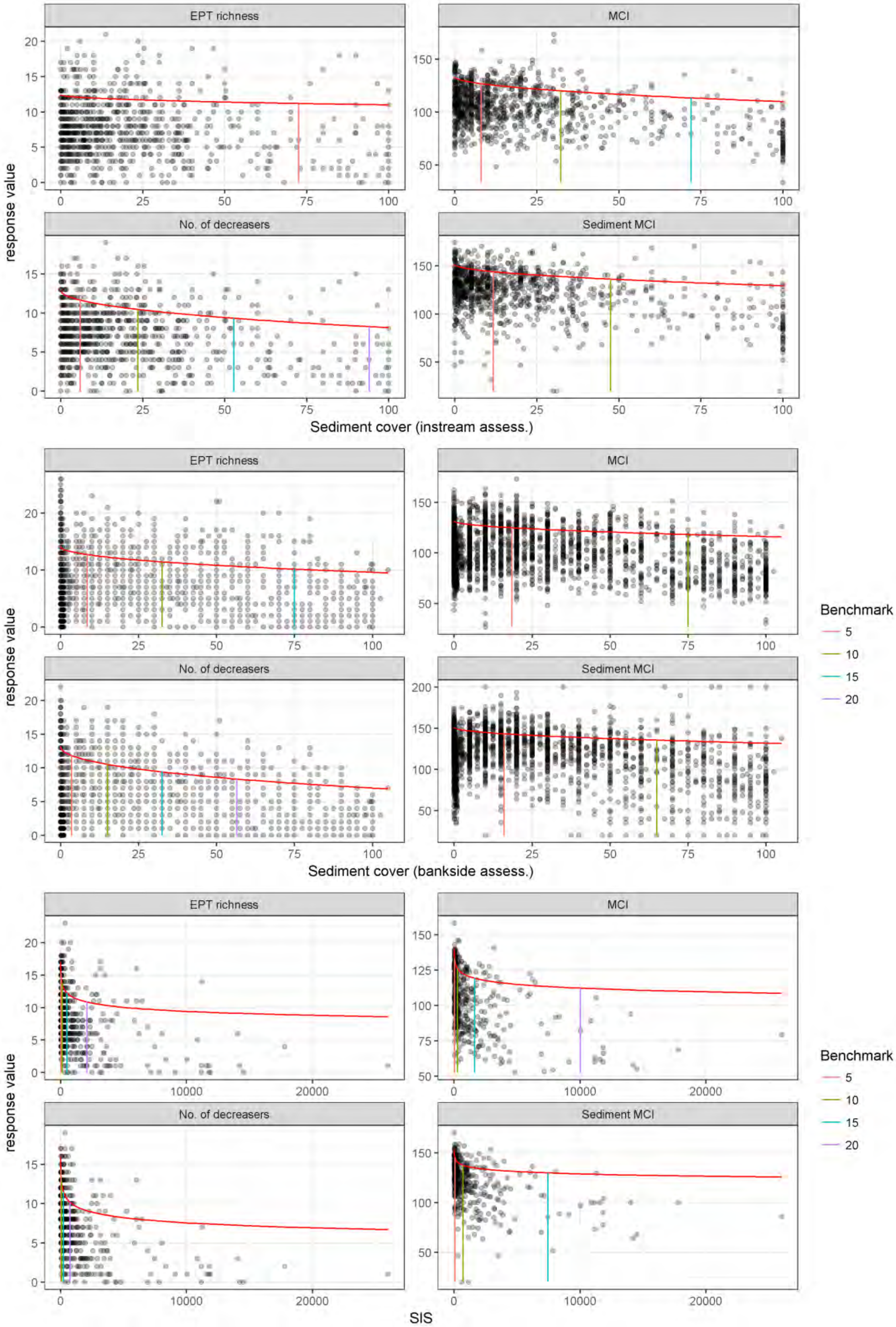


Figure O-1: The 85th quantile regression models plotted for raw sediment and metric values along with the sediment thresholds (vertical lines). Calculated at 5, 10, 15 and 20% benchmark effect. Note that data has been back-transformed to help comprehension of the data distribution and sediment thresholds.

Appendix P Additional results from gradient forest (GF, method 3) analyses of deposited sediment and macroinvertebrate response

Table P-1: Summary of taxa present and analysed for GF approach for each deposited sediment measure.

	% sediment cover (instream)	*% sediment cover (bankside)*	SIS (Quorer)
Number of taxa present	112	114	107
Number of taxa analysed ≥10 occurrences)	64	60	45
Number of taxa RF model R^2>0	13	37	17
Mean R^2 (range)	0.11 (0.00-0.47)	0.19 (0.01-0.46)	0.21 (0.03-0.53)

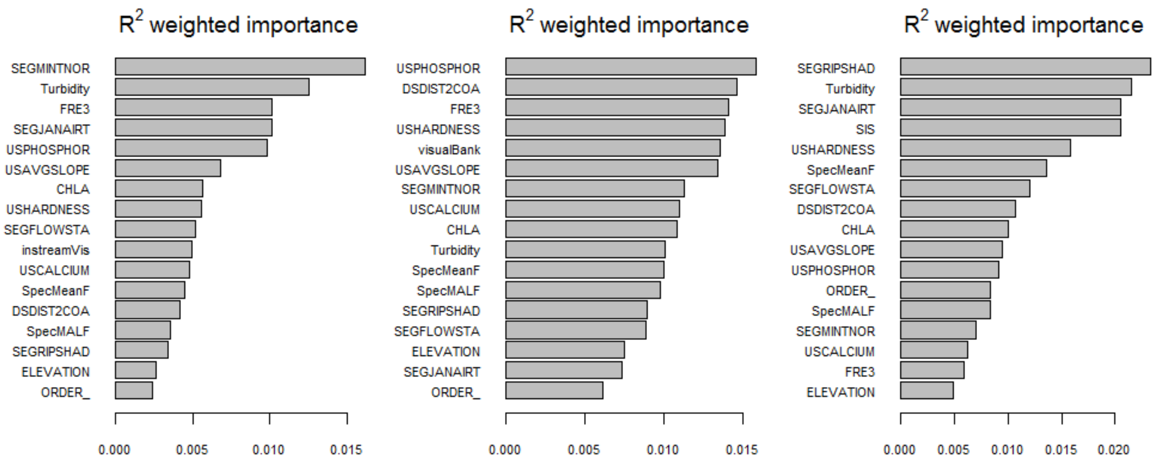


Figure P-1: Overall predictor importance (in R2 units) for taxon distribution. Calculated by gradient forest (GF) analysis allowing assessment of the relative importance of the predictors for *% sediment cover (instream)* (left), *% sediment cover (bankside)* (middle), and SIS (right).

Appendix Q Analysis of temporal variation to inform attribute frequency criteria

Methods

We explored temporal variation in SIS and *% sediment cover* metrics to determine the minimum/recommended number of sampling events required to assess compliance against proposed deposited sediment attribute thresholds.

For SIS, the only suitable data available was from the Whatawhata integrated catchment management research project (NIWA, unpublished data). Eight sites were sampled quarterly or biannually between 1995 and 2013, providing between 22 and 53 replicate samples per site. Five of the 8 sites were treatment sites and subject to a change in catchment vegetation (predominantly due to pine afforestation) in 2000/2001, and the remaining 3 sites were controls (with greater than 69% native vegetation), as described in Quinn et al. (2009). For each site, we calculated the mean and standard deviation in SIS pre-and post-treatment, checked for autocorrelation among samples, and determined the sample size required to estimate the mean within a 99% confidence interval (2.58 x standard error of the mean). To do this we calculated the rolling mean, using 2 to 30 replicate measures, and determined the number of times the rolling mean fell within the previously-calculated 99% CI for the population mean.

For the % cover sediment dataset, we identified sites from the collated database where the % cover of fine sediment had been repeatedly measured over time using either bankside (SAM 1) or instream (SAM2) visual assessment methods. 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 rather than the 99% confidence interval approach used for SIS to account for experimental error; visual assessments can only assess % deposited sediment cover in 5% intervals, at best (more realistically at 10% intervals).

Results

The site mean SIS value observed at Whatawhata ranged from 515 g/m^2 (site NW5) to 1404 g/m^2 (site PR2), and the site standard deviation ranged from 340 g/m^2 (site PKR) to 1430 g/m^2 (site PW3). Durbin-Watson testing of serial temporal autocorrelation showed no significant violation of independence in samples over time. No change in variability was detected (Levene's test for homogeneity in variance) before and after catchment treatment (Figure Q-1). At only one site (DB4), was there a significant difference in mean SIS before and after catchment treatment (Student's t test, $p = 0.03$).

To obtain greater than 80% chance of a rolling mean providing an estimate within the 99% CI required a window width encompassing ~13 consecutive measurements. The number varied from site to site depending on the variance of that site, ranging between samples of 7 measurements to 25 measurements. This equates to 2 years of quarterly measurements at a reference site and 6 years of quarterly measurements at a non-reference site, although there was no significant difference ($p=0.67$) between the number of measurements needed at reference and non-reference sites. On average, 3 years of quarterly measurements are required to accurately estimate SIS at Whatawhata sites.

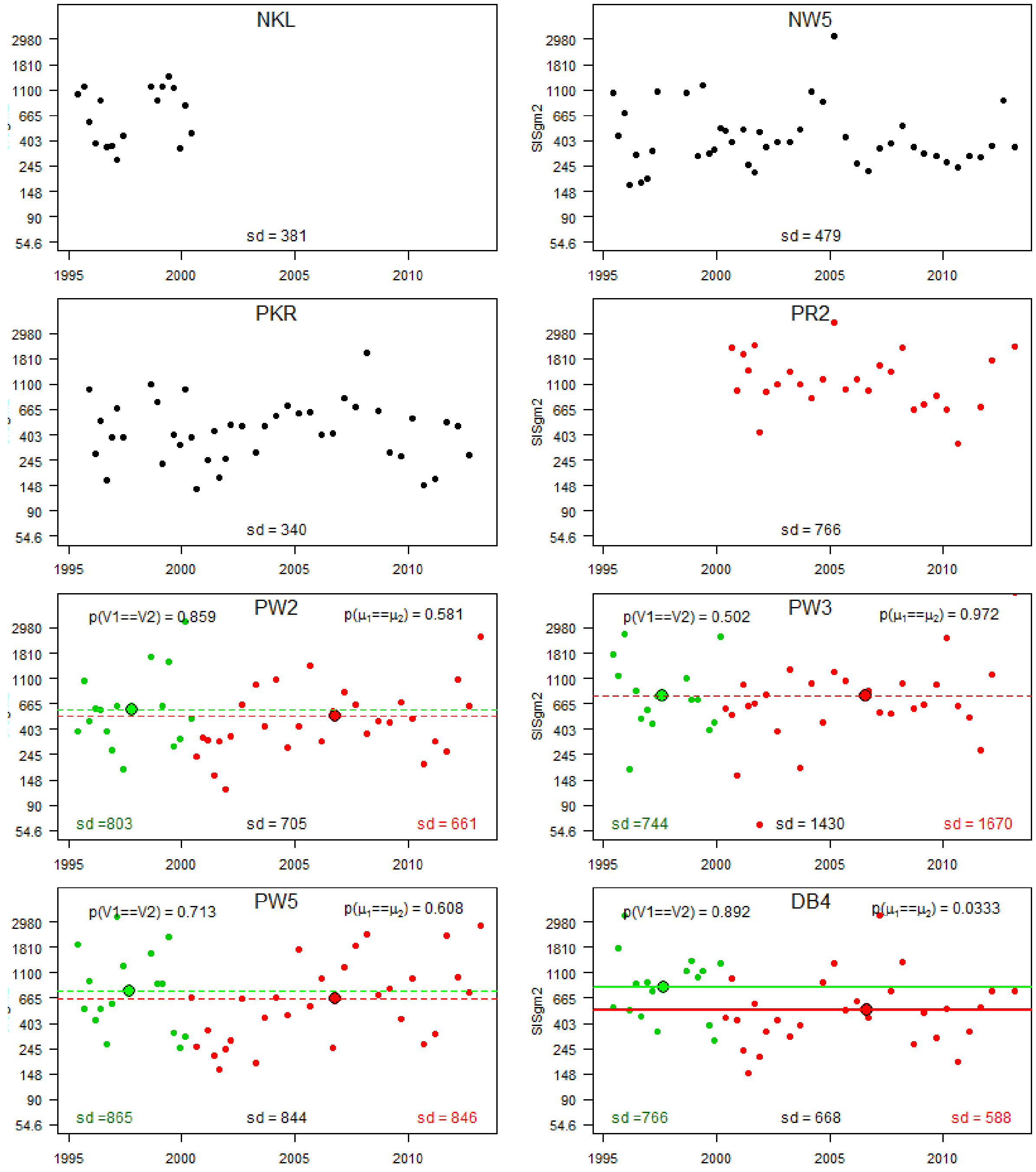


Figure Q-1: Temporal variation in suspendable inorganic sediment (SIS) measured at 8 Whatawhata research sites. See Quinn et al. 2009 for description of sites. The means are indicated by horizontal dashed lines. Green and red dots and green and red solid lines are pre- and post-catchment treatment, respectively.

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 Q-2) and SAM2 data (Welch's t-test $p=0.007$).

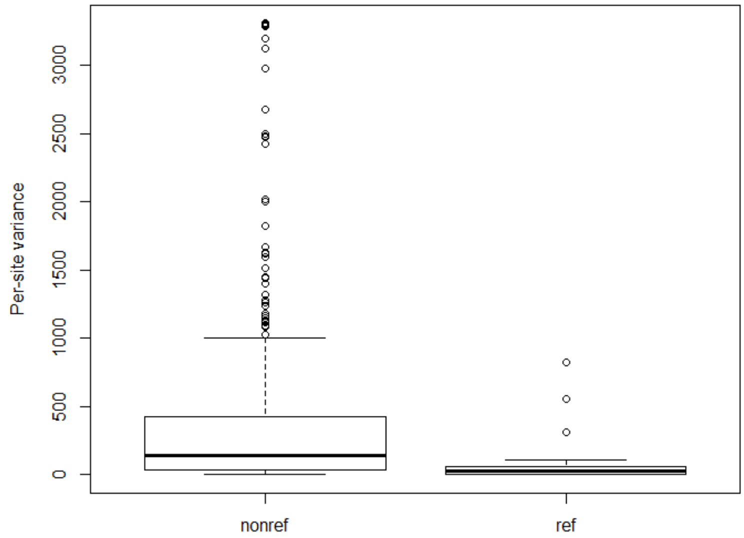


Figure Q-2: Average temporal variation in % cover of deposited fine sediment measured using the visual bankside (SAM1) method at reference sites and non-reference sites.

The variance was strongly dependent on the mean sediment cover: sites with mean sediment closer to 50% had higher variance (Figure Q-3). 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.

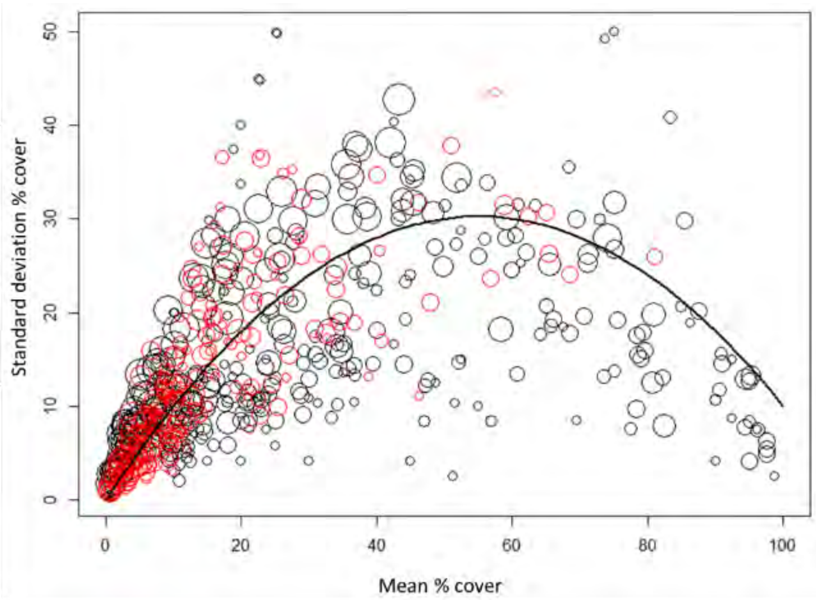
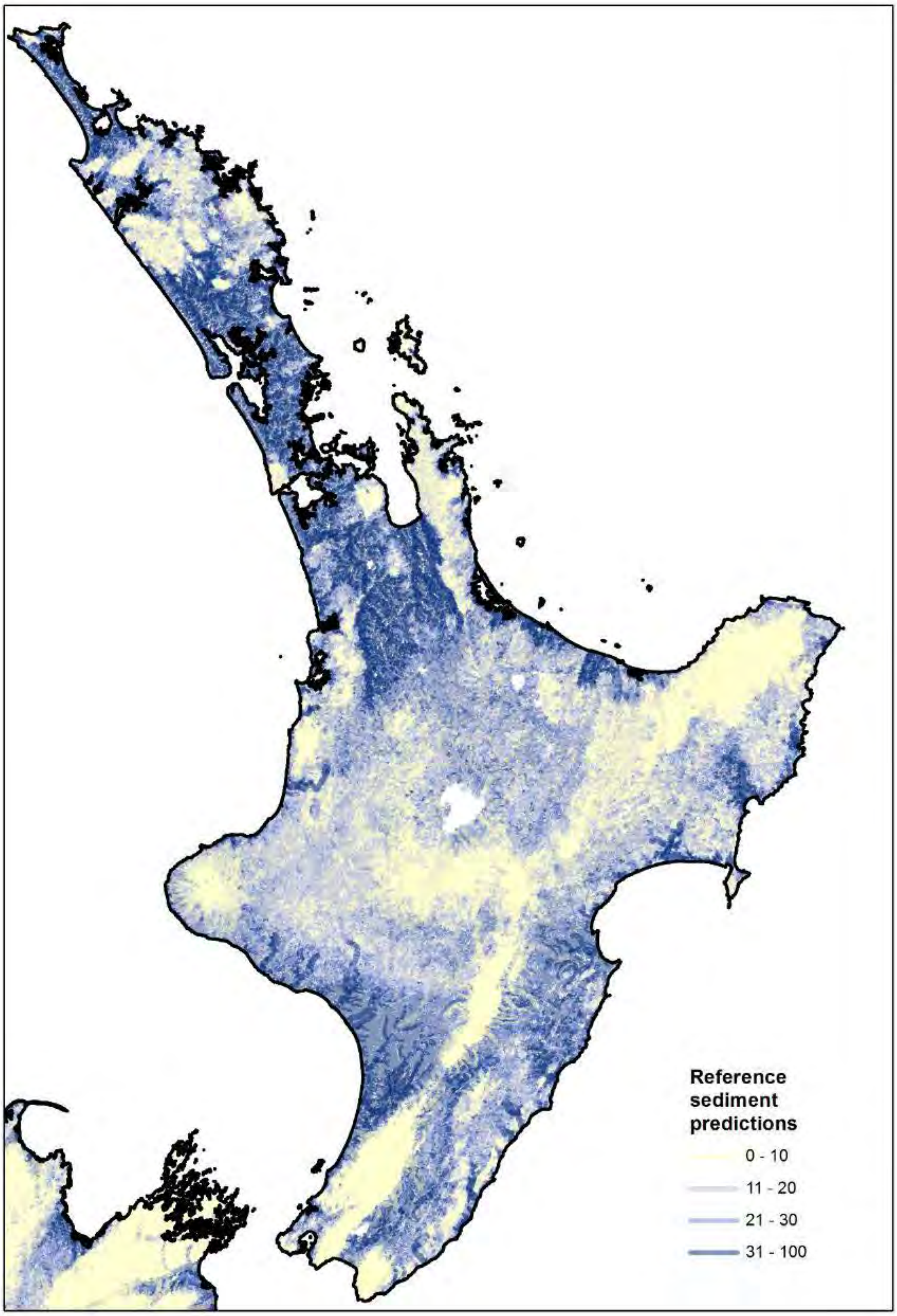


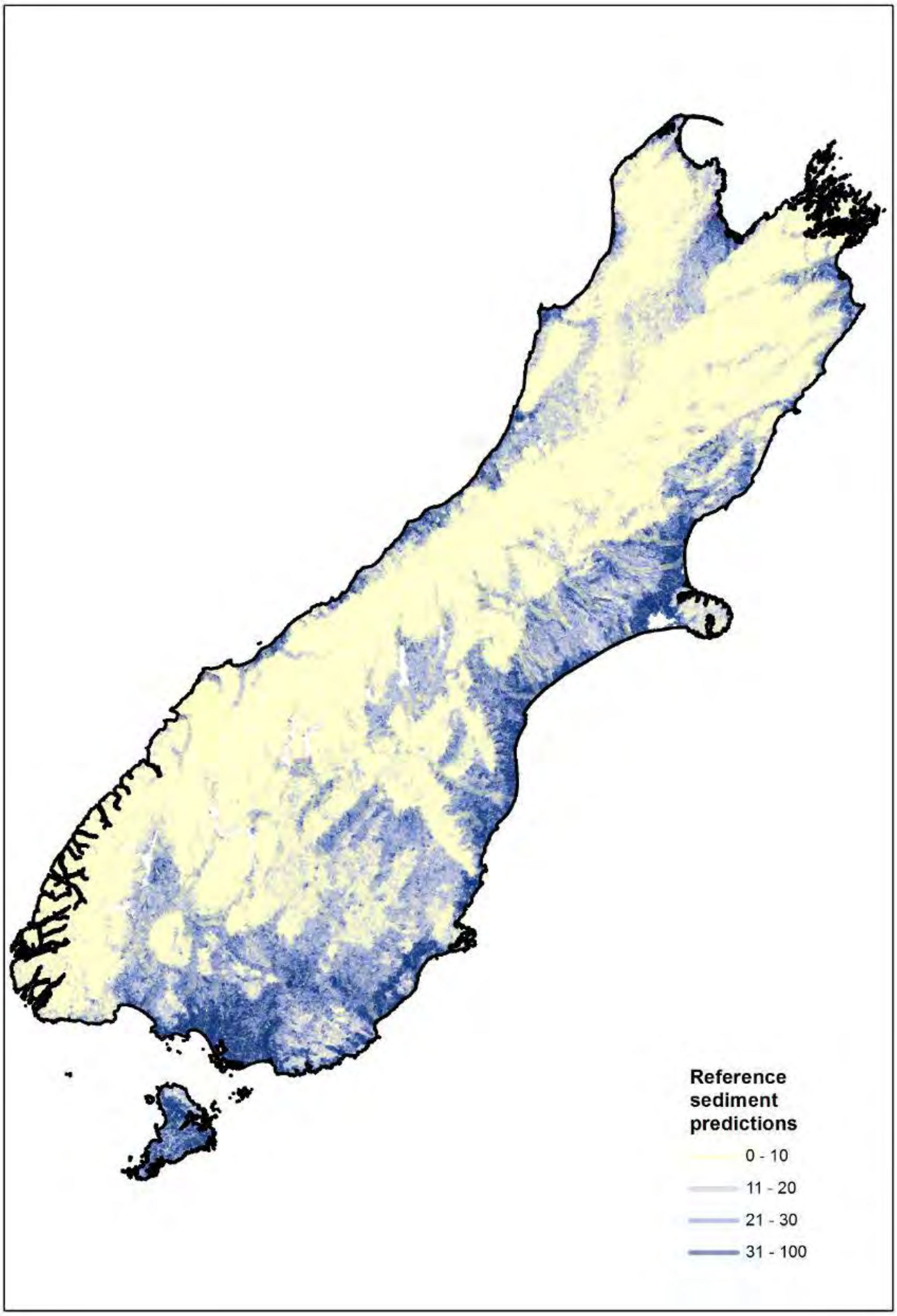
Figure Q-3: The linear relationship (with a quadratic term; R^2 = 0.85) between mean and standard deviation in % sediment cover. Red and black circles show % cover bankside and % cover instream data respectively. Circle size is relative to sample number for each site.

Summary

We explored temporal variation in SIS data at 8 sites and concluded that up to 6 years of quarterly measurements are needed to accurately estimate mean values. In contrast as few as 24 monthly samples are required to accurately estimate the cover of fine sediment based on % cover instream measurements, taking into account the likely experimental error of visual estimates.

Appendix R Maps showing predictions of deposited fine sediment cover modelled from environment variation at 2022 reference sites.




Reference
sediment
predictions
0 - 10
11 - 20
21 - 30
31 - 100

APPENDICES FOR SECTION 5

Appendix S Quantile regression background

(modified from (US EPA 2017))

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 & Noon 2003). As with mean regression, the relationship is often assumed to be a straight line (Figure S-1).

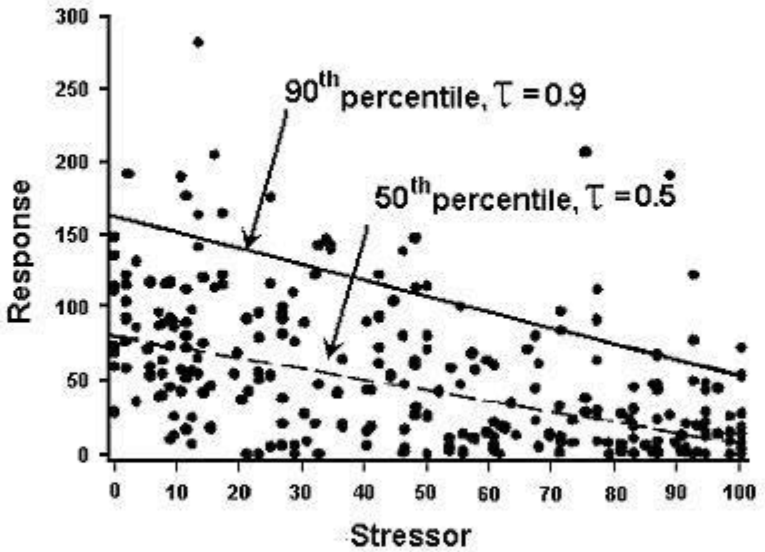


Figure S-1: Quantile regression of matched data for a stressor and a response with the 50th and 90th percentiles noted.

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). For example, modelling the 50th quantile of a response variable produces the median line under which 50% of the observed responses are located, and modelling the 90th quantile produces a line under which 90% of the observed responses are located (Figure S-1).

Like conventional linear regression, a common functional form that is assumed for a quantile regression analysis may be a linear model:

$$y_\tau = \beta_0 + \beta_1 x_1$$

where y_τ denotes the τ^{th} quantile of y, β_1 are constant coefficients, x_1 is the explanatory variable.

Quantile regressions can have more than one explanatory variable, but we limit the following discussion to the univariate case.

Quantile regression as illustrated in Figure S-1 is a semi-parametric method (i.e., combines parametric and nonparametric methods). A given quantile may be assumed related linearly to the independent (response, Y-axis) variable, but the distribution of Y at a given X is not assumed to have a normal distribution. Moreover, other specific assumptions, such as homogeneity of variance do not apply.

Quantile regression is robust to outliers in dependent (response, Y-axis) variables, but is sensitive to points sparsely distributed toward the extremes of the independent (explanatory, X-axis) variables. In cases where such leverage points are present, one may do a weighted quantile regression. The influence of outliers, censored data, data clusters, and leverage points may be evaluated by comparing plots after removing (or, in the case of leverage points, weighting) these points. Any data pruning of this nature must be transparently described. In general, the points should remain on the plot with flags indicating whether they were weighted or omitted from the model.

Is the assumed functional form appropriate?

Although non-linear quantile regression analyses are available, a simplifying assumption is that the relationships being modelled are linear with respect to the explanatory variables. In Figure S-1, the relationships between response variable and the explanatory variable is assumed to be linear. In many cases, a linear relationship is a reasonably accurate estimate of the actual relationship, if there is no reason to believe differently. Most biotic metrics are generally considered to change linearly or log-linearly in relation to stressor gradients, but ecological knowledge of the underlying processes may help one select alternate functional forms.

Are there other assumptions with quantile regression?

An assumption for using the 90*th* percentile is that the data wedge often observed in scatter plots of biological metrics is the result of other stressors co-occurring with the modelled stressor which cause additional decline in biological response over the stressor gradient.

If data from the impaired site are located far outside the upper boundary determined from regional data, it may be an indication that the comparison to the regional data is not valid. This situation can arise for a variety reasons. For example, field sampling methods applied at the impaired site may differ significantly from those applied to collect the regional data. In general, large outliers should be inspected carefully to determine whether they can be usefully compared to regional data.

How do I run a quantile regression analysis?

Unlike regular linear regression, tools for quantile regression are less readily available, although algorithms are available in specific software packages and in R (Koenker & Hallock 2001; Koenker 2013).

What do quantile regression results mean?

As with mean regression, programs generally provide estimated values for the coefficients along with their standard errors and *p*-values. A measure of the degree the model accounts for observed variability in the response relative to a constant null model that is like r^2 in mean regression may also be calculated. It is generally useful to plot the data and superimpose the fitted line (Figure S-1).

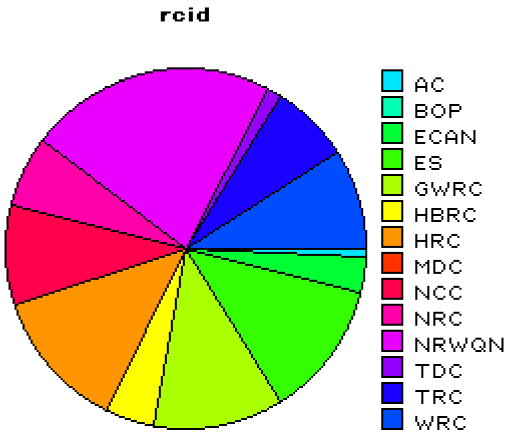
How do I use quantile regression in causal analysis?

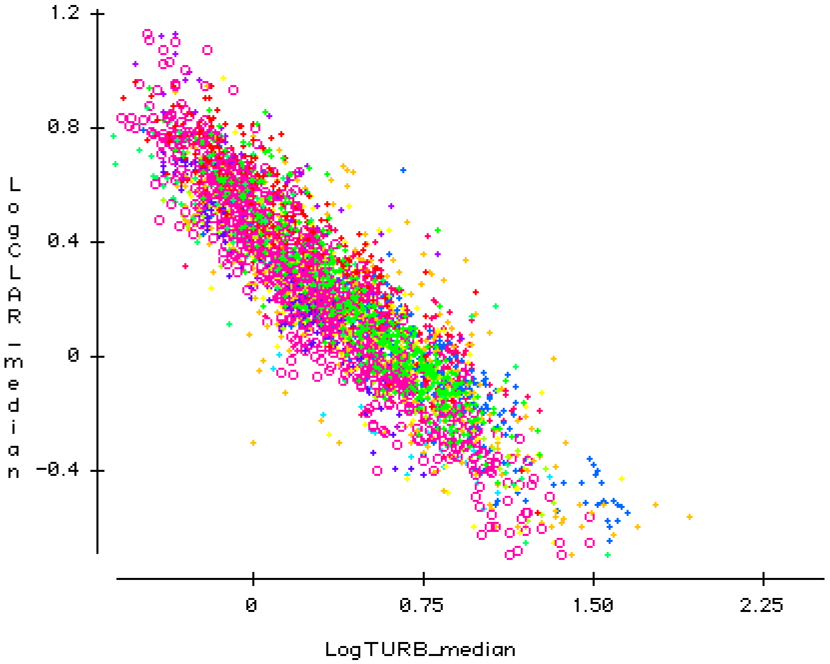
Quantile regression can be used to help describe stressor-response relationships. Quantile regression provides a means of estimating the location of the upper boundary of a scatter plot (e.g., the 90^{th} percentile line in Figure S-1). An assumption for using this upper boundary is that the wedge shape often observed in scatter plots of biological metrics results from the effects of other stressors co-occurring with the modelled stressor that cause additional negative effects on the biological response.

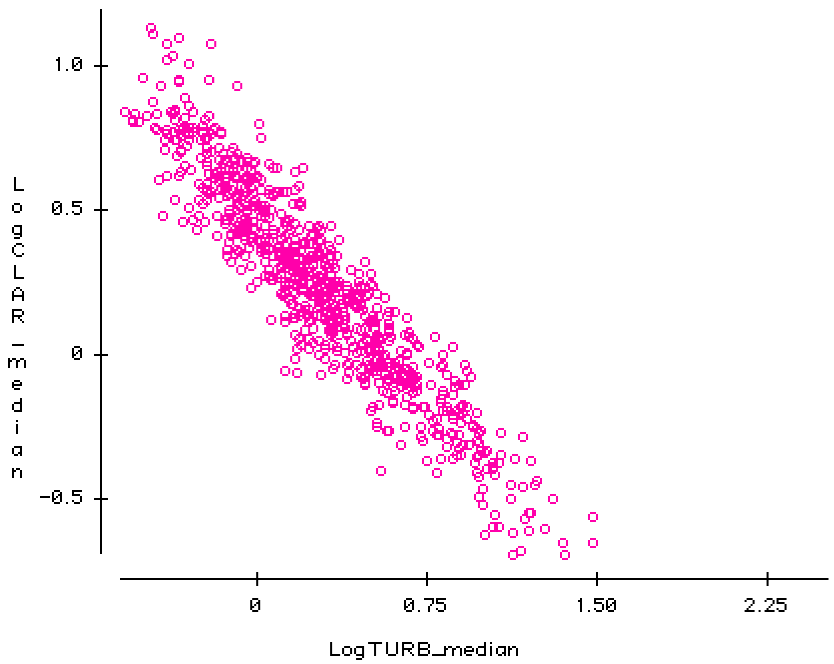
Interpretation of the results of quantile regressions in causal analysis is based on the proximity of observations from the site of the impairment to this upper boundary. These interpretations are qualitative and comparative. Evaluation of the potential contribution of other candidate stressors and a process of stressor elimination.

Appendix T Turbidity and visual clarity relationships for New Zealand river sites

Median turbidity and visual clarity relationships (Note: NRWQN sites = o; shade portion of histograms are NRWQN sites)







Dependent variable is: **LogCLAR_median**
cases selected according to **CLAR&TURB_exclWCRC**
14264 total cases of which 10621 are missing
R squared = 81.5%   R squared (adjusted) = 81.5%
s = 0.1408 with 3643 - 2 = 3641 degrees of freedom

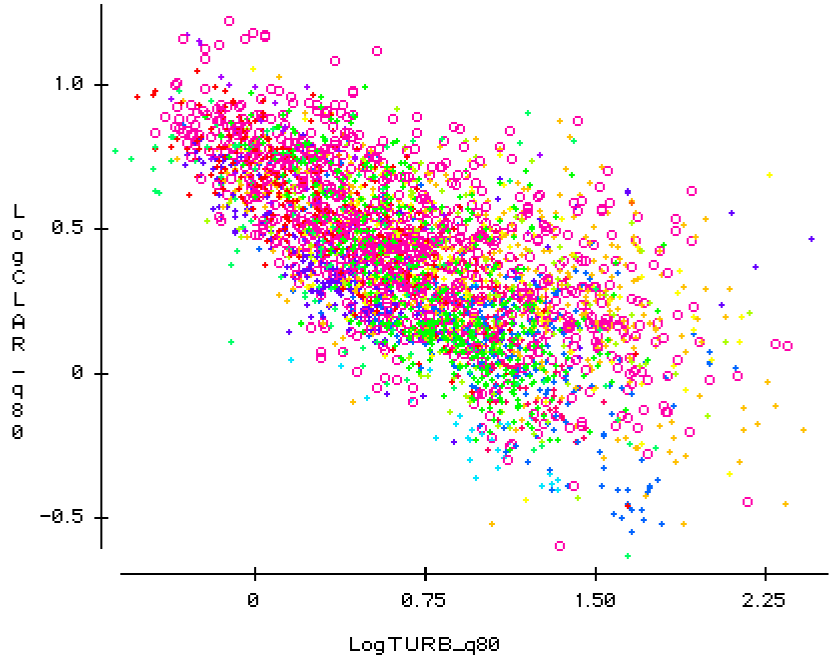
Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	317.108	1	317.108	16e3
Residual	72.1504	3641	0.0198161	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.473679	0.003166	150	≤ 0.0001
LogTURB_m...	-0.709564	0.005609	-127	≤ 0.0001

Dependent variable is: **LogCLAR_median**
cases selected according to **NRQWN_SEL**
14264 total cases of which 13465 are missing
R squared = 87.7%   R squared (adjusted) = 87.6%
s = 0.125 with 799 - 2 = 797 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	88.3892	1	88.3892	5.66e3
Residual	12.449	797	0.0156198	

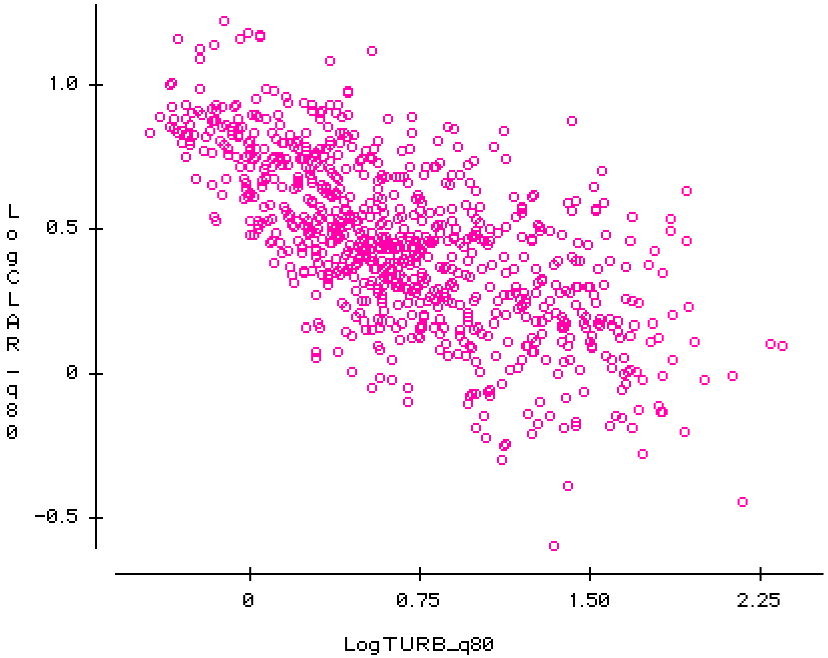
Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.464412	0.005398	86	≤ 0.0001
LogTURB_m...	-0.814271	0.01082	-75.2	≤ 0.0001



Dependent variable is: **LogCLAR_q80**
cases selected according to **CLAR&TURB_exclWCRC**
14264 total cases of which 10577 are missing
R squared = 44.9%   R squared (adjusted) = 44.9%
s = 0.2131 with 3687 - 2 = 3685 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	136.615	1	136.615	3.01e3
Residual	167.36	3685	0.0454165	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.648964	0.006056	107	≤ 0.0001
LogTURB_q...	-0.374262	0.006824	-54.8	≤ 0.0001

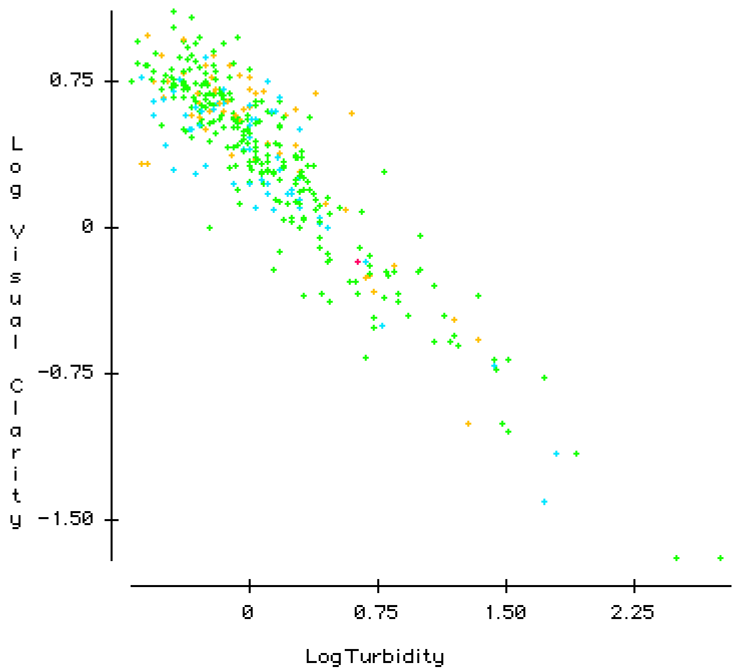


Dependent variable is: **LogCLAR_q80**
cases selected according to **NRQWN_SEL**
14264 total cases of which 13452 are missing
R squared = 44.5%   R squared (adjusted) = 44.5%
s = 0.2157 with 812 - 2 = 810 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	30.2746	1	30.2746	651
Residual	37.6875	810	0.0465278	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.699025	0.01221	57.3	≤ 0.0001
LogTURB_q...	-0.35669	0.01398	-25.5	≤ 0.0001

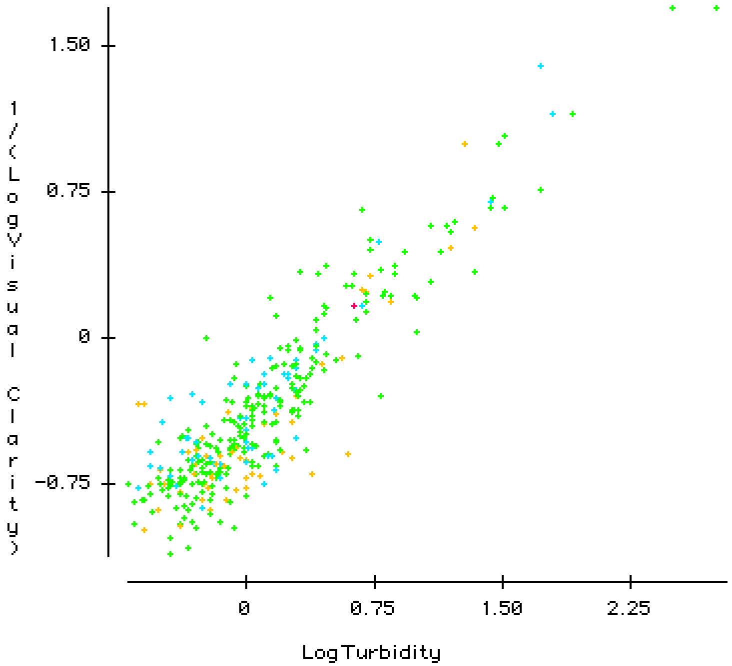
Concurrent measurement regression for turbidity and visual clarity from NRWQN



Dependent variable is: **Log Visual Clarity**
cases selected according to **SELECTOR**
1275 total cases of which 914 are missing
R squared = 82.6%   R squared (adjusted) = 82.6%
s = 0.1922 with 361 - 2 = 359 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	62.9953	1	62.9953	1.71e3
Residual	13.2614	359	0.0369399	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.43677	0.0103	42.4	≤ 0.0001
LogTurbidity	-0.802688	0.01944	-41.3	≤ 0.0001



Dependent variable is: **1/(LogVisual Clarity)**
cases selected according to **SELECTOR**
1275 total cases of which 914 are missing
R squared = 82.6%   R squared (adjusted) = 82.6%
s = 0.1922 with 361 - 2 = 359 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	62.9953	1	62.9953	1.71e3
Residual	13.2614	359	0.0369399	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	-0.43677	0.0103	-42.4	≤ 0.0001
LogTurbidity	0.802688	0.01944	41.3	≤ 0.0001

NRWQN dataset as used for quantile regression analysis:

Coloured by stream order

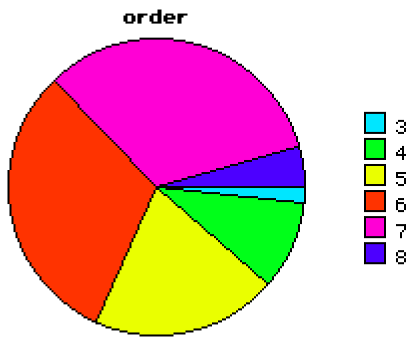
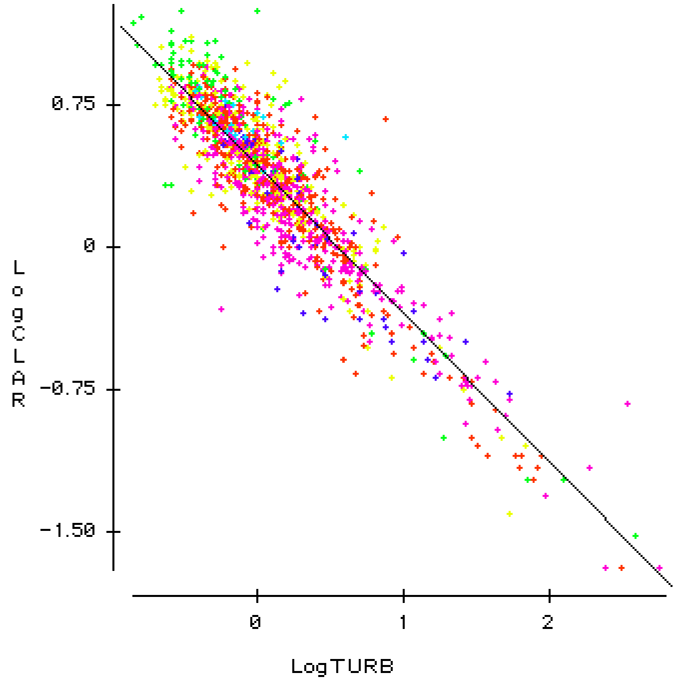
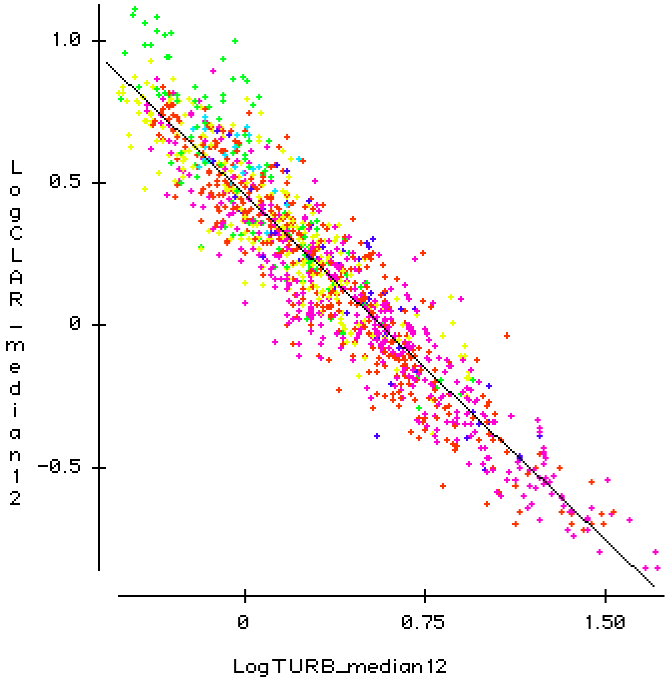


Table 0-1: Summary statistics for annual median visual clarity and turbidity for NRWQN dataset.

Variable	units	Number	Mean	Median	Minimum	Maximum	5th percentile	95th percentile
Visual clarity	m	1237	2.2	1.7	0.14	13.1	0.33	6.0
Turbidity	NTU	1263	3.7	1.9	0.30	53.0	0.48	13.6





Dependent variable is: **LogCLAR**
No Selector
1275 total cases of which 9 are missing
R squared = 81.9%   R squared (adjusted) = 81.9%
s = 0.182 with 1266 - 2 = 1264 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	189.853	1	189.853	5.73e3
Residual	41.8837	1264	0.0331358	

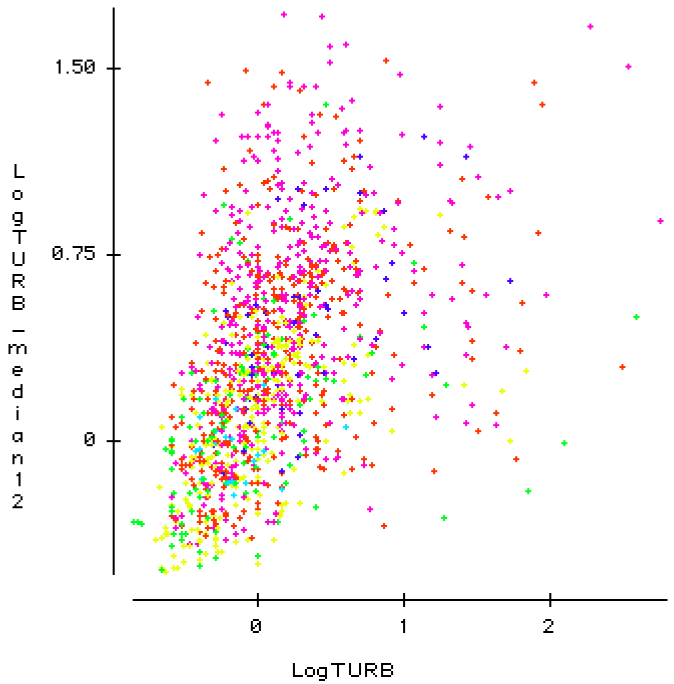
Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.438767	0.00528	83.1	≤ 0.0001
LogTURB	-0.780569	0.01031	-75.7	≤ 0.0001

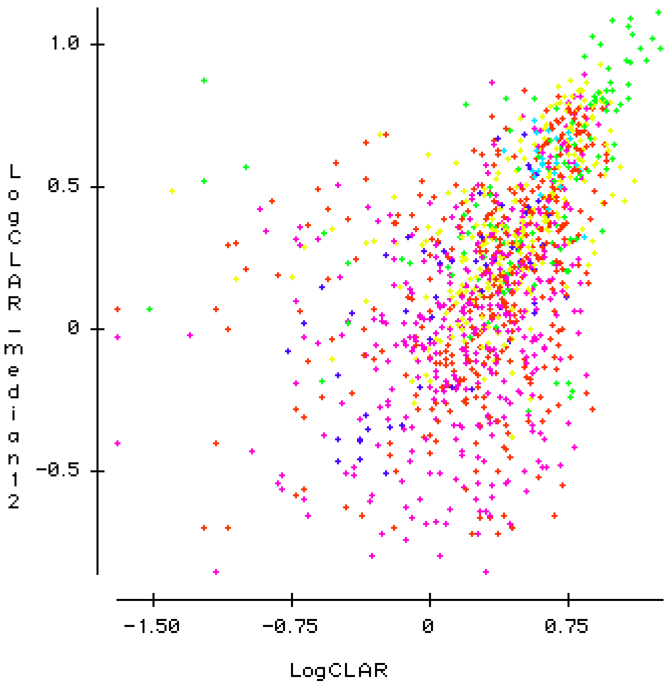
Dependent variable is: **LogCLAR_median12**
No Selector
1275 total cases of which 50 are missing
R squared = 87.6%   R squared (adjusted) = 87.6%
s = 0.1315 with 1225 - 2 = 1223 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	149.066	1	149.066	8.62e3
Residual	21.1395	1223	0.0172849	

Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.459967	0.00469	98.1	≤ 0.0001
LogTURB_m...	-0.808887	0.00871	-92.9	≤ 0.0001

Comparison of instantaneous (i.e., at time of macroinvertebrate sampling) and median clarity and turbidity data for NRWQN dataset





Dependent variable is: **LogTURB_median12**
No Selector
1275 total cases of which 13 are missing
R squared = 20.5% R squared (adjusted) = 20.5%
s = 0.3851 with 1262 - 2 = 1260 degrees of freedom

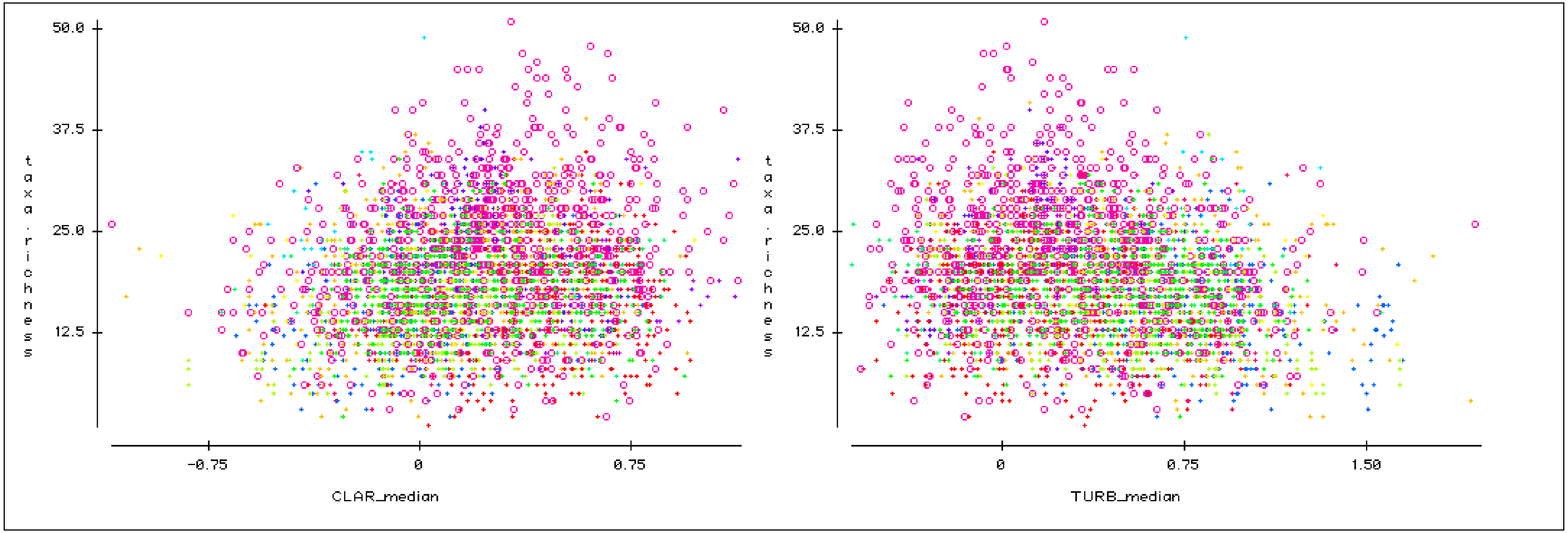
Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	48.2615	1	48.2615	326
Residual	186.818	1260	0.148268	

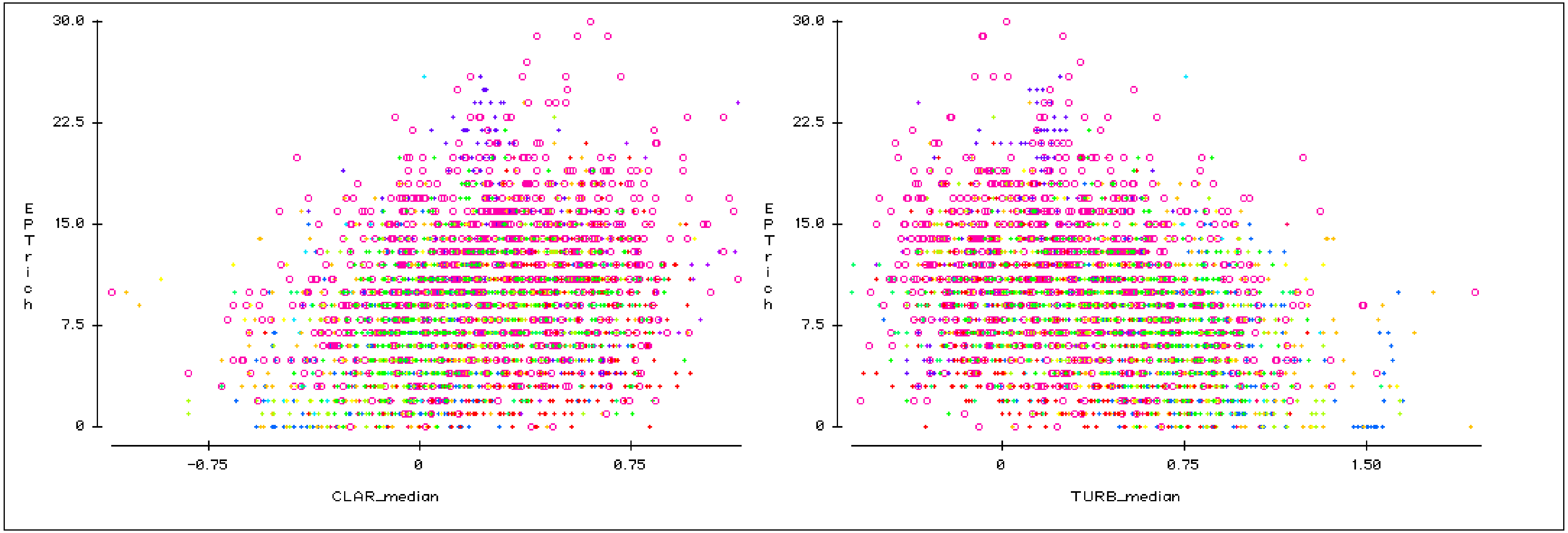
Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.272696	0.01118	24.4	≤ 0.0001
LogTURB	0.397707	0.02204	18	≤ 0.0001

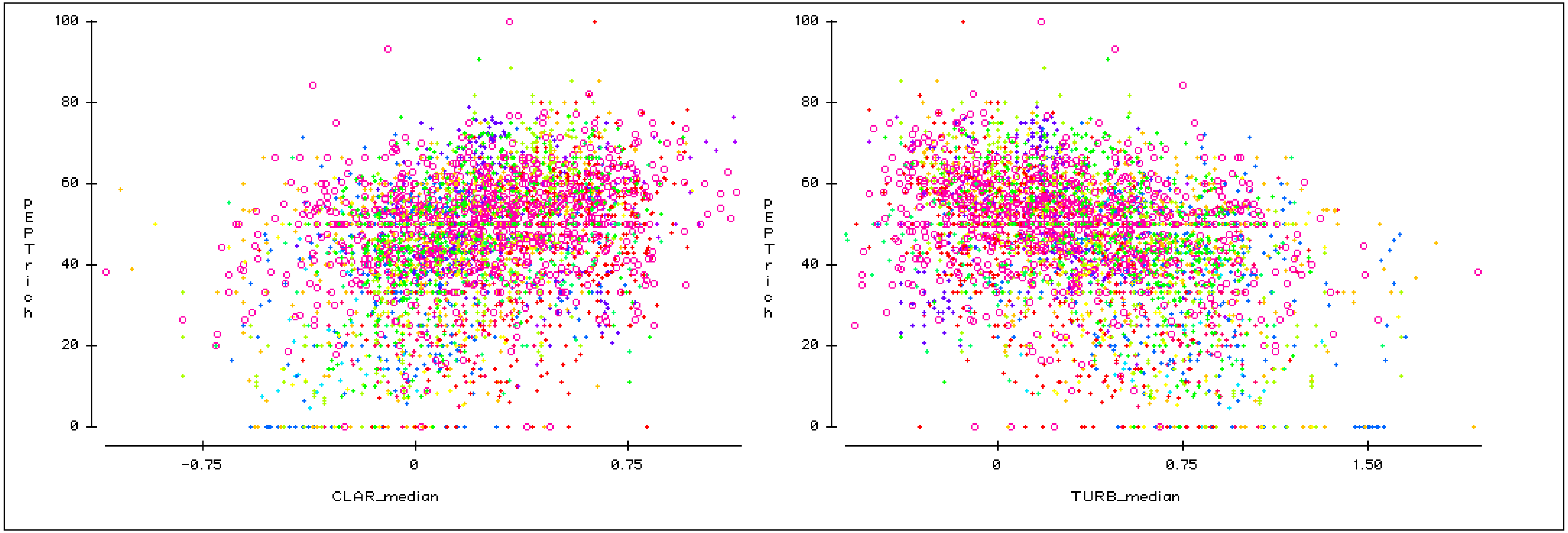
Dependent variable is: **LogCLAR_median12**
No Selector
1275 total cases of which 46 are missing
R squared = 23.7% R squared (adjusted) = 23.6%
s = 0.3261 with 1229 - 2 = 1227 degrees of freedom

Source	**Sum of Squares**	**df**	**Mean Square**	**F-ratio**
Regression	40.5444	1	40.5444	381
Residual	130.491	1227	0.10635	

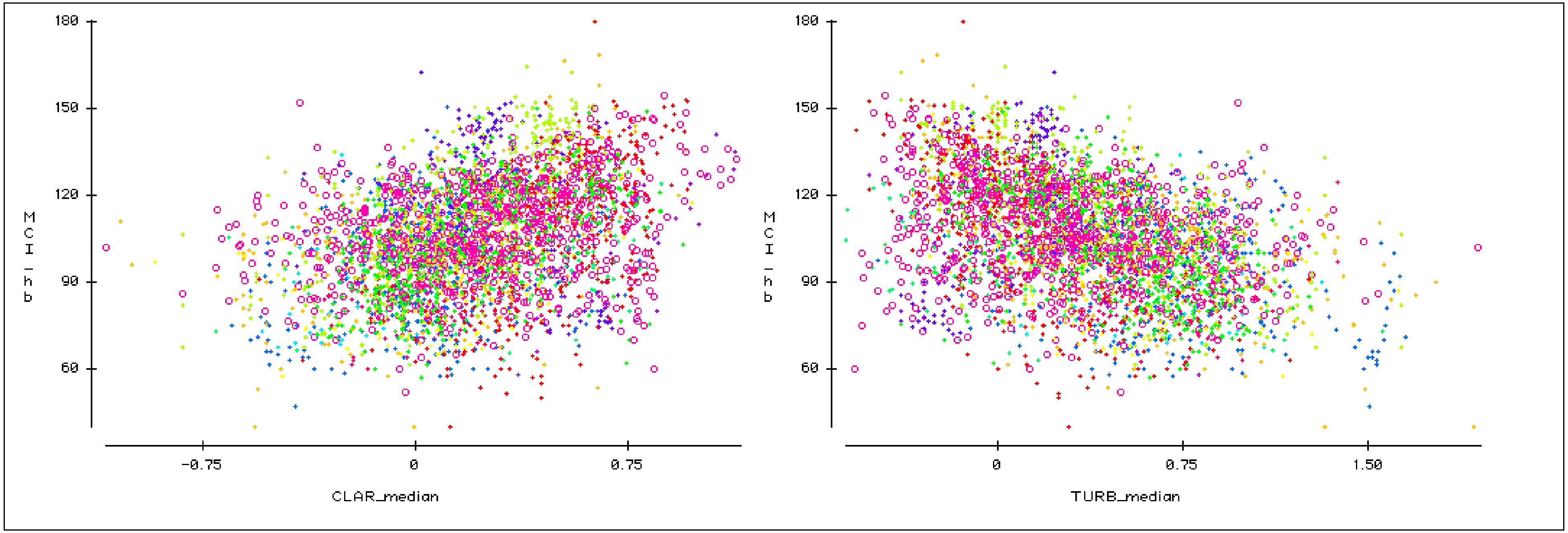
Variable	**Coefficient**	**s.e. of Coeff**	**t-ratio**	**prob**
Constant	0.0557644	0.01187	4.7	≤ 0.0001
LogCLAR	0.424428	0.02174	19.5	≤ 0.0001

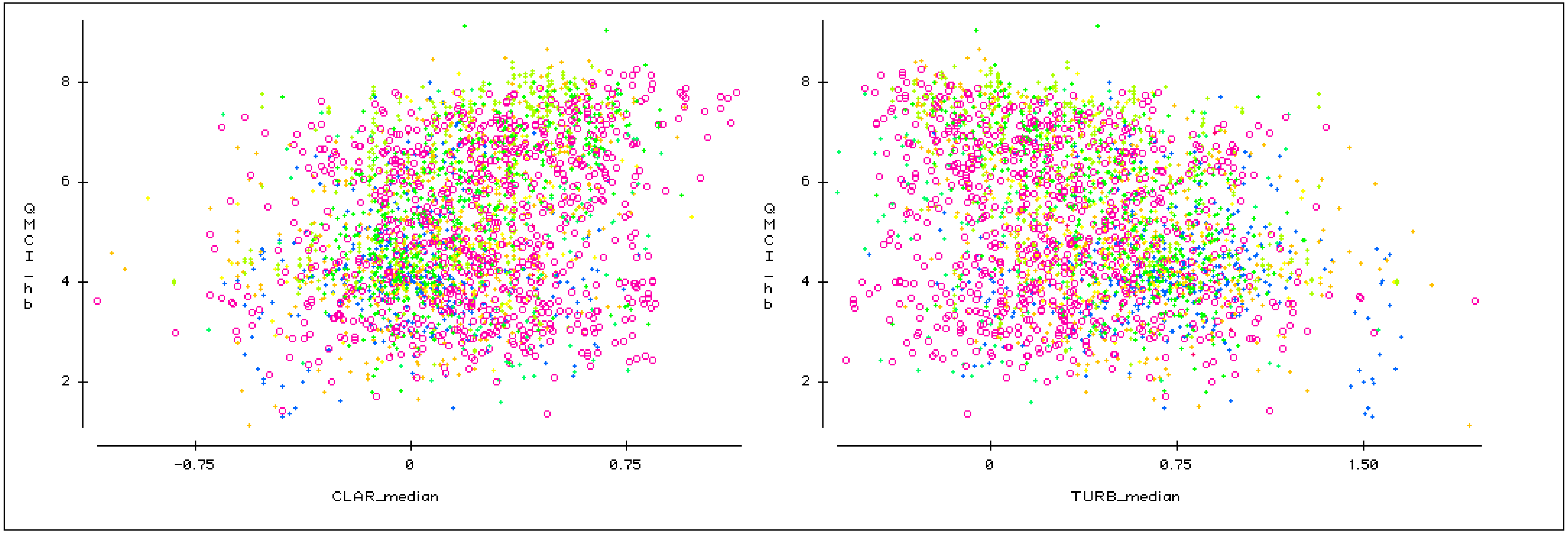

50.0
37.5
25.0
12.5
taxa.richness
-0.75
0
0.75
CLAR_median
50.0
37.5
25.0
12.5
taxa.richness
0
0.75
1.50
TURB_median


30.0
22.5
15.0
7.5
0
EPTrich
-0.75
0
0.75
CLAR_median
30.0
22.5
15.0
7.5
0
EPTrich
0
0.75
1.50
TURB_median



100
80
60
40
20
0
PEPTrich
-0.75
0
0.75
CLAR_median
100
80
60
40
20
0
PEPTrich
0
0.75
1.50
TURB_median


180
150
120
90
60
MCI_hb
-0.75
0
0.75
CLAR_median
180
150
120
90
60
MCI_hb
0
0.75
1.50
TURB_median



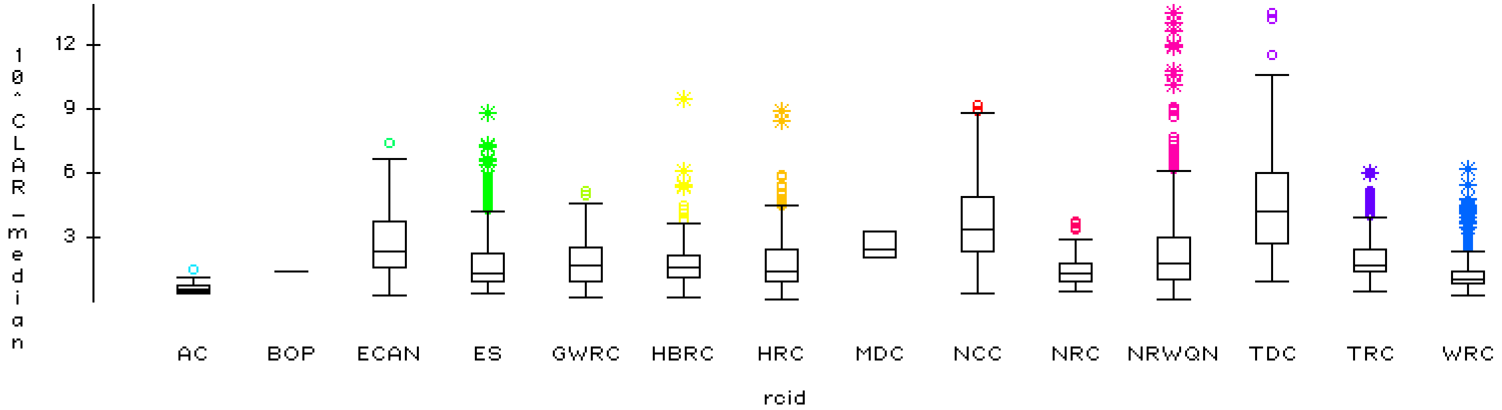
QMCI_hb
8
6
4
2
-0.75
0
0.75
CLAR_median
QMCI_hb
8
6
4
2
0
0.75
1.50
TURB_median

Appendix U Comparison of turbidity, clarity and biotic distributions in New Zealand rivers

Frequency breakdown of rcid
cases selected according to CLAR&TURB_exclWCRC
14264 total cases of which 10564 are missing

Total Cases	3700
Number of Categories	17

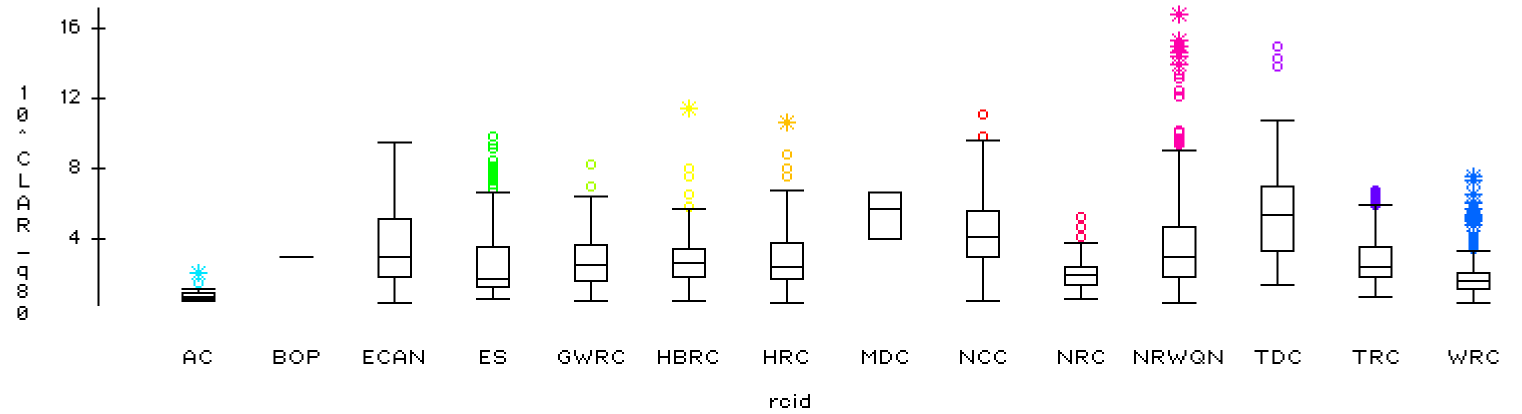
Group	Count
AC	30
BOP	1
ECAN	117
ES	446
GDC	0
GWRC	438
HBRC	162
HRC	464
MDC	3
NCC	341
NRC	235
NRWQN	817
ORC	0
TDC	52
TRC	259
WCRC	0
WRC	335



Summary of 10^CLAR_median
cases selected according to CLAR&TURB_exclWCRC
14264 total cases of which 10584 are missing

Percentile 5

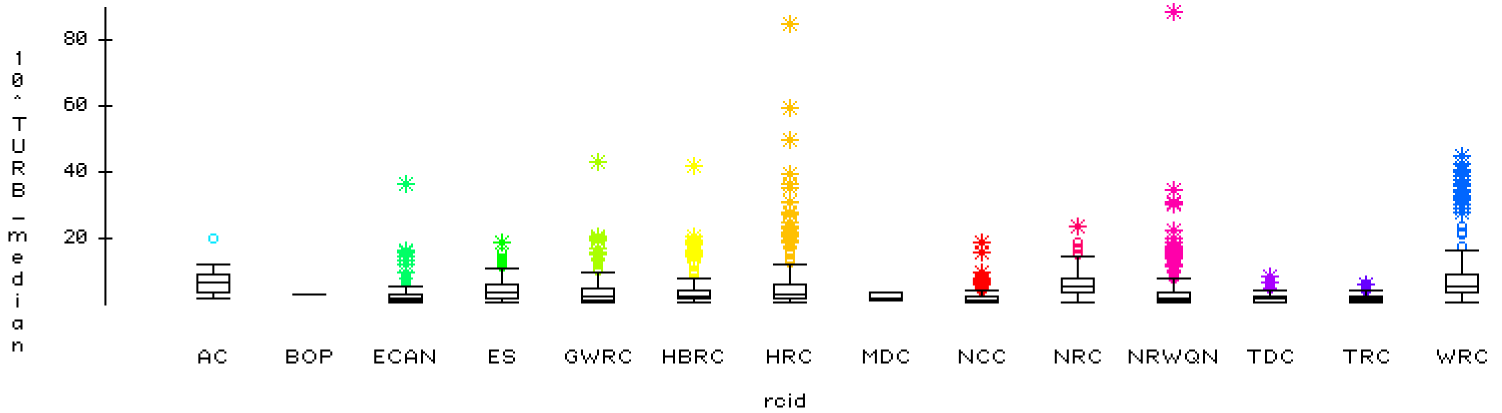
Count	3680
Mean	2.10106
Median	1.6
Min	0.08
Max	13.59
Lower ith %tile	0.452465
Upper ith %tile	5.44489



Summary of 10^CLAR_q80
cases selected according to CLAR&TURB_exclWCRC
14264 total cases of which 10569 are missing

Percentile 5

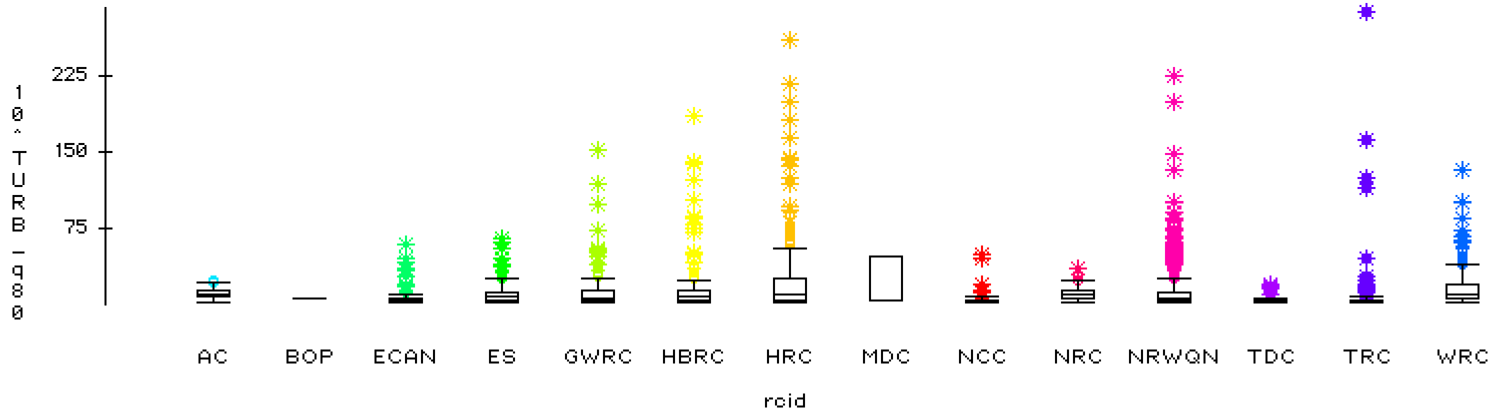
Count	3695
Mean	2.95382
Median	2.38
Min	0.23
Max	16.9063
Lower ith %tile	0.8
Upper ith %tile	6.7739



Summary of 10^TURB_median
cases selected according to CLAR&TURB_exclWCRC
14264 total cases of which 10601 are missing

Percentile 5

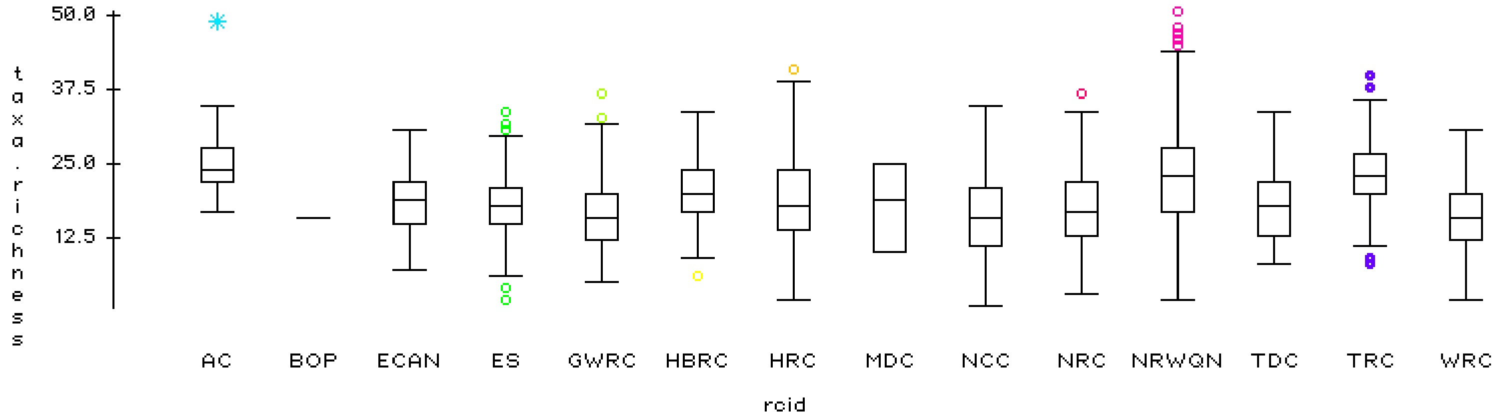
Count	3663
Mean	3.91955
Median	2.32
Min	0.24
Max	89
Lower ith %tile	0.549494
Upper ith %tile	11.4891



Summary of 10^TURB_q80
cases selected according to CLAR&TURB_exclWCRC
14264 total cases of which 10572 are missing

Percentile 5

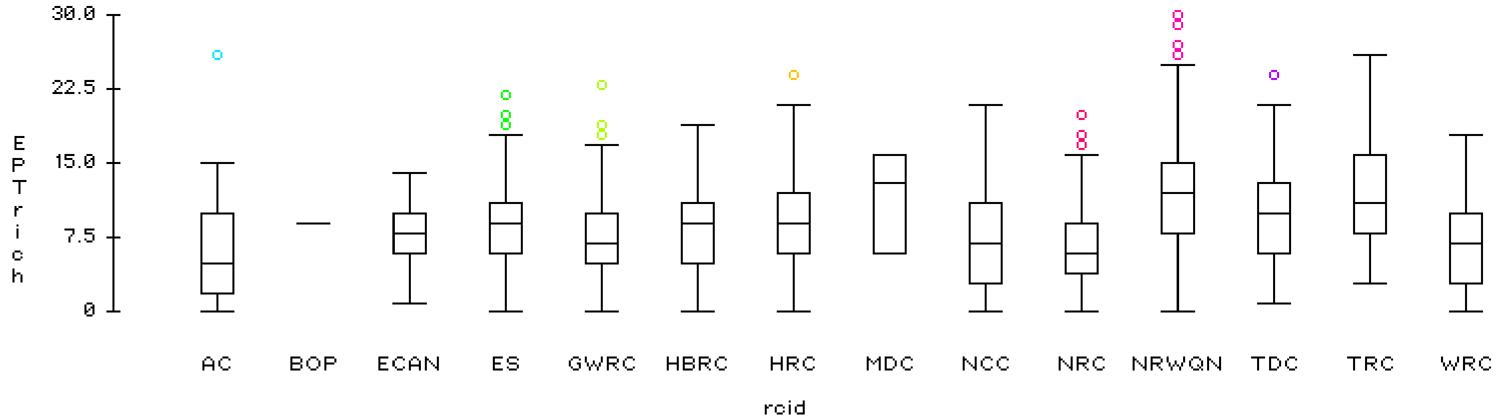
Count	3692
Mean	11.0428
Median	5.0831
Min	0.24
Max	289.002
Lower ith %tile	0.81418
Upper ith %tile	41.3015



Summary of **taxa.richness**
cases selected according to **CLAR&TURB_exclWCRC**
14264 total cases of which 10564 are missing

Percentile 5

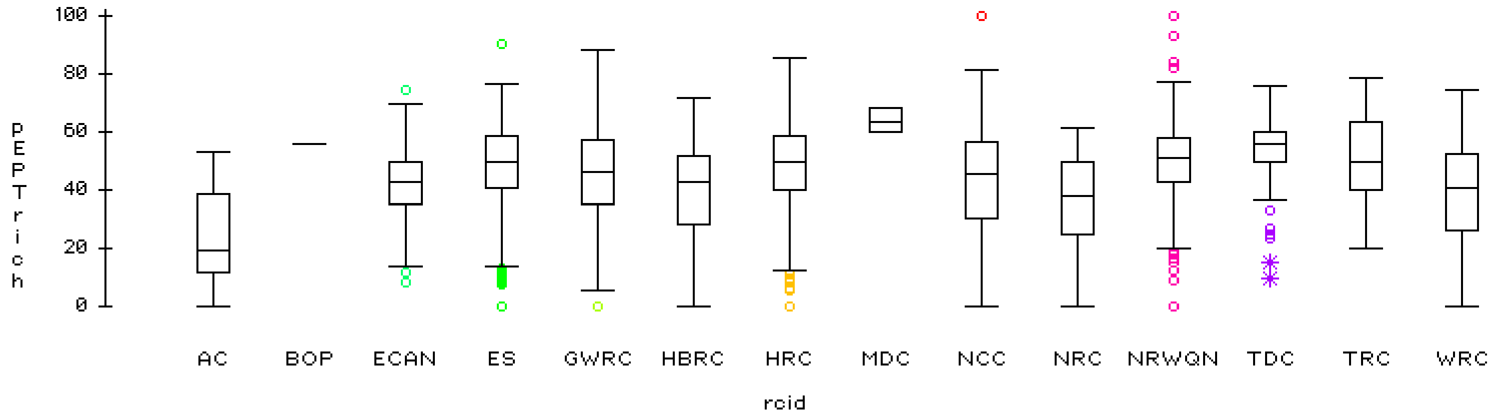
Count	3700
Mean	19.1065
Median	19
Min	1
Max	51
Lower ith %tile	8
Upper ith %tile	31



Summary of **EPTrich**
cases selected according to **CLAR&TURB_exclWCRC**
14264 total cases of which 10564 are missing

Percentile 5

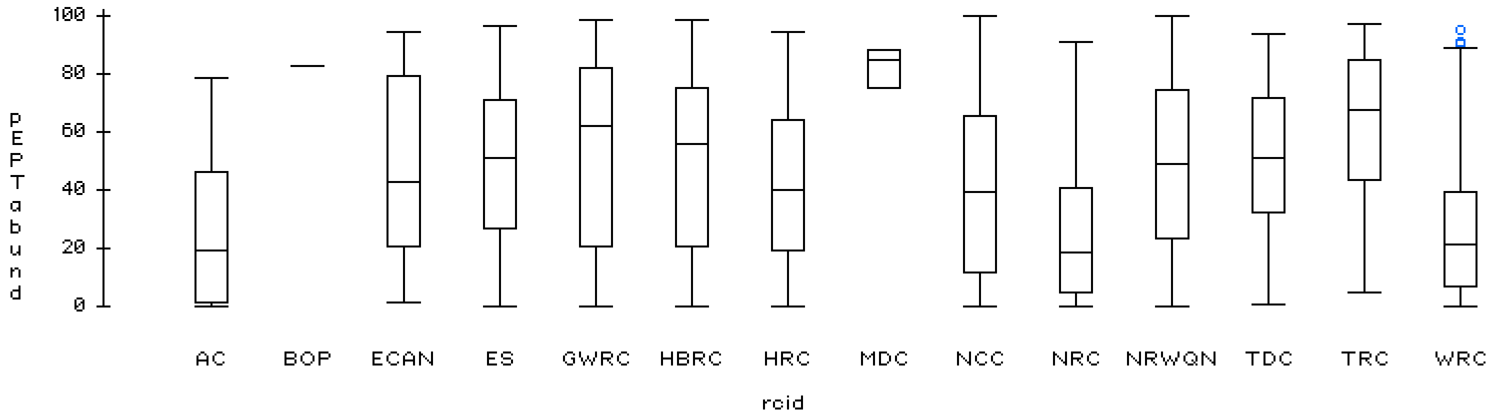
Count	3700
Mean	9.11378
Median	9
Min	0
Max	30
Lower ith %tile	1
Upper ith %tile	18



Summary of pEPTrich
cases selected according to CLAR&TURB_exclWCRC
14264 total cases of which 10564 are missing

Percentile 5

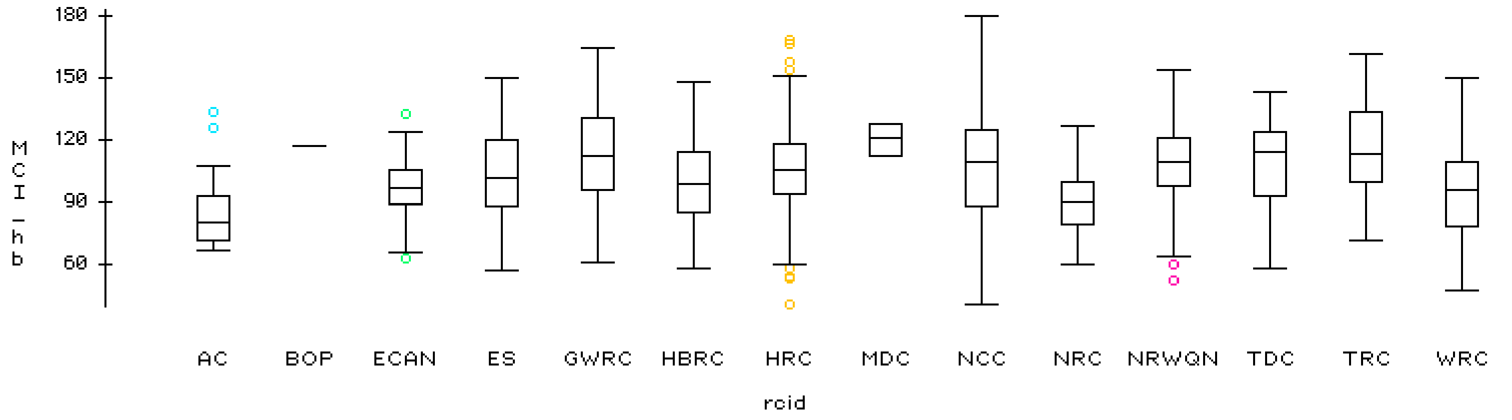
Count	3700
Mean	45.7332
Median	47.619
Min	0
Max	100
Lower ith %tile	13.961
Upper ith %tile	69.5048



Summary of **pEPTabund**
cases selected according to **CLAR&TURB_exclWCRC**
14264 total cases of which 10564 are missing

Percentile 5

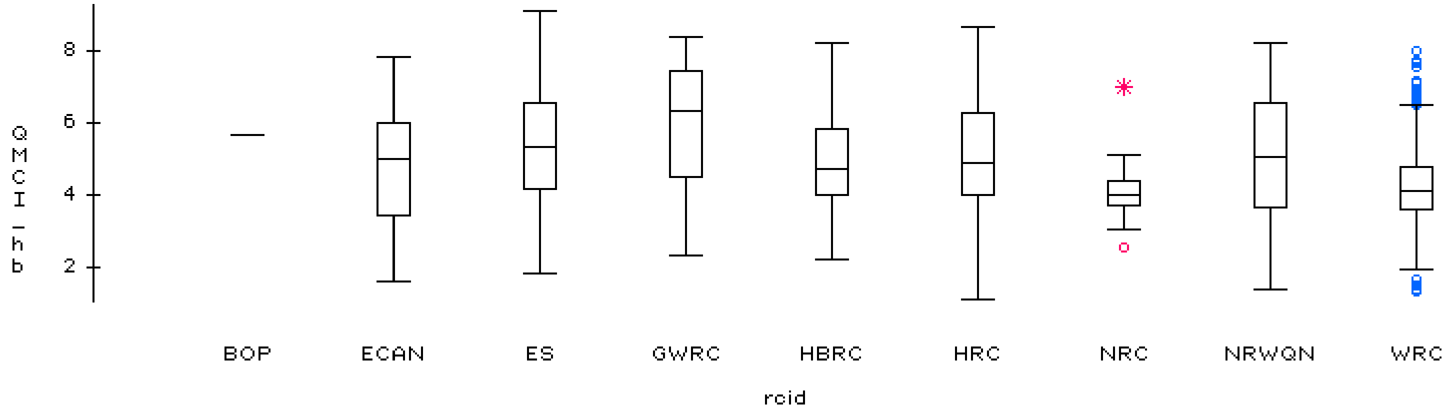
Count	3700
Mean	45.2047
Median	45.156
Min	0
Max	100
Lower ith %tile	1.16963
Upper ith %tile	90.7325



Summary of **MCI_hb**
cases selected according to **CLAR&TURB_exclWCRC**
14264 total cases of which 10564 are missing

Percentile 5

Count	3700
Mean	105.307
Median	105
Min	40
Max	180
Lower ith %tile	72
Upper ith %tile	140



Summary of QMCI_hb
cases selected according to CLAR&TURB_exclWCRC
14264 total cases of which 11484 are missing

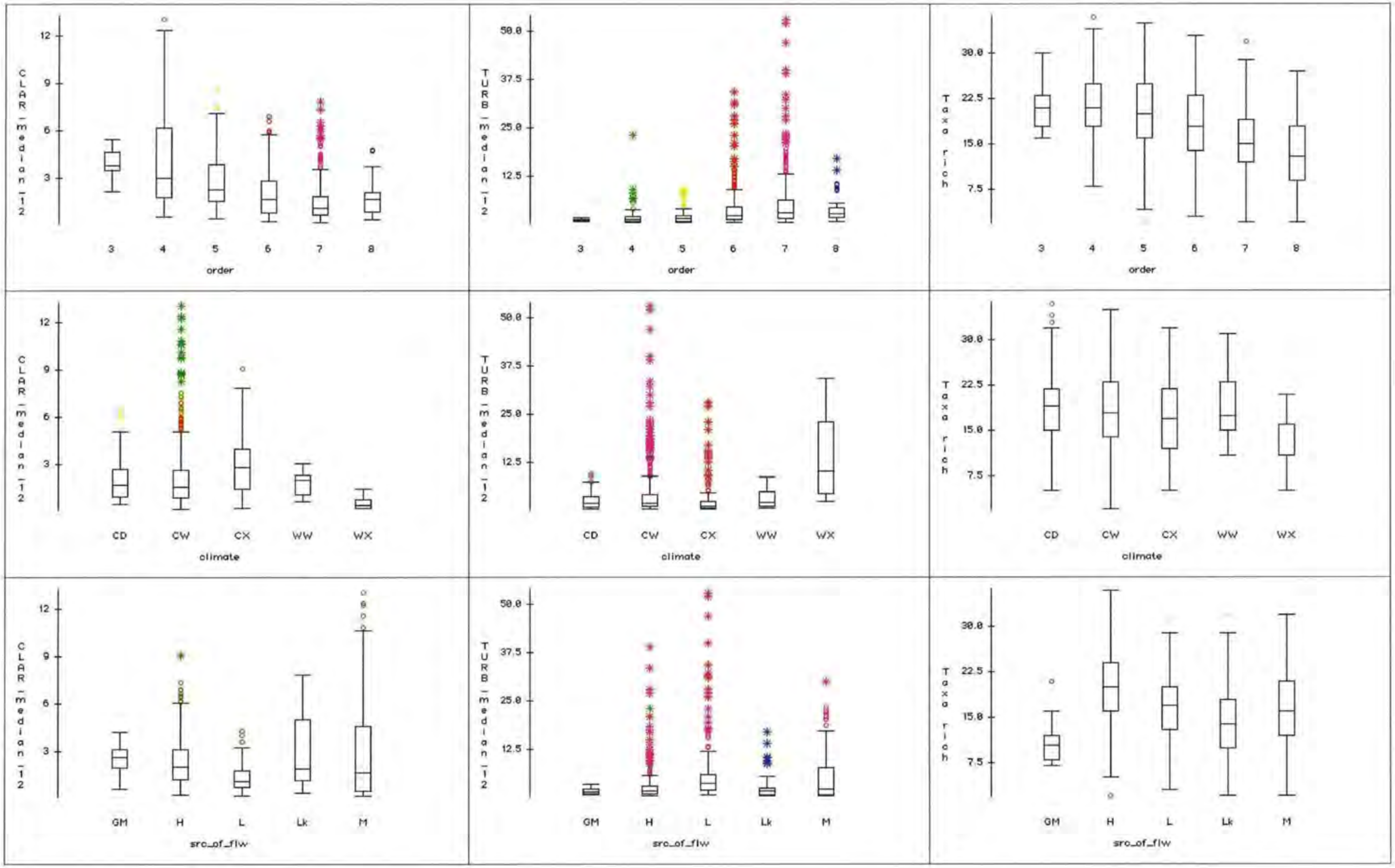
Percentile 5

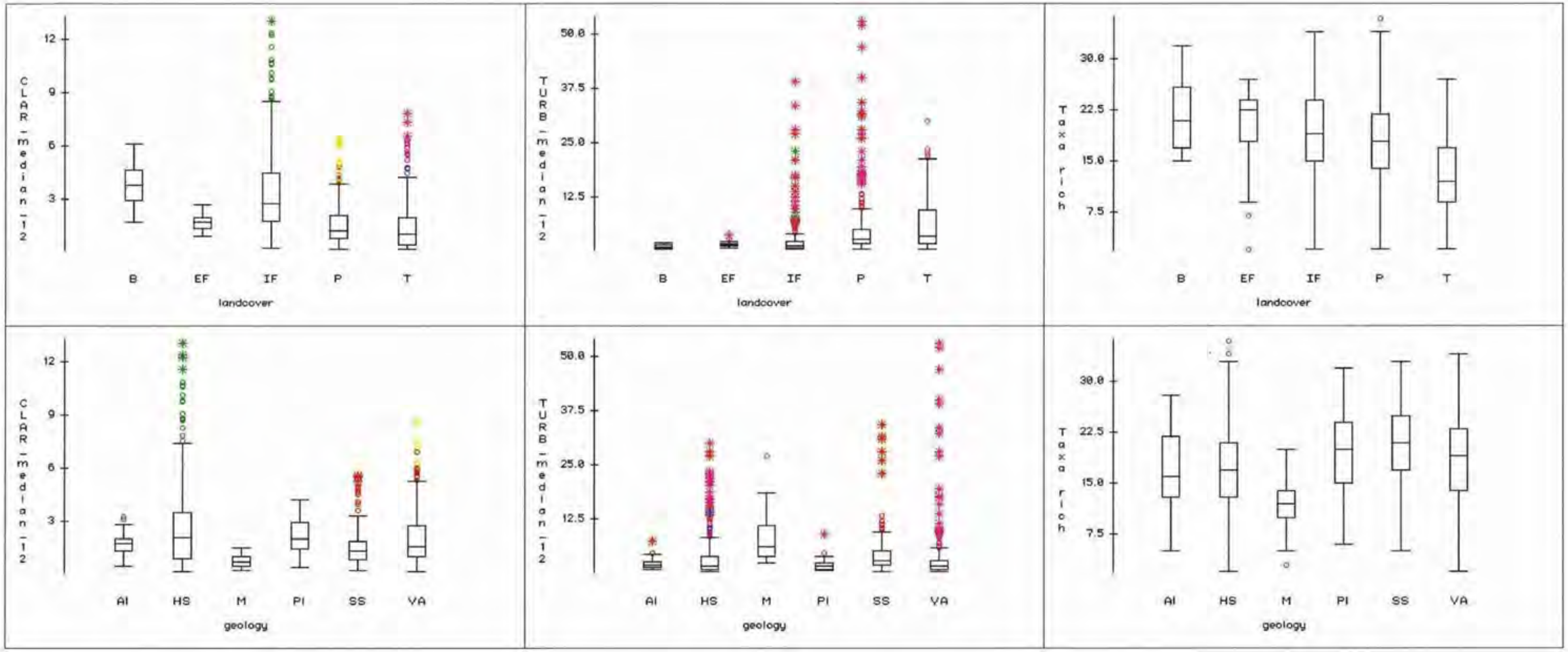
Count	2780
Mean	5.19487
Median	4.96806
Min	1.10811
Max	9.1446
Lower ith %tile	2.84505
Upper ith %tile	7.72625

Appendix V NRWQN sites in relation to REC classification classes

Abbreviations (from)(Snelder & Biggs 2002b):

Classification Level	Classes	Notation	Mapping Characteristics	Category Assignment Criteria
1. Climate	Warm Extremely Wet Warm Wet Warm Dry Cool Extremely Wet Cool Wet Cool Dry	WX WW WD CX CW CD	Mean annual precipitation, mean annual potential evapotranspiration, and mean annual temperature.	Warm: Mean annual temperature ≥ 12°C. Cool: Mean annual temperature < 12°C. Extremely Wet: Mean annual effective precipitation ≥ 1,500 mm. Wet: 500 > mean annual effective precipitation < 1,500 mm. Dry: Mean annual effective precipitation < 500 mm.
2. Source of Flow	Mountain Hill Low Elevation Lake	M H L Lk	Catchment rainfall volume in elevation categories, lake influence index	M: > 50 percent annual rainfall volume above 1,000 m ASL. H: 50 percent rainfall volume between 400 and 100 m ASL. L: 50 percent rainfall below 400 m ASL. Lk: Lake influence index > 0.033.
3. Geology	Alluvium Hard Sedimentary Soft Sedimentary Volcanic Basic Volcanic Acidic Plutonic	Al HS SS Vb Va Pl	Proportions of each geological category in section catchment	Class = The spatially dominant geology category unless combined soft sedimentary geological categories exceed 25 percent of catchment area, in which case class = SS..
4. Land Cover	Bare Indigenous Forest Pasture Tussock Scrub Exotic Forest Wetland Urban	B IF P T S EF W B	Proportions of each land cover category in section catchment	Class = The spatially dominant land cover category unless P exceeds 25 percent of catchment area, in which case class = P, or unless U exceed 15 percent of catchment area, in which case class = U.
5. Network Position	Low Order Middle Order High Order	LO MO HO	Stream order of network section	Stream order 1 and 2. Stream order 3 and 4. Stream order ≥ 5.
6. Valley Landform	High Gradient Medium Gradient Low Gradient	HG MG LG	Valley slope of section based on Euclidian length	Valley slope > 0.04. 0.02 ≥ Valley slope ≤ 0.04. Valley slope < 0.02.

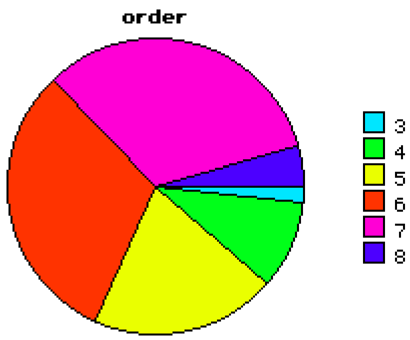

CLAR_median_12
TURB_median_12
Taxa rich
order
climate
src_of_flw
CD CW CX WW WX
GM H L Lk M

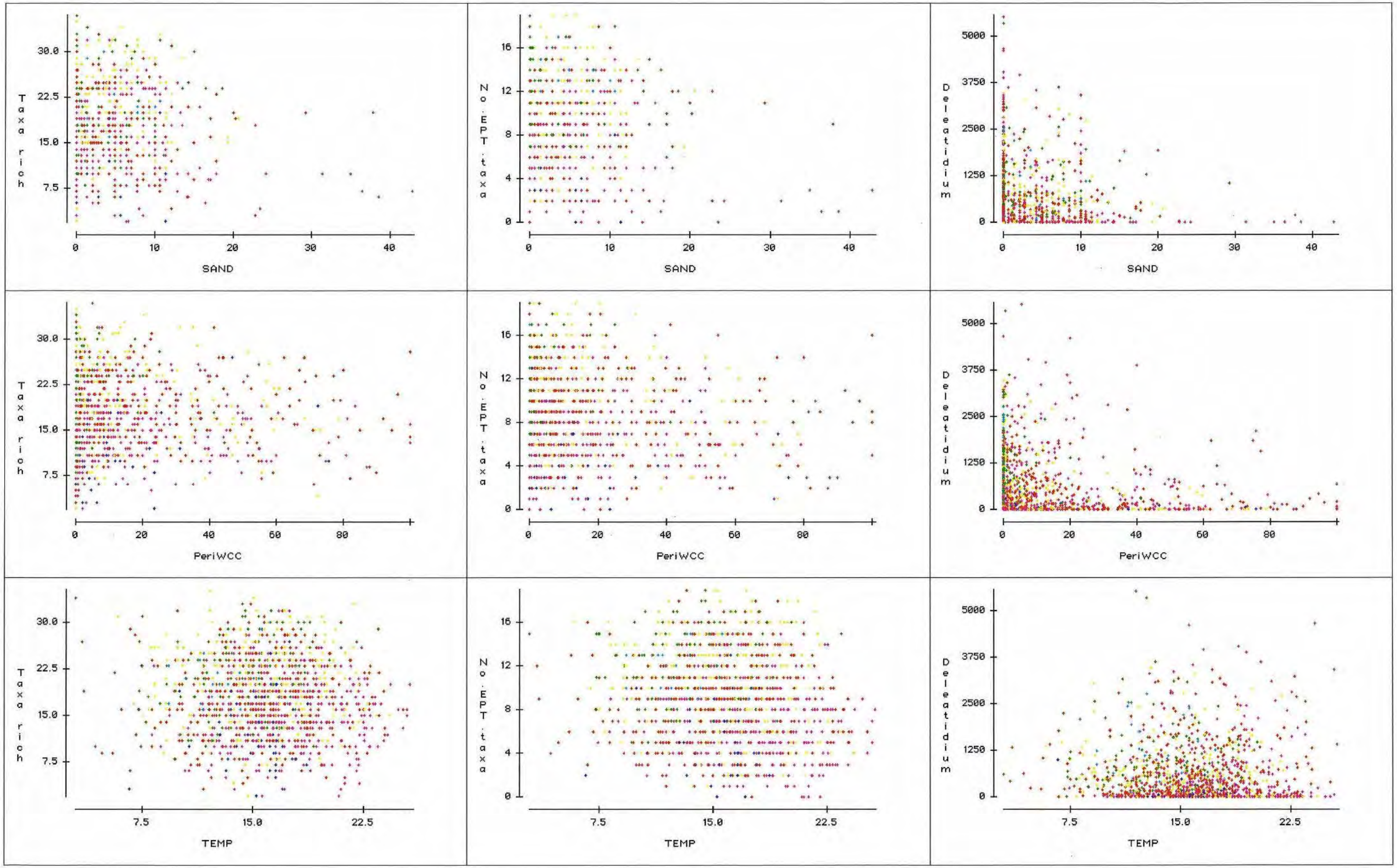

CLAR_median_12
12
9
6
3
B
EF
IF
P
T
landcover
TURB_median_12
50.0
37.5
25.0
12.5
B
EF
IF
P
T
landcover
Taxa rich
30.0
22.5
15.0
7.5
B
EF
IF
P
T
landcover
CLAR_median_12
12
9
6
3
Al
HS
M
Pl
SS
VA
geology
TURB_median_12
50.0
37.5
25.0
12.5
Al
HS
M
Pl
SS
VA
geology
Taxa rich
30.0
22.5
15.0
7.5
Al
HS
M
Pl
SS
VA
geology

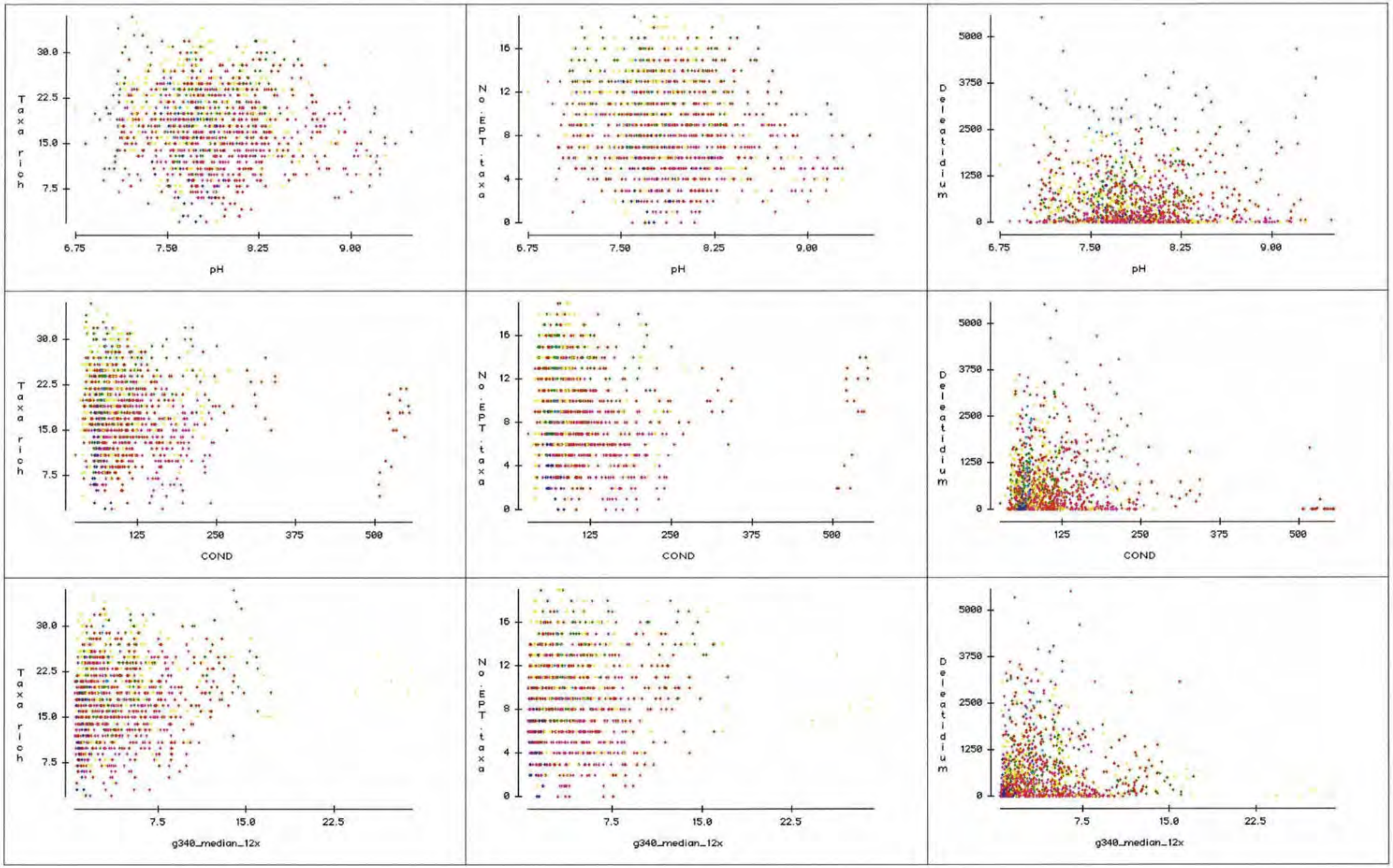
Appendix W NRWQN sites in relation to other stressors

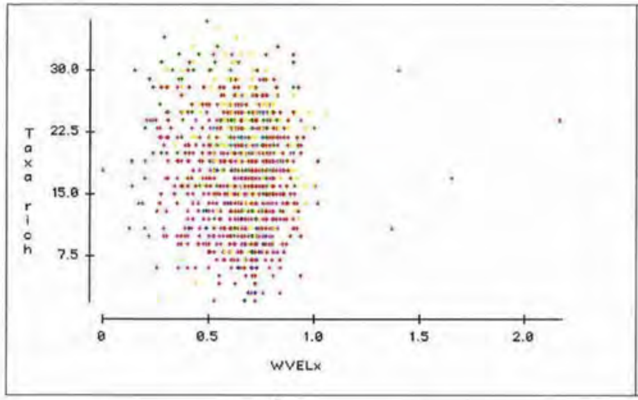
Plots shown below are for Taxa richness, number of EPT species and mayfly *Deleatidium* sp. Abundance.

Points coloured relative to stream order:

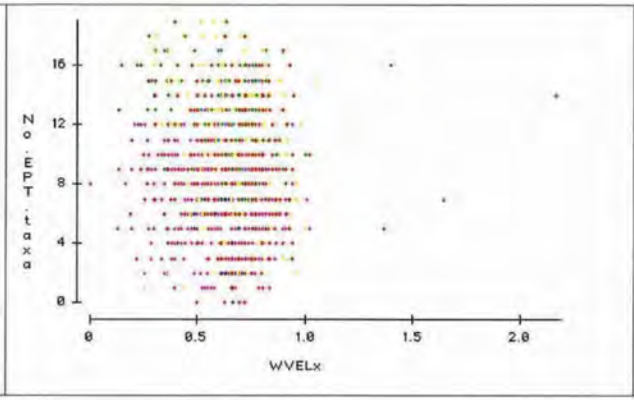



Taxa rich
No.EPT.taxa
Deleatidium
SAND
PeriWCC
TEMP

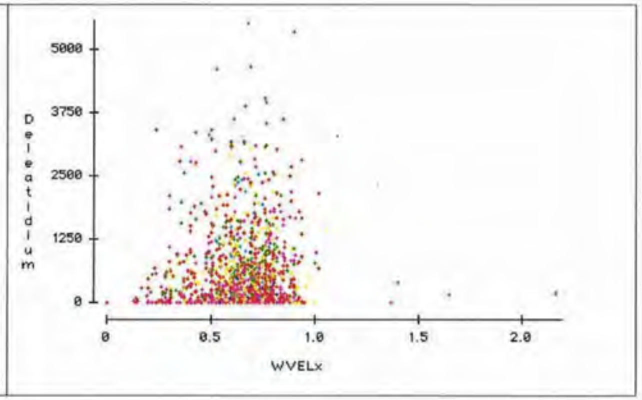

Taxa rich
No.EPT.taxa
Deleatidium
pH
COND
g340_median_12x
6.75
7.50
8.25
9.00
125
250
375
500
7.5
15.0
22.5
30.0
0
4
8
12
16
1250
2500
3750
5000



30.0
22.5
15.0
7.5
Taxa rich
0
0.5
1.0
1.5
2.0
WVELx



16
12
8
4
0
No.EPT.taxa
0
0.5
1.0
1.5
2.0
WVELx



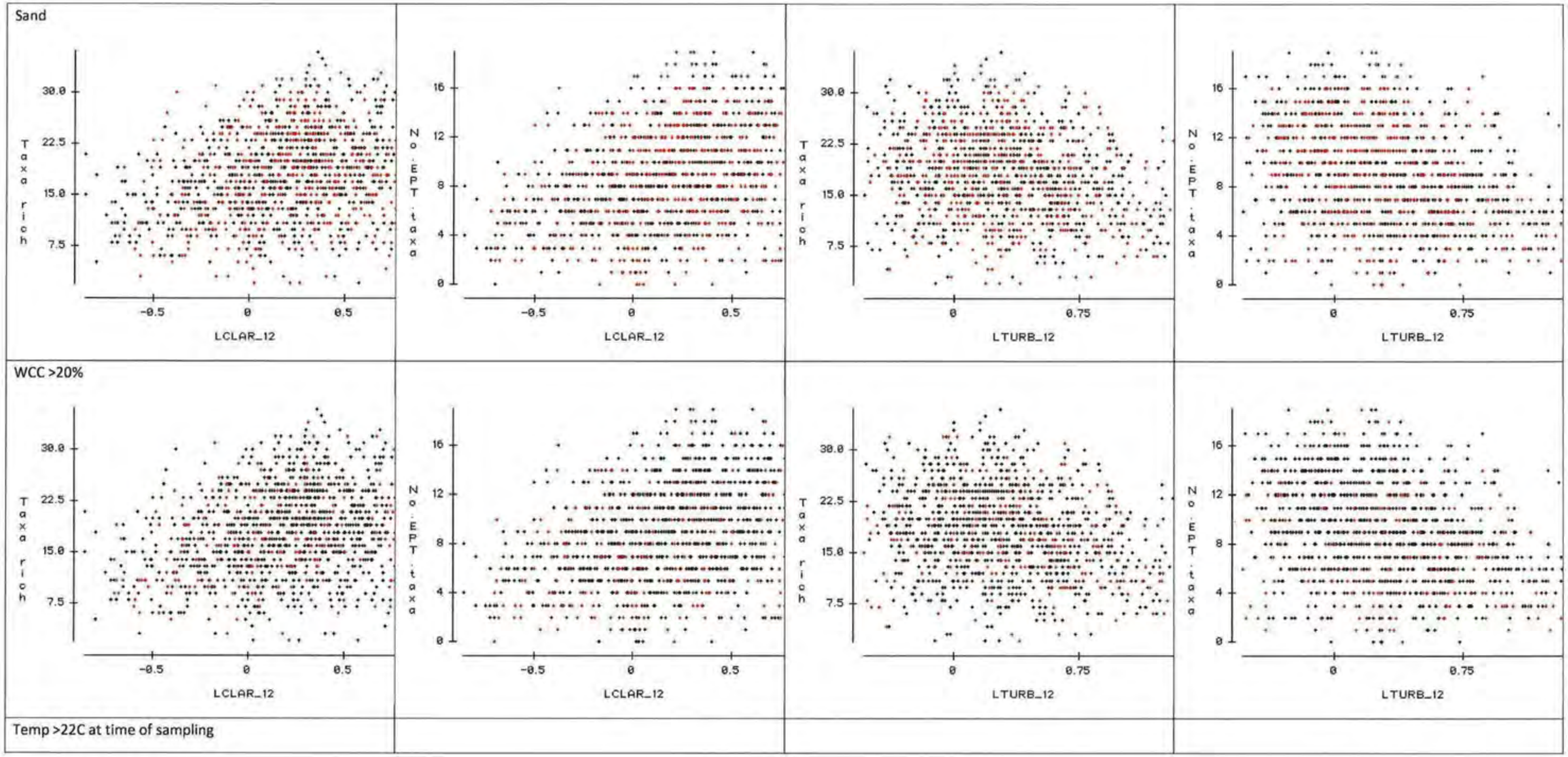
5000
3750
2500
1250
0
Deleatidium
0
0.5
1.0
1.5
2.0
WVELx

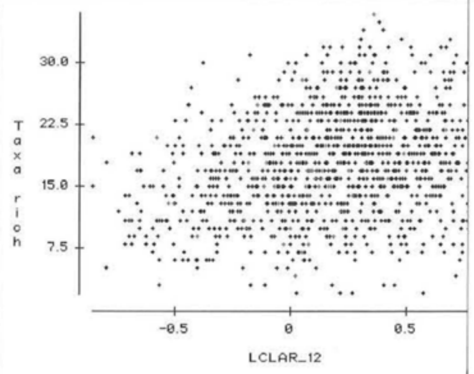
Appendix X NRWQN sites in relation to turbidity and clarity with potential stressors highlighted

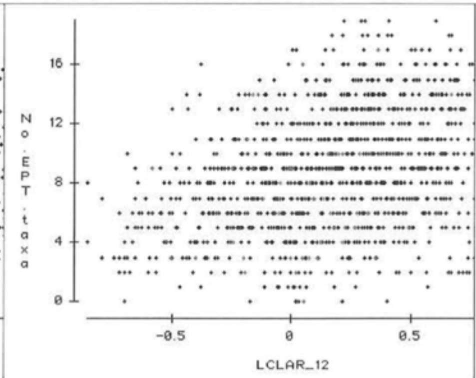
Plots shown below with red indicating data that would be excluded if selection criteria for that stressor was applied.

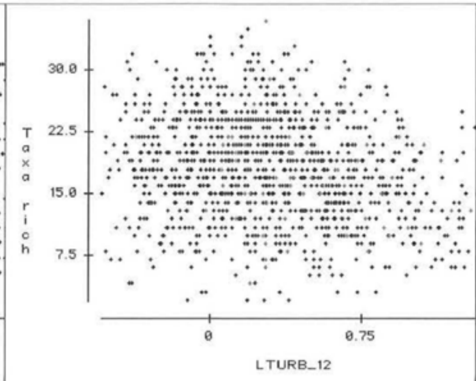
Stressor criteria as highlighted in plots:

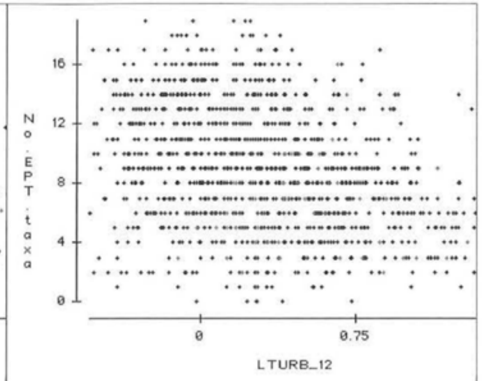
1. Sand content in bed >0%.
2. Weighted periphyton composite cover (WCC) >20%.
3. Water temperature at time of macroinvertebrate sampling >22^{o}C.
4. pH at time of macroinvertebrate sampling >8.5.
5. Absorption at 340 nm (‘g340x1000) >7.5.
6. Average river velocity <0.6 m/s or >0.8 m/s.


Sand
WCC >20%
Temp >22C at time of sampling
Taxa rich
No. EPT taxa
LCLAR_12
LTURB_12
30.0
22.5
15.0
7.5
16
12
8
4
0
-0.5
0
0.5
0.75

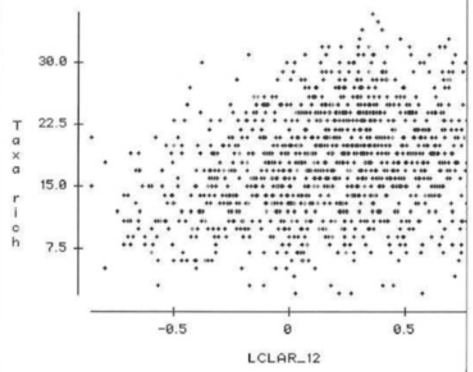


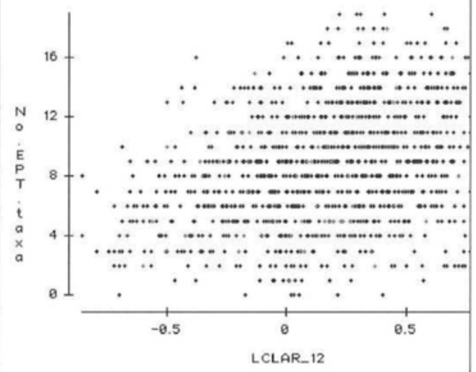


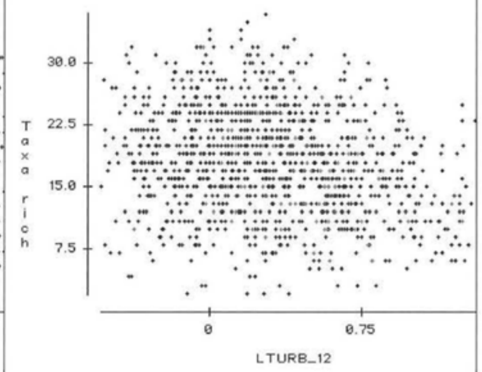


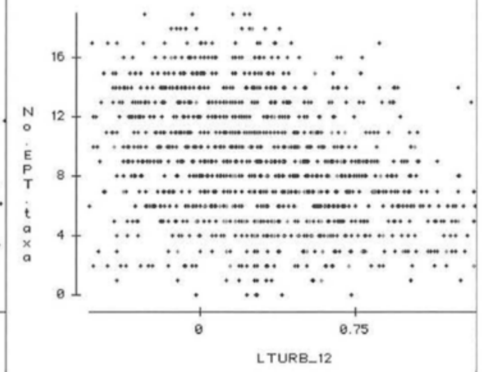


pH >8.5 at time of sampling









Conductivity - no Tarawera River

Appendix Y Summary of quantile regression results using the ‘rq’ function in R (Koenker 2013)

Ricker equation summary statistics for 95th percentile distributions.

Linear equation form:
#exponentiate to recover non-linear form of the eqn “y=beta0 * x^beta1 * e^(beta2*x)”
Beta0 = exp(INT)
Beta1 = B1
Beta2 = B2

Visual clarity quantile regressions and based on inverse visual clarity data.

Taxa richness

- Visual clarity
- tau: [1] 0.95
-

```
Coefficients:
                 Value    Std. Error t value  Pr(>|t|)
(Intercept)       3.49840  0.05234   66.83353  0.00000
log(aa[, 2] + 1) -0.04657  0.03362   -1.38527  0.16622
aa[, 2]          -0.00577  0.00227   -2.54468  0.01106
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                 Value     Std. Error t value   Pr(>|t|)
(Intercept)        3.48612   0.02760  126.30147   0.00000
log(aa[, 2] + 1)  -0.10194   0.03236   -3.15002   0.00167
aa[, 2]            0.00022   0.00365    0.05977   0.95235
```

Total Invert Density

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                 Value    Std. Error t value  Pr(>|t|)
(Intercept)       8.55303  0.23130   36.97788  0.00000
log(aa[, 2] + 1)  0.65038  0.14427    4.50808  0.00001
aa[, 2]          -0.05243  0.00786   -6.66630  0.00000
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                 Value    Std. Error t value  Pr(>|t|)
(Intercept)       9.23081  0.23239   39.72175  0.00000
log(aa[, 2] + 1)  0.33048  0.22160    1.49131  0.13613
aa[, 2]          -0.05762  0.02171   -2.65419  0.00805
```

EPT Taxa

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        2.94432  0.05144   57.23446  0.00000
log(aa[, 2] + 1)  -0.04722  0.03715   -1.27081  0.20404
aa[, 2]           -0.00763  0.00305   -2.50488  0.01238
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value     Std. Error t value   Pr(>|t|)
(Intercept)         2.93220   0.02469  118.77967   0.00000
log(aa[, 2] + 1)   -0.12190   0.03175   -3.83952   0.00013
aa[, 2]            -0.00203   0.00370   -0.54745   0.58416
```

EPT indiv

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        8.41927  0.20549   40.97153  0.00000
log(aa[, 2] + 1)   0.30889  0.13441    2.29806  0.02173
aa[, 2]           -0.05456  0.00920   -5.93338  0.00000
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        8.61668  0.15698   54.89020  0.00000
log(aa[, 2] + 1)   0.14868  0.15008    0.99065  0.32205
aa[, 2]           -0.06090  0.01605   -3.79531  0.00015
```

Deleatidium

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        7.63741  0.21985   34.73875  0.00000
log(aa[, 2] + 1)   0.17863  0.17368    1.02851  0.30391
aa[, 2]           -0.03021  0.01700   -1.77723  0.07578
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        7.64132  0.12982   58.86241  0.00000
log(aa[, 2] + 1)   0.31633  0.14560    2.17259  0.03000
aa[, 2]           -0.08794  0.01859   -4.72926  0.00000
```

Aoteapsyche

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        6.51694  0.41763   15.60438  0.00000
log(aa[, 2] + 1)   1.00954  0.25708    3.92699  0.00009
aa[, 2]           -0.08310  0.01541   -5.39295  0.00000
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        7.79212  0.42489   18.33926  0.00000
log(aa[, 2] + 1)   0.26849  0.39018    0.68811  0.49151
aa[, 2]           -0.06431  0.03730   -1.72434  0.08489
```

Potamopyrgus

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        5.50649  0.84614    6.50775  0.00000
log(aa[, 2] + 1)   0.91310  0.53355    1.71135  0.08727
aa[, 2]           -0.05599  0.04085   -1.37081  0.17069
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        6.58295  0.41915   15.70545  0.00000
log(aa[, 2] + 1)   0.47256  0.37887    1.24730  0.21252
aa[, 2]           -0.06372  0.04082   -1.56097  0.11878
```

Pycnocentria

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
```

```
(Intercept)        5.28743  1.01318    5.21868  0.00000
log(aa[, 2] + 1)  0.40093  0.59469    0.67418  0.50033
aa[, 2]           -0.09806  0.02474   -3.96301  0.00008
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        6.34374  0.49833   12.73010  0.00000
log(aa[, 2] + 1) -0.77752  0.47954   -1.62138  0.10519
aa[, 2]           -0.06117  0.04806   -1.27298  0.20326
```

Zelandobius

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        1.88657  0.63976    2.94886  0.00325
log(aa[, 2] + 1)  0.87752  0.43634    2.01108  0.04454
aa[, 2]           -0.07426  0.01944   -3.81989  0.00014
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        2.75740  0.44838    6.14964  0.00000
log(aa[, 2] + 1)  0.44350  0.55898    0.79341  0.42769
aa[, 2]           -0.06981  0.07049   -0.99035  0.32219
```

Oxyethira

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        3.02651  0.68588    4.41257  0.00001
log(aa[, 2] + 1)  0.75324  0.50868    1.48077  0.13893
aa[, 2]           -0.03963  0.03610   -1.09768  0.27256
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        3.68123  0.37238    9.88561  0.00000
log(aa[, 2] + 1)  0.41954  0.43215    0.97082  0.33182
aa[, 2]           -0.00716  0.05598   -0.12790  0.89825
```

Zelandoperla

- Visual clarity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        8.16143  0.45053   18.11534  0.00000
log(aa[, 2] + 1) -2.69143  0.34117   -7.88872  0.00000
aa[, 2]            0.08122  0.02745    2.95897  0.00315
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                  Value    Std. Error t value  Pr(>|t|)
(Intercept)        6.65022  0.50104   13.27275  0.00000
log(aa[, 2] + 1) -2.95243  0.66277   -4.45469  0.00001
aa[, 2]            0.15782  0.09561    1.65070  0.09905
```

Quantile log-transformed linear correlations

MCI

- Visual clarity:

```
tau: [1] 0.95

Coefficients:
                           Value     Std. Error t value   Pr(>|t|)
(Intercept)                119.77656   1.42896   83.82074   0.00000
CLAR_median_12               4.02930   0.51711    7.79195   0.00000
```

- Turbidity :

```
tau: [1] 0.95

Coefficients:
                           Value     Std. Error t value   Pr(>|t|)
(Intercept)                133.54472   0.94334  141.56567   0.00000
TURB_median_12              -0.68952   0.11631   -5.92819   0.00000
```

QMCI

- Visual clarity:

```
tau: [1] 0.95

Coefficients:
                           Value    Std. Error t value  Pr(>|t|)
(Intercept)                 7.10567  0.07916   89.76022  0.00000
CLAR_median_12              0.14314  0.02051    6.97801  0.00000
```

- Turbidity

```
tau: [1] 0.95

Coefficients:
                          Value     Std. Error t value   Pr(>|t|)
(Intercept)                 7.66216   0.05690  134.66634   0.00000
TURB_median_12             -0.03739   0.01397   -2.67735   0.00752
```

% EPT

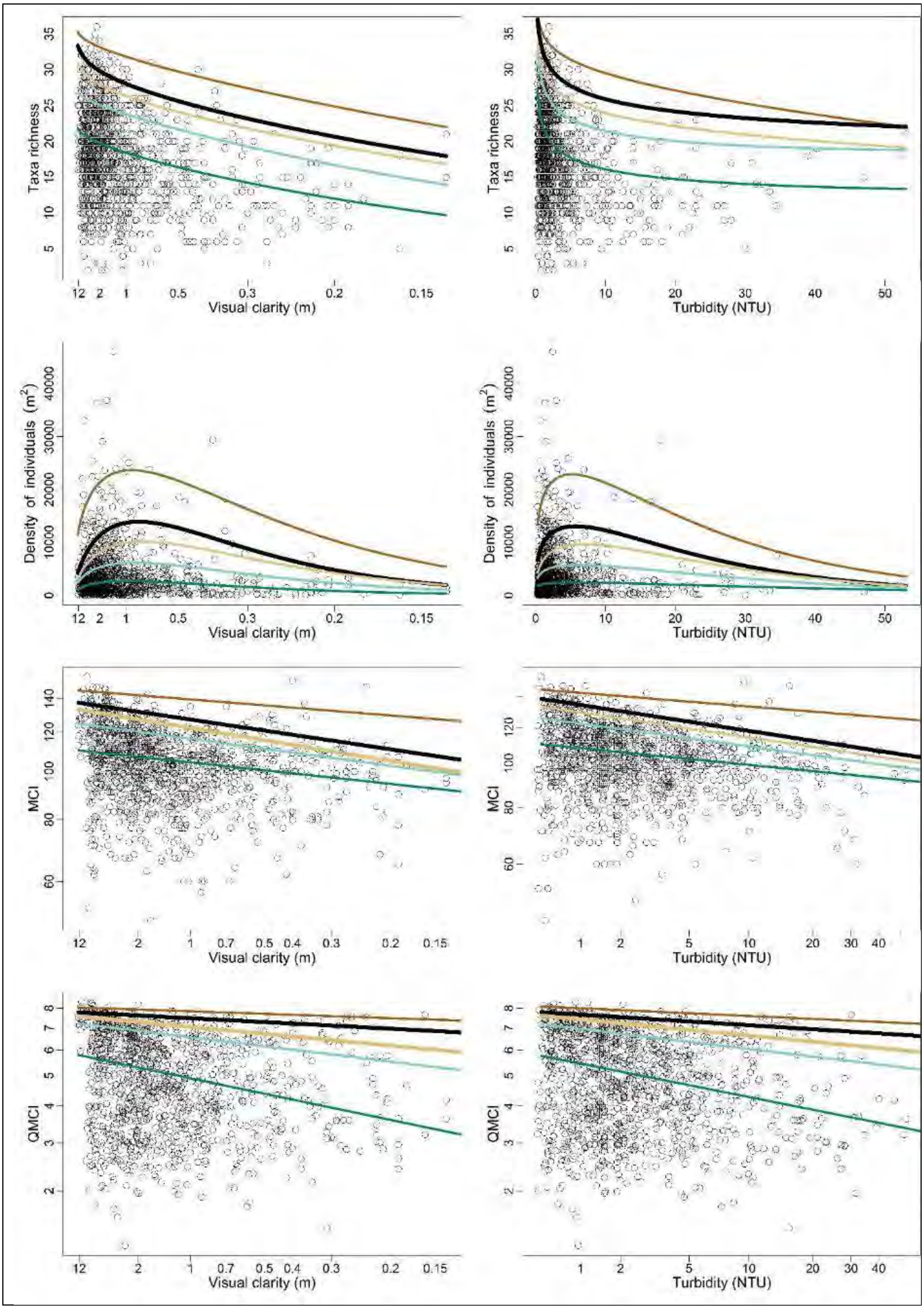
- Visual clarity:

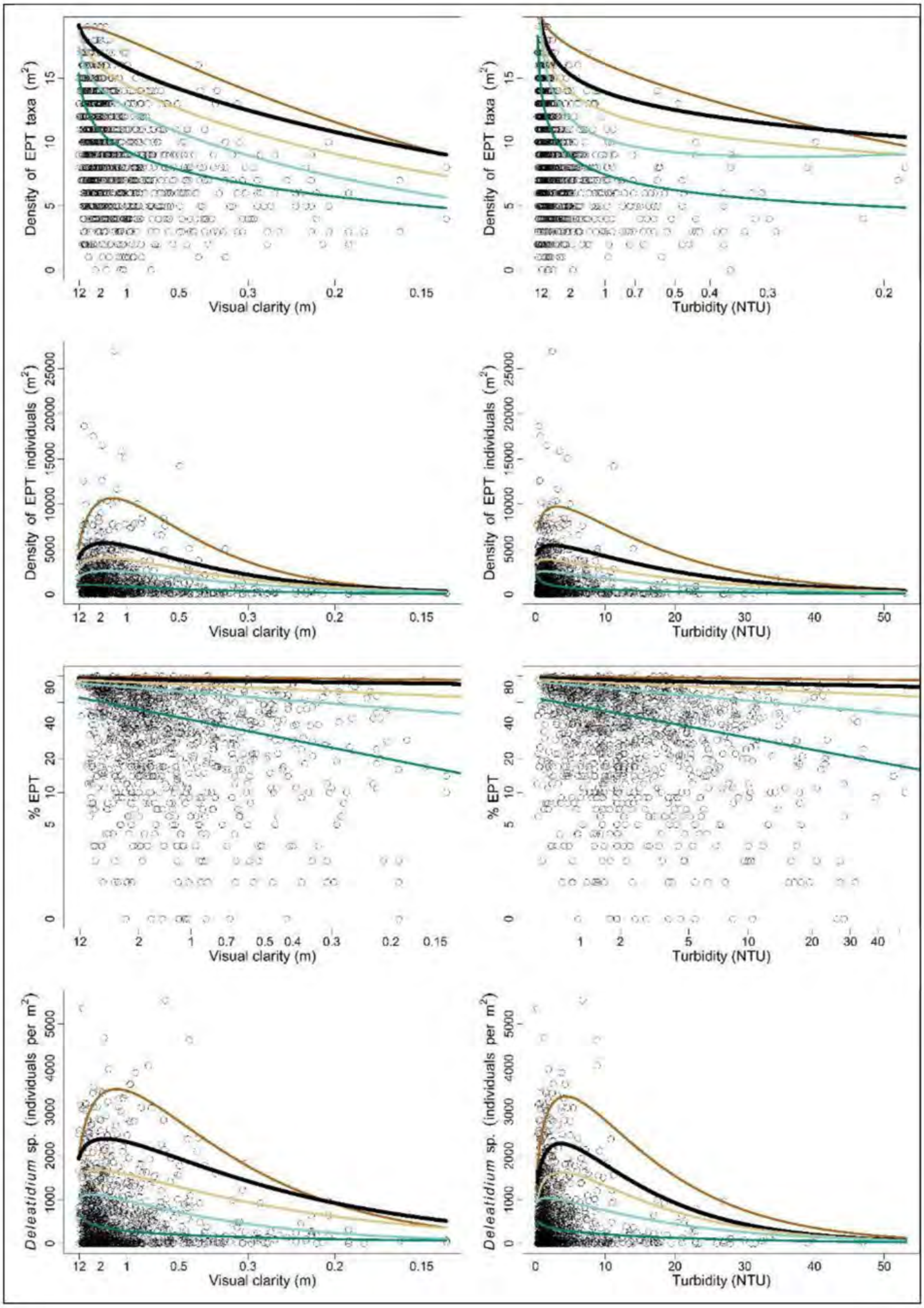
```
tau: [1] 0.95

Coefficients:
                          Value    Std. Error t value  Pr(>|t|)
(Intercept)               87.64812  1.41728   61.84240  0.00000
CLAR_median_12             1.35338  0.29611    4.57048  0.00001
```

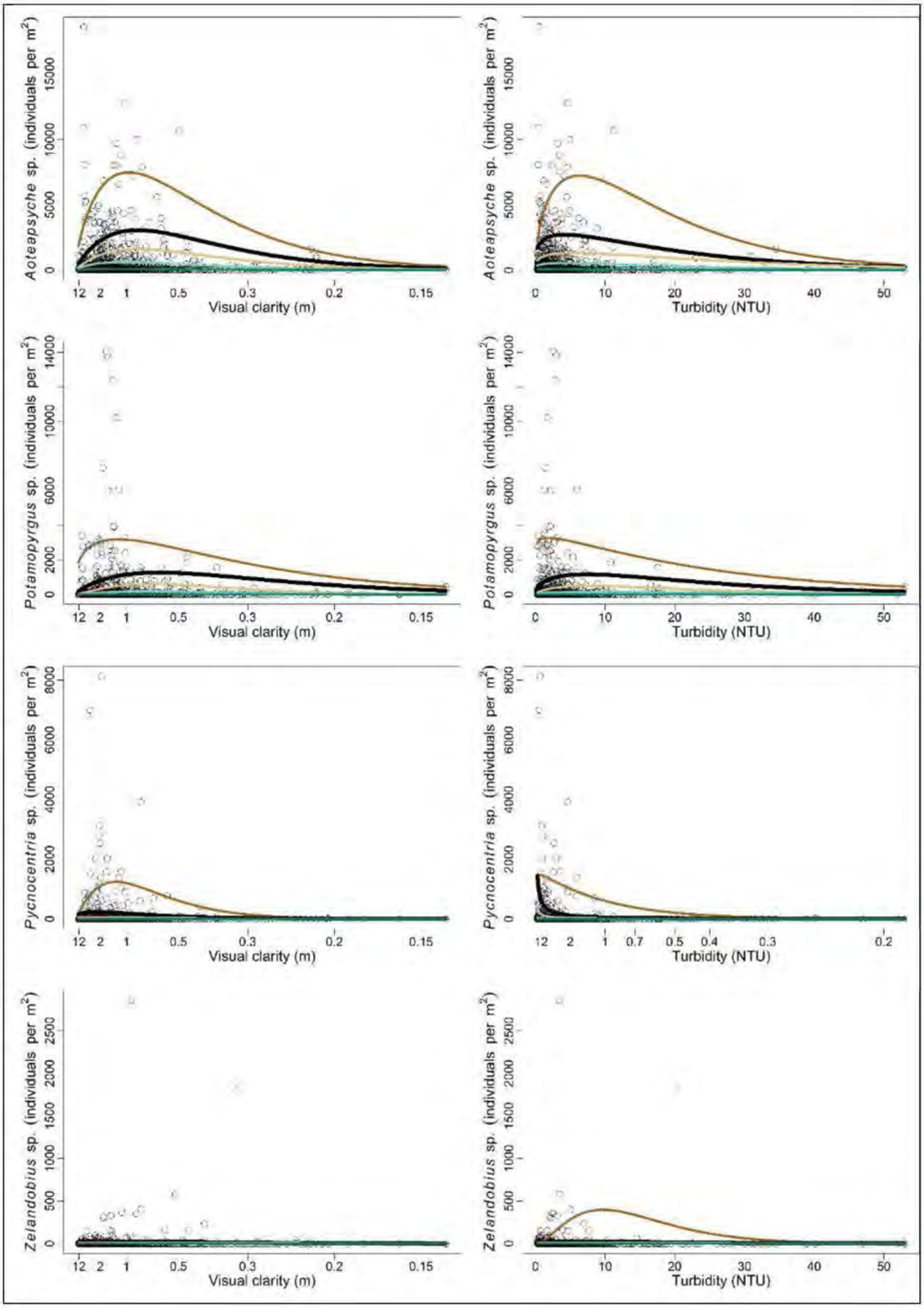
- Turbidity

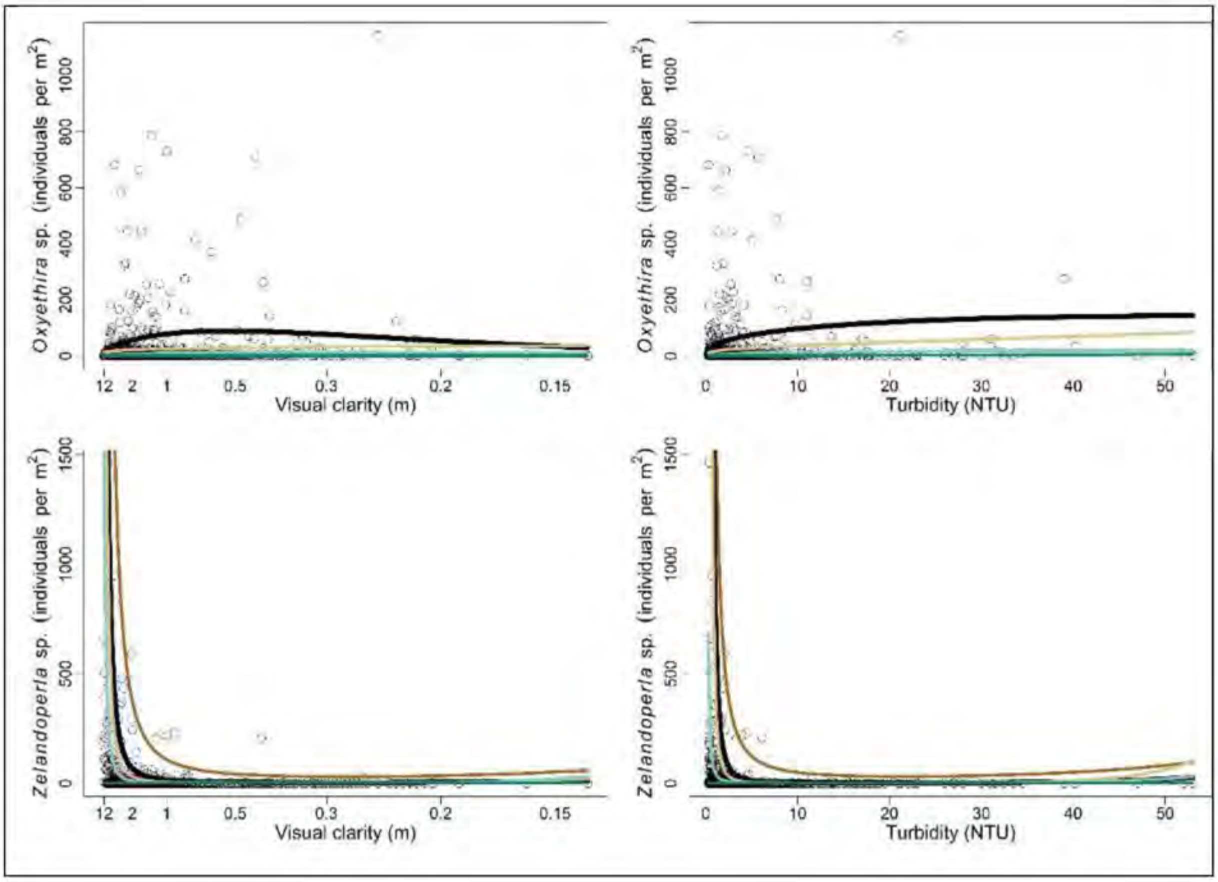
```
tau: [1] 0.95

Coefficients:
                          Value    Std. Error t value  Pr(>|t|)
(Intercept)               94.23306  0.97714   96.43785  0.00000
TURB_median_12            -0.54201  0.29616   -1.83010  0.06747
```

Appendix Z Quantile regression curves fitted to macro-invertebrate responses for visual clarity (left) and turbidity (right)




Density of EPT taxa (m^2)
Visual clarity (m)
Turbidity (NTU)
Density of EPT individuals (m^2)
Visual clarity (m)
Turbidity (NTU)
% EPT
Visual clarity (m)
Turbidity (NTU)
Deleatidium sp. (individuals per m^2)
Visual clarity (m)
Turbidity (NTU)


Aoteapsyche sp. (individuals per m²)
Potamopyrgus sp. (individuals per m²)
Pycnocentria sp. (individuals per m²)
Zelandobius sp. (individuals per m²)
Visual clarity (m)
Turbidity (NTU)



Oxyethira sp. (individuals per m2)
0
200
400
600
800
1000
12
2
1
0.5
0.3
0.2
0.15
Visual clarity (m)
Oxyethira sp. (individuals per m2)
0
10
20
30
40
50
Turbidity (NTU)
Zelandoperla sp. (individuals per m2)
0
500
1000
1500
12
2
1
0.5
0.3
0.2
0.15
Visual clarity (m)
Zelandoperla sp. (individuals per m2)
0
10
20
30
40
50
Turbidity (NTU)

Appendix AA Summary of 30% suspended sediment (visual clarity and turbidity) effect thresholds for the 14 macroinvertebrate response metrics

Table AA-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. ND indicates model fit not suitable for use in effects determination.

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
Deleatidium	2,383	1.69	1668	0.38
Aoteapsyche	3,067	0.82	2147	0.39
Potamopyrgus	1,264	0.61	885	0.28
Pycnocentria	206	6	144	0.71
Zelandobius	24	0.85	17	1.2
Oxyethira	89	0.52	62	1.4
Zelandoperla	1114	6	710	5.2

Table AA-2: Summary of 30% effect thresholds for turbidity based on the 95th percentile quantile relationships. All variables show variable maximum and corresponding 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 corresponding to a low turbidity condition. ND 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	13063	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	5435	2.4	3805	12.2
%EPT	93.9	0.5	65.8	ND
Deleatidium	2275	3.6	1593	12
Aoteapsyche	2717	4.2	1902	15
Potamopyrgus	1161	7.4	812	14.8
Pycnocentria	946	0.5	662	0.8
Zelandobius	24	11.8	16.8	8.2
Oxyethira	ND			
Zelandoperla	6477	0.5	3239	0.6

APPENDICES FOR SECTION 6

Appendix BB Literature: deposited sediment and fish

Table BB-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 day	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 day	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 day	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 day	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 day	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 day	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 day	UK, experimental stream channel	60% volume of fine sediment (peat material) in stream gravel	Olsson & Persson (1986)
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 day	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 day	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 (2014a, b)
Banded kōkopu	Juvenile & Adult	Size-based microhabitat selection; juvs associated with fine (<2mm) substrates, adults associated with coarse (>2mm 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 (2014b)

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)

Appendix CC Literature: suspended sediment and fish

Table CC-1: 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 – Nephelometric Turbidity Units) 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 & Merkens (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. (2002)
Smelt	Adult	No significant effect on feeding rate		1 hrs	NZ, lab tank	160 NTU	Rowe et al. (2002)
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 day	NZ, lab tank	15 NTU	Cavanagh et al. (2014) tech. report
Kōaro	Juvenile (assumed)	Growth slowed, no effect on weight	Reduced feeding efficiency	21 day	NZ, lab tank	50 NTU	Cavanagh et al. (2014) tech. report
Eel sp.	Juvenile (assumed)	No effect on growth, no effect on weight	Reduced feeding efficiency	21 day	NZ, lab tank	200 NTU	Cavanagh et al. (2014) tech. report
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 day	NZ, lab tank	15 NTU	Cavanagh et al. (2014) tech. report
Kōaro	Juvenile (assumed)	No mortality		21 day	NZ, lab tank	50 NTU	Cavanagh et al. (2014) tech. report
Eel sp.	Juvenile (assumed)	No mortality		21 day	NZ, lab tank	200 NTU	Cavanagh et al. (2014) tech. report

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. (2002) tech report)
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. (2002) tech. report
Kōura	Adult	No mortality		24 hrs	NZ, lab tank	20,000 NTU	Rowe et al. (2002) tech. report
Inānga	Juvenile	LC50	Gill damage	24 hrs	NZ, lab tank	20,235 NTU	Rowe et al. (2002) tech. report
Redfin bully	Adult	No mortality		24 hrs	NZ, lab tank	40,000 NTU	Rowe et al. (2002) tech. report
Banded kōkopu	Juvenile	No mortality		24 hrs	NZ, lab tank	40,000 NTU	Rowe et al. (2002) tech. report
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	Boubée 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	Boubée 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)
Inānga	Juvenile	Avoidance response (50%)		20 min	NZ, lab tank	420 NTU	Boubée et al. (1997)
Redfin bully	Juvenile	No avoidance		20 min	NZ, lab tank	1,110 NTU	Boubée et al. (1997)
Shortfin eel	Juvenile	No avoidance		20 min	NZ, lab tank	1,110 NTU	Boubée et al. (1997)
Brown trout	Juvenile	Reduction in abundance (85%)		361 day	England, in-stream	5,838 g m^{-3}	Herbert & Merkens (1961)

Appendix DD Technical methods: fish-sediment ESV responses

Introduction

This section sets out the full details of the technical methods used in the process of characterising fish-sediment ESV responses as part of this project. It expands on the methodology described in the main report in Section **Error! Reference source not found.**, but is set out in a way that it can be read a s a standalone document. The objective of this component of the project was to test for, and characterise, relationships between fish 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 for fish were to:

1. determine the availability of suitable datasets
2. derive reference state for the sediment ESVs as a function of landcover
3. model fish probability of capture as a function of sediment ESVs within landscape settings
4. evaluate fish community change in response to deviation of ESV state from reference conditions, and
5. 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 national river network version 1), 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 (version 1) 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.

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.

It should be noted that these observations of % cover of 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). % cover of total fines from the NZFFD were, therefore, compared with the % cover of total fines data observed using the visual instream method (SAM2), the closest equivalent measure in (Clapcott *et al.* 2011a), collated in Stage 1a (Figure DD-1 and Figure DD-2). Visual inspection of these 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. However, further investigation indicated a large discrepancy between the size of rivers within the two data sets (Figure DD-3 and Figure DD-4). NZFFD observations of % cover of total fines 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 instream visual observations.

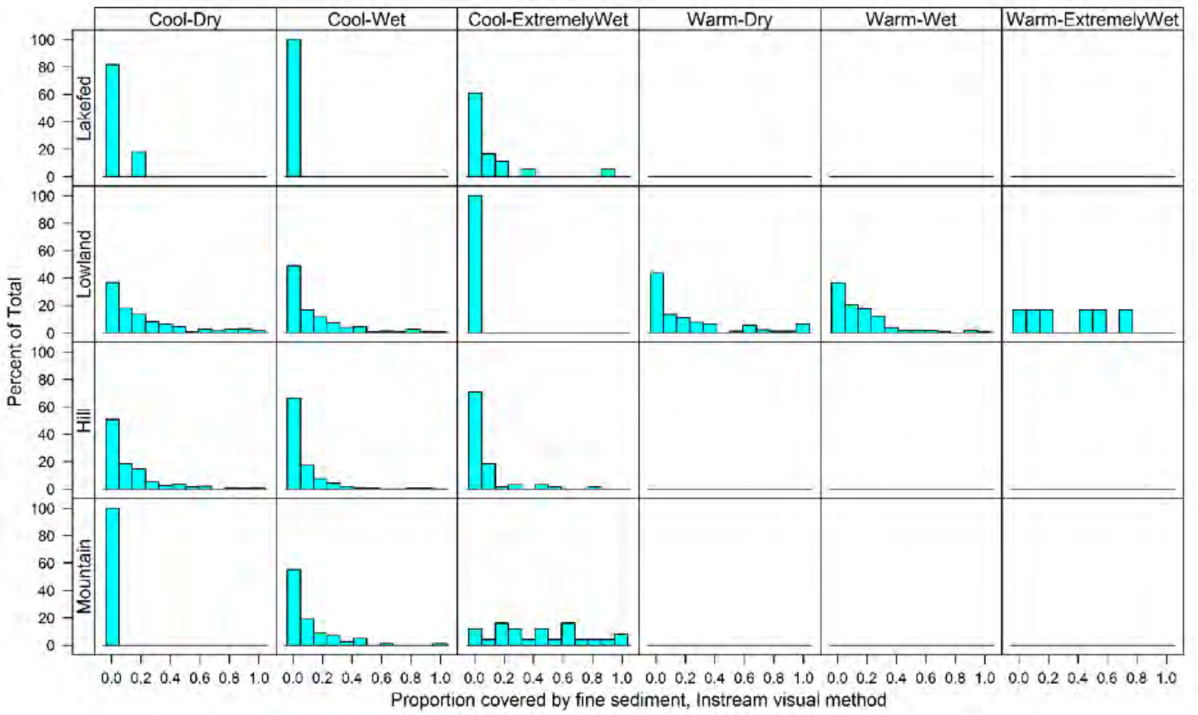


Figure DD-1: Histograms of areal cover of deposited fine sediment observed using the instream visual method (SAM2) by REC climate and topography classes. These data are from the deposited sediment dataset assembled during Stage 1a of this project.

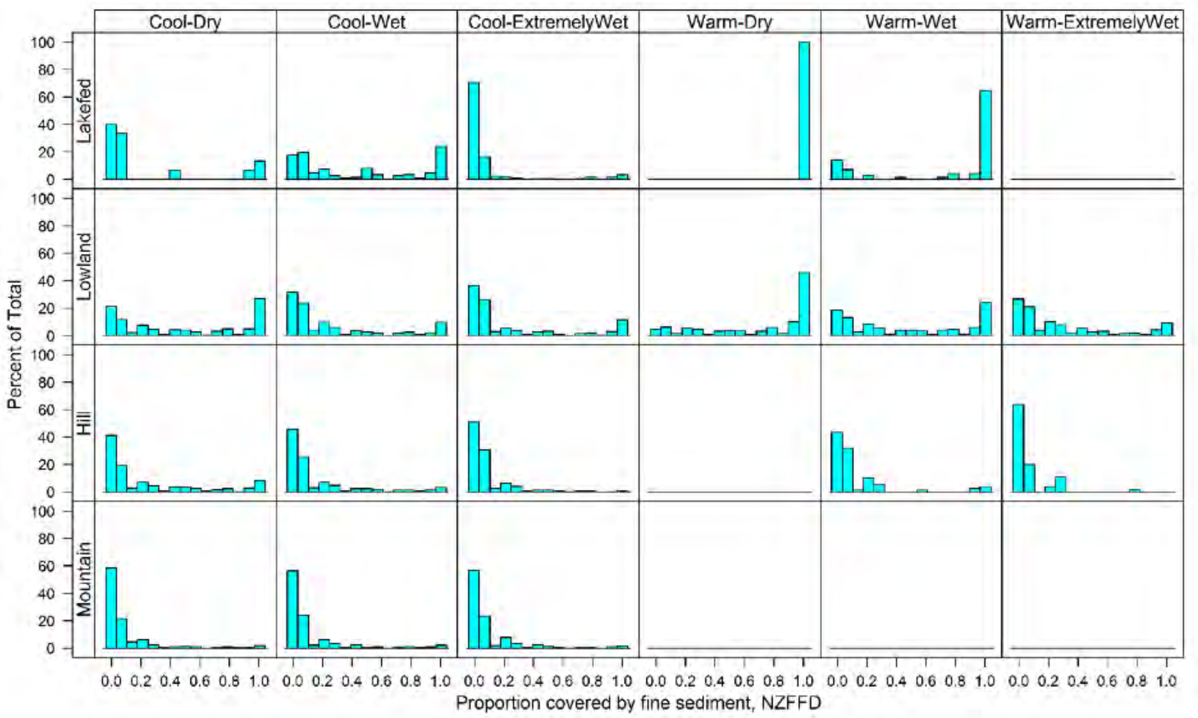


Figure DD-2: Histograms of areal cover of deposited fine sediment recorded in the NZFFD by REC climate and topography classes.

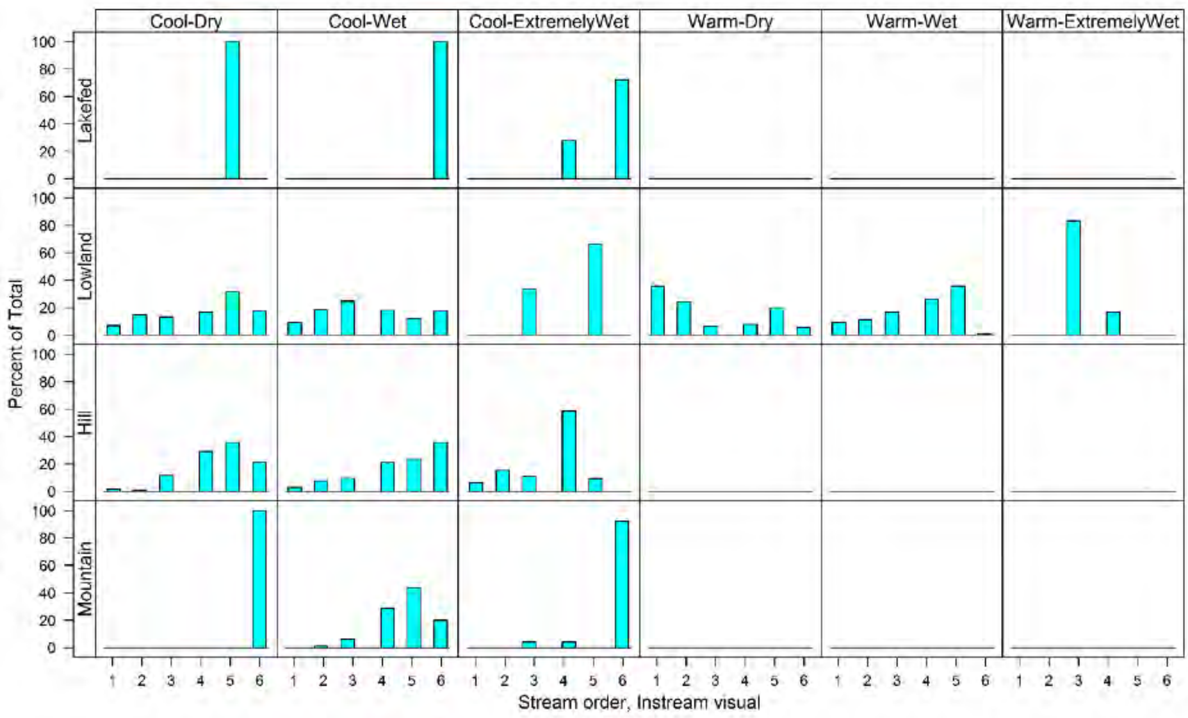


Figure DD-3: Histogram of stream orders from which deposited fine sediment cover has been observed independently using the SAM2 instream visual method. Stream order 6 represents orders 6 and above.

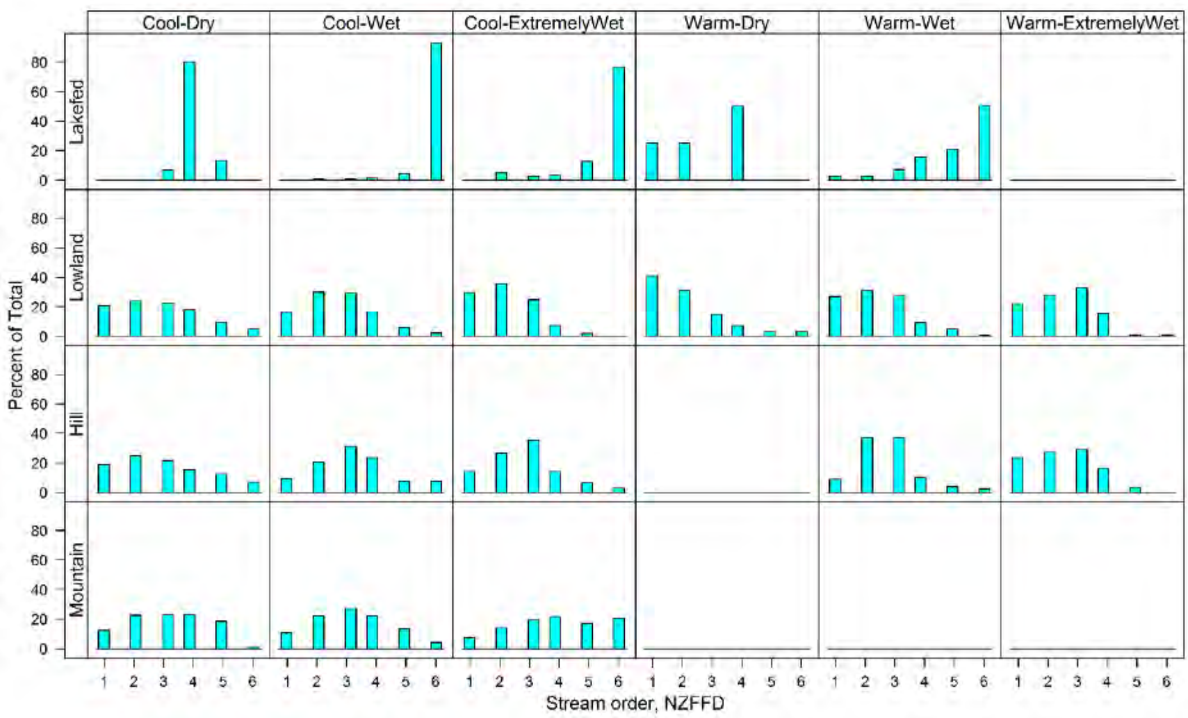


Figure DD-4: Histogram of stream orders from which deposited fine sediment cover has been observed in the NZFFD records. Stream order 6 represent orders 6 and above.

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 DD-1). All other variables were significant. This indicated that it was legitimate to employ all available NZFFD deposited sediment data to investigate the relationships between deposited total fine sediment and heavy pasture.

Table DD-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

For the purposes of this project we utilised the long-term site median suspended sediment ESV values derived from the dataset compiled by (Unwin & Larned 2013). This consisted of data assembled from regional council State of the Environment monitoring and NIWA's National River Water Quality Network programme. Site medians for visual clarity were available for 722 sites (590 with > 5years, all sites had at least one full year of data) and site medians for turbidity were available for 833 sites (808 with > 5years, all sites had at least one full year of data).

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 2002a). (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 section 4.3.1).

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 (Stage 1a deposited sediment dataset and (Unwin & Larned 2013) suspended sediment dataset).

Several spatial matches between independent ESV observations and NZFFD records on the same reach, but on different dates, were found (instream visual % cover of total fines = 260, visual bankside % cover of total fines = 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 DD-5). Only three of the paired sediment and fish observations also coincided by sampling date. Comparison between values of % cover of total fines recorded in the NZFFD and the deposited sediment ESVs independently observed on the same NZ reach, but at different times, showed very weak patterns (Figure DD-6). 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. It was, therefore, considered inappropriate to pair the fish and deposited sediment ESV observations by spatial match alone.

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 DD-7), and there were very few time-series of paired observations in the same catchment.

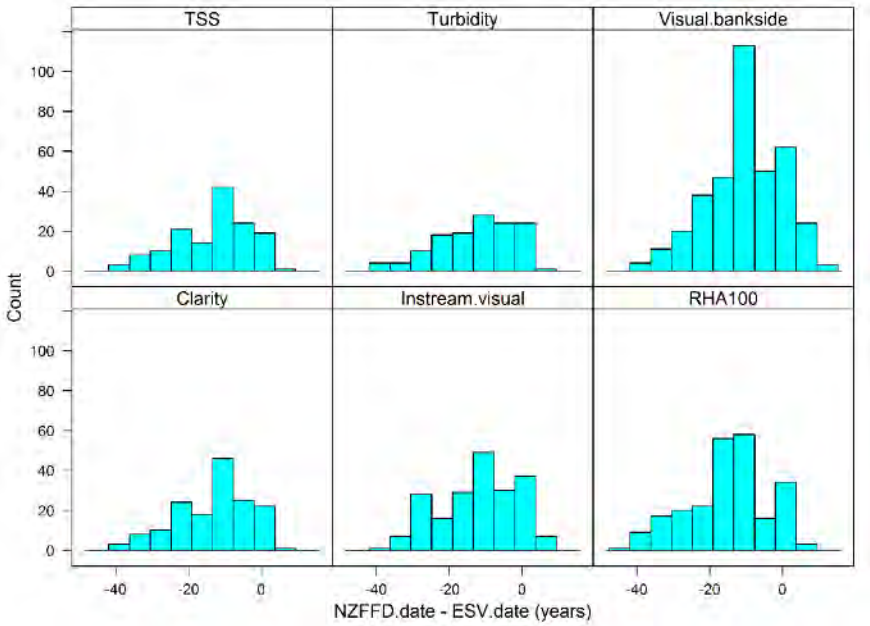


Figure DD-5: Summary of temporal separation between spatially paired fish and sediment ESV observations. Negative numbers mean that the fish observation was prior to the sediment observation. Positive numbers mean that the sediment observation was prior to the fish observation. TSS, turbidity and visual 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.* 2011a) for deposited sediment.

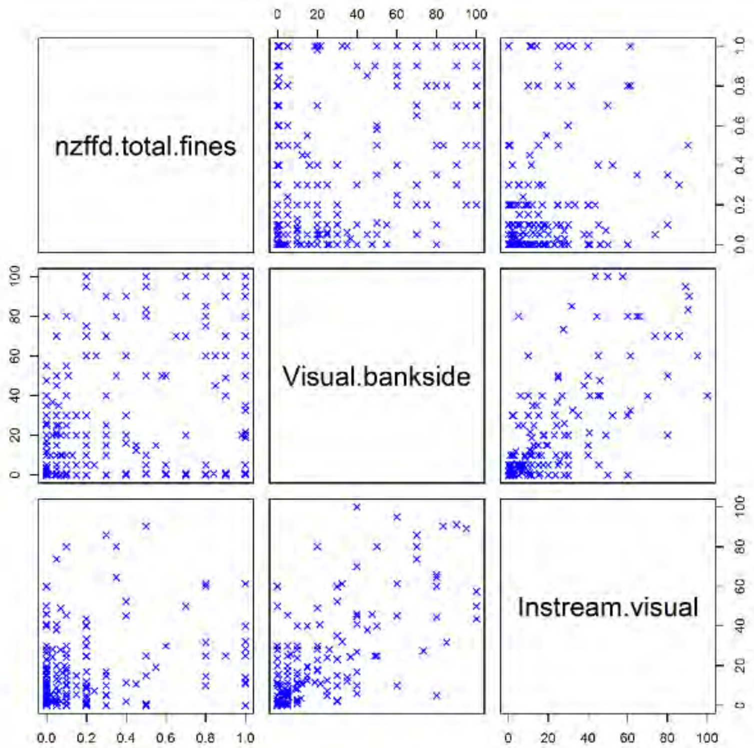


Figure DD-6: Comparison between non-synchronously observed sediment ESVs. Comparison of % cover of total fines measurements from the NZFFD (nzffd.total.fines) at the same NZ reaches, using visual bankside method (Visual.bankside, SAM1) and using the instream visual method (Instream.visual, SAM2).

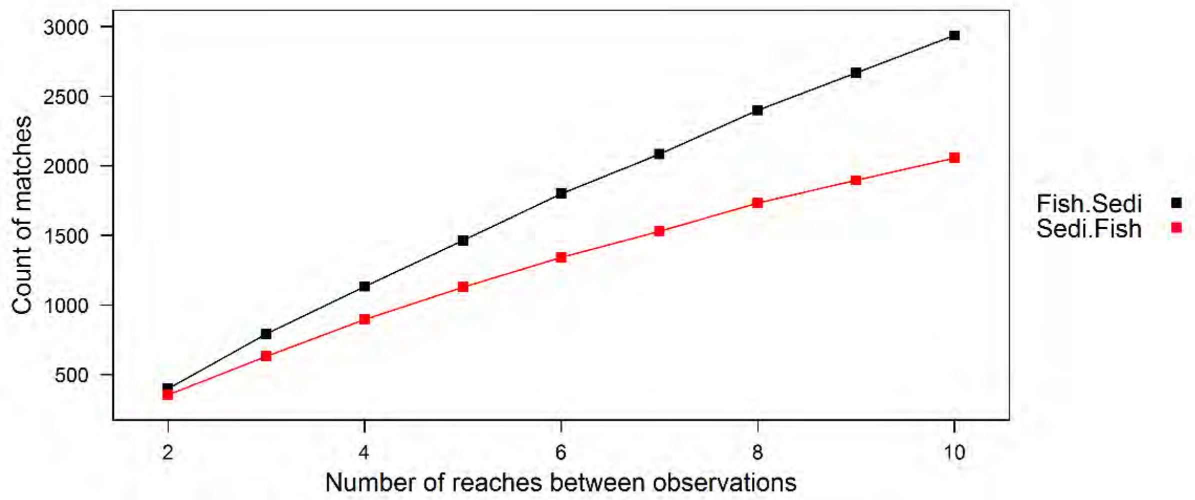


Figure DD-7: Count of spatial matches between NZFFD records located upstream (Fish.Sedi) or downstream (Sedi.Fish) of an independent sediment ESV observation.

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 ESV observations making this approach unsuitable for this project. In the case of deposited sediment, it was decided to use the deposited sediment data (% cover of fine sediment) associated with the NZFFD records. To advance the analyses for suspended sediment, modelled median visual clarity and turbidity derived by (Unwin & Larned 2013) 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. Predicted medians of visual clarity and turbidity provided by (Unwin & Larned 2013) were calculated by fitting random forest models to observed medians for these two variables. Because the summary statistic of these ESVs was the long-term median, results would be compatible with existing state of the environment monitoring strategies for these variables.

Deriving reference state for sediment ESV as a function of landcover

The method applied in this project essentially replicated that of (McDowell *et al.* 2013), which investigated relationships between the ESV and the proportion of the upstream catchment in heavy pasture (Figure DD-8 and Figure DD-9), although the influence of exotic vegetation (e.g., exotic forestry) and urban landcover was also incorporated into the analysis. The equivalent plot for observed % cover of total fines from the NZFFD is not show due to the large number (22,946) of observations.

Following the method of (McDowell *et al.* 2013), the response of each sediment ESV was modelled as a function of the proportion of the upstream catchment covered by artificial landcovers using mixed-effects models. Landcover proportions were incorporated by including heavy pasture, exotic vegetation and urban landcover as fixed-effects. Landcover proportions were calculated by joining several LCDB3 classes. The influence of landscape setting was incorporated by including the first two levels of the REC (Snelder & Biggs 2002a) as random-effects (topography class within climate class). Mountain and Glacial Mountain classes were amalgamated. These amalgamations were applied to avoid having classes with very small numbers of sites and to maintain balance within random-effects. Crossed random-effects were applied such that lower level random-effects were not independent between their respective higher levels (e.g., hill topography class observations within each climate class were related to hill topography observations within other climate classes).

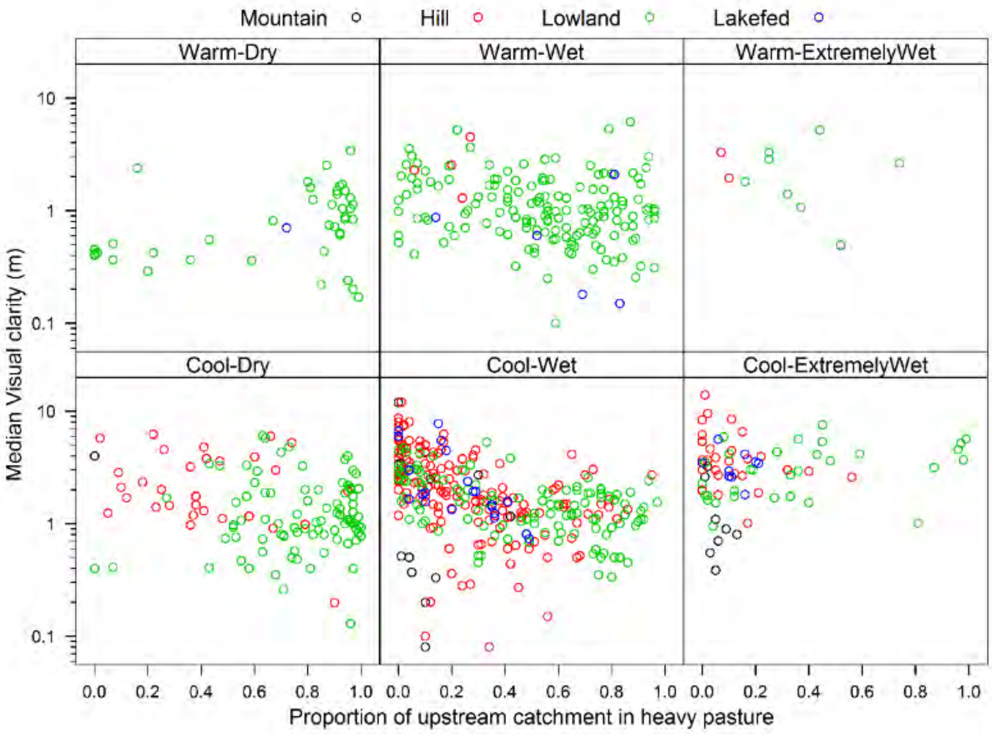


Figure DD-8: Median visual clarity plotted against proportion of upstream catchment in heavy pasture. Data are the site median visual clarity from (Unwin & Larned 2013). Observations are split by REC climate and topography classes.

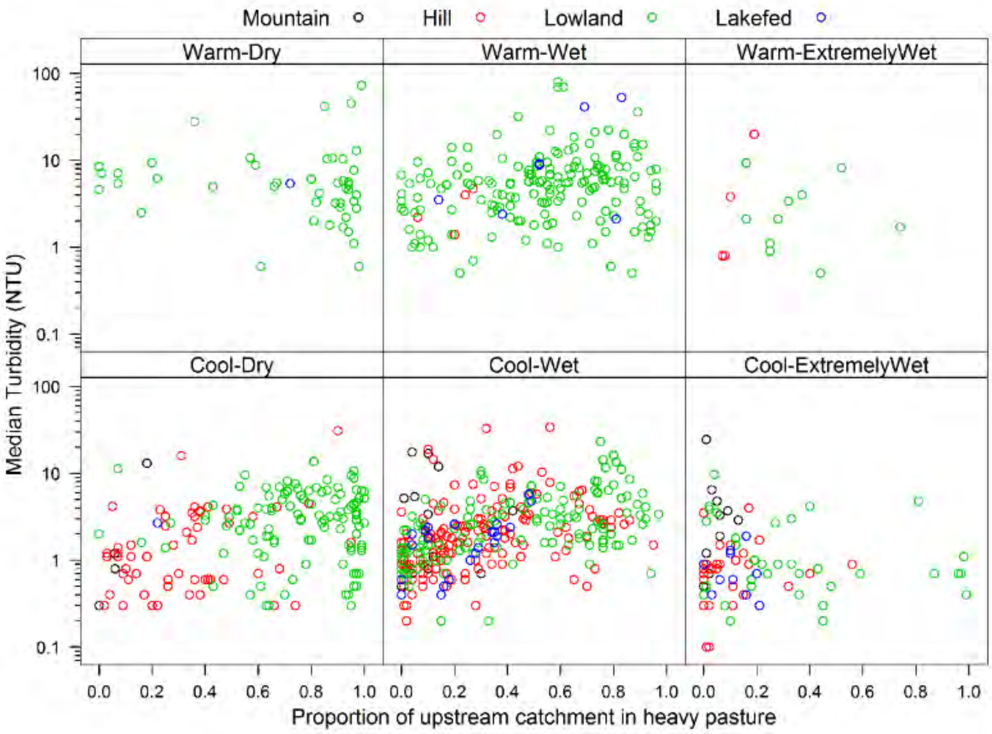


Figure DD-9: Median turbidity plotted against proportion of upstream catchment in heavy pasture. Data are the site median turbidity from (Unwin & Larned 2013). Observations are split by REC climate and topography classes.

For clarity and turbidity, generalised additive mixed models (gamms) were applied (Equations (5) and (6)). A smoother term was included to account for non-linearity in the relationships between the ESV and each of heavy pasture and exotic vegetation separately using the method of (Wood 2004). A smoother term was not applied for urban because there was not a good spread of urban cover from 0 to 1. The state of the environment data describing observed median clarity and observed median turbidity from (Unwin & Larned 2013) were used to fit these models.

For % cover of total fines, generalised linear mixed models (glmm) were applied with a binomial family as is appropriate for proportion data (Equation (7)). Observations of the proportion of river bed covered by total fines (sum of mud/silt and sand categories) extracted from the NZFFD were used to fit these models. Network position (high, medium or low Strahler stream orders) was included as a fixed-effect on % cover of total fines as it was expected that network position would influence patterns in deposition of fine sediment, and because a sufficient spread of data across network positions was contained within the NZFFD observations.

Following the method of (McDowell *et al.* 2013), inclusion of random-effects of climate on the slope of the ESV-landcover relationships was assessed by comparing Akaike Information Criteria (AIC) values for competing models. Random-effects on topography were not included as visual inspection of the data indicated that there were insufficient data across the possible range of landcover proportions to characterise slopes at these lower levels of the mixed-effects models.

Models that included interactions between fixed-effects (e.g., an interaction between heavy pasture and exotic forest) were also fitted. Inspection of the models showed that, whilst some interactions were statistically significant, their inclusion had a negligible influence on predictions due to very low coefficients on the interactions terms.

Following (McDowell *et al.* 2013), the intercept on the y-axis of each mixed-effects model was used to obtain the ESV value that would be expected on average under natural landcover within each landscape setting. This is the predicted ESV value when upstream heavy pasture, exotic vegetation and urban landcover are zero. The same models can also be used to obtain predictions of ESV values for any combination of heavy pasture, exotic forest and urban landcovers within each landscape setting.

The following models were selected as the most appropriate for predicting sediment ESV reference state for the purposes of this project:

$$\text{Log}_{10}(\text{Clarity}) \sim s(\text{heavy pasture}) + s(\text{exotic veg}) + \text{urban} + \text{heavy pasture}|\text{climate} + {+1}|\text{climate/topography} \quad (5)$$

$$\text{Log}_{10}(\text{Turbidity}) \sim s(\text{heavy pasture}) + s(\text{exotic veg}) + \text{urban} + \text{heavy pasture}|\text{climate} + 1|\text{climate/topography} \quad (6)$$

% fine sediment cover ~ heavy pasture + exotic veg + urban + network position + heavy pasture|climate +1|climate/topography (7)

One advantage of the mixed-effects modelling approach is that predictions can be obtained for a landscape setting that is not present in the fitted dataset (e.g., an alluvial, cool-wet, mountain), by providing predictions at the next available level in the model (e.g., cool-wet, mountain).

Another advantage of this method is that ability to distinguish conditions between different levels of random-effects can also be investigated to assess uncertainty within the regression model. Empirical Bayes estimates of coefficient on random-effects showed that there were differences between landscape settings for each ESV (Figure DD-10, Figure DD-11 and Figure DD-12) as indicated by non-overlapping standard errors. However, these plots also suggest that there is some overlap between ESV conditions between some landscape settings; indicating some form of amalgamation may be feasible. Inspection of the number of samples revealed that smaller sample numbers were responsible for wider standard errors for some landscape settings (e.g., warm-wet lakefed).

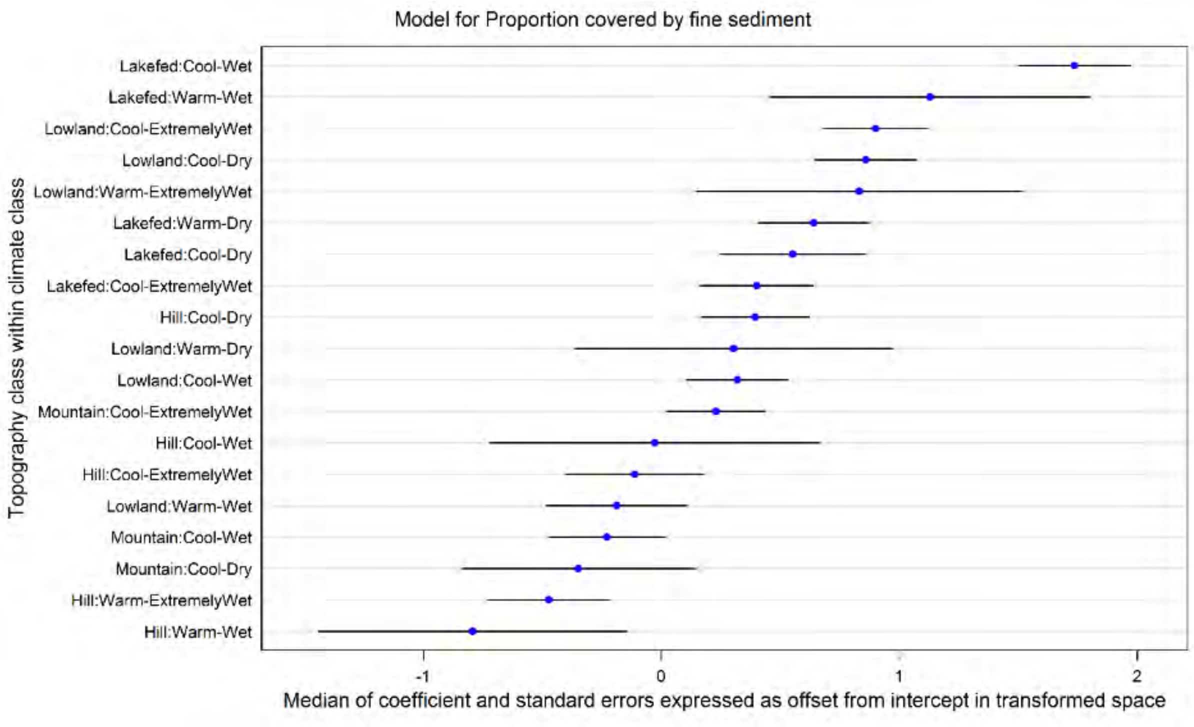


Figure DD-10:Median values and standard errors on random-effects for the % cover of total fines regression model calculated using empirical Bayes estimates.

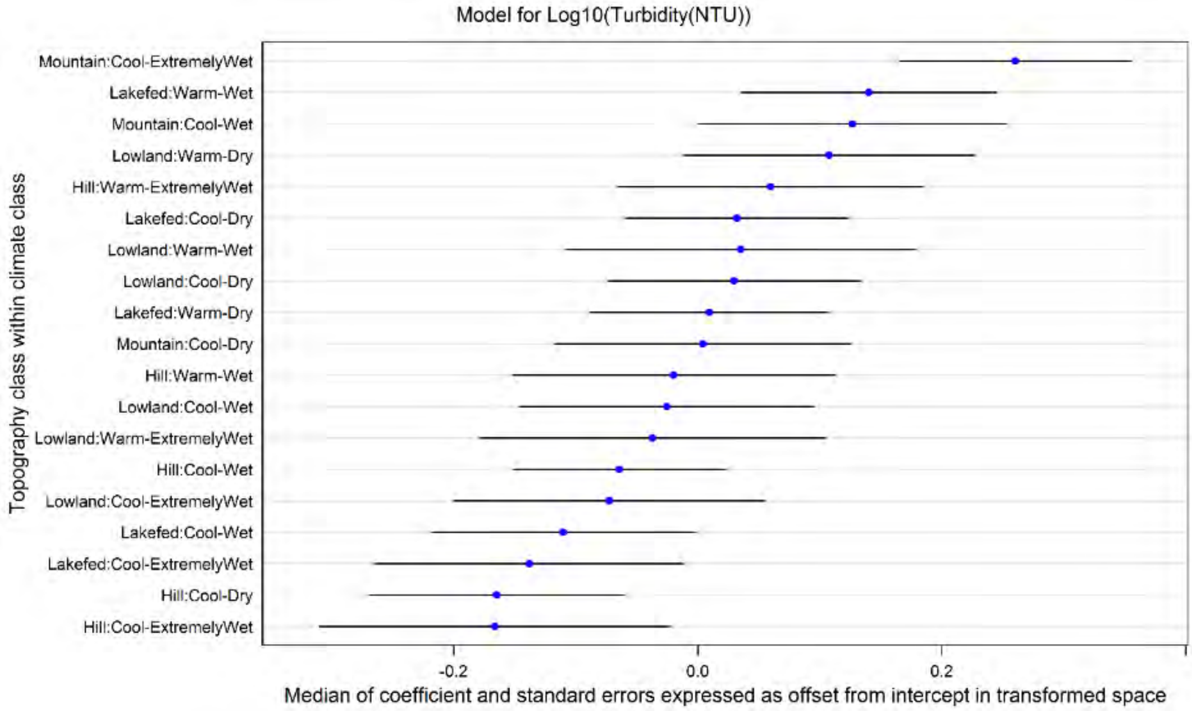


Figure DD-11:Median values and standard errors on random-effects for the turbidity regression model calculated using empirical Bayes estimates.

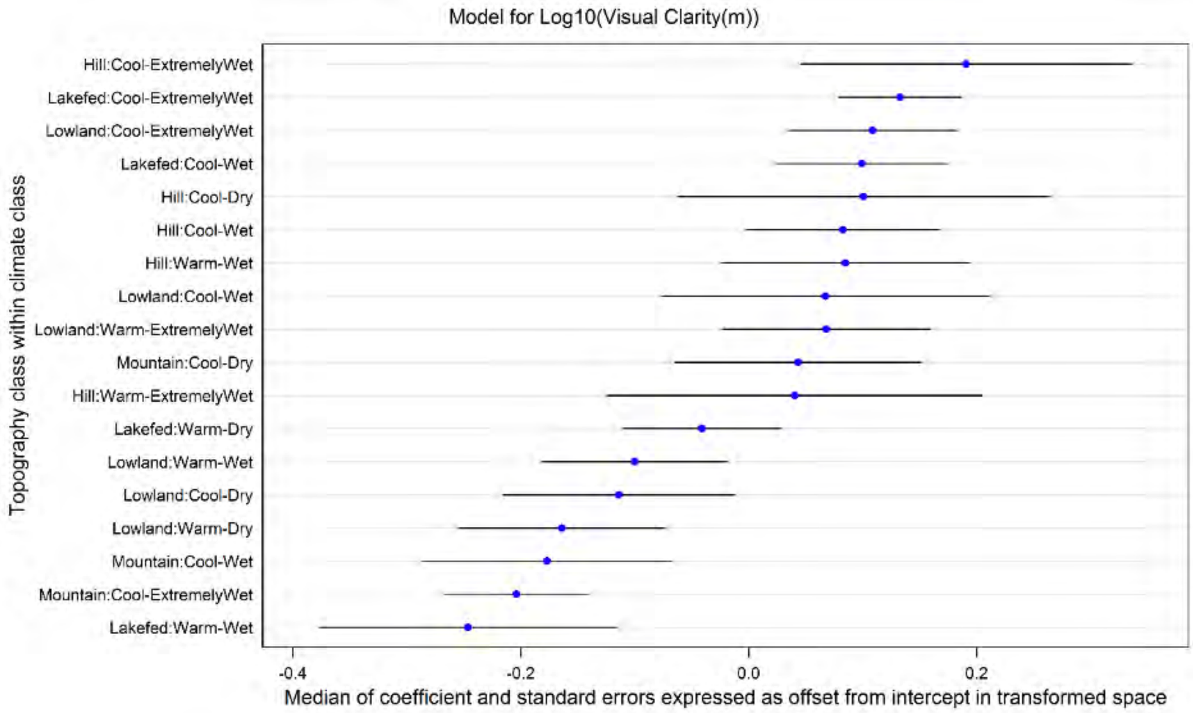


Figure DD-12:Median values and standard errors on random-effects for the visual clarity regression model calculated using empirical Bayes estimates.

Fish probability of capture as a function of ESV within landscape settings

Eleven species were selected for this analysis. These species were included in the analysis because each was found across New Zealand and was present in a reasonable proportion of samples in the NZFFD (at least 7%). Ten species were natives. Despite not being a native species, brown trout was also included in the analysis due to the strong likelihood of this species showing a response to the ESVs and because of its high recreation value. 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 DD-13).

Fish probability of capture (FPC) for each of the 11 species was modelled as a function of each sediment ESV using a generalised linear mixed-effects model (Equation (8)). The response of each FPC was modelled as a function of each ESV by including the ESV as a fixed-effect. The proportion of the stream bed observed to be covered by total fines in the NZFFD records was used as the deposited sediment ESV. The modelled median clarity and turbidity values of (Unwin & Larned 2013) were used to relate FPC to these ESVs.

Network position (low, medium and high river orders) was included as a fixed-effect because it was expected to affect FPC because of its influence on physical habitats, such as cover for predators and spawning habitat. Distance to sea (Log to the base 10 transformed) was also included as a fixed-effect because it was expected to influence FPC. This is because many fish species spend some part of their life-cycle at sea, and differ in ability to penetrate inland due to migration speeds and climbing abilities, meaning some species are more likely to be found close to the sea. Fishing method was included as an additional fixed-effect to account for changes in FPC caused by different fishing methods (combination, trapping, visual, netting and electric fishing; Figure DD-14).

Landscape setting (from a fish distribution perspective) was incorporated by including the first two levels of the REC (topography class within climate class) as random-effects. Crossed random-effects were applied such that lower level random-effects were not independent between their respective higher levels (e.g., hill topography class observations within each climate class were related to hills topography observations within other climate classes). It was assumed that these random-effects encapsulated many aspects of hydrology, geomorphology and climate that may influence FPC through knock-on-effects from temperature, hydraulic conditions, food supply and natural barriers to migration such as steep slopes in hill and mountain settings.

Interactions between ESV and fishing method were 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).

The following model was selected as the most appropriate for describing the response of fish probability of capture to changes in sediment ESVs:

$$FPC \sim ESV + \text{fishing method} + \text{distance to sea} + \text{network position} + 1|\text{Climate/topography} \qquad (8)$$

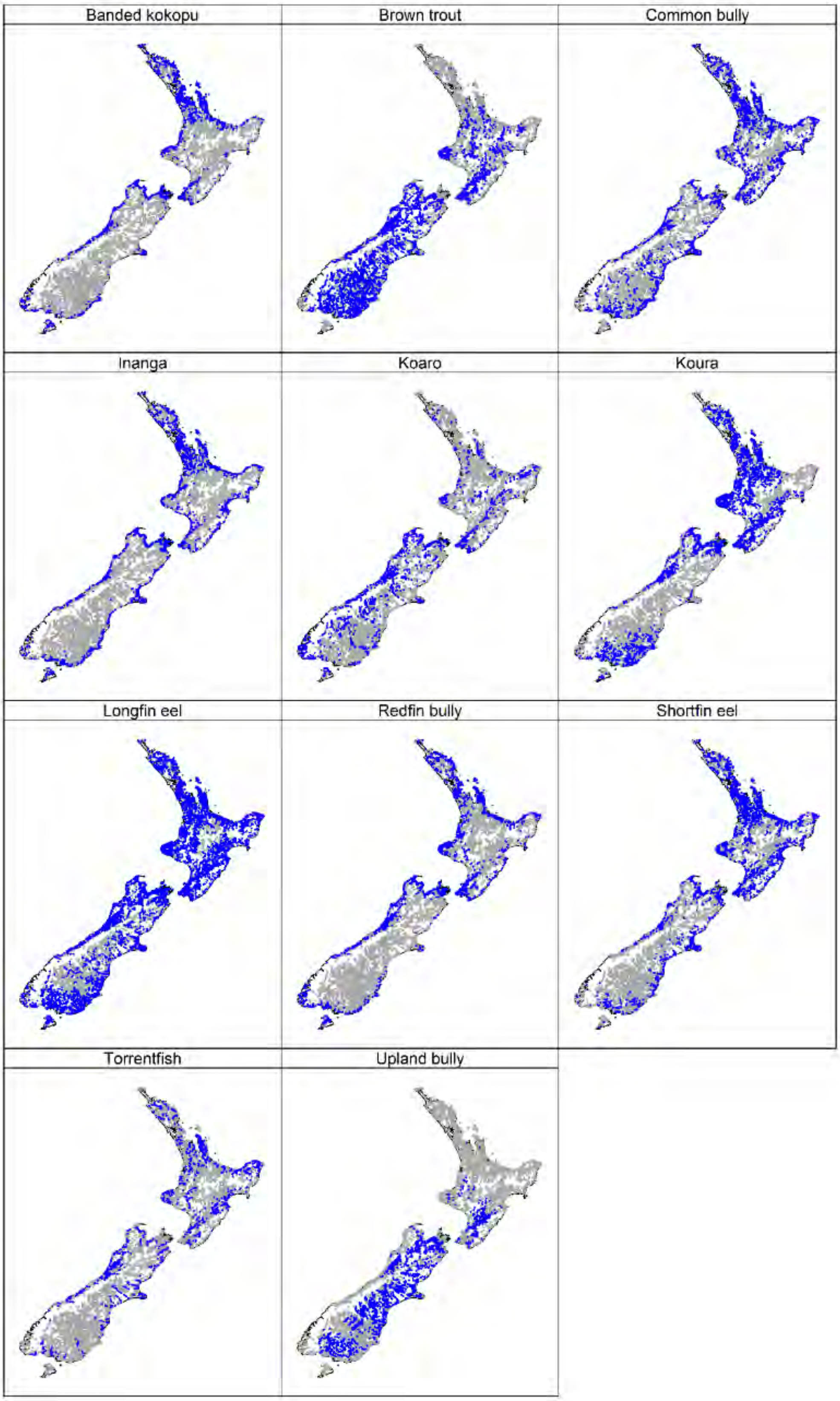

Banded kokopu
Brown trout
Common bully
Inanga
Koaro
Koura
Longfin eel
Redfin bully
Shortfin eel
Torrentfish
Upland bully

Figure DD-13:Maps of presence (blue) and absence (grey) in the NZFFD records for the eleven species used in these analyses.

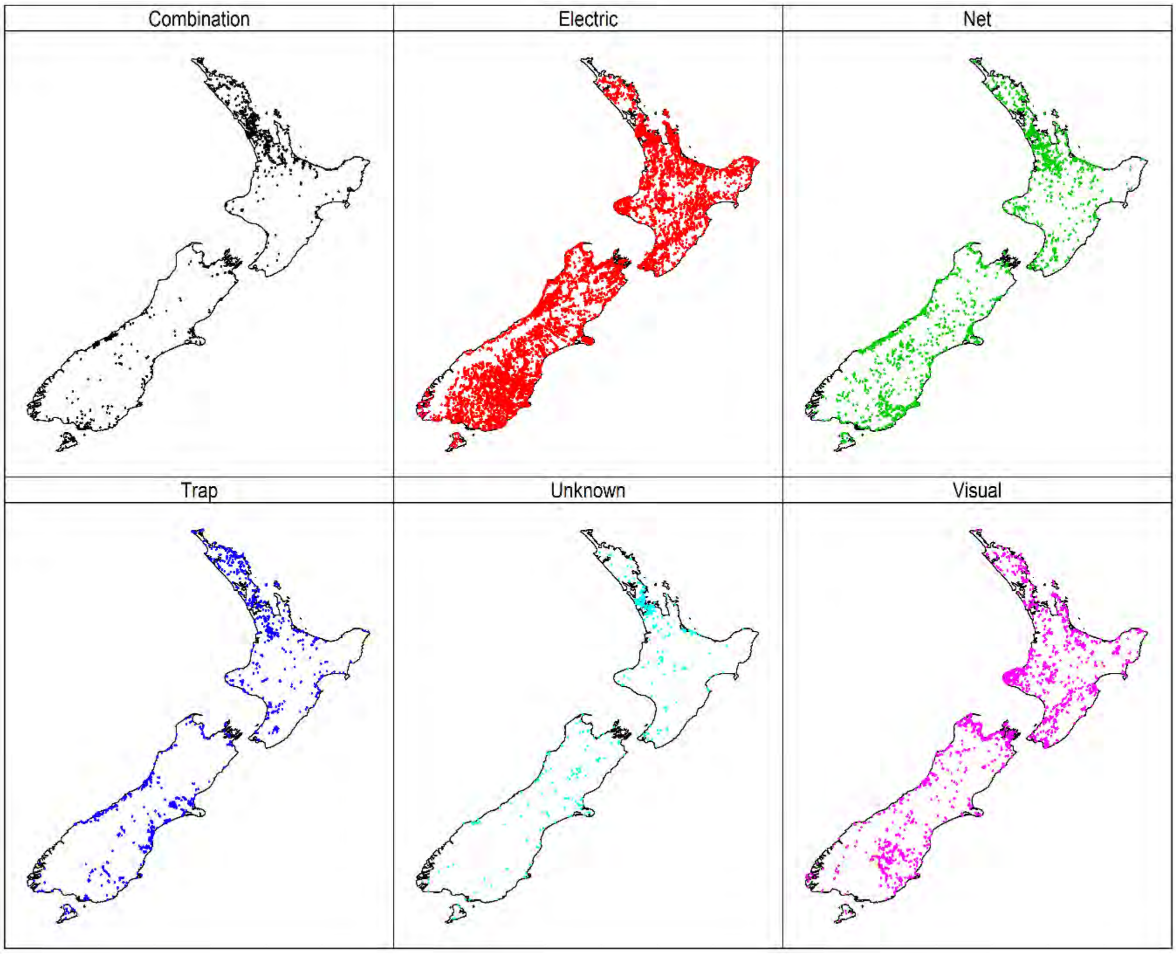


Figure DD-14:Map showing NZFFD sample locations by fishing method. Combinations included combinations of electric fishing and nets, combinations of nets and traps, and combinations of nets, traps and electric fishing; electric fishing included backpack and mains set methods; netting methods included 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 FPC model (Equation (8)) provides an estimate of the probability of capturing a species in a particular fish setting (climate/topography/network position/distance inland) at a given 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 FPC threshold at which Cohen's kappa was maximised (maxKappa) was calculated in R using the 'PresenceAbsence' package for each species. In effect, if FPC > maxKappa the species is more likely present than absent, and if FPC < maxKappa the species is more likely absent than present.

Assessing fish community change resulting from ESV state

Several steps were required to translate the predicted FPC ESV responses for individual species into a metric of expected fish community change at different ESV states. In simple terms this first involved determining the FPC at reference ESV state and an array of different ESV states for each individual species in each fish setting. These values were then combined into a metric (ΔC) describing the overall expected change in fish community relative to the community that might be expected at reference ESV state.

The information on predicted ESV reference state for each landscape setting, and FPC for each species in each fish setting, were subsequently combined to evaluate the potential impacts on fish community composition resulting from changes in ESV state. In all cases, the FPC predictions were made for electric fishing only as this gave the highest probability of capture for all species except kōura, banded kōkopu and inānga (Figure DD-15).

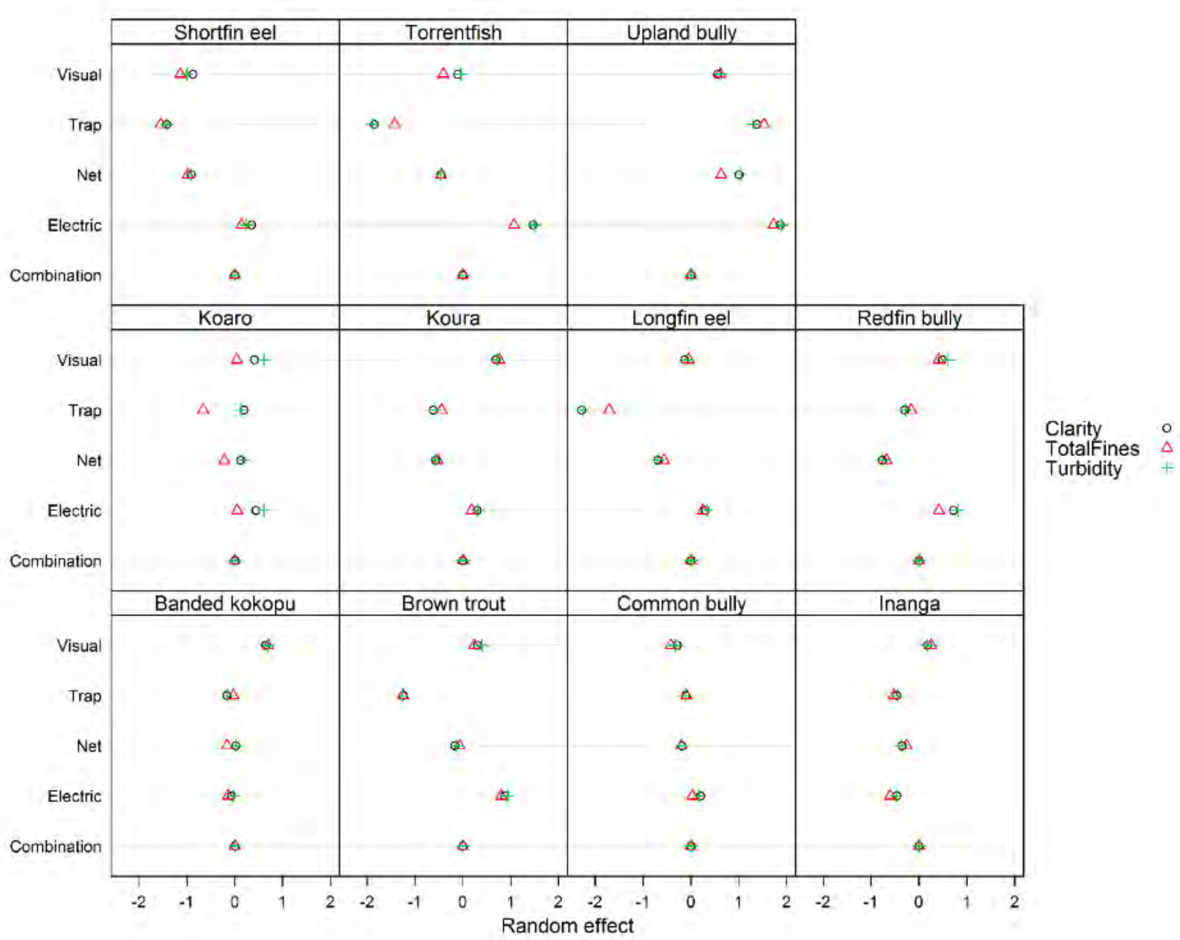


Figure DD-15:Model coefficients for different fishing methods for each of three ESVs.

The first step was to calculate FPC under predicted reference ESV state (FPC_{ref}) for each fish setting. FPC was then also calculated at different ESV values (FPC_{ESV}) for each fish setting. The ratio of these FPCs to maxKappa was then calculated:

$$P_{ESV} = FPC_{ESV} / maxKappa \quad (9)$$

$$P_{ref} = FPC_{ref} / maxKappa \quad (10)$$

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:

$$\Delta P_{ESV} = P_{ESV} - P_{ref} \quad (11)$$

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 ($\sum P_{ref}$).

$$\sum \Delta P_{ESV} / \sum P_{ref} = \Delta C \quad (12)$$

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., FPC_{ref} > maxKappa for more species), would always show more change in the expected fish 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 fish community composition relative to reference conditions. Positive values in ΔC represent net gains in fish community composition across species relative to reference conditions. ΔC, therefore, represents a deviation in fish 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 fish community integrity. Consequently, increasing departure from reference state was considered to increase the risk of negative outcomes for fish communities.

For the purposes of this study a 20% departure in fish community integrity relative to average reference condition (i.e., ΔC = -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 (ΔC = -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 clarity) for each fish setting.

To help reduce the number of different landscape settings for which bands had to be derived, two landscape settings which each had one NZ reach in the NZ river network were removed. Of a possible 24 potential combinations (climate/topography), 22 were included in the analysis. This amounted to 560,715 of 560,717 reaches, after having removed all reaches that intersected with lakes by cross-referencing the river network with a GIS description of all lakes with areas greater than one hectare.

Appendix EE Results: fish-sediment ESV responses

Predicted sediment ESV reference state

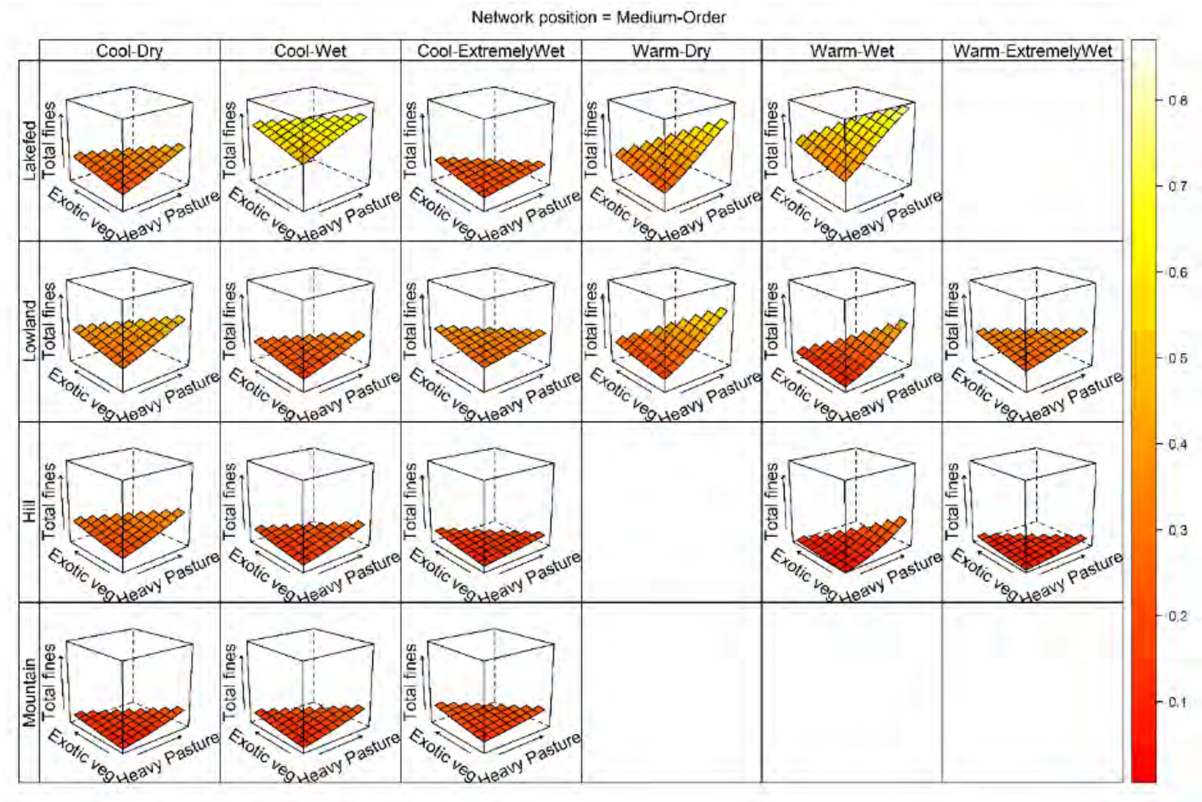


Figure EE-1: Relative effects of % cover of heavy pasture and % cover of exotic vegetation on % cover of total fines in different REC climate/topography classes. This plot represents the relationships for medium Strahler stream order rivers. Similar patterns were observed for other stream sizes.

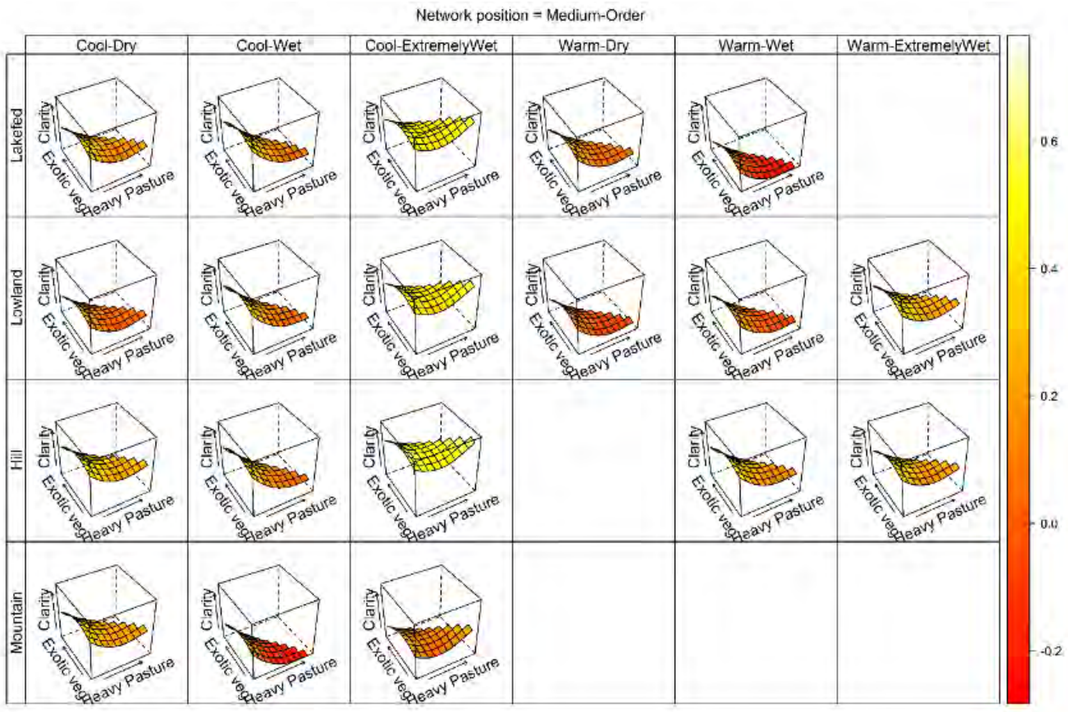


Figure EE-2: Relative effects of % cover of heavy pasture and % cover of exotic vegetation on visual clarity in different REC climate/topography classes. This plot represents the relationships for medium Strahler stream order rivers. Similar patterns were observed for other stream sizes.

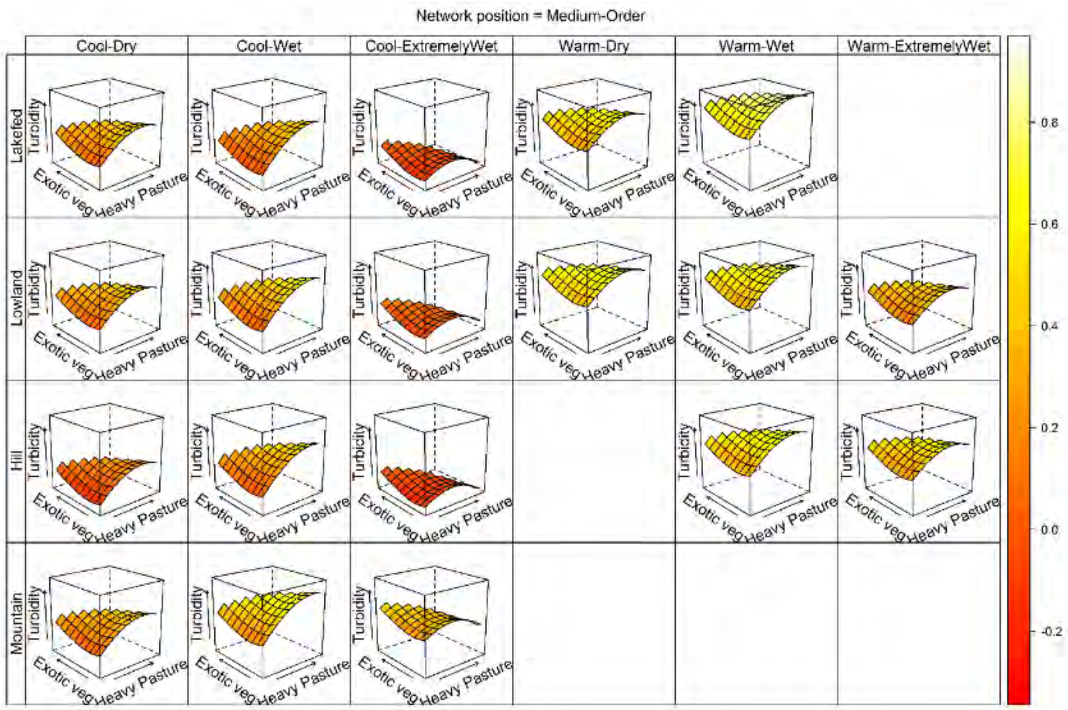


Figure EE-3: Relative effects of % cover of heavy pasture and % cover of exotic vegetation on turbidity in different REC climate/topography classes. This plot represents the relationships for medium Strahler stream order rivers. Similar patterns were observed for other stream sizes.

Predicted fish probability of capture as a function of sediment ESVs

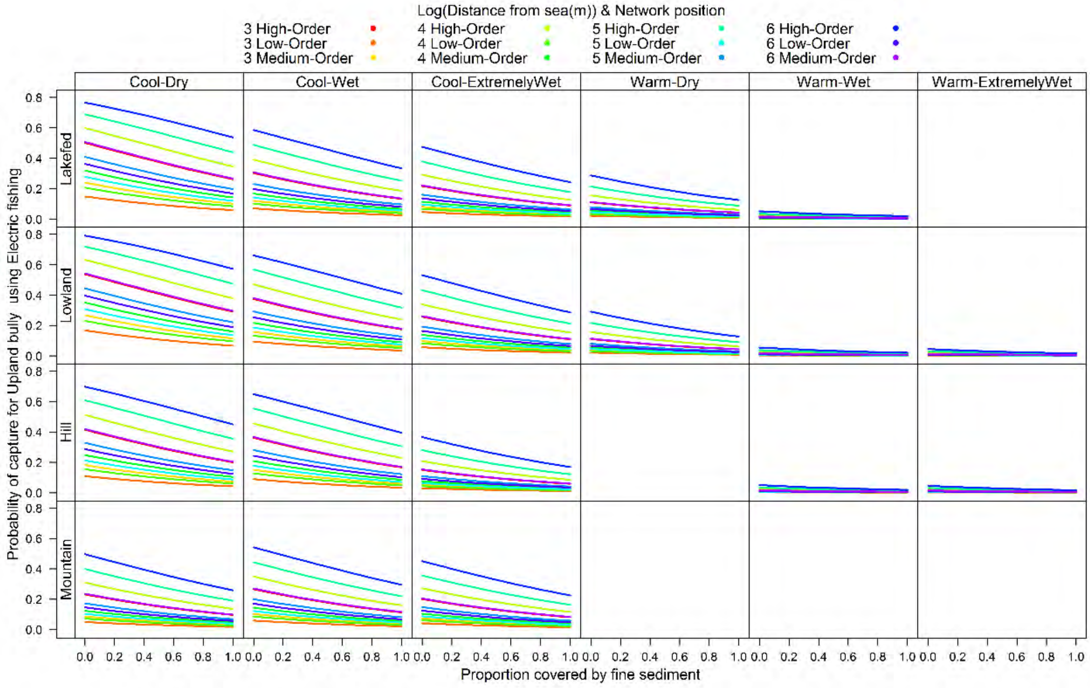


Figure EE-4: Predicted fish probability of capture (FPC) by electric fishing for upland bully relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

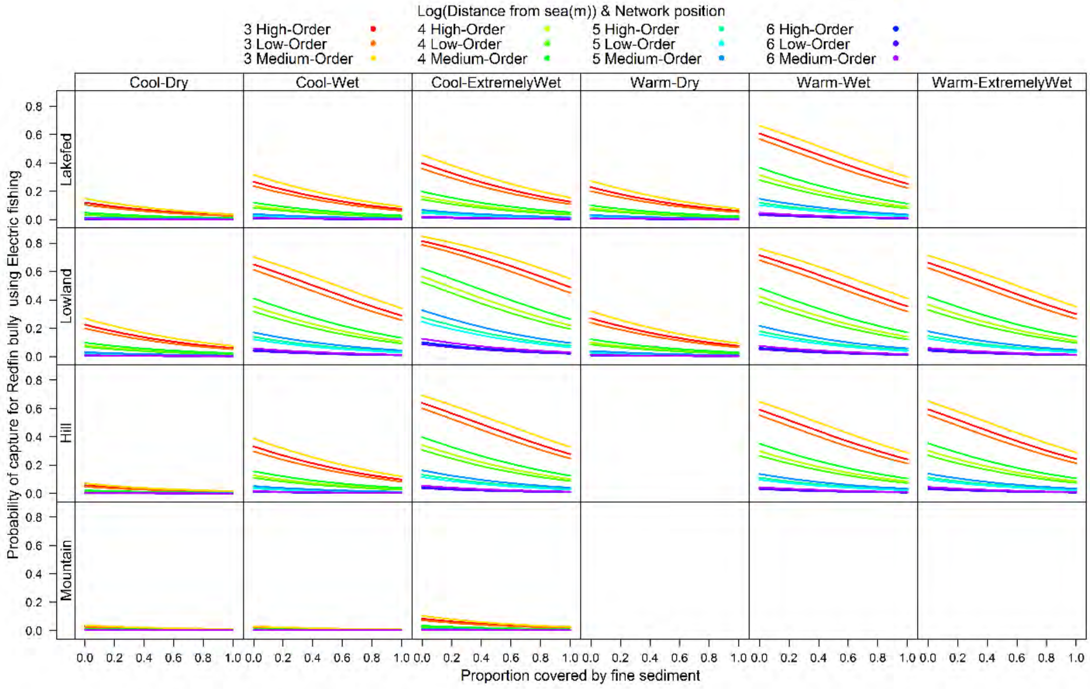


Figure EE-5: Predicted fish probability of capture (FPC) by electric fishing for redfin bully relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

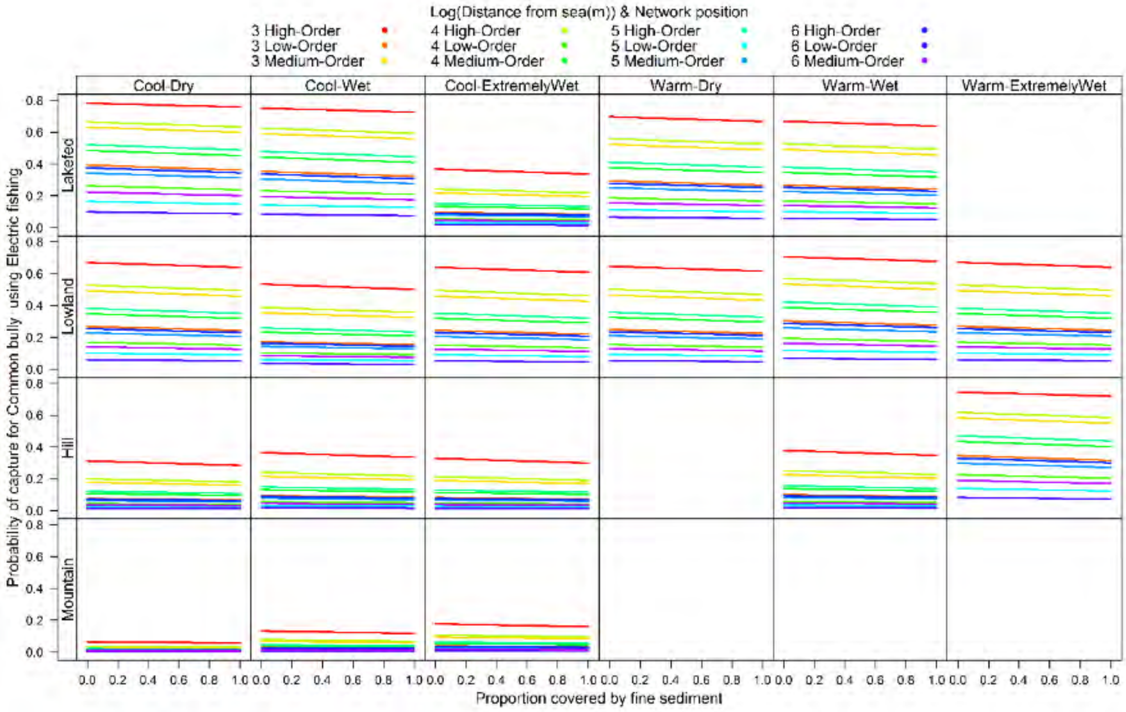


Figure EE-6: Predicted fish probability of capture (FPC) by electric fishing for common bully relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

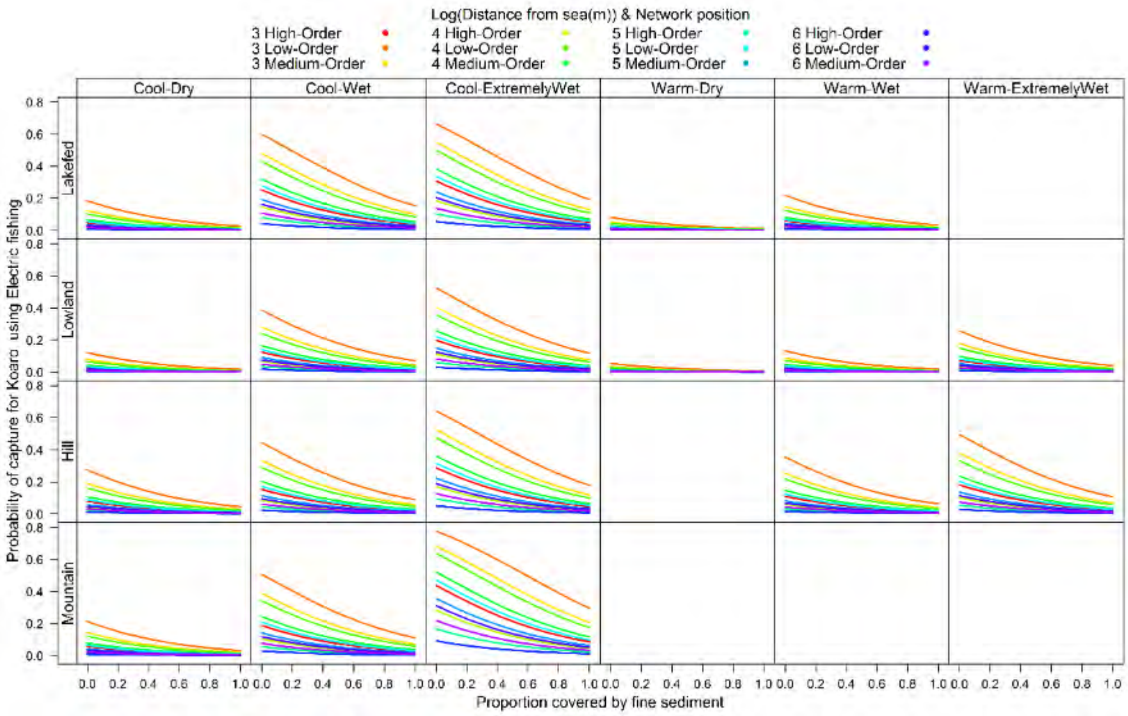


Figure EE-7: Predicted fish probability of capture (FPC) by electric fishing for kōaro relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

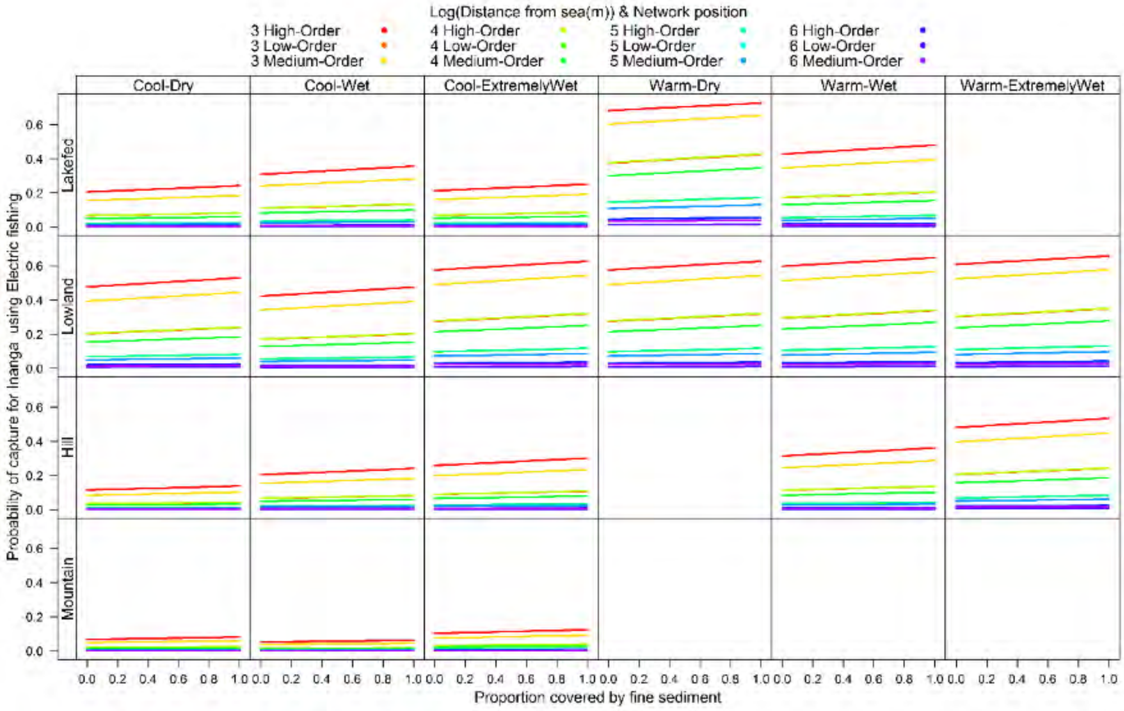


Figure EE-8: Predicted fish probability of capture (FPC) by electric fishing for inānga relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

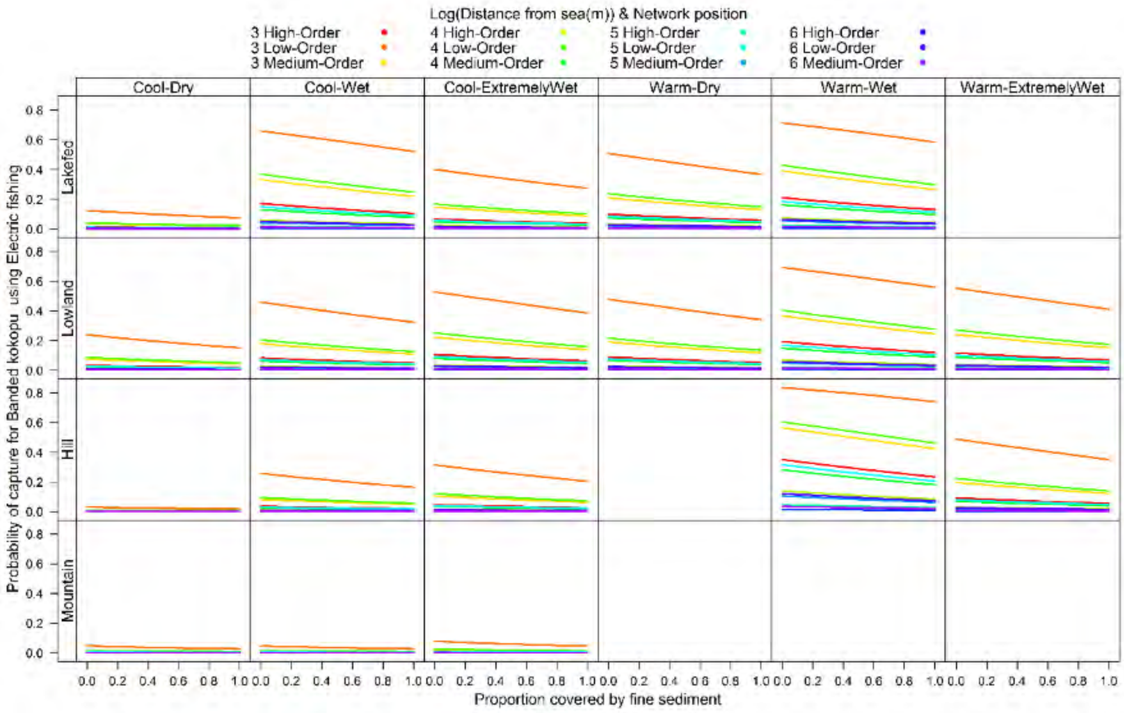


Figure EE-9: Predicted fish probability of capture (FPC) by electric fishing for banded kōkopu relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

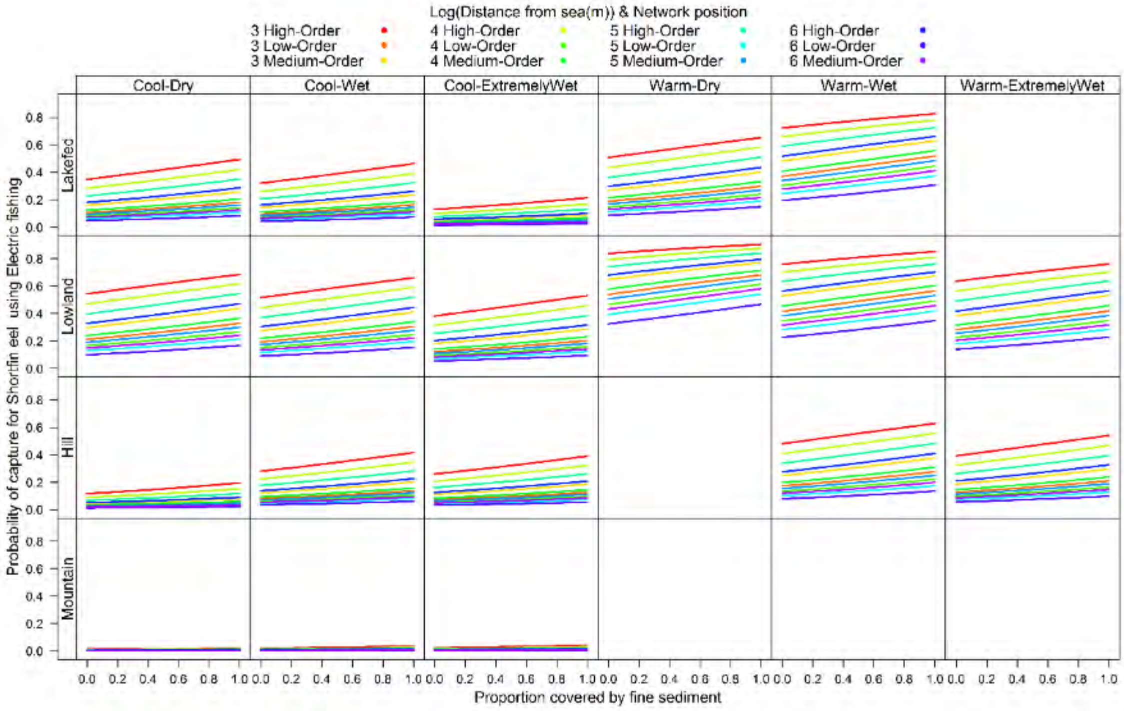


Figure EE-10: Predicted fish probability of capture (FPC) by electric fishing for shortfin eel relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

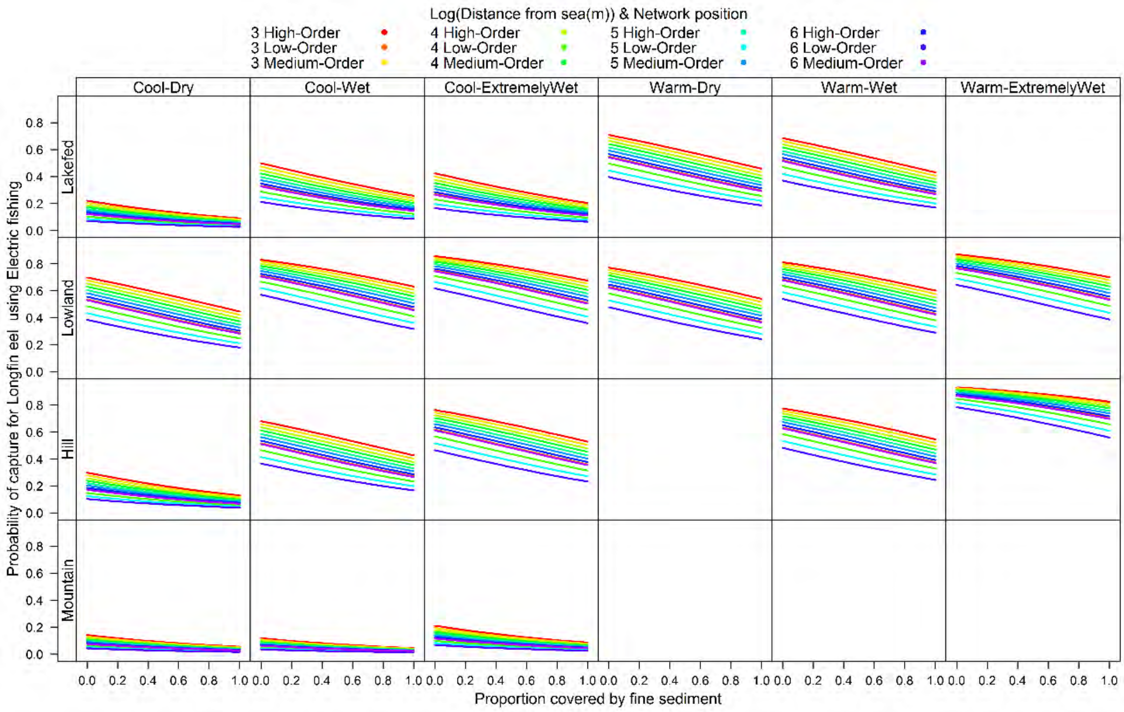


Figure EE-11: Predicted fish probability of capture (FPC) by electric fishing for longfin eel relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

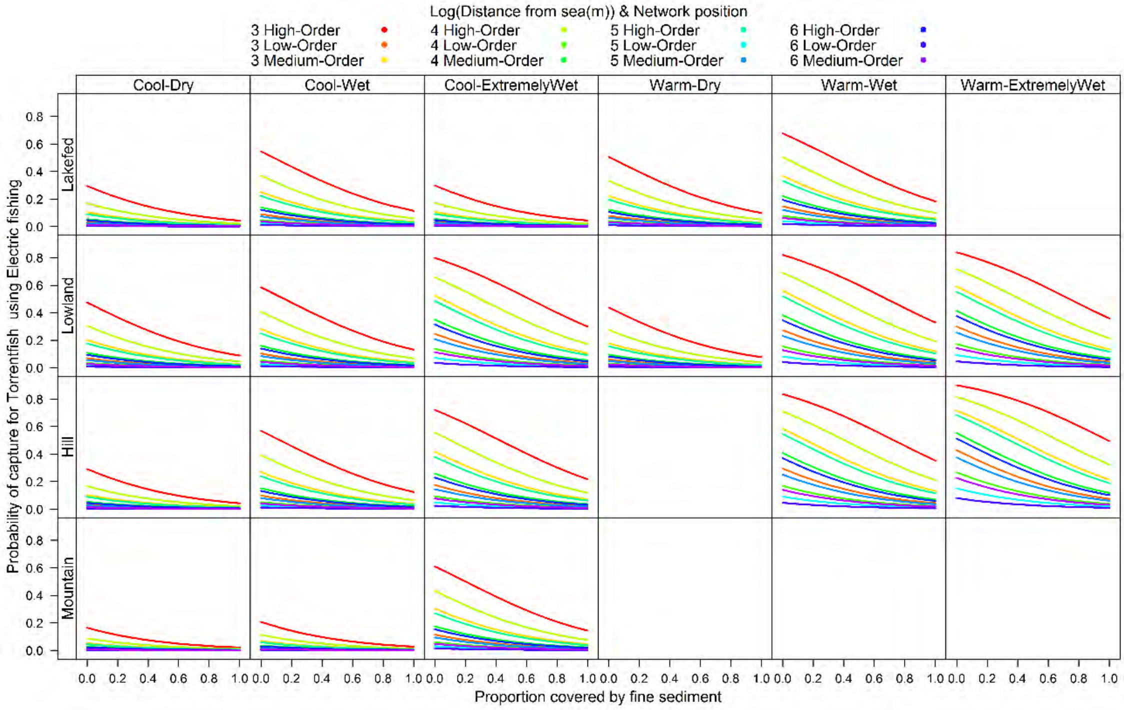


Figure EE-12: Predicted fish probability of capture (FPC) by electric fishing for torrentfish relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

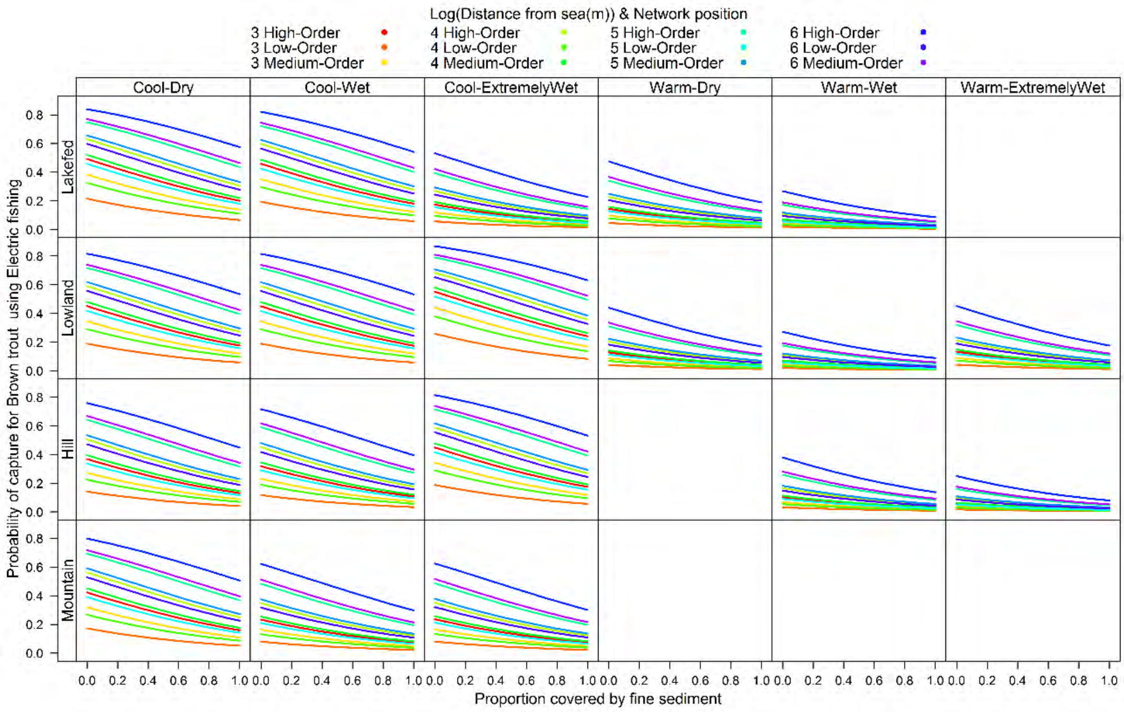


Figure EE-13: Predicted fish probability of capture (FPC) by electric fishing for brown trout relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

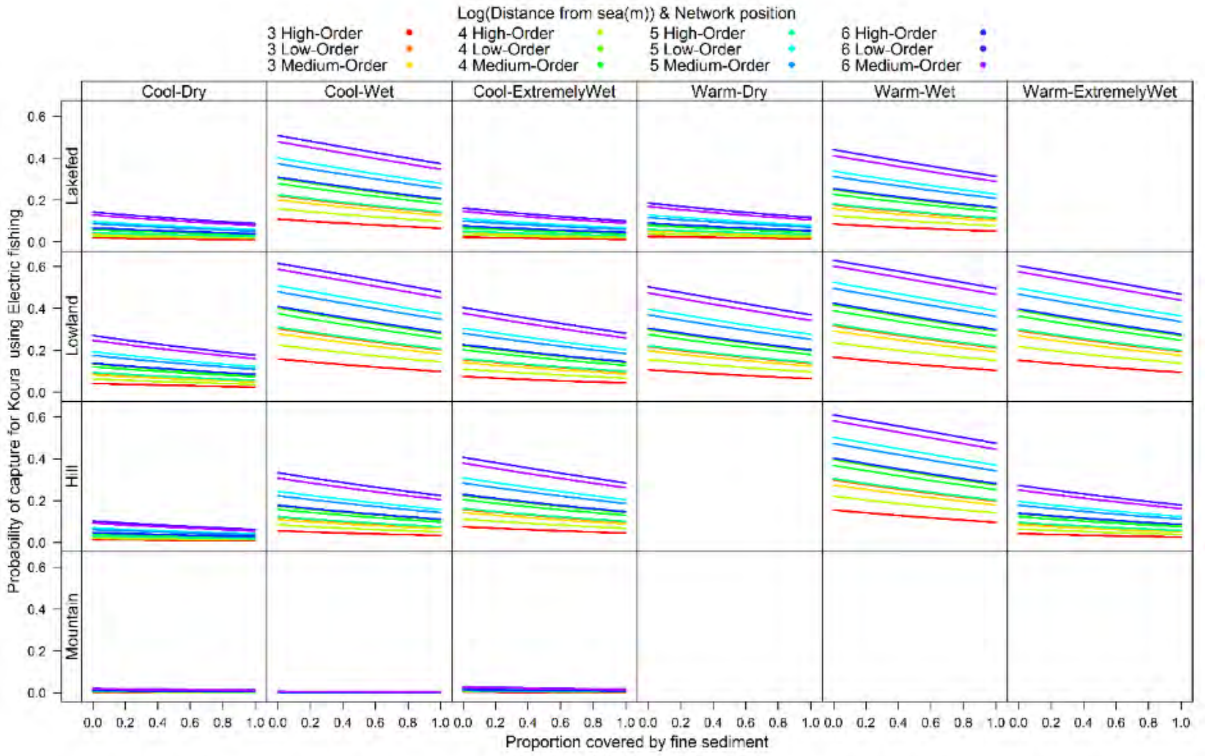


Figure EE-14: Predicted fish probability of capture (FPC) by electric fishing for kōura relative to proportional cover of total fines. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

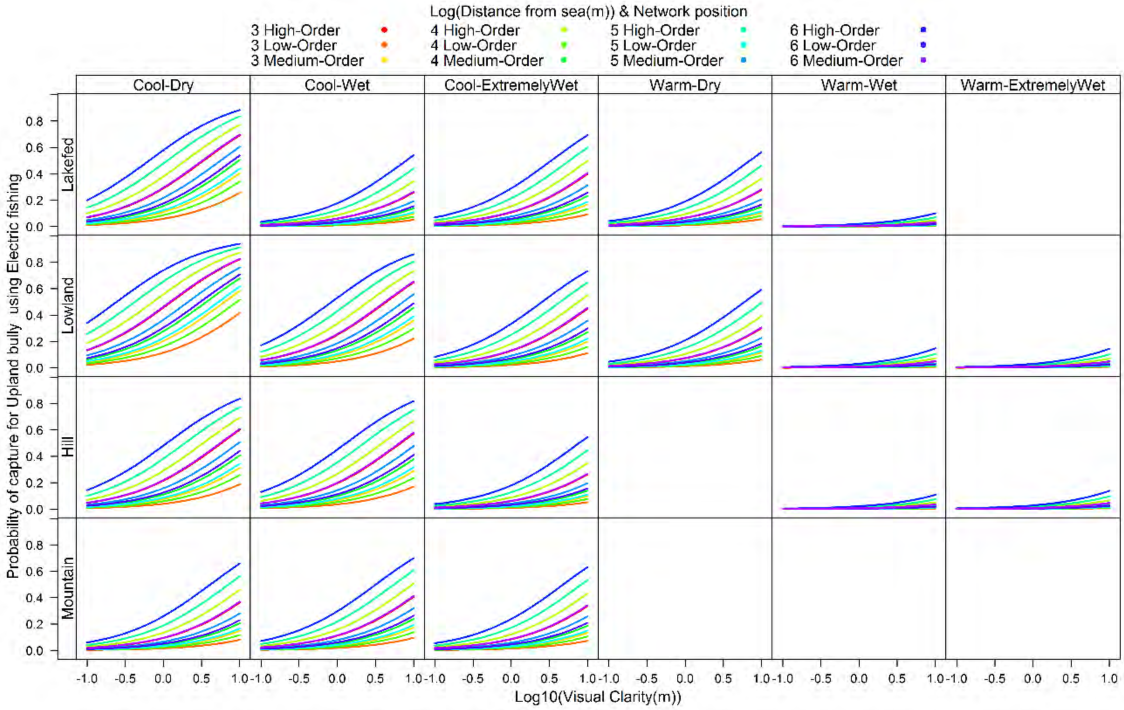


Figure EE-15: Predicted fish probability of capture (FPC) by electric fishing for upland bully relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

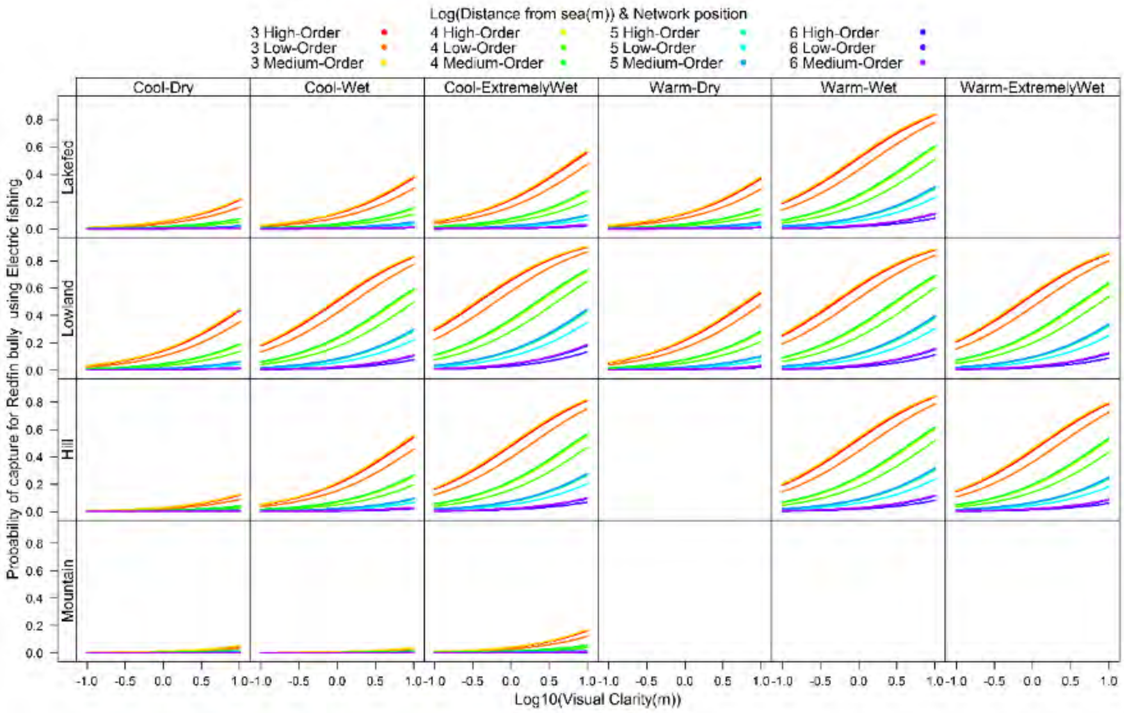


Figure EE-16: Predicted fish probability of capture (FPC) by electric fishing for redfin bully relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

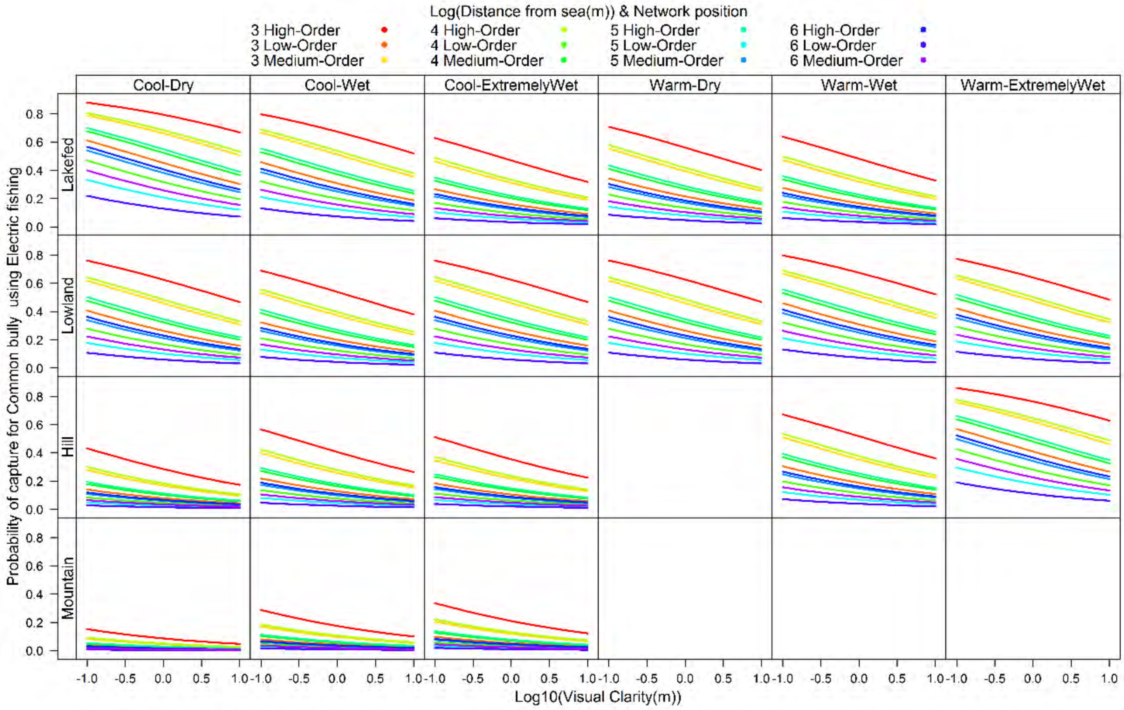


Figure EE-17: Predicted fish probability of capture (FPC) by electric fishing for common bully relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

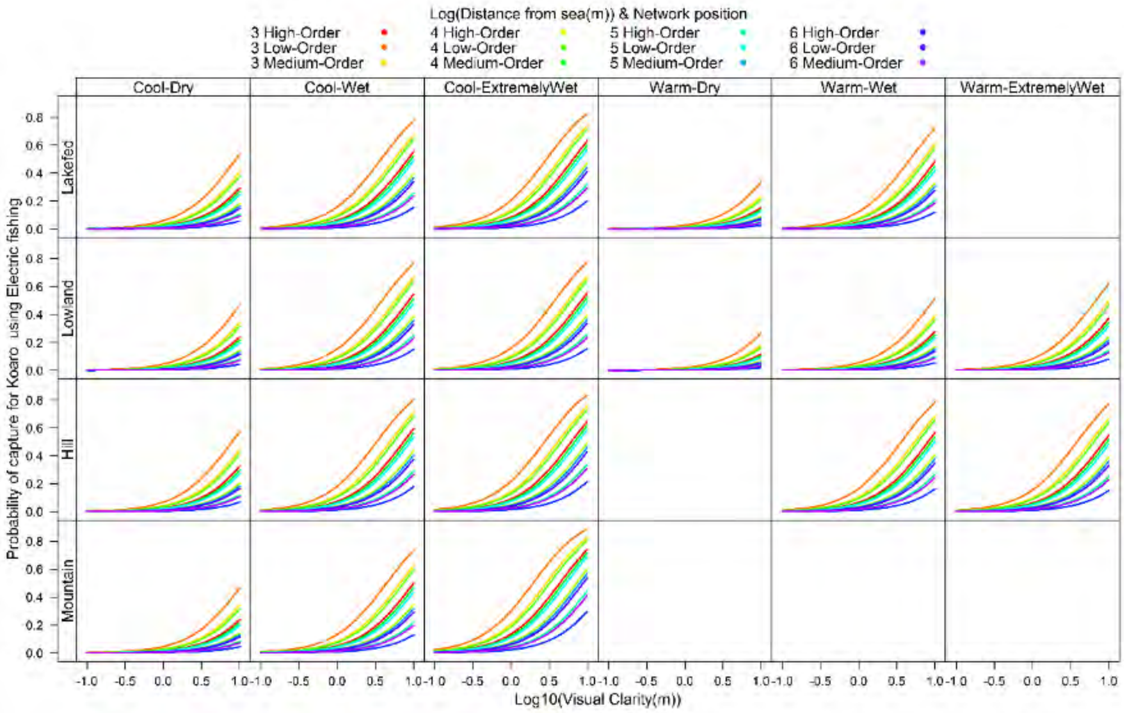


Figure EE-18: Predicted fish probability of capture (FPC) by electric fishing for kōaro relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

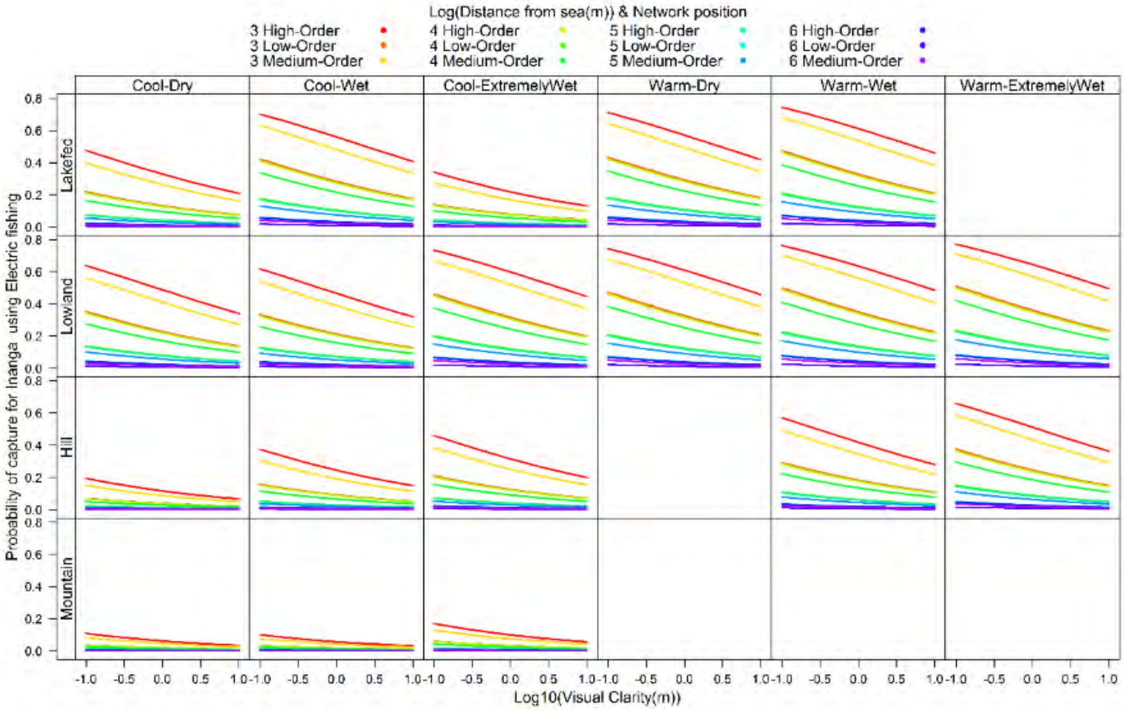


Figure EE-19: Predicted fish probability of capture (FPC) by electric fishing for inānga relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

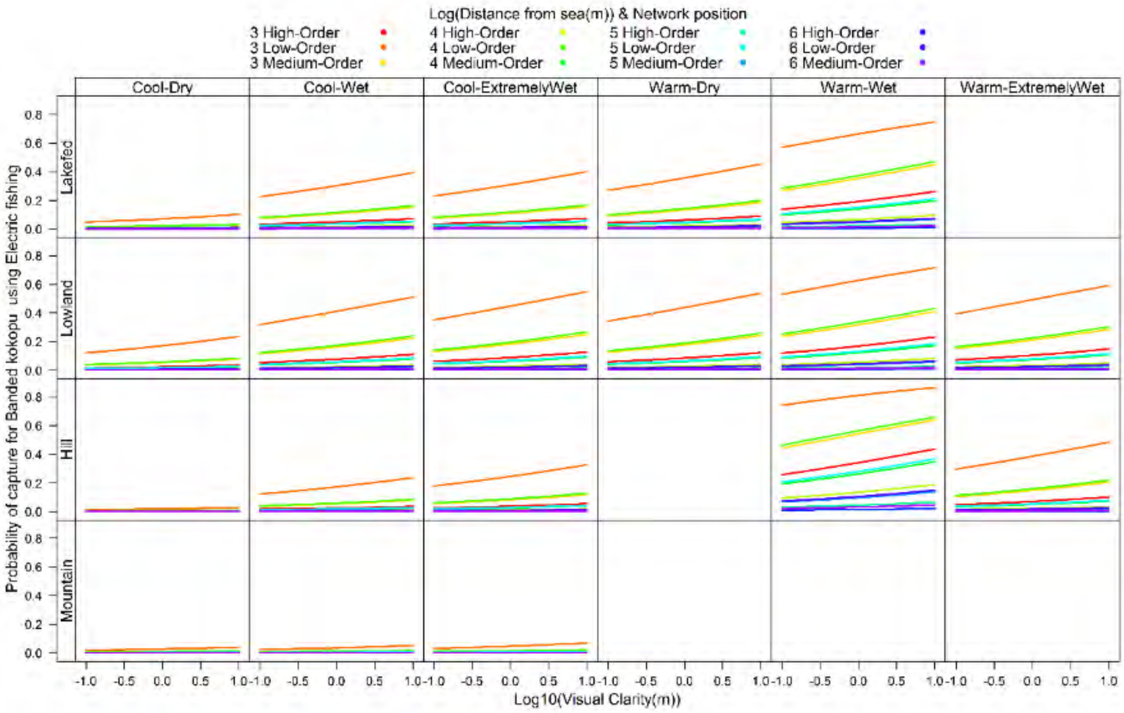


Figure EE-20: Predicted fish probability of capture (FPC) by electric fishing for banded kōkopu relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

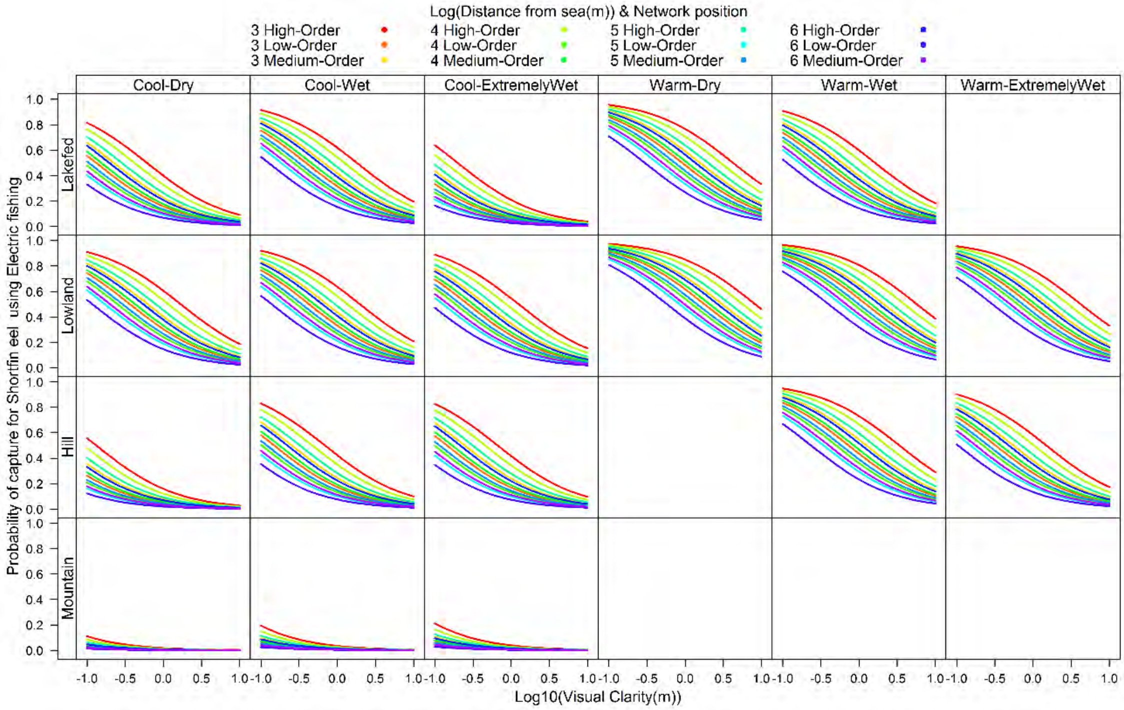


Figure EE-21: Predicted fish probability of capture (FPC) by electric fishing for shortfin eel relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

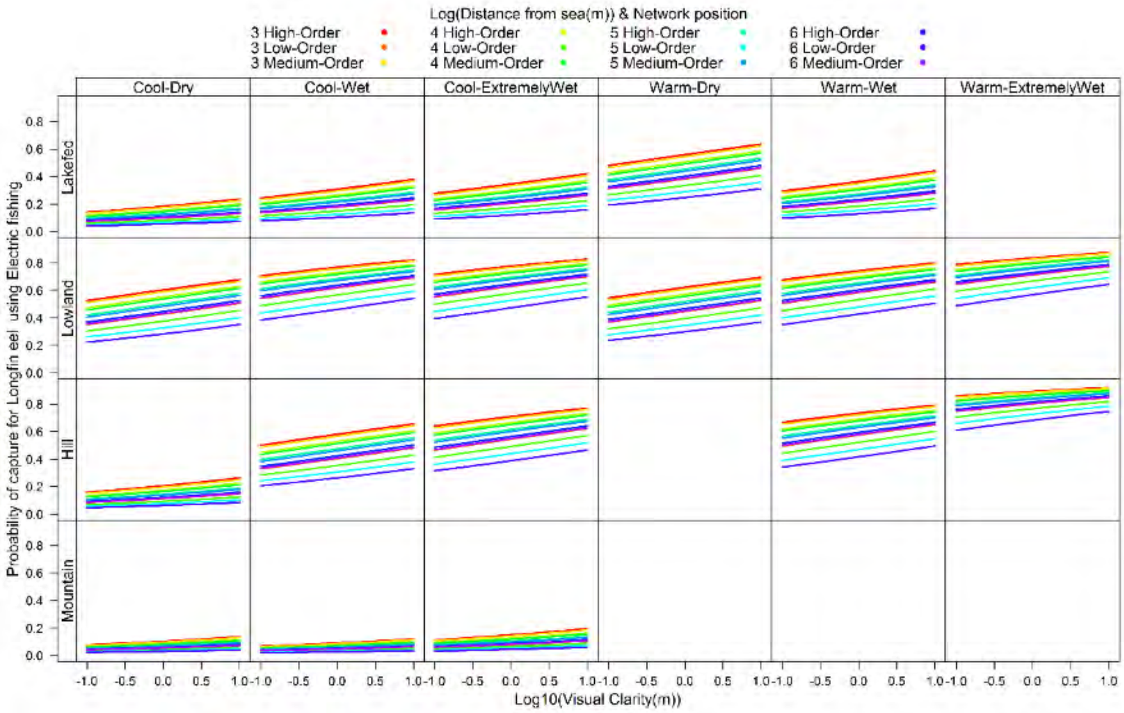


Figure EE-22: Predicted fish probability of capture (FPC) by electric fishing for longfin eel relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

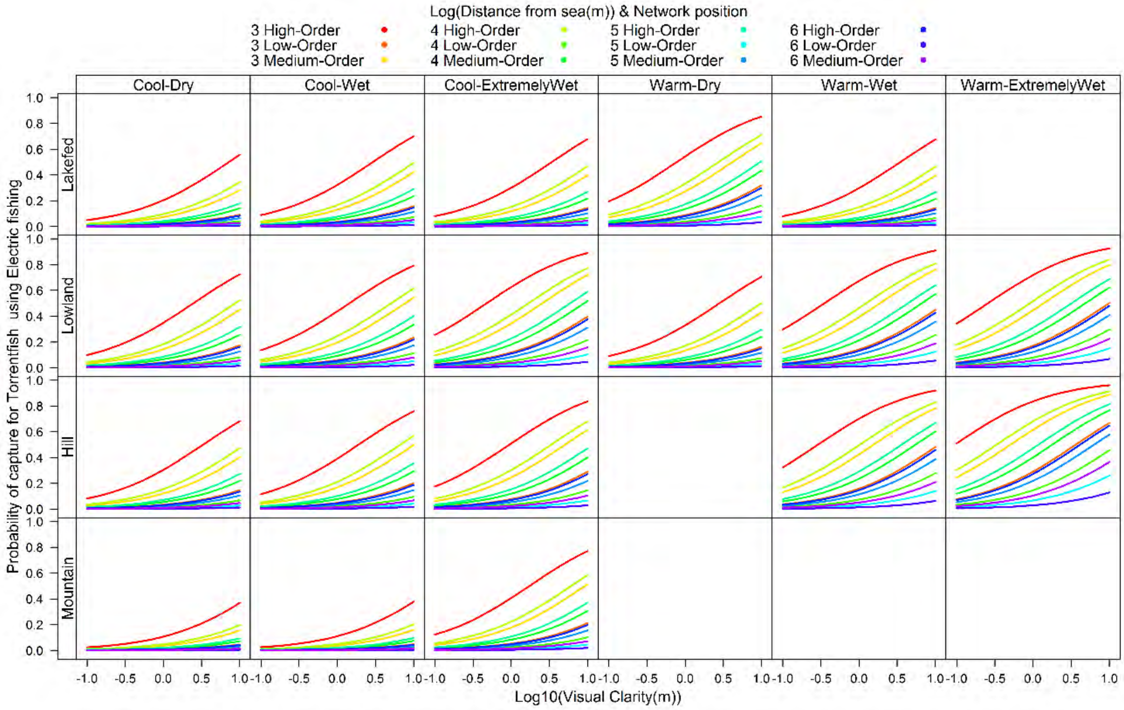


Figure EE-23: Predicted fish probability of capture (FPC) by electric fishing for torrentfish relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

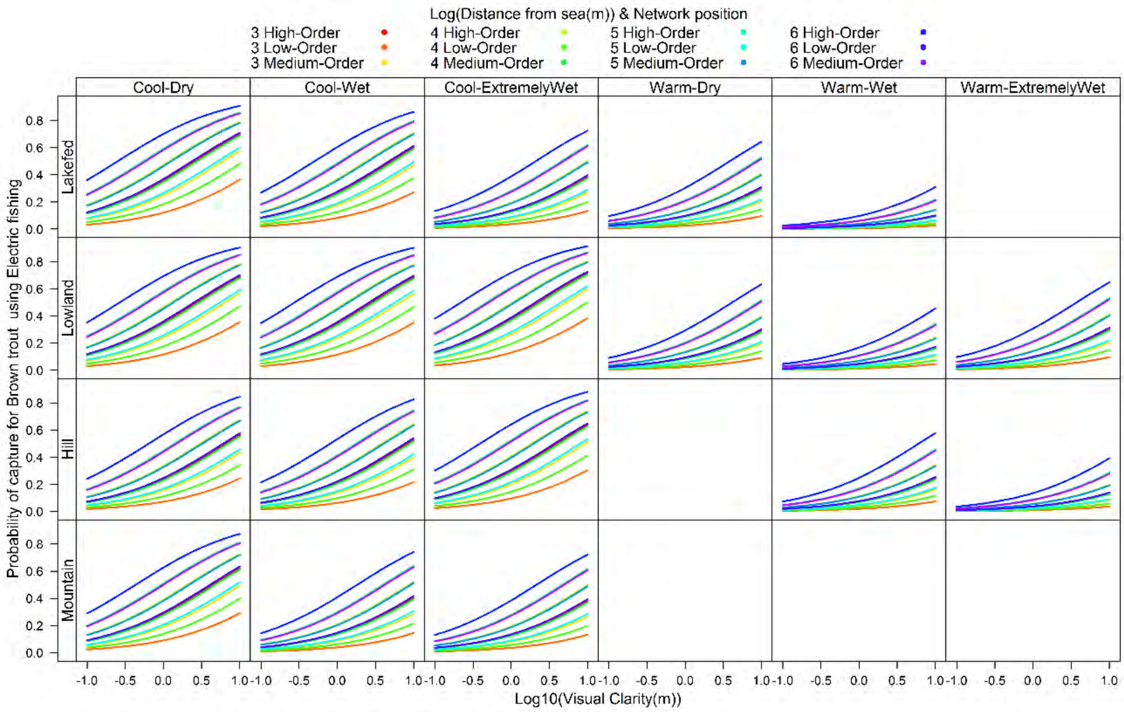


Figure EE-24: Predicted fish probability of capture (FPC) by electric fishing for brown trout relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

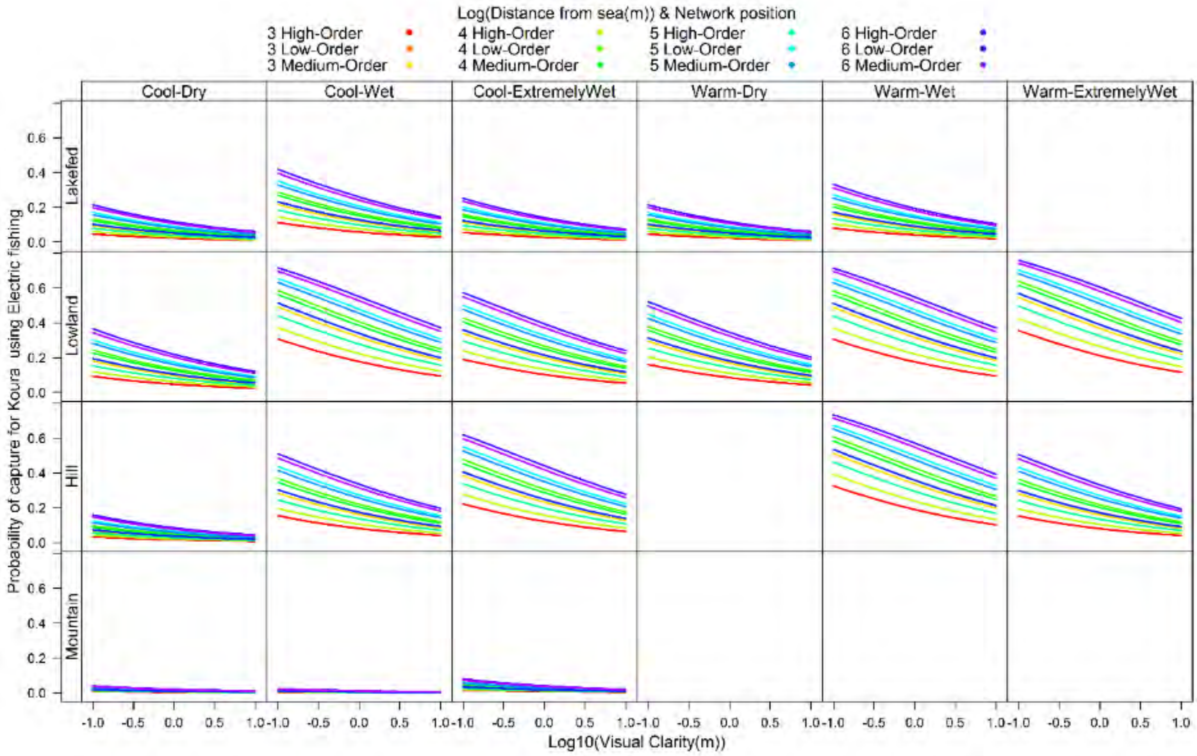


Figure EE-25: Predicted fish probability of capture (FPC) by electric fishing for kōura relative to visual clarity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

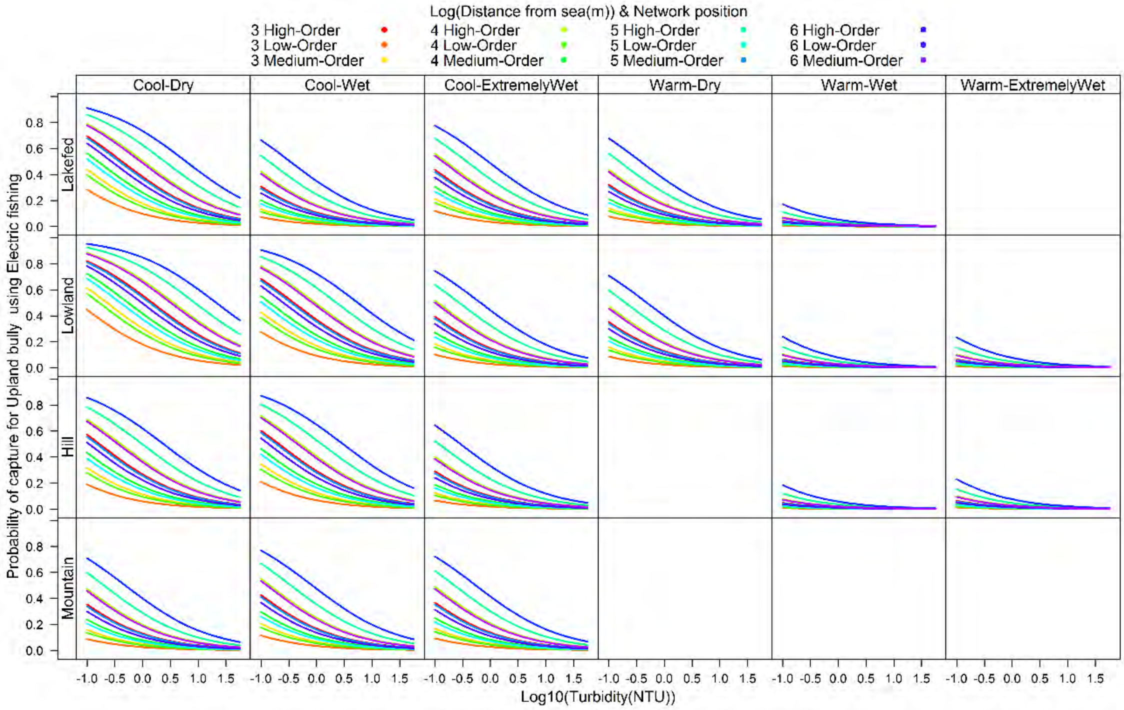


Figure EE-26: Predicted fish probability of capture (FPC) by electric fishing for upland bully relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

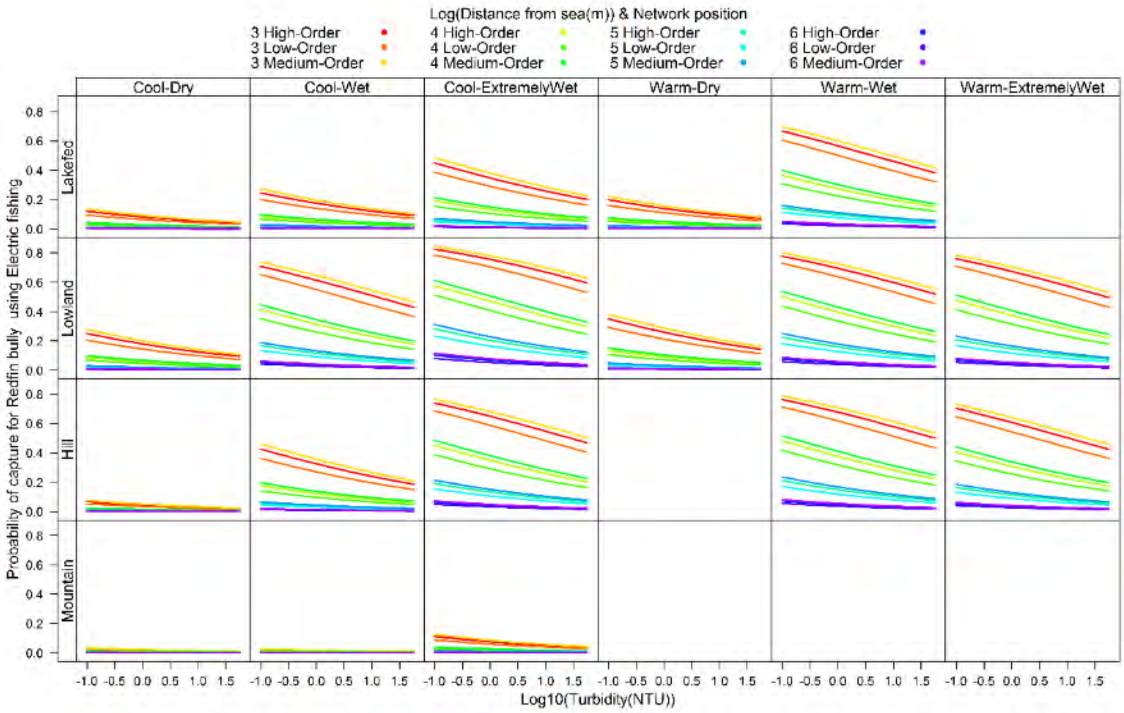


Figure EE-27: Predicted fish probability of capture (FPC) by electric fishing for redfin bully relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

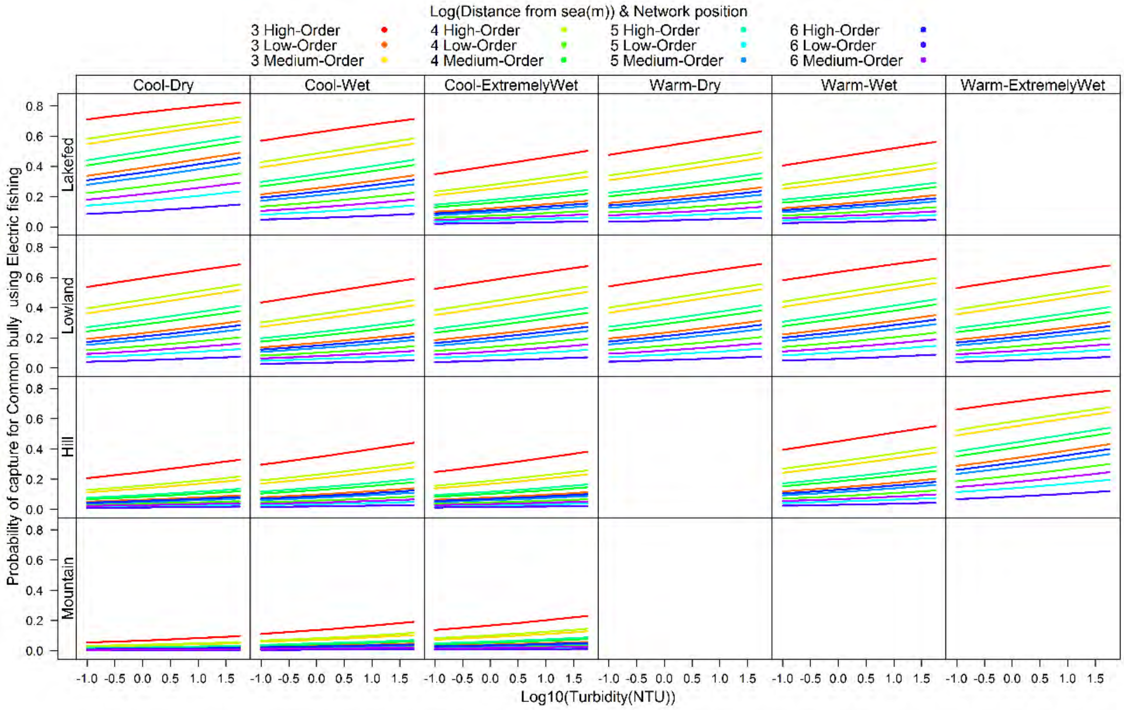


Figure EE-28: Predicted fish probability of capture (FPC) by electric fishing for common bully relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

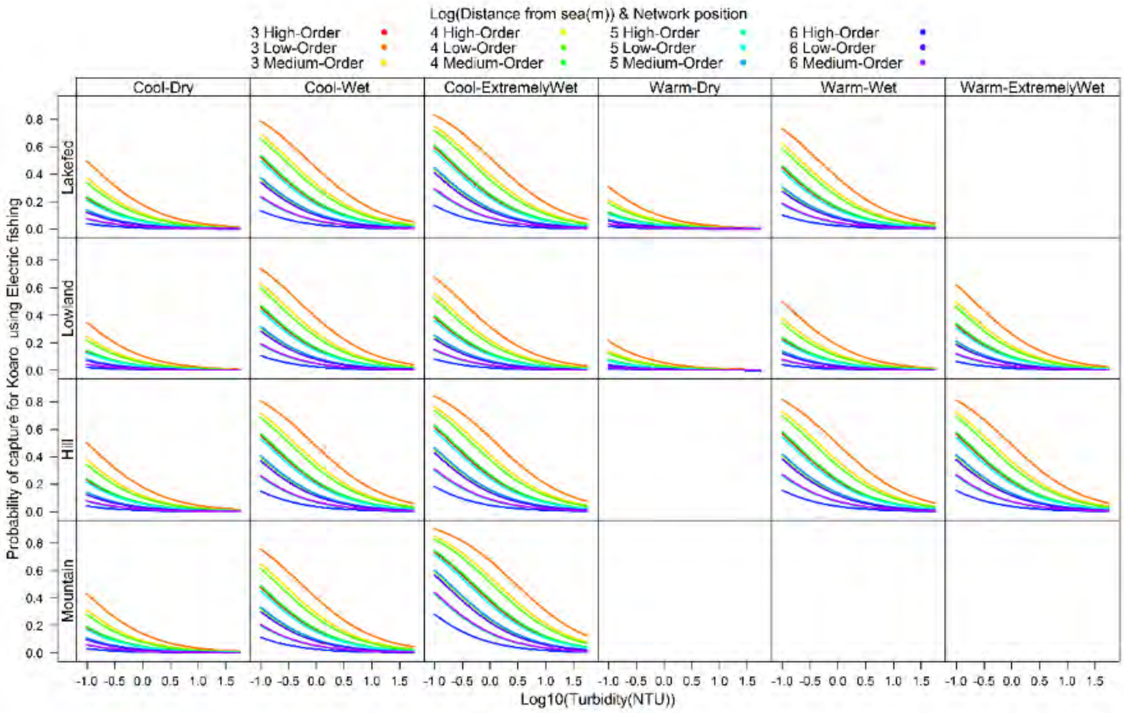


Figure EE-29: Predicted fish probability of capture (FPC) by electric fishing for kōaro relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

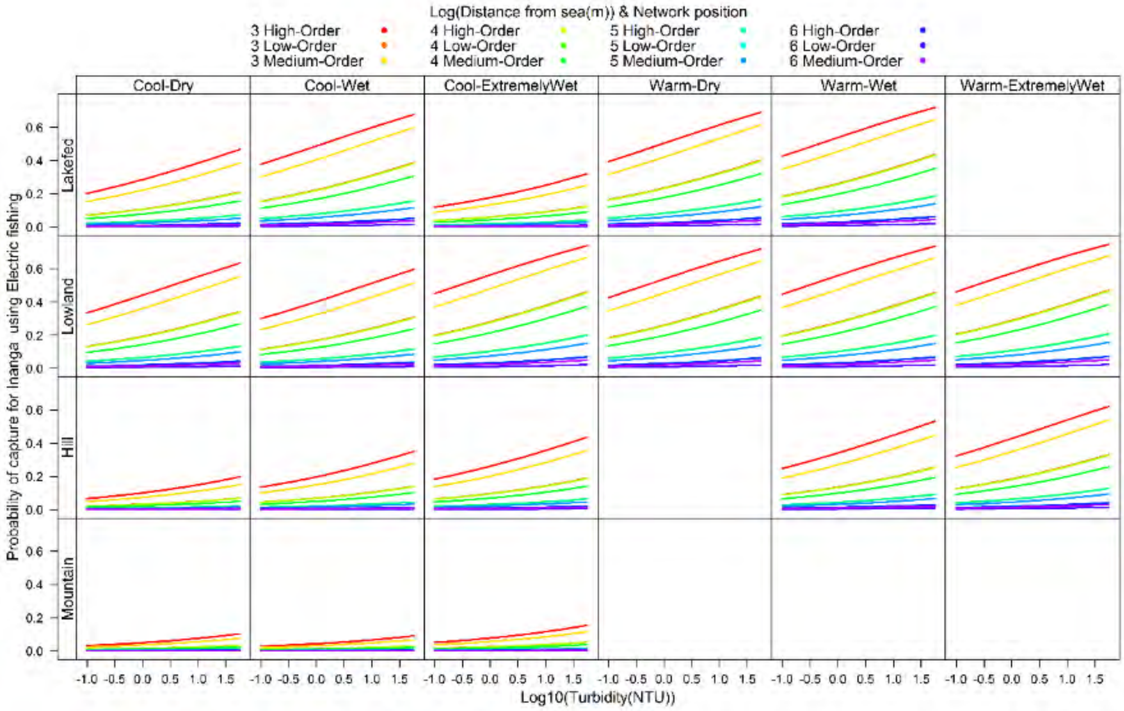


Figure EE-30: Predicted fish probability of capture (FPC) by electric fishing for ināanga relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

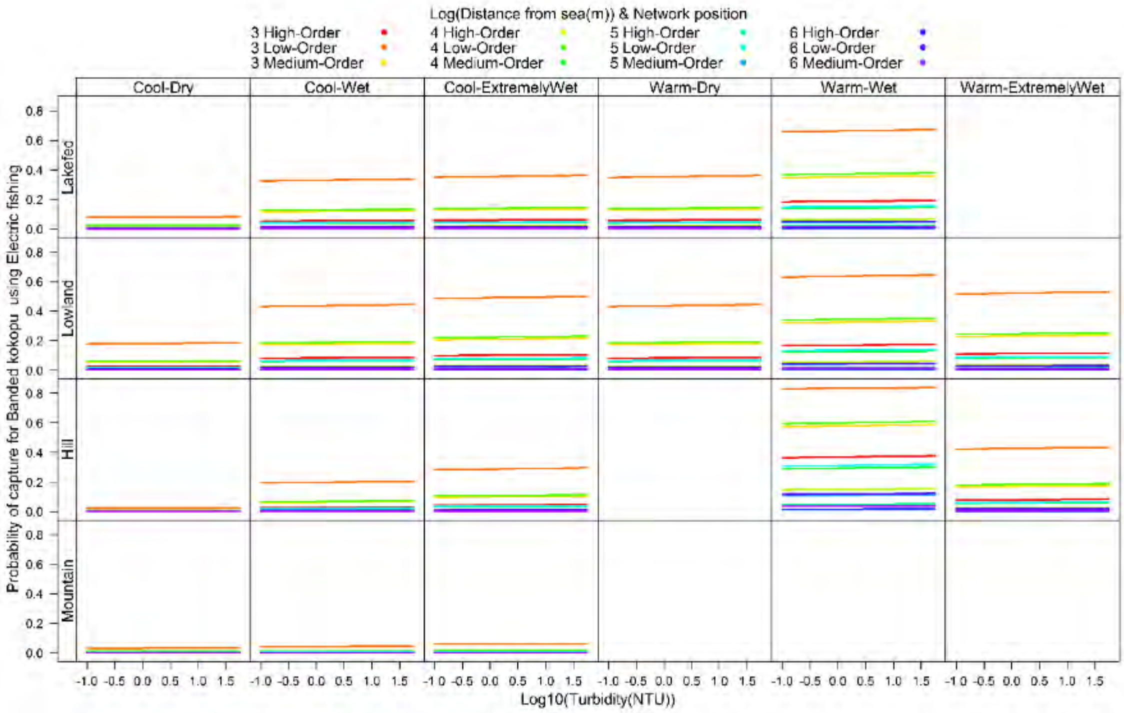


Figure EE-31: Predicted fish probability of capture (FPC) by electric fishing for banded kōkopu relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

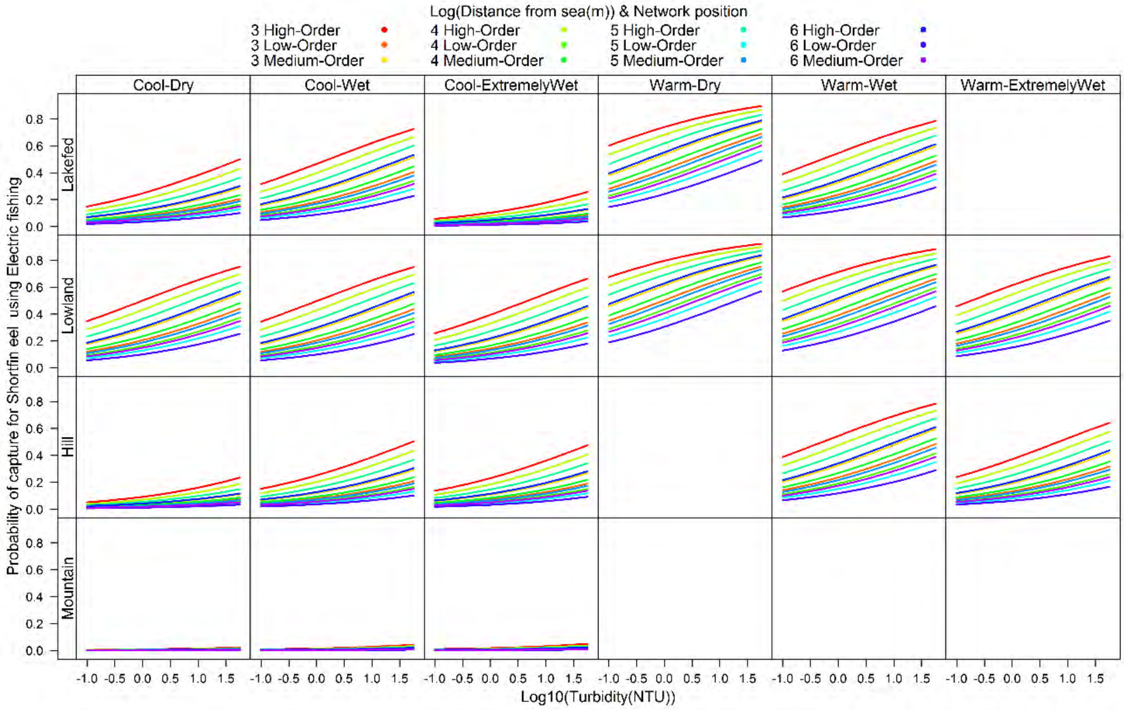


Figure EE-32: Predicted fish probability of capture (FPC) by electric fishing for shortfin eel relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

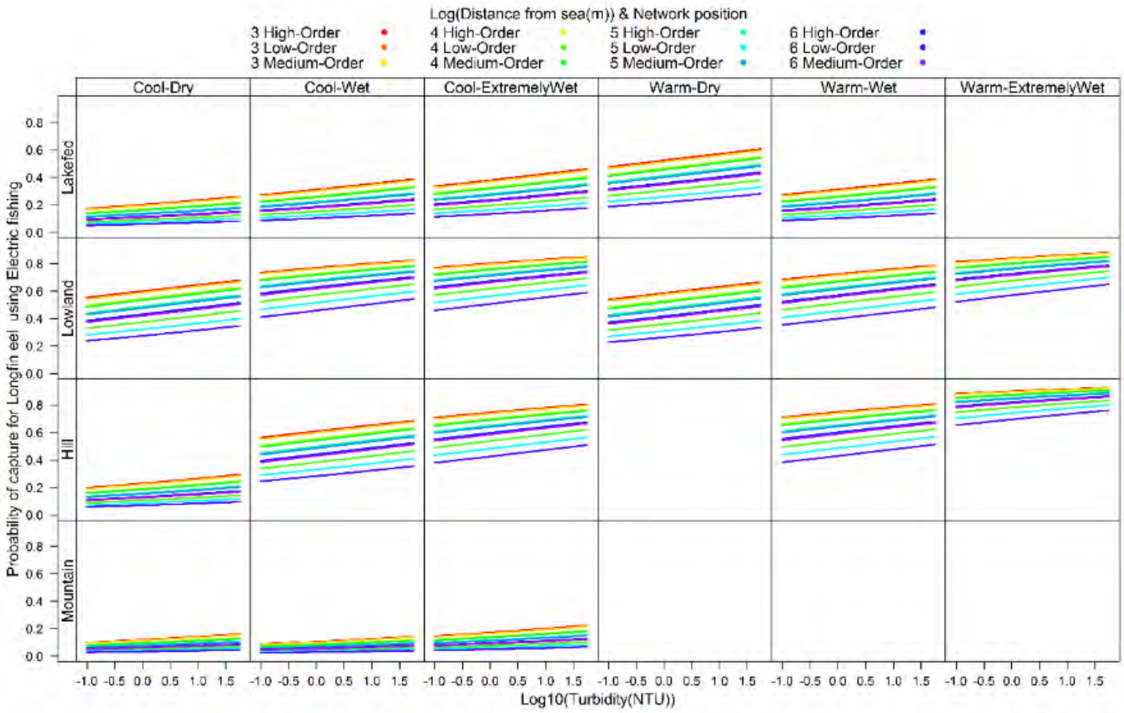


Figure EE-33: Predicted fish probability of capture (FPC) by electric fishing for longfin eel relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

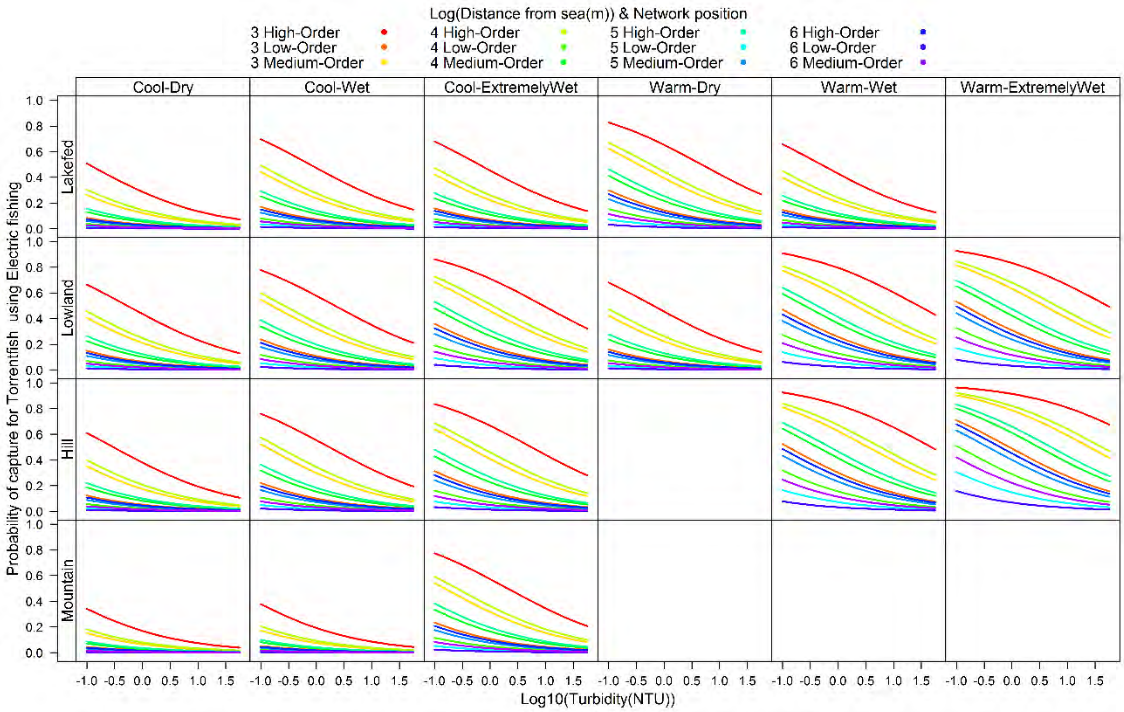


Figure EE-34: Predicted fish probability of capture (FPC) by electric fishing for torrentfish relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

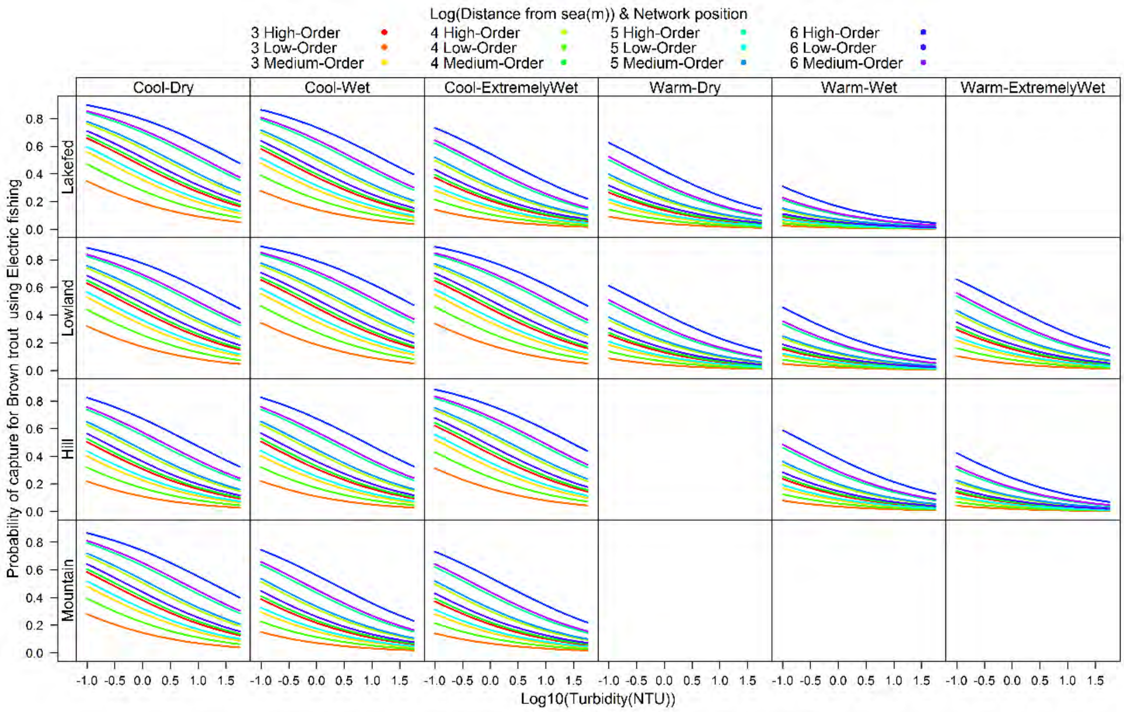


Figure EE-35: Predicted fish probability of capture (FPC) by electric fishing for brown trout relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

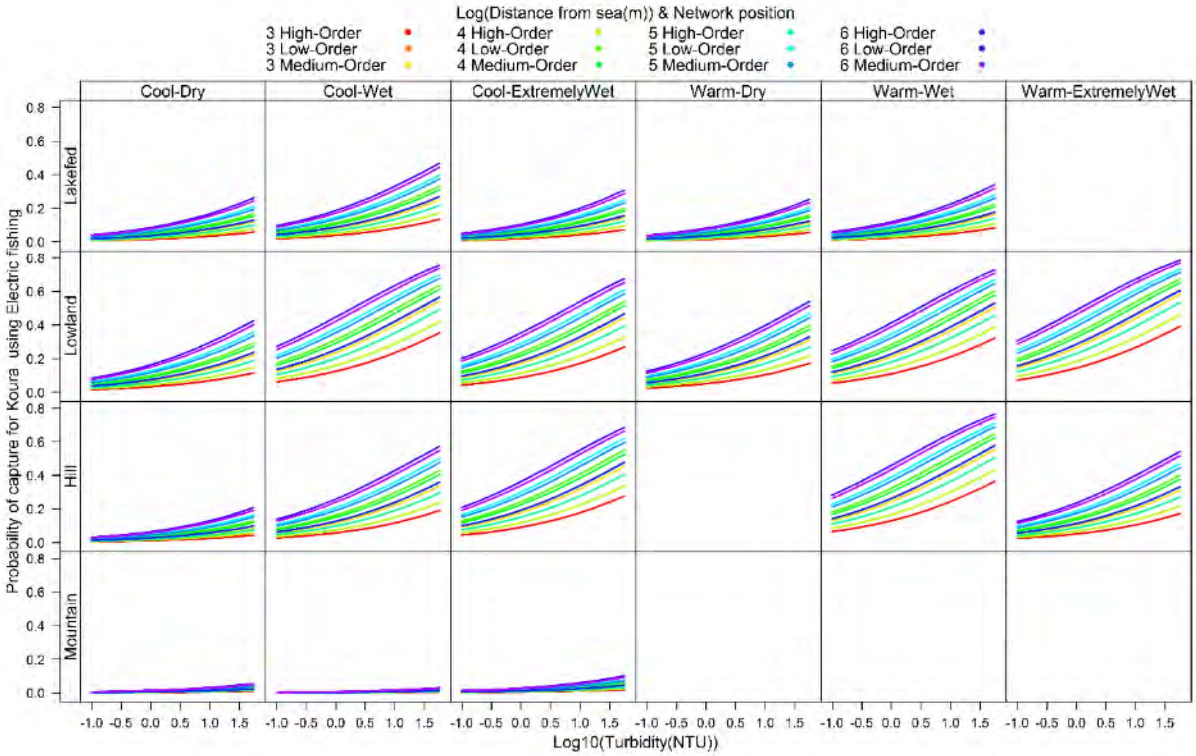


Figure EE-36: Predicted fish probability of capture (FPC) by electric fishing for kōura relative to turbidity. Predictions are made for each fish setting (climate/topography/network setting (stream order)/distance to sea). Distance to sea (Log10 transformed) is represented by four discrete classes (3, 4, 5, 6) at 1 km, 10 km, 100 km and 1000 km from the sea respectively.

Potential sediment ESV bands

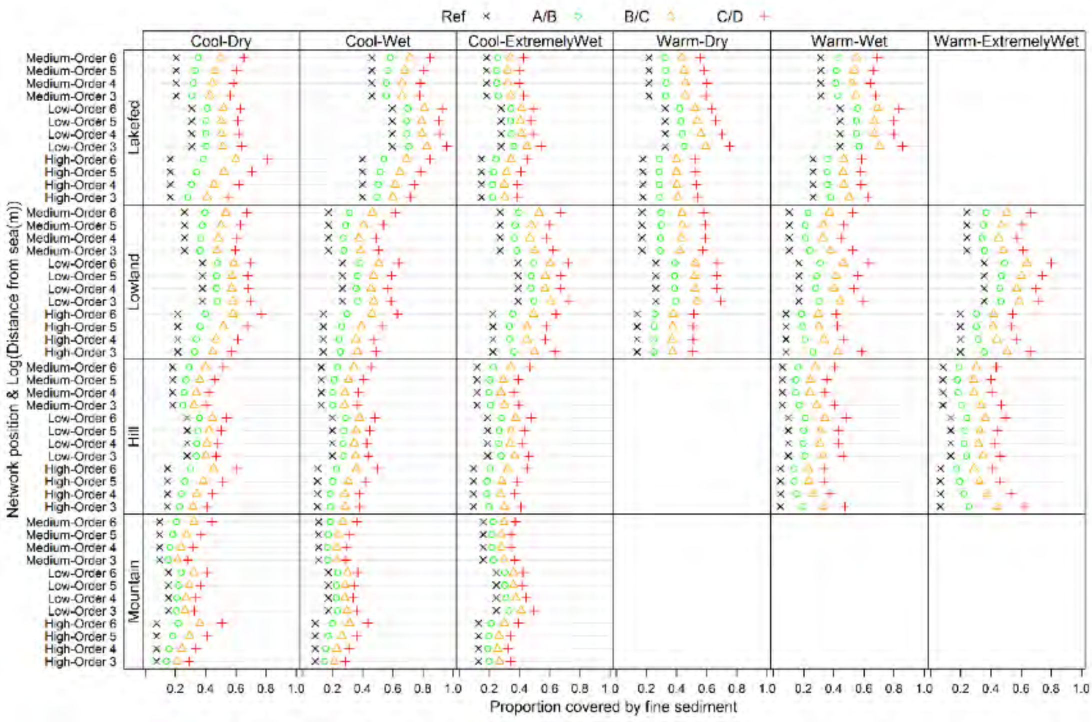


Figure EE-37: Potential attribute band thresholds for proportional cover of total fines across landscape settings. Values are provided for each climate/topography/network position/distance inland setting. Ref = predicted average reference ESV state, AB = A/B band threshold, BC = B/C band threshold, CD = C/D band bottom-line.

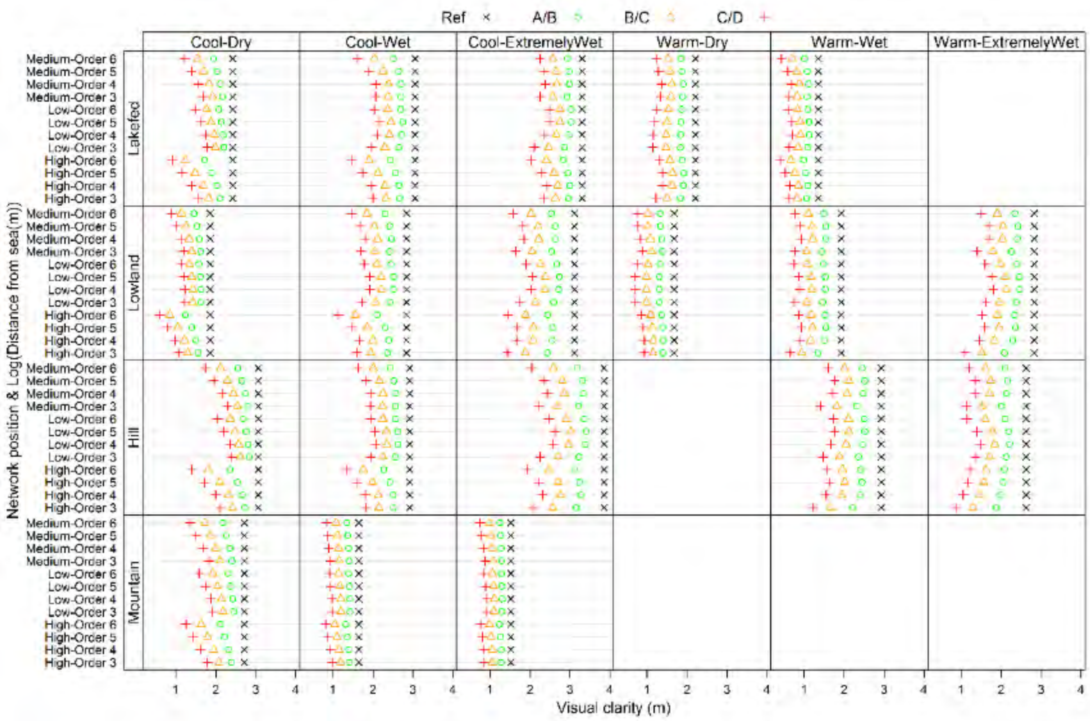


Figure EE-38: Potential attribute band thresholds for proportional visual clarity across landscape settings. Values are provided for each climate/topography/network position/distance inland setting. Ref = predicted average reference ESV state, AB = A/B band threshold, BC = B/C band threshold, CD = C/D band bottom-line.

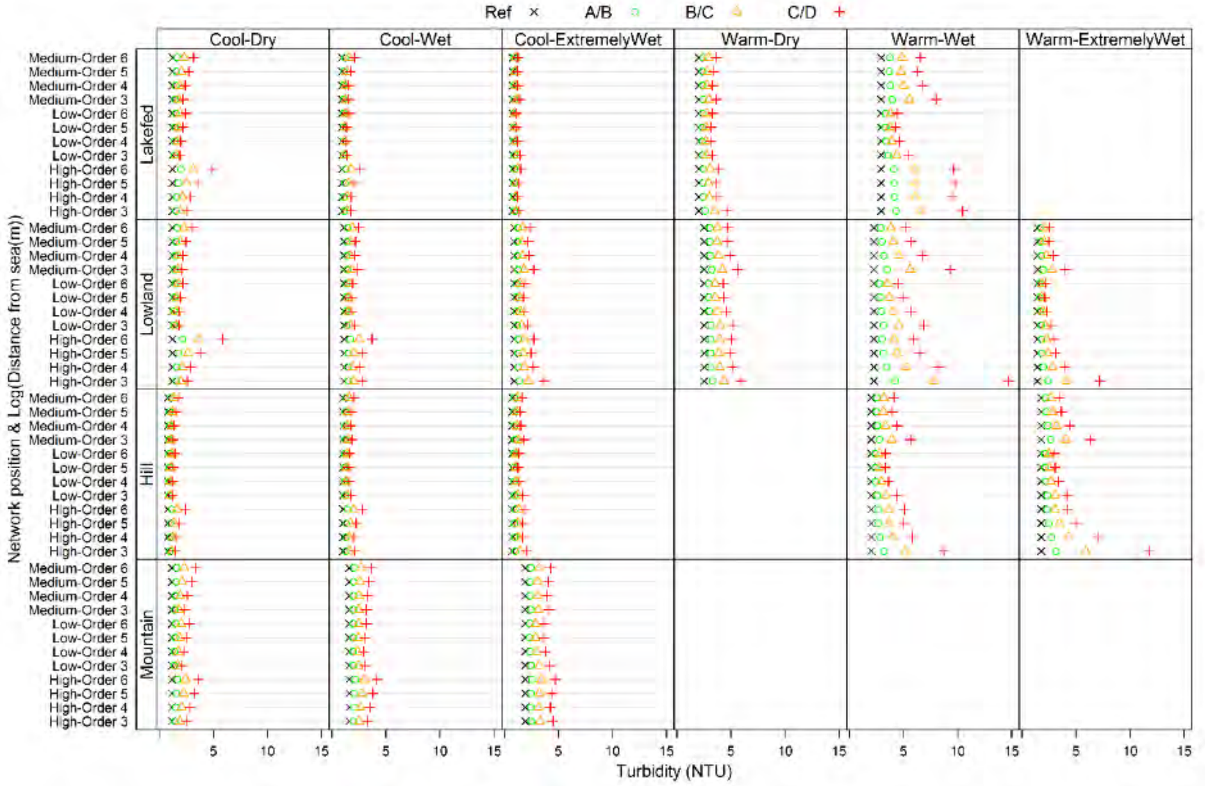


Figure EE-39: Potential attribute band thresholds for proportional turbidity across landscape settings. Values are provided for each climate/topography/network position/distance inland setting. Ref = predicted average reference ESV state, AB = A/B band threshold, BC = B/C band threshold, CD = C/D band bottom-line.

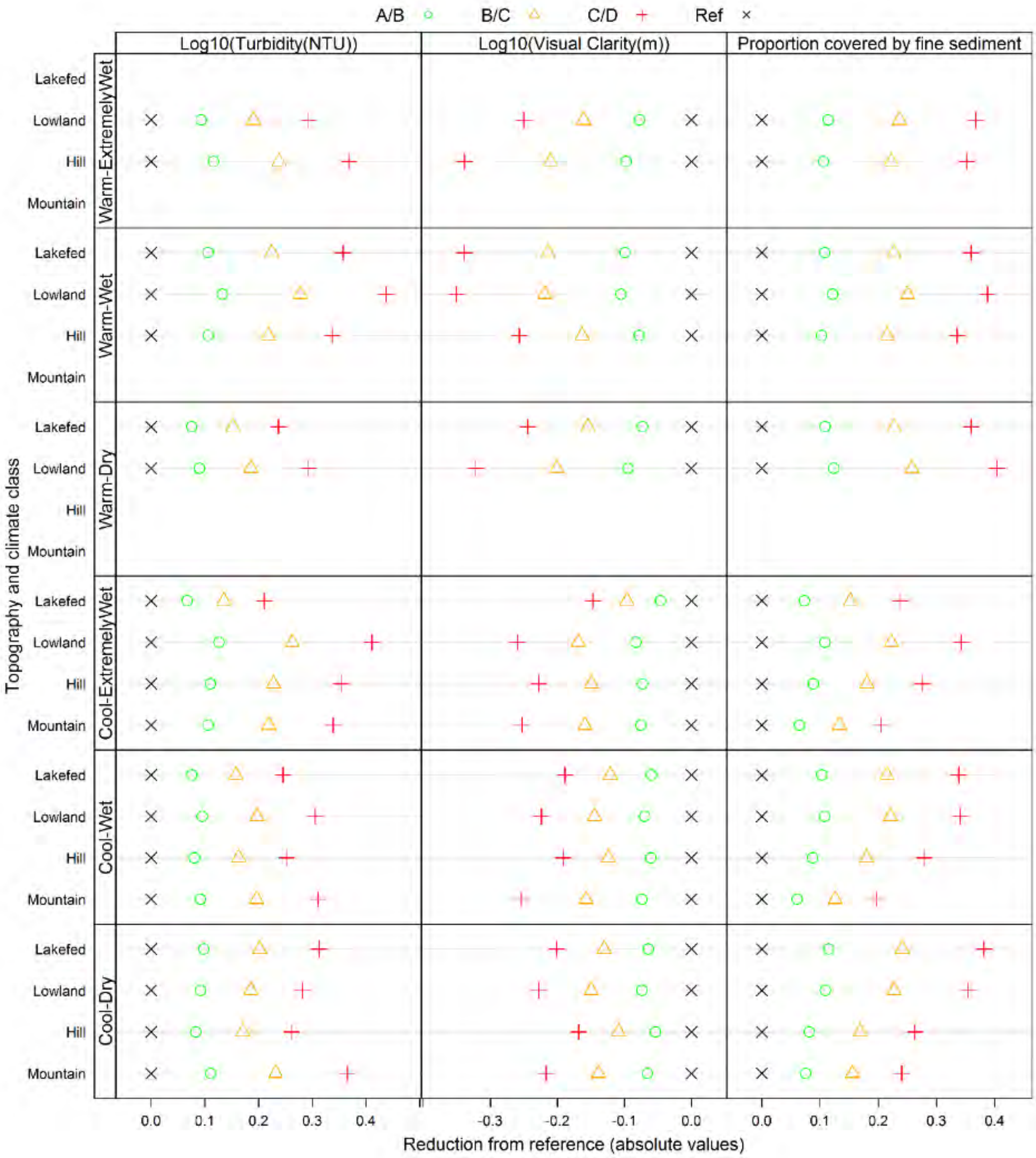


Figure EE-40: Potential attribute band thresholds expressed as reduction from reference ESV state. While the absolute values of the different sediment ESV thresholds varied somewhat between landscape settings, the relative change was similar. This allowed collapsing of the multiple levels in Figure EE-37 - Figure EE-39 to REC level two. Expressing limits in this way also allows an alternative method for deriving reference condition to be used in place of the method implemented here. The absolute values of these reductions are provided in Table EE-1.

Table EE-1: Potential attribute band thresholds expressed as reduction from reference ESV state at REC level 2. Reductions are expressed as absolute changes. The units are Log10(visual clarity (m)), Log10(turbidity (NTU)) and proportion of the stream bed. These values should be added to the estimates of refefence condition ESV state for the relevant landscape setting to derive the absolute values for each threshold (e.g., Figure EE-41).

REC source of flow	Clarity A/B	Clarity B/C	Clarity C/D	Turbidity A/B	Turbidity B/C	Turbidity C/D	Total Fines A/B	Total Fines B/C	Total Fines C/D
Cool-Dry.Hill	-0.05	-0.11	-0.17	0.08	0.17	0.26	0.08	0.17	0.26
Cool-Dry.Lakefed	-0.06	-0.13	-0.20	0.10	0.20	0.31	0.11	0.24	0.38
Cool-Dry.Lowland	-0.07	-0.15	-0.23	0.09	0.19	0.28	0.11	0.23	0.35
Cool-Dry.Mountain	-0.07	-0.14	-0.22	0.11	0.23	0.37	0.08	0.16	0.24
Cool-ExtremelyWet.Hill	-0.07	-0.15	-0.23	0.11	0.23	0.35	0.09	0.18	0.28
Cool-ExtremelyWet.Lakefed	-0.05	-0.10	-0.15	0.07	0.14	0.21	0.07	0.15	0.24
Cool-ExtremelyWet.Lowland	-0.08	-0.17	-0.26	0.13	0.26	0.41	0.11	0.22	0.34
Cool-ExtremelyWet.Mountain	-0.08	-0.16	-0.25	0.11	0.22	0.34	0.06	0.13	0.20
Cool-Wet.Hill	-0.06	-0.12	-0.19	0.08	0.16	0.25	0.09	0.18	0.28
Cool-Wet.Lakefed	-0.06	-0.12	-0.19	0.08	0.16	0.25	0.10	0.21	0.34
Cool-Wet.Lowland	-0.07	-0.14	-0.22	0.10	0.20	0.31	0.11	0.22	0.34
Cool-Wet.Mountain	-0.07	-0.16	-0.25	0.09	0.20	0.31	0.06	0.13	0.20
Warm-Dry.Lakefed	-0.07	-0.15	-0.24	0.08	0.15	0.24	0.11	0.23	0.36
Warm-Dry.Lowland	-0.09	-0.20	-0.32	0.09	0.19	0.29	0.12	0.26	0.40
Warm-ExtremelyWet.Hill	-0.10	-0.21	-0.34	0.12	0.24	0.37	0.11	0.22	0.35
Warm-ExtremelyWet.Lowland	-0.08	-0.16	-0.25	0.09	0.19	0.29	0.11	0.24	0.37
Warm-Wet.Hill	-0.08	-0.16	-0.26	0.11	0.22	0.34	0.10	0.21	0.34
Warm-Wet.Lakefed	-0.10	-0.21	-0.34	0.11	0.22	0.36	0.11	0.23	0.36
Warm-Wet.Lowland	-0.11	-0.22	-0.35	0.13	0.28	0.44	0.12	0.25	0.39

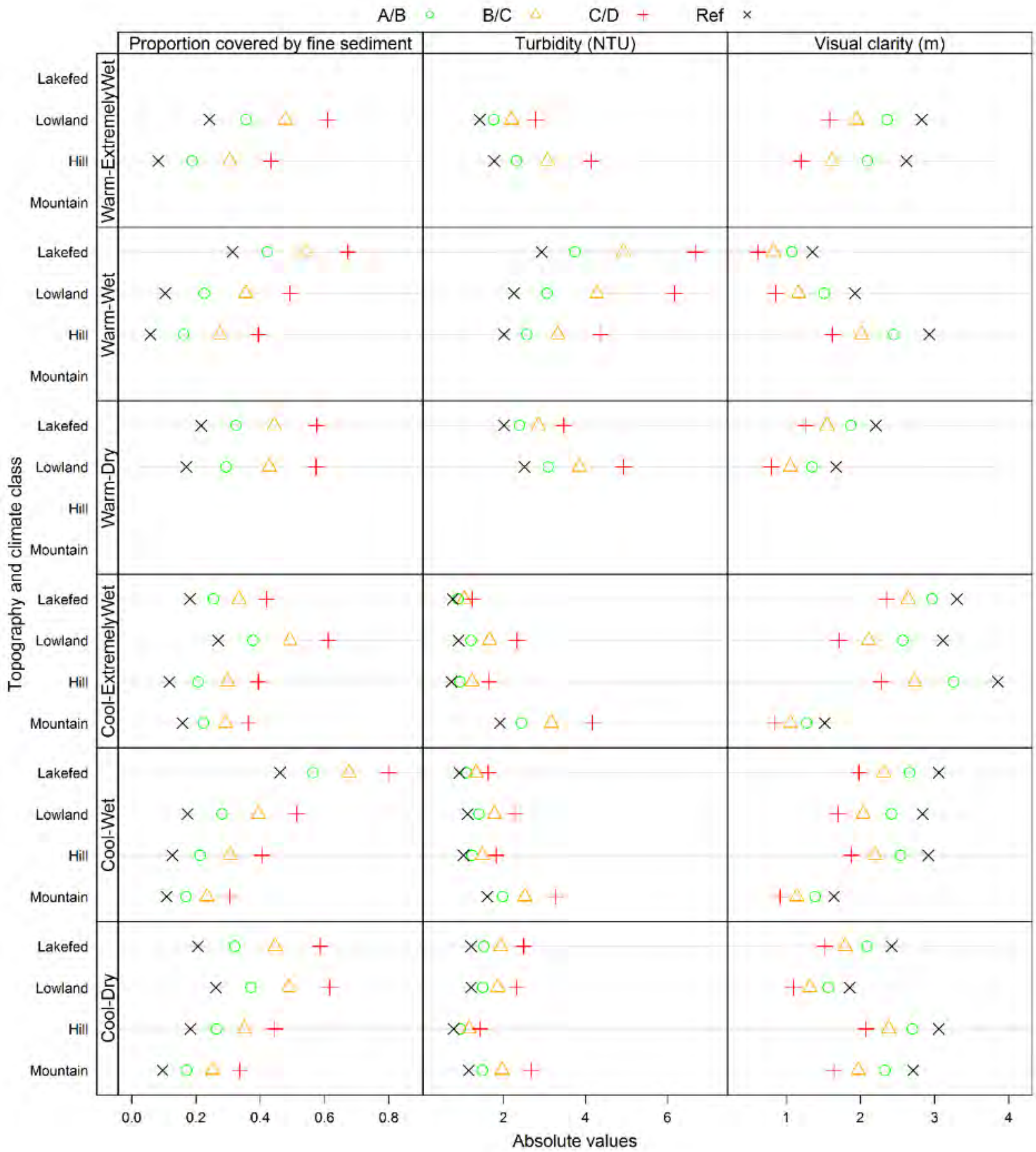
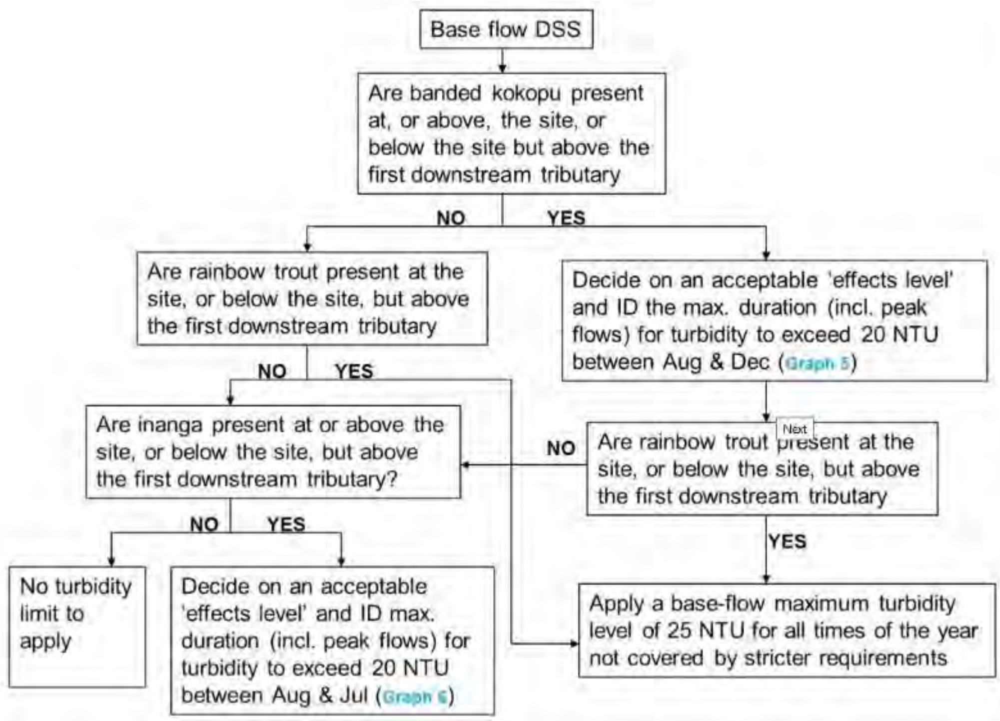
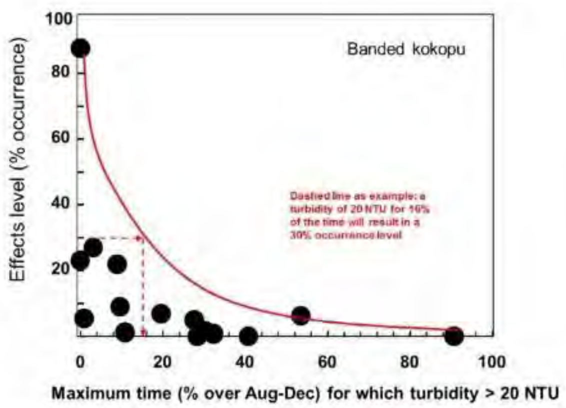


Figure EE-41: Potential ESV attribute bands at the second level (source of flow) of the REC classification based on fish community change. These values are based on the reference state model developed in this project.

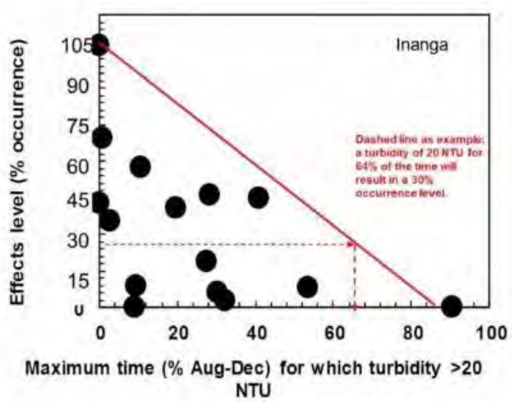
Appendix FF Rowe suspended sediment decision support tool

Available from: https://www.niwa.co.nz/our-science/freshwater/tools/turbidity/base





Graph 5



Graph 6