Development of macroinvertebrate response models for the FWMT

Prepared for Auckland Council

November 2023

Prepared by:
Richard White, Rick Stoffels

For any information regarding this report please contact:

Richard White
Quantitative Freshwater Ecologist
Freshwater modelling
+64 3 343 8033
richard.white@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 8602
Riccarton
Christchurch 8440

Phone +64 3 348 8987

NIWA CLIENT REPORT No: 2023332CH
Report date: November 2023
NIWA Project: ARC23501

Revision	Description	Date
Version 1.0	Final version sent to client	28 November 2023

Quality Assurance Statement		
MGreenwood	Reviewed by:	Michelle Greenwood
	Formatting checked by:	Rachel Wright
	Approved for release by:	Scott Larned

Contents

Tables

Figures

Executive summary

Management of water quality and ecological state is a responsibility of regional authorities under the National Policy Statement for Freshwater Management (NPSFM). To address this responsibility, Auckland Council's (AC) stormwater department 'Healthy Waters' is developing the Freshwater Management Tool (FWMT), a process-based model, that simulates daily loads of nutrients, heavy metals, *E. coli*, and sediment for 5465 sub-catchments in the Auckland region from instream and catchment processes.

The FWMT does not predict ecological responses to contaminants and could better support decision-making if it could. A review, conducted by NIWA, previously concluded that models of macroinvertebrate responses to the stressors: dissolved inorganic nitrogen (DIN), dissolved reactive phosphorous (DRP) and visual clarity (VC[1]), could be developed using national state of the environment (SoE) monitoring datasets. The objective of this study was to develop predictive macroinvertebrate models focussing on effects that are manageable and outputs of the FWMT – DIN, DRP and VC.

Two options to model macroinvertebrate stressor responses were proposed:

- Macroinvertebrate index models (MIMs), which model macroinvertebrate indices (Macroinvertebrate Community Index [MCI], Quantitative MCI [QMCI] and Average Score Per Metric [ASPM]) as functions of the stressors.
- Stacked species distribution models (SSDMs), which model the probability of occurrence (PO) of macroinvertebrate taxa as functions of the stressors.

Macroinvertebrate index models are parameterised to explicitly link the response of NPSFM macroinvertebrate attributes to water quality stressors. This gives MIMs direct relevance to the NPSFM. However, the link between macroinvertebrate indices and specific water quality stressors is unclear owing to a) ambiguity concerning what types of water quality pollution the indices are designed to respond to, and b) the use of expert opinion to quantify taxon pollution tolerances used in index calculations, which are subject to confounding and bias.

In contrast, SSDMs yield taxon tolerances from empirical relationships in data, not expert opinion. Provided the model is well specified to—as much as is possible—partition confounding factors, SSDMs could more accurately capture the response of communities to water quality stressors. However, this benefit of SSDMs needs to be evaluated before they can be applied to the FWMT.

The objectives of this report were:

1. Using national SoE data, comprehensively evaluate both MIMs and SSDMs as options for modelling macroinvertebrate responses to DIN, DRP and VC.
2. Produce a series of lookup tables that specify thresholds of DIN, DRP and VC linked to a probability of occurrence for macroinvertebrate taxa using suitable models identified by under Objective 1.

---

[1] Here we use VC as a proxy for suspended fine sediment and consider it a stressor to macroinvertebrate communities *sensu* NPSFM (2020), which worsens as VC decreases, and improves as VC increases.

Modelling approach

The general modelling framework used was Generalised Additive Mixed-Effects Models (GAMMs). The mixed-effects (also known us multi-level) modelling approach enabled us to:

- Isolate the effects of primary, management interest—effects of DIN, DRP and VC—from the effects of numerous other spatial and temporal contingencies on those effects, but without having to resort to highly complex models that run a greater risk of being overfitted to training data.
- Generate predictions of stressor effects that are more generalisable than those arising from models that include numerous parameters that explicitly account for all variables that (supposedly) capture spatial and temporal contingencies.
- Account for correlations among observations within sites, climate zones and years—correlations within sampling units than can bias model outputs.
- Control for spatial and temporal imbalances in sampling effort that can bias model outputs—imbalances that are prevalent in the monitoring data used in this study.

Steps involved in SSDM and MIM fitting were identical, with only minor adjustments to accommodate different distributions of response variables (e.g. Gaussian vs. Bernoulli). The MIMs and SSDM were then compared using multiple criteria:

1. Model fit – did the models meet the statistical assumptions underlying them? Was there any bias in the estimated stressor effects?
2. Model accuracy – how much uncertainty was associated with the model predictions?
3. Sensitivity – how sensitive were model predictions to each stressor?
4. Ecological plausibility – were model consistent with our ecological understanding?

Stacked species distribution model

SSDM evaluation

GAMMs were parameterised for 104 macroinvertebrate taxa comprising the SSDM. Following our four criteria above we concluded:

- No major violations of statistical assumptions underlying the SSDM were observed, and the predicted stressor effects were unbiased.
- Stressor effects were selected for the final models of 68 taxa, with most involving additive or interactive multi-stressor effects.
- Predictive accuracy of GAMMs for individual taxa of the SSDM was generally low. On average, <10% of the internal variance in taxon PO was attributed to the combined effects of each stressor.
- Sensitivity of the SSDM to each stressor was high. The PO for some taxa varied 100% over observed nutrient and sediment gradients.

- The direction and form of stressor responses of taxa predicted by the SSDM were consistent with our understanding of the ecology of individual taxa. Stressor responses for many taxa involved complex multi-stressor interactions, consistent with empirical observations.

Stressor limit lookup tables derived from the SSDM

Biological Extirpation Analysis using Monte-Carlo simulations of the SSDM was used to derive stressor limits required to achieve several macroinvertebrate targets. The analysis estimated thresholds of DIN, DRP and VC required to curtail extirpation to different target states (extirpation of ≤1%, 2.5%, 5% and 10% of the taxa in the modelled macroinvertebrate community).

Multi-stressor effects had significant and strong effects on taxon extirpation. Single-stressor thresholds required to achieve specific macroinvertebrate targets were strongly dependent on levels of other stressors.

Given the strength of these multi-stressor effects, stressor thresholds were calculated for three multi-stressor scenarios (Figure 1-1):

- An *optimistic scenario*, which estimated the thresholds of each stressor required to achieve macroinvertebrate targets when levels of the other two stressors are at the low end of the observed distributions of stressor data (either through effective stressor reductions, or naturally low concentrations of those stressors).

- A *pessimistic scenario*, which estimated the thresholds of each stressor required to achieve macroinvertebrate targets when levels of the other two stressors are high and of the observed distributions of stressor data (either as a consequence of anthropogenic pollution, or naturally high concentrations of those stressors).

- A *typical scenario*, which estimated the thresholds of each stressor required to achieve macroinvertebrate targets when levels of the other two stressors are close to the medians of the observed stressor data (and are therefore characteristic of a ‘typical’ river within the training data).

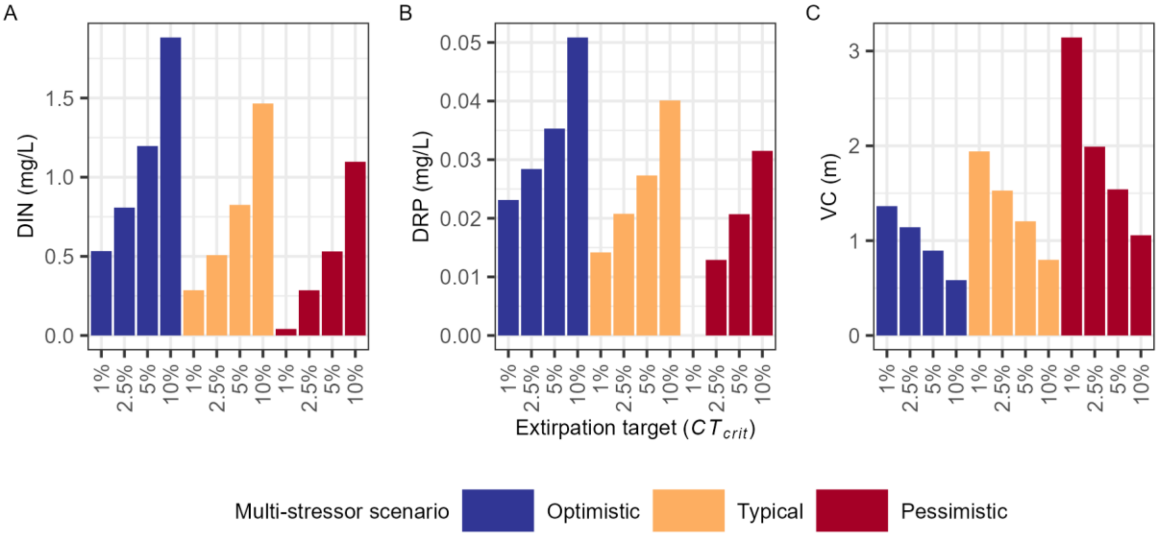


Figure 1-1: Thresholds for a) DIN, b) DRP and c) VC required to limit community extirpation to 1%, 2.5%, 5%, or 10% ($CT_{X,crit}$), for optimistic (blue), pessimistic (red) and typical (orange) multi-stressor scenarios. The typical, pessimistic and optimistic multi-stressor scenarios show thresholds for each stressor while remaining stressors vary about their median, worst or least stressful conditions, respectively. Thresholds are repeated in table form in Appendix G.

To achieve a specific macroinvertebrate target state, thresholds placed on one water quality stressor may have to be much more stringent (i.e., *pessimistic*), when/where the states of the other stressors are also poor. However, such thresholds may be relaxed (i.e., *optimistic*) if the states of all stressors are improved in concert.

Macroinvertebrate index model fitting and evaluation

MIMs were developed and evaluated for MCI, QMCI and ASPM. Following our four criteria above we concluded:

- Predictive accuracy of the MIMs was also low. Less than 10% of the internal variance in MCI, QMCI and ASPM were attributed to the combined stressor effects.
- Macroinvertebrate indices were insensitive to all stressors considered. Index insensitivity was such that MCI and QMCI did not shift beyond a single NPSFM band as stressors worsened. This indicates it would not be possible to achieve macroinvertebrate objectives – as defined in the NPSFM – by reducing DIN, DRP or increasing VC.
- The insensitivity of macroinvertebrate attributes was likely a result of the taxon-specific tolerance scores – derived from expert opinion – being poorly related to observed taxon-specific tolerances derived from data (our SSDM).
- MCI, QMCI and ASPM are not suitable indicators of the effects that DIN, DRP and VC have on macroinvertebrate communities. Stressor thresholds were therefore not developed using the MIMs.

Conclusions

Despite the relevance of MIMs to central policy, we found no evidence that they are effective for understanding or managing the effects of DIN, DRP and VC on macroinvertebrate communities. By contrast, we presented a biological extirpation analysis based on an SSDM that, while arguably being less directly relevant to central policy, should be more effective for managing multiple water quality stressors to meet ecological objectives.

Next steps should focus on application of the SSDM to meet ACs requirements for simulating macroinvertebrate responses to contaminant scenarios arising from the FWMT:

1. Whether SSDM forecasts should consider only a subset of taxa most relevant to the Auckland region.
2. What type of simulation of the SSDM would best capture whole-community response to alternative water quality scenarios arising from application of the FWMT.
3. How to calibrate VC as a function of total suspended solids (TSS; output of the FWMT) such that we may simulate macroinvertebrate response to TSS?
4. Determine the amount of SSDM uncertainty that is acceptable, given the intended use of the FWMT, and whether additional FWMT variables should be included into the SSDM.

1 Introduction

1.1 Background

Management of water quality and ecological state is a responsibility of regional authorities in New Zealand under the National Policy Statement for Freshwater Management (NPSFM; MFE 2020). Pre-emptive of NPSFM regulatory implementation, Auckland Council's stormwater department, 'Healthy Waters', is developing a regional-scale catchment management and operational decision-support tool, the 'Freshwater Management Tool' (FWMT) (Auckland Council 2021a-c). The FWMT simulates daily loads of nutrients, heavy metals, *E. coli*, and sediment for 5465 sub-catchments in the Auckland region from instream and catchment processes.

The FWMT is used to forecast how changes in catchment hydrology and contaminant sources affect contaminant concentrations in the water column of rivers. Contaminant sources include point sources (e.g., engineered overflow contributions from reticulated wastewater networks) and nonpoint sources (e.g., discharge from engineered stormwater networks including various residential, commercial and industrial surfaces, rooves, roads, unpaved urban surfaces, septic areas, horticultural, pastoral, exotic and native forested areas, broader open space and streambank erosion).

The FWMT will allow Healthy Waters to model the movement of water and operationally-relevant contaminants through rural and urban catchments, identify critical contaminant sources, estimate the loads and concentrations of contaminants in receiving environments, infer their likely impacts on ecosystem health, simulate the effectiveness of different management options for mitigating those impacts, estimate the cost of achieving desired management outcomes, and determine where those costs are likely to fall.

This information will guide Healthy Waters as it develops action plans for achieving water quality targets in Auckland's rural and urban catchments, and as it makes operational, asset management, and investment decisions within the context of its stormwater network discharge consent.

Importantly, the FWMT does not currently predict instream ecological responses to contaminants. Healthy Waters has identified a need to be able to contextualise and predict changes to ecological measures changes in operational contaminants ("stressors").

A critical review of options to model ecological responses to contaminant loads was conducted by NIWA (White et al. 2022). The review concluded that:

- There are limited time-series pairing ecology with contaminant loads within the Auckland region. Ecological models must therefore be built using national data sets.
- Linking FWMT output to models of periphyton and fish is not possible due to data limitations, and/or poor predictive performance of existing models in the Auckland region.
- Sufficient national data are available to model macroinvertebrate responses to the stressors dissolved inorganic nitrogen (DIN), dissolved reactive phosphorous (DRP), and visual clarity (VC).

Two options to model macroinvertebrate responses to those stressors with sufficient data (DIN, DRP and VC) were identified:

- Macroinvertebrate index models (MIMs) , which model macroinvertebrate indices (e.g., macroinvertebrate community index [MCI], Quantitative MCI [QMCI] or Average Score Per Metric [ASPM]) as statistical functions of the stressors.
- Stacked species distribution models (SSDMs), which model the probability of occurrence (PO) of individual macroinvertebrate taxa as statistical functions of the stressors.

Trade-offs were identified in both approaches (Table 1-1), which will now be discussed.

Table 1-1: Strengths and weaknesses of macroinvertebrate index models (MIMs) and stacked species distribution models (SSDMs).

Strengths/Weaknesses	MIMs	SSDMs
Strengths	● Policy-relevant out-of-the-box ● Simple to implement ● Easy to communicate to stakeholders	● Stressor responses not confounded ● Flexible applications
Weaknesses	● Taxon tolerance scores may confound stressor responses ● Limited applications	● Very rare taxa cannot be modelled ● Complex to implement

1.2 Macroinvertebrate index models

The main strength of MIMs is that they are parameterised to explicitly link the response of NPSFM macroinvertebrate attributes with water quality stressors. The MCI, QMCI and ASPM (Stark and Maxted 2007, Collier 2008) are attributes in the National Policy Statement for Freshwater

Management (NPSFM) requiring actions. Specifically, Clause 3.13 of the NPSFM requires councils to set exceedance thresholds for DIN and DRP that achieve target index states. Macroinvertebrate index models could therefore be directly used to determine how catchment planning scenarios affect NPSFM invertebrate attributes.

Macroinvertebrate index models are also simple. They aggregate a complex multivariate system – the ecological community - into one metric. Each taxon in the community is assigned a tolerance score representing its pollution tolerance. These scores are averaged among taxa, and in the case of ASPM, weighted by EPT (Ephemeroptera, Plecoptera, Trichoptera) membership, to calculate the indices. Communities comprised of taxa with high pollution tolerance are assumed to indicate environmental degradation and poor ecosystem health.

However, this simplicity is also a drawback. Taxon tolerance was broadly defined as tolerance to 'pollution', 'organic pollution or enrichment' and 'nutrient enrichment' (Stark 1985, Stark and Maxted 2007). However, it is unclear how tolerance scores relate to particular contaminants (Clapcott et al. 2017) (e.g., DIN, DRP, VC). If tolerance scores do not relate well to a particular stressor, the indices may not accurately represent the effect of the stressor on the community. Because the tolerance scores were based on expert opinion and not quantitatively linked to any particular stressor[2], the risk that these indices fail to characterise community responses to specific stressors is high.

A second issue is that each taxon is assigned only a single tolerance score. This assumes that each taxon has the same tolerance to all stressors that the index is applied to. However, research shows that nutrients and sediment often affect macroinvertebrate communities in complex, interactive ways (Matthaei et al. 2010; Piggott et al. 2012; Piggott et al. 2015), suggesting this assumption is too simplistic. Failure to distinguish tolerance to different stressors could render indices insensitive to multi-stressor interactions.

1.3 Stacked species distribution models

The strengths and weaknesses of SSDMs directly contrast those for MIMs (Table 1-1). Stacked species distribution models yield taxon tolerances from empirical relationships in data. Provided the model is well specified to—as much as is possible—partition confounding factors, SSDMs can disentangle multi-stressor effects on communities (Schipper et al. 2014). These features of SSDMs may mean they more accurately capture the response of communities to water quality stressors. By extension, SSDMs arguably present a more scientifically-defensible means of predicting the response of communities to changes in water quality.

However, SSDMs are complex. Rather than fitting a single model to an index-stressor data set, SSDMs require separate models for each taxon, of which there may be many. Additional algorithms are required to aggregate taxon-specific responses into a whole-community response. Several approaches are available, including Biological Extirpation Analysis (Cormier and Suter 2013), Gradient Forests (Wagenhoff et al. 2017), or simple model stacking (Schipper et al. 2014). However, these algorithms and outputs may be difficult to communicate.

---

[2] Tolerance scores were originally assigned by Stark (1985) according to distribution of taxa amongst sites ranked into three pollution classes in the Taranaki ring plain. However, pollution classes were assigned to each site based on expert opinion, and not according to how sites differed quantitatively in some aspect of pollution (e.g., DIN). Subsequent changes/additions to tolerance scores were based on expert opinion or by reconstitution of those original scores (Stark and Maxted 2007), such that they remain informed entirely by expert opinion.

Nevertheless, this flexibility, is also a key SSDM strength. Using SSDMs we could:

- More objectively estimate stressor-specific taxon tolerances from data and develop stressor-specific indices.
- Simulate the effects of stressors on macroinvertebrate community dynamics.
- Quantify the interactive effects of multiple stressors.
- Isolate taxa of cultural, functional, or regional importance to base decisions on.

Despite the strengths of SSDMs, they remain rarely applied in freshwater management in New Zealand. The benefits of SSDMs for linking with the FWMT are potentially great, but need testing. This requires testing of both the SSDM itself, as well as ways to aggregate the SSDM to predict stressor responses of the whole community.

1.4 Report objectives

The objectives of this report are two-fold. The first objective was to comprehensively evaluate both MIMs and SSDMs as options for modelling macroinvertebrate community responses to changes in operationally relevant stressors: DIN, DRP, and VC (noting only DIN and DRP are simulated by the FWMT Stage 1 – the brief from HW was to focus on sediment and nutrient operational contaminants [and forms thereof] but without constraining the exercise only to those simulated currently by the FWMT).

The models were evaluated using multiple criteria:

- Model fit – did the models meet the statistical assumptions underlying them? Was there any bias in estimated stressor effects?
- Model accuracy – how much uncertainty was associated with model predictions?
- Sensitivity – how sensitive were model predictions to each stressor?
- Ecological plausibility – were models consistent with our ecological understanding?

Our second objective was to use the best model identified from objective one, to produce a series of lookup tables that specify how changes in DIN, DRP and VC change macroinvertebrate state, which can be compared against FWMT scenario outputs for changing contaminant regimes instream.

1.5 Report structure

Sections of this report are structured as follows:

Section 2 details the data used for modelling. It also details the rationale for spatial and temporal partitioning of variance, used to control for confounding processes in MIMs and SSDMS.

Section 3 details the methods and results used to fit and evaluate the SSDM.

Section 4 develops and evaluates an approach to aggregate and predict community stressor responses using the SSDM. We use it to evaluate macroinvertebrate community extirpation responses to stressors and multi-stressor interactions predicted by the SSDM. We use the approach to develop stressor thresholds required to limit macroinvertebrate extirpation to specific targets, under different multi-stressor management scenarios.

Section 5 presents the methods and results used to fit and evaluate index-stressor responses predicted by the MIMs. We also evaluate stressor tolerance assumptions made by MCI and QMCI using the SSDM.

Section 6 summarises sections 3-5 with respect to the report objectives and recommends next steps.

2 Data

2.1 Data sources

2.1.1 State of the environment monitoring data

The data used in this report come from two sources. The first consists of national State of the Environment (SOE) monitoring data collected by regional councils. These data were compiled previously by Depree et al. (2017) to develop suspended sediment thresholds for the NPSFM. They comprised 1616[3] paired annual stressor-macroinvertebrate observations taken between 2006-2016 among 300 sites. Macroinvertebrate samples were collected using semi-quantitative kick-net methods as detailed in Stark et al. (2001). DIN, DRP and VC samples were collected monthly using standardised protocols and were summarised such that they represented the median observation of the 12 months prior to each macroinvertebrate sample.

The spatial design of SOE monitoring programmes varied between each regional council but are biased towards wadeable rivers and catchments dominated by pasture. Meanwhile variable availability of VC among regions and years mean that SOE data are poorly balanced (Figure A-1). In particular, Auckland, Gisborne, Tasman, Nelson and Marlborough were underrepresented as these councils did not monitor VC over the period in Depree et al. (2017). The FWMT produces forecasts of total suspended solids (TSS) as a measure of suspended sediment. However, TSS data are not routinely monitored by Regional Councils (Dupree et al. 2017). Nevertheless, as noted by Montgomery et al. (2022), FWMT forecasts of TSS could be utilised provided the appropriate models describing relationships between VC and TSS were developed. We therefore considered VC essential for this study.

2.1.2 National River Water Quality Network monitoring data

The second source consisted of data collected by the NIWA National River Water Quality Network (NRWQN). These data comprised 1602[3] annual stressor-macroinvertebrate observations for 63 sites taken between 1990-2019. Macroinvertebrate samples were collected using quantitative surber sampling protocols (Smith and McBride 1990). Stressor samples were taken monthly using standardised protocols (Smith and McBride 1990). These data were processed following the same procedures described by Depree et al. (2017) and then combined with the SOE data. We restricted NRWQN data to 1990-2019 as there was a change to resolution of taxa recorded in samples after 2019, which resulted in a spurious increase in new taxa.

The NRWQN was designed by NIWA to measure water quality status and trend in nationally important rivers and is therefore slightly biased towards larger rivers (Smith and McBride 1990). However, the data are better balanced among regions and years than the SOE data (Figure A-1). The sites are also better balanced among catchments dominated by upstream pasture and indigenous forest.

2.2 Spatial and temporal partitioning of data

Our objective was to partition correlated variance among sites, years, and regions into random parameters in MIMs and SSDMs. We had several reasons for doing so. Firstly, this would account for non-independence and correlations of observations at each grouping level (Pinheiro and Bates 2000). The data mostly consisted of annual repeated measures within each site of up to 30 years, with only 43 out of 363 sites containing single observations. Including random parameters for each site would ensure the assumption of sample independence was met in subsequent models (Pinheiro and Bates

---

[3] The number of complete samples including DIN, DRP, VC and macroinvertebrate observations after censored data treatments.

2000). This would also help account for effects associated with spatial differences among sites, such as altitude, that might otherwise confound stressor effects (White et al. 2022).

Secondly, we wished to control for interannual changes in climate, and subsequent effects on water temperature and flow that might confound stressor effects. For this purpose, we partitioned all sites into six different climate zones as defined by Mullen (1995) (Figure 2-1). Climate zones (hereafter, CZ) are defined by how they differ in response to the El Nino Southern Oscillation (ENSO). Climate zone boundaries are largely determined by the interaction of mountain chains with mean westerly flow, which drive differences in temperature and precipitation among zones during ENSO phases (Mullan 1995, Kidson and Renwick 2002). Including random parameters for each CZ, and year within each CZ, would ensure that confounding interannual and periodic climate effects were accounted for in subsequent models.

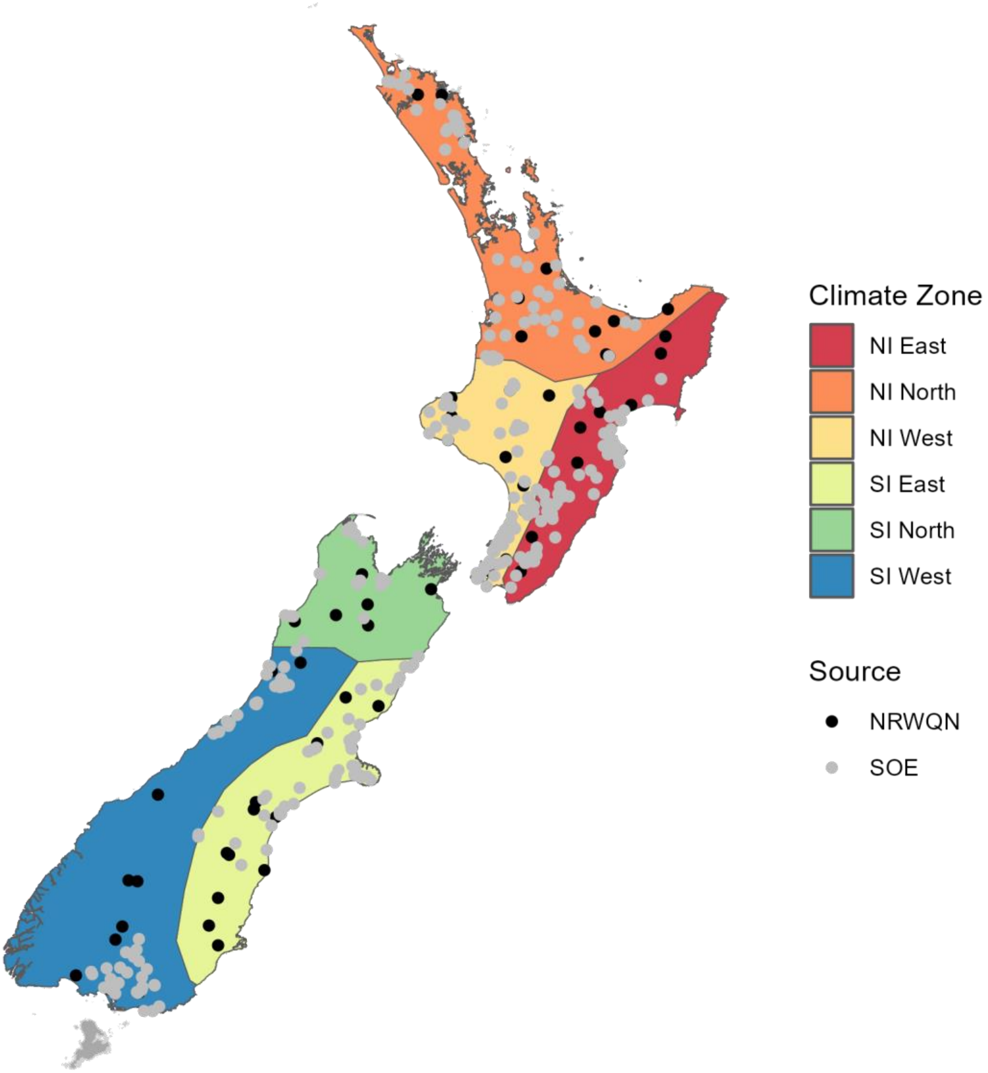


Figure 2-1: Distribution of sites among climate zones and sources.

An additional reason for partitioning data by site, CZ and year was to account for spatial and temporal sample biases. Although the data were better balanced when partitioned by CZ, there remained strong spatio-temporal biases in the data (Figure A-2). This was primarily caused by differences in data availability among regional councils. Given the potential biogeographic differences in taxon distribution, these sample biases might confound stressor effects if not otherwise controlled for.

2.3 Macroinvertebrate indices

We followed the National Environmental Monitoring Standards (NEMS 2022) macroinvertebrate guidelines for calculating MCI, QMCI and ASPM for each sample. The MCI was calculated as:

$$MCI = \frac{\sum_{i=1}^{T} a_i}{T} \times 20 \tag{1}$$

where, T is the total number of scoring taxa in the sample and a_i is the tolerance score of taxon i. The MCI is therefore simply the tolerance score of the average taxon in the sample.

The QMCI was calculated as:

$$QMCI = \sum_{i=1}^{T} (n_i \times a_i) \tag{2}$$

where, n_i is the relative abundance of taxon i in the sample, calculated as the total abundance of taxon i divided by the total abundance of all taxa in the sample. Note that equation (2) differs slightly from that presented in NEMS (2022) which uses taxon counts as opposed to relative taxon abundance used here, however both formulae yield the same values for QMCI. For MCI and QMCI we used the hard bottomed (HB) tolerance scores and taxa as defined in the macroinvertebrate NEMS (2022) guidelines.

The ASPM was calculated as the average of the metrics MCI, EPT richness and percent EPT abundance, whereby EPT richness is the total number of taxa belonging to the orders Ephemeroptera, Plecoptera, and Trichoptera (EPT), and percent EPT abundance is the number of EPT individuals divided by the total number of individuals in the sample. Each metric was range-standardised as $X_s = (X - X_{min})/(X_{max} - X_{min})$, where X_s and X were scaled and unscaled metric values, respectively, and X_{min} and X_{max} the minimum and maximum unscaled metric values among all samples. This range-standardisation results in each metric having range [0,1].

2.4 Treatment of censored stressor data

Censored stressor values – measurements that are below or above detection limits – were present in the water quality data. Treatment of such data in New Zealand environmental reporting is often done by imputation. This involves approximating the values for censored data that are consistent with the uncensored data using various procedures (Helsel 2012). However, these imputation procedures do not preserve the covariance that occurs between uncensored data. Failure to preserve the (uncensored) covariance may bias mathematical relationships between stressors and dependent variables at the upper and lower domains of each stressor. Rather than risk introducing such artefacts, we chose to remove all samples with censored values for any stressor prior to model fitting (245 samples). We also restricted any predictions made from models to the lower 2.5% and upper 97.5% quantiles of the stressor distributions, to reduce leverage effects in the tails of covariate distributions.

3 Stacked species distribution model fitting and evaluation

3.1 Summary

The objective of this section was to fit and evaluate the stacked species distribution model (SSDM).

Generalised additive mixed-effects models were fitted to the probability of occurrence of 104 taxa and the fitted stressor responses of each taxon were evaluated.

Conclusions from this section were:

- No major violations of statistical assumptions underlying the SSDM were observed and predicted stressor effects were unbiased as evinced by residual examination.
- Predictive accuracy of models of individual taxa of the SSDM was generally low. This means there was a large amount of uncertainty associated with predicting taxon PO based solely on the effects of DIN, DRP and VC.
- Sensitivity of the SSDM to each stressor was high and indicated that observed variation in nutrient and sediment enrichment would be sufficient to cause extirpation of sensitive taxa.
- The direction and form of taxon-specific responses to stressors predicted by the SSDM were consistent with our understanding of the ecology of each taxon. Many taxon-specific stressor responses involved complex multi-stressor interactions consistent with empirical observations.

3.2 Methods

3.2.1 SSDM parameterisation and selection

Generalised additive mixed effects models (GAMMs) were used to model spatial and interannual changes in probability of occurrence (PO)[4] of each macroinvertebrate taxa in the SSDM. The taxonomic resolution used to define a taxon was equivalent to that used to calculate macroinvertebrate indices (Section 2.3). All GAMMs were fitted using the gamm4 package (Wood and Scheipl 2017).

We sought relatively simple model parameterisations of how stressors drove variation in PO of each taxon. The stressors DIN, DRP and VC entered each model as fixed effects, while site, climate zone (CZ) and year of observation nested within CZ entered as random intercepts. Following Dormann et al. (2013), we checked for collinearity among fixed effects, but did not find any correlations > 0.7 (Figure B-1).

We used an information-theoretic approach with Akaike's information criterion (AIC; Akaike, 1974) to select the most parsimonious model from a series of 12 candidate models (Burnham & Anderson, 2002) for each taxon. The models were mostly nested and assessed hypotheses on the additive and interactive effects of each stressor (Table 3-1). The final model for each taxon was selected as that minimising AIC, which is the most parsimonious model that minimised under- and over-fitting.

---

[4] Probability of occurrence here being defined as the probability of a species being present in a sample (which could be a composite of multiple kick-net or surber sub-samples) taken during a specific year from a stream reach of undefined spatial extent, among all such samples used for analysis. Models fitted for all species used the full national complement of samples.

Prior to model selection and testing, we scaled all stressor variables, S, such that $0 \leq S \leq 1$, according to $S = (S - \min[S])/(\max[S] - \min[S])$, whereby min[S] and max[S] are the minimum and maximum values of S, respectively, in the data. This allowed us to use isotropic smoother interactions for stressor variables in the GAMMs. These are less computationally intensive but yield responses equivalent to anisotropic smoothers (Wood 2017).

We set the basis dimension used to represent each GAMM smooth term (i.e., number of smoother knots) to $1K$ for univariate smoothers, $2K$ for bivariate smoother interactions, and $4K$ for multivariate smoother interactions, where $K = 6$. The basis dimension restricts the maximum complexity of smoother non-linearity considered during model cross-validation. Preliminary model selection of five representative taxa indicated that setting K=6 was sufficient to capture the complexity of smoother non-linearity (Appendix C).

Models were only fitted to macroinvertebrate taxa with at least 6% prevalence (i.e., present in at least 6% of sites monitored). This prevalence cut-off was based on research showing model predictive performance is overestimated for taxa below 6% prevalence (Barbet-Massin et al., 2012). This rule resulted in 104 taxa being eligible for modelling out of a total of 182 taxa present in the data.

Table 3-1: Description of GAMM models used to model PO of each taxon in the SSDM.

Number	Name	Formula	Description
1	NULL	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+\varepsilon$	Probability of occurrence (PO_{ijt}) of a taxon during Y t within CZ j at SID i is a function of the population-level PO (α) + deviation induced by SID i (α_i) + deviation induced by CZ j (α_j) + deviation induced by Y t within CZ j (α_{jt}) + error (ε). The factor α is fixed while α_i, α_j and α_{jt} are random.
2	DIN	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DIN_{it})+\varepsilon$	As for 1, but including the effect of DIN at SID i during Y t entering as a fixed effect smoother.
3	DRP	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DRP_{it})+\varepsilon$	As for 1, but including the effect of DRP at SID i during Y t as a fixed effect smoother.
4	VC	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(VC_{it})+\varepsilon$	As for 1, but including the effect of VC at SID i during Y t as a fixed effect smoother.
5	DIN+DRP	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DIN_{it})+ f(DRP_{it})+\varepsilon$	As for 2, but including DRP as an additive smoother.
6	DIN+VC	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+ f(DIN_{it})+ f(VC_{it})+\varepsilon$	As for 2, but including VC as an additive smoother.

Number	Name	Formula	Description
7	DRP+VC	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DRP_{it})+ f(VC_{it})+\varepsilon$	As for 3, but including VC as an additive smoother.
8	DIN:DRP	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DIN_{it},DRP_{it})+\varepsilon$	As for 5, but including the interaction between DIN and DRP as a smoother.
9	DIN:VC	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DIN_{it},VC_{it})+\varepsilon$	As for 6, but including the interaction between DIN and VC as a smoother.
10	DRP:VC	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DRP_{it},VC_{it})+\varepsilon$	As for 7, but including the interaction between DRP and VC as a smoother.
11	DIN+DRP+VC	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DIN_{it})+ f(DRP_{it})+f(VC_{it})+\varepsilon$	As for 5, but including effect of VC at SID *i* during Y *t* as a smoother.
12	DIN:DRP:VC	$PO_{ijt}=\alpha+\alpha_i+\alpha_j+\alpha_{jt}+f(DIN_{it},DRP_{it},\mathrm{VC}_{it})+\varepsilon$	As for 11, but including the interaction between DRP, DIN and VC as a smoother.

3.2.2 SSDM fit

The SSDM makes several key assumptions regarding the effects of each stressor on PO of each taxon. These assumptions include a) that the data arise from a Bernoulli distribution (i.e., the response is either a 0 or 1), b) that the model residuals are independent (i.e., independent in space and time) and c) that the fixed effects are not confounded by collinearity. These assumptions were addressed by fitting models with a Bernoulli distribution, using random effects to account for correlations and non-independency within sites, years, and CZs, and by only including fixed effects that were uncorrelated.

If these assumptions were correctly addressed, the quantile residuals of the SSDM should be normally distributed and homogeneous with respect to model predictions and the fixed covariates (Dunn and Smyth 1996). We calculated and inspected model quantile residuals using the DHARMa package (Hartig 2022). Problems with model fit were identified as departures from linearity in residual QQ plots of the quantile residuals, and heterogeneity in the quantile residuals with regards to the predicted values (residual versus predicted plots) and observed values of each stressor (residual versus stressor plots).

3.2.3 SSDM predictive accuracy

We used 10-fold cross-validation to examine the accuracy of predictions of the final model for each taxon. This involved splitting the full data set into 10 testing folds, then iteratively re-training the model after withholding each fold, and testing the predictive accuracy on the withheld fold. We wished to test model transferability in space, so we used the groupdata2 package (Olsen, 2017) to ensure no leakage of data from individual sites across folds.

We calculated the predictive accuracy of the final model on each fold using the area under the receiver operator characteristic curve (AUC). This statistic measures the accuracy of the model predictions relative to the observed data. An AUC of 0.5 indicates the model predicts taxon presences and absence no better than random chance, while AUC approaching 1 indicates perfect discrimination. Values <0.5 indicate the model performs worse than the NULL under cross-validation. Predictions made on each fold used only the fixed effects parameters thus ensuring predictive accuracy was evaluated for DIN, DRP and VC effects only, excluding any random parameters.

We also calculated pseudo-R^2 for each taxon as a measure of the amount of variance explained by the fixed stressor effects, relative to the random effects (Tjur, 2009). Pseudo-R^2 is the difference between the average predicted PO for observed presences and that of the observed absences of the training data and is interpreted as the coefficient of discrimination for models of Bernoulli distributed data. We calculated pseudo-R^2 from the final model of each taxon fit to the full data set (i.e., internal pseudo-R^2), while iteratively excluding each random effect from the prediction. This allowed us to calculate the contribution of each model term on pseudo-R^2.

3.2.4 SSDM stressor sensitivity

As a final step, we calculated the absolute and relative sensitivity of each taxon's PO to each stressor as predicted by the SSDM. Absolute sensitivity was defined as the range in predicted PO of a taxon across the domain of a stressor (i.e., lower 2.5% and upper 97.5% stressor observations). Relative sensitivity was defined as absolute sensitivity divided by the maximum PO for that taxon.

Relative sensitivity was used to account for differences in rarity among taxa. For example, the rarest taxon modelled may have low maximum PO, and therefore low absolute sensitivity, but still show high sensitivity relative to their maximum PO. Meanwhile, absolute sensitivity may indicate a sensitive taxon that has high influence on overall community dynamics (i.e., a sensitive common taxon). Both relative and absolute sensitivity was calculated from each taxon's partial response to each stressor while holding all other stressors at their median observed values and excluding random effects. These partial stressor responses were also inspected to assess the general direction and form of taxa-specific responses, as well as ecological consistency.

Note that the sensitivity analyses were used as an additional step to analyse the absolute and relative magnitude of stressor effects on taxon PO. However, these analyses did not inform model selection (i.e., selection of fixed effects described in Section 3.2.1), nor selection of taxa to be considered in the SSDM. All taxa meeting the 6% prevalence threshold (n=104), whose final models passed residual diagnostic inspections, were included for all analyses throughout this report, including those whose selected final models were the NULL.

3.2.5 Assumptions regarding stressor effects

A key assumption made by the SSDM is that stressor effects reflect both indirect and direct effects of the stressor on taxon presence. Direct effects include the stressors' effect on physiological homeostasis (i.e., toxicity). Indirect effects refer to the stressors' effect on taxon PO through some indirect pathway, such as via its' effect on a key resource (e.g., periphyton, prey). Following, MFE (2022), we assumed that nutrient effects measured here reflected primarily indirect effects via stimulation of periphyton growth. Increased VC is also likely to have similar indirect effects on periphyton growth by increasing solar exposure and stimulating photosynthesis. However, VC, as a proxy for suspended sediment, may also have strong direct effects on certain taxa, such as filter feeders, whose ability to feed may be impeded by reduced VC. Nevertheless, the SSDM cannot distinguish between such mechanisms. Indirect pathways involve more ecological contingencies and are therefore less certain. We therefore ensured that stressor effect uncertainty is fully factored into all simulations made by the SSDM to account for this.

3.3 Results

Given the large number of models comprising the SSDM (104 taxa), we present only the key summaries of results here or in the appendices. Detailed results for each taxon in the SSDM, including model selection, residual diagnostics, model accuracy and partial plots are presented in an html file accompanying this report (SSDM_Evaluation.html), which can be opened using any internet browser.

3.3.1 SSDM parameterisation and selection

The most frequently selected final model among macroinvertebrate taxa was the NULL, which was selected for 36 taxa (Figure 3-1, Table D-1). The remaining 68 taxa included 1-3 stressors in their final models. If two or more stressors were included in the same model, they were usually involved in interactions (:) as opposed to additive effects (+). Model selection results for individual taxa are presented in the SSDM_Evaluation.html file.

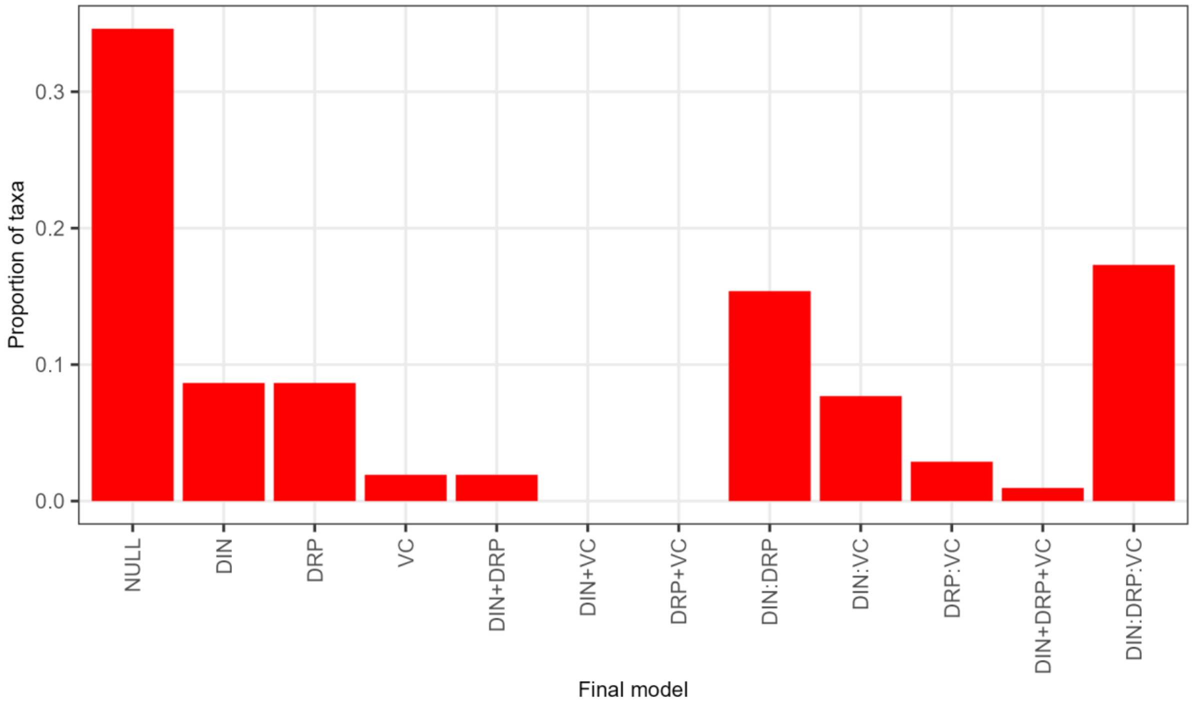


Figure 3-1: Distribution of final models among taxa modelled.

3.3.2 SSDM fit

Residual diagnostics indicated that the models were well fitted for most taxa in the SSDM (i.e., residual QQ plots were linear, and residuals were homogeneously distributed about 0.5 in residual versus predicted plots (see the SSDM_Evaluation.html fille for details). This means that the assumptions made by the SSDM generally met, and stressor effects were unbiased for most taxa in the SSDM.

Non-linear residual QQ plots were observed for some taxa, with observed residuals being either above (e.g., Aoteapsyche, Deleatidium) or below (e.g., Maoridiamesa, Ostracoda) that expected. This indicates that model fits were biased slightly higher (i.e., residuals >0.5) or lower (i.e., residuals <0.5) than observations in the training data, respectively. However, for most of those taxa (e.g., Maoridiamesa), these biases were not related to any stressor (residuals versus stressor plots). This means the stressor effects were generally unbiased, and well-fitted.

For a minority of taxa (e.g., *Aoteapsyche*, *Deleatidium*, Tanytarsini), there were some patterns in the residuals with respect to the stressors. Such patterns may reflect underfitting of the stressor smoother parameters at subsets of the stressor domains (i.e., the smoother non-linearity was not sufficiently complex) (Zuur et al. 2009, Landwehr et al. 1984). However, given that smoother complexity was determined via cross-validation, such biases likely reflect noise in the data, rather than some true effect of the stressor that was not captured by the smoother.

3.3.3 SSDM predictive accuracy

Cross-validated predictive accuracy of models of individual taxa was generally low for most taxa (AUC < 0.6), but good for some taxa (AUC>0.7) (Figure 3-2a). Pseudo-R^2 of stressor effects was also low, reaching up to only 0.12 for one taxon (Figure 3-2b). More variation was attributed to the random effects particularly site, followed by year within CZ, and CZ. Pseudo-R^2 of the full model (including fixed and random effects) ranged 0.02 – 0.67 and was generally comparable to previous random forest SSDMs on similar data (Wagenhoff et al. 2017). This means that SSDM predictions to each site were associated with very high uncertainty.

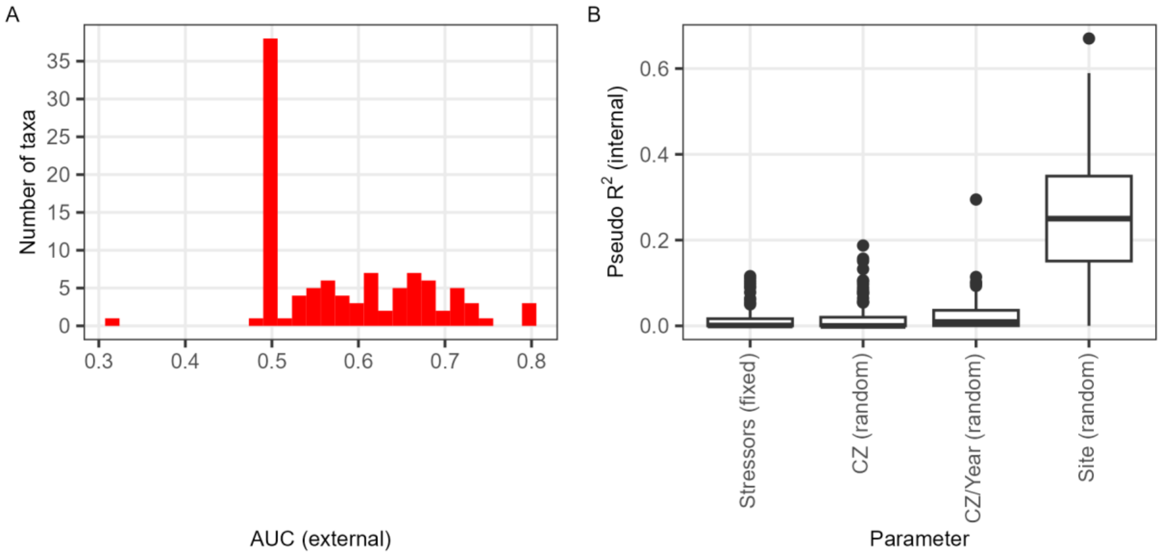


Figure 3-2: Distribution of a) AUC scores among taxa and b) pseudo-R2 among parameters in the SSDM. External indicates the statistic was calculated on novel data withheld from model training. Internal indicates the statistic was made on data used to train the model. Both panels include taxa whose final model was the null.

3.3.4 SSDM stressor sensitivity

Absolute sensitivity of taxa to certain stressors was often quite large, reaching >0.5 change in PO across stressor domains for some taxa (Figure 3-3). For many taxa the change in PO was nearly 100% relative to their maximum PO (i.e., relative sensitivity). Such large changes in PO could be sufficient to cause extirpation of those taxa at some point along each stressor (note – we test this assumption in Section 4).

The direction of taxon stressor responses was also broadly consistent with expectations. Taxa characteristic of high nutrient and/or sediment pollution typically increased in PO over stressor domains (e.g., *Physa*, Ostracoda, Oligochaeta, Chironomidae, Platyhelminthes), while other taxa decreased in PO (Appendix E). The form of these relationships was also consistent with expectations, being mostly simple non-linear increases or decreases, with occasional quadratic (i.e., hump-shaped) forms typical of nutrient enrichment (e.g., *Costachorema* response to DIN).

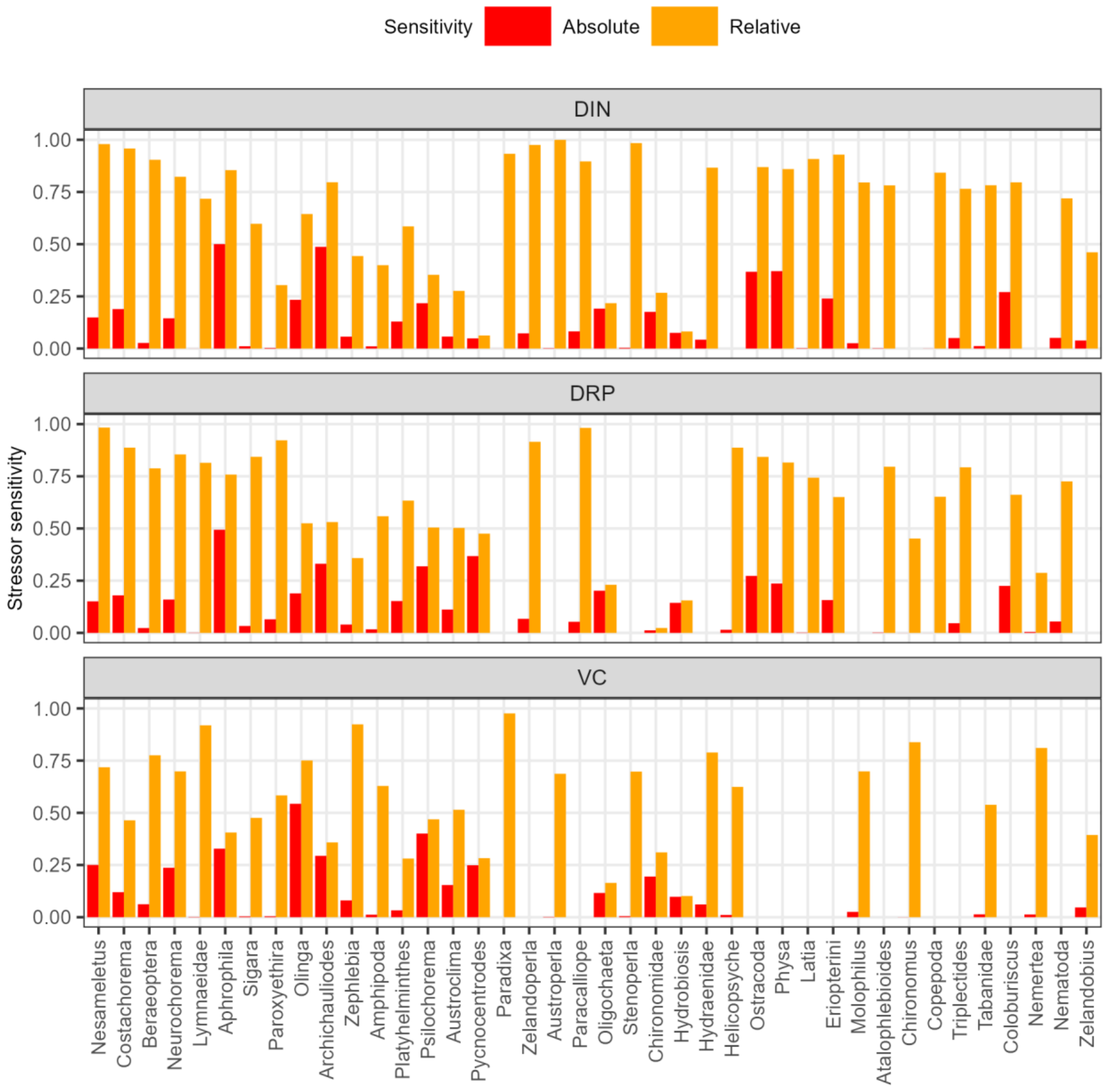


Figure 3-3: Absolute (red) and relative (orange) sensitivity of taxa to each stressor. Absolute sensitivity was defined as the range in predicted PO of a taxon across the domain of a stressor. Relative sensitivity was defined as absolute sensitivity divided by the maximum PO for that taxon. Taxa are ordered left to right by the mean rank of relative sensitivity among all three stressors. Only the 40 most sensitive taxa among all three stressors are shown (see Table D-1 for sensitivity of all taxa modelled).

3.4 Conclusions

- The SSDM was well-fitted and showed minimal signs of bias for most taxa.

- Predictive accuracy of models of individual taxa in the SSDM was generally low. This means there was a large amount of uncertainty associated with predicting taxon PO based solely on the effects of DIN, DRP and VC.
- Sensitivity of individual taxa in the SSDM to each stressor was often very high, and likely sufficient to extirpate some taxa as stressors worsened.
- The direction and form of taxon responses to stressors predicted by the SSDM were biologically sensible and consistent with expectations.

4 Predicting macroinvertebrate community stressor responses from the SSDM

4.1 Summary

The objectives of this section were to develop and evaluate an approach to predict community stressor responses using the SSDM. Our approach was based on Biological Extirpation Analysis (BEA) designed to simulate macroinvertebrate community extirpation in response to stressors. We integrated the SSDM into the BEA to:

- Assess the sensitivity of community extirpation to each stressor.
- Assess the sensitivity of community extirpation to different multi-stressor scenarios.
- Provide lookup tables that specify thresholds of DIN, DRP and VC required to minimise extirpation to certain levels (e.g., of ≤1%, 2.5%, 5% and 10% extirpation of taxa in a macroinvertebrate community) under different multi-stressor scenarios.

The simulations showed that:

- Individual stressors had significant and strong effects on community extirpation. Depending on the stressor, between 20-30% of the modelled community was extirpated as stressors worsened.
- Relative to the effects of DRP and VC, DIN had the strongest detrimental effects on community extirpation.
- Extirpation was strongly influenced by multi-stressor effects. The community extirpated by each single stressor increased approximately 10-fold when additional multi-stressor effects were taken into account.

Implications and outcomes:

- In order to achieve a specific invertebrate community extirpation state, thresholds placed on one water quality stressor may have to be much more stringent when/where the values of other water quality stressors are high.
- Regional plans aiming to reduce and/or control multiple stressors in concert will likely result in better environmental outcomes than plans focused on reduction of single stressors in isolation.

4.2 Methods

4.2.1 Biological extirpation analysis algorithm

We used the SSDM presented in Section 3 to estimate the proportion of macroinvertebrate taxa that may be locally extirpated as each stressor state worsened. With respect to DIN and DRP, states are assumed to worsen as their values increase. In contrast, VC state worsens as its value decreases. Our approach was based on biological extirpation analysis (BEA), developed for the USEPA by Cormier and Suter (2013), and now commonly used internationally (Zhao et al. 2016, Griffith et al. 2018, Shackleton et al. 2019).

Broadly, the analysis involved two key steps. First, for each stressor, we identify the taxon tolerance limit to that stressor likely to extirpate that taxon. This value is defined here as TT_{crit} (i.e., TT = taxon tolerance). TT_{crit} is the value of the stressor (in units of the stressor) above which the taxon is considered extirpated. This step is completed for all taxa (n=104) in the SSDM (subsequent sections will detail TT_{crit} estimation).

Second, for each stressor, the TT_{crit} values of taxa are ranked from the lowest (least tolerant) to highest (most tolerant) and transformed to yield a community tolerance distribution (CTD) for the stressor. The CTD is a cumulative probability distribution describing the proportion of taxa extirpated as a function of stressor state. The CTD is used to estimate $CT_{X,crit}$, the critical value of the stressor, at which some target proportion, *X,* of the macroinvertebrate community, *C*, cannot tolerate the stressor and are likely to become extirpated (i.e., CT = community tolerance). We estimated $CT_{X,crit}$ for *X*=1%, 2.5%, 5% and 10% community extirpation.

BEA assumptions

The assumptions of BEA as applied in this report are as for the SSDM (Section 3.2.5). A key additional assumption made by BEA is that taxa have an upper stressor tolerance threshold (i.e., TT_{crit}), but not a lower threshold[5]. It is therefore assumed that taxa can tolerate any value of the stressor below TT_{crit}. This assumption could be broken if a taxon's PO shows an increasing or hump-shaped stressor response. While hump-shaped responses were rare among taxa modelled here, increasing responses were common among pollution tolerant taxa. Such taxa may in fact be extirpated as stressors improve; however, taxa were always assumed to increase in abundance as the state of the stressor improved (lower nutrients/higher clarity).

Given the purpose of BEA, and national policy (NPSFM 2020), is to manage for excessive stressor effects, this assumption is arguably reasonable. However, application of the BEA should be carefully considered in light of this assumption to both management and regulatory decision-making.

Monte-Carlo simulation of CTDs, TT_{crit} and $CT_{x,crit}$ from the SSDM

The original BEA developed by Cormier and Suter (2013) used a weighted cumulative frequency distribution of taxon presences to estimate TT_{crit}. However, this does not account for a) sample non-independence, b) confounding processes (Roark et al. 2013, Franklin et al. 2019), and c) the interactive effects multiple stressors which could additively or synergistically alter community sensitivity to a stressor. We resolved these problems using Monte-Carlo simulations of the SSDM to estimate CTDs, TT_{crit} and $CT_{X,crit}$ as follows:

1. The final model of each taxon was used to simulate 1000 realisations of the change in PO of each taxon over the domain of a 'target stressor' (either DIN, DRP or VC).
    a. The domain of the target stressor was limited to between the 2.5-97.5% quantiles of the stressor observations to avoid making PO predictions in stressor space with limited observations.
    b. For each Monte Carlo realisation the predicted PO values were random draws from a normal distribution whose mean and standard error were, respectively, the fitted value, and standard error about the fitted value, of each taxon's best model.
    c. When simulating PO for the target stressor, the values for the 'other stressors' in the multi-stressor model were fixed at values deemed most relevant to management

---

[5] Reminder: By 'upper limit' we mean tolerance of high levels of stress. This does not mean that the stressor value is high in numerical value. For example, high levels of stress may be imposed by low visual clarity, but high levels of nutrients.

and regulation. We first ran the Monte-Carlo simulation while fixing the other two stressors at specific values shown in Table 4-1. This was done to analyse the sensitivity of CTDs to different environmental conditions likely encountered by managers and sense-check the algorithm. However, final CTDs and stressor thresholds were quantified while allowing the other two stressors to vary about the observed multi-stressor domain as described in Section 4.2.2.

2. The direction of change in PO (*increasing* or *decreasing*) with respect to the target stressor was classified for each realisation.
   - The direction of change was estimated from the sign of the median slope resulting from a Sen slope analysis on PO as a function of the target stressor (Sen, 1968). Clarity was inverse-transformed at this step such that all three stressors worsen as their numerical values increase; this simplified the encoding of the algorithm.
3. For each Monte Carlo realisation all estimates of PO along the target stressor gradient were converted into binary presences/absences.
   - Following Liu et al. (2005), we used the mean PO fitted to the training data as the threshold to determine presence/absence for a given taxon.
4. Compute TT_{crit} for each realisation as the maximum value of the target stressor associated with a predicted presence.
   - TT_{crit} is the value of the stressor (in units of the stressor) above which the taxon is considered extirpated, which according to step 3, occurs when the predicted PO for the taxon falls below its mean PO[6].
   - If, for a specific realisation, no presences were predicted throughout the domain of the target stressor, TT_{crit} was assigned *-Inf*. This result occurred when the effects of the entire domain of the target stressor (e.g. VC) were being simulated within particularly stressful regions of the domains of the other stressors (e.g. high values of DIN and/or DRP) with which the target stressor is interacting. In such a case, the high values of the other stressors may extirpate the taxon, irrespective of the value of the target stressor. Assignment of a value of *-Inf* to TT_{crit} is logical in a case like this, as it explicates that no reduction in the value of the target stressor can be sufficient to reverse extirpation, given the high values of the other stressors.
   - The final model of some taxon was NULL or did not include the target stressor. Some taxa were classified as *increasers* with respect to the target stressor (step 2). For all such cases we assumed the taxa concerned were either insensitive to the target stressor or found such environmental conditions favourable. To these taxa we assigned an TT_{crit} value of *+Inf* (i.e., undefined given the data and model set, following Cormier and Suter 2013).
5. Compute $CT_{X,crit}$ for each Monte Carlo realisation as follows:

---

[6] Using the mean species' PO as a threshold to infer species presence/absence and therefore extirpation, ensures that extirpation is not biased towards very rare or very common species (Liu et al. 2005).

a. Rank taxa by ascending TT_{crit}, such that taxa with the lowest TT_{crit} are ranked 1 (i.e., most sensitive to the stressor). Taxa tied for the same rank were assigned the maximum cumulative rank.

b. Calculate the CTD from the ranks of TT_{crit}. The CTD is a cumulative probability distribution whose y-values give the proportion of taxa locally extirpated as the stressor increases.

c. Calculate $CT_{X,crit}$ via linear interpolation of the CTD axes. The $CT_{X,crit}$ is the value of the stressor leading to extirpation of *X*% of the community. We calculated $CT_{X,crit}$ for *X*=1%, 2.5%, 5% and 10%.

Table 4-1: Fixed values of stressors used to sense-check the multi-stressor Monte-Carlo simulations. The fixed values of DRP, VC and DIN were chosen to span the relevant NPSFM attribute bands (i.e., those of VC, DRP, nitrate and ammonia), while remaining within the 95 percentile range of the observed data. Monte-Carlo simulations for any target stressor involved fixing the other two stressors at all possible pairings, yielding 4 x 4 = 16 multi-stressor scenarios for each target stressor.

Stressor	Fixed values for multi-stressor scenarios			
DRP (mg/L)	0.006	0.010	0.018	0.054
DIN (mg/L)	0.03	0.24	1.3	2.2
VC (m)	3	2.2	1.4	0.6

4.2.2 Optimistic-pessimistic multi-stressor extirpation scenarios

Our objective was to provide final stressor CTDs, TT_{crit} and $CT_{X,crit}$ estimates that characterised both the uncertainty in the model, as well as the range of multi-stressor conditions that were most likely given the data. To achieve this, we re-ran the Monte-Carlo simulation for each stressor, while allowing values of the other two stressors to vary randomly within multi-stressor domain of the observed data.

Algorithmically, this involved two steps conducted at Step 1 on each realisation of the of the simulation. First, the stressor of interest was binned into 50 equally sized bins spanning its domain. Second, for each stressor bin, we sampled (with replacement) one random row from the observed data containing observations of the stressor lying within the range of the bin. For each realisation, this resulted in 50 observations of the focal stressor, spaced equally across its domain, with the observations of the other two stressors sampled from the same rows in the observed data.

This approach resulted in 1000 simulated CTDs for each stressor, the variation of which characterised the range of multi-stressor conditions observed in the data. We then summarised this variation into three CTDs for each stressor. The upper 97.5 %, median, and lower 2.5 % quantile of the CTDs were defined as the *optimistic, typical,* and *pessimistic* CTD, respectively, for a given stressor. The *optimistic* CTD characterised the community extirpated by the stressor while other stressors varied about their least stressful state. The *typical* and *pessimistic* CTD characterised the community extirpated by the stressor while other stressors varied about their median or most stressful states respectively.

4.3 Results

4.3.1 Sensitivity of extirpation to multiple stressors

Differences in CTDs among multi-stressor conditions were consistent with our ecological understanding. The percent of the community extirpated increased as DIN (Figure 4-1a), DRP (Figure 4-1b) and VC (Figure 4-1c) conditions worsened (i.e., DIN/DRP increased, VC decreased) up to a maximum of approximately 30% of the modelled community. The level of extirpation for any value of the target stressor was generally higher when the other stressors were fixed at more stressful values.

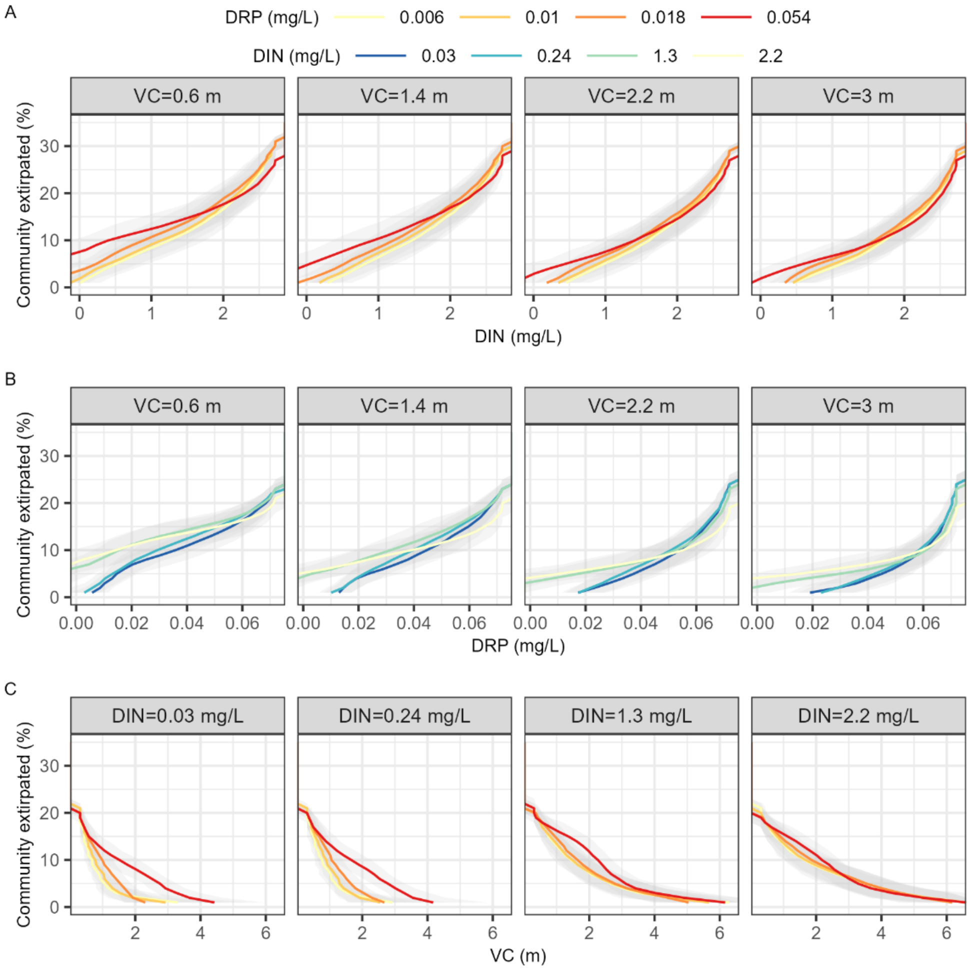


Figure 4-1: Community tolerance distributions (CTDs) for a) DIN, b) DRP and c) VC, under different multi-stressor scenarios. Each panel shows the CTD for a stressor on the x-axis, as other stressors (denoted by colour or panel headings) are fixed at different combinations of values. The y-axis indicates the % of taxa extirpated among those modelled in the SSDM, including those whose best models were the NULL (i.e., n=104 taxa). Because, TT_{crit} values for NULL taxa were assumed to be beyond the domain of the observed stressor data (i.e., Inf on the x-axis), the CTDs terminate on the y-axis at values <100%. Note that the x-axes, and fixed stressor levels (colour/panels), remain within the 2.5-97.5% quantiles of the stressor observations to avoid interpolating in poorly sampled stressor space.

Multi-stressor effects were strong. Consider the CTDs for DIN (Figure 4-1a). When DIN reached 0.24 mg/L, extirpation may vary between 0% and 10%, depending on levels of other stressors (0% extirpation under 0.006 mg/L DRP and 3m VC; 10% extirpation under 0.054 mg/L DRP and 0.6m VC;

Figure 4-1a). In this example, approximately 10 additional taxa were extirpated as a result of multi-stressor effects.

Stressor thresholds required to achieve different macroinvertebrate extirpation targets ($CT_{X,crit}$) are provided for each set of multi-stressor conditions tested in Appendix F.

4.3.2 Optimistic-pessimistic multi-stressor extirpation scenarios

Throughout our multi-stressor domain community extirpation ranges from 0% to 30% (Figure 4-2). Extirpation under the *pessimistic* scenario was approximately 10% (i.e., 10 taxa) higher than that observed under the *optimistic* scenario (Figure 4-2).

The differences between $CT_{X,crit}$ thresholds resulting from the *optimistic* and *pessimistic* CTDs were substantial. As an example, limiting DIN to <0.5 mg/L according to the *optimistic* scenario would minimise extirpation to <1% (Figure 4-3a). However, this same DIN concentration would lead to >5% extirpation under the *pessimistic* scenario (Figure 4-3a). It follows that to achieve a specific invertebrate community target, thresholds placed on one water quality stressor may have to be much more stringent when/where the values of other water quality stressors are high. Regional plans that aim to reduce and/or limit multiple stressors in concert will likely result in better environmental outcomes than plans focused on reduction of single stressors in isolation.

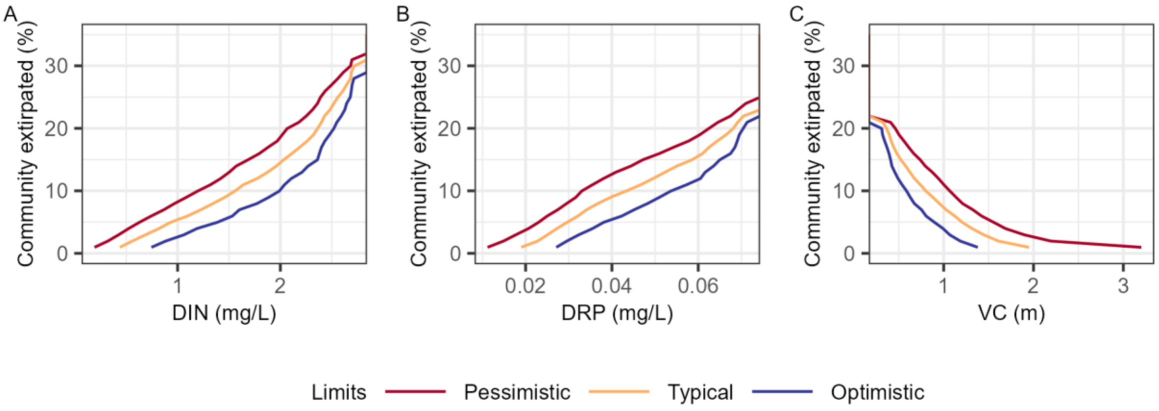


Figure 4-2: Community tolerance distributions (CTDs) for a) DIN, b) DRP and c) VC, for different multi-stressor scenarios (colour). The typical (orange), pessimistic (red) and optimistic (blue) multi-stressor scenarios show CTDs for each stressor while remaining stressors vary about their median, most or least stressful conditions, respectively, according to the observed data. Note that the y-axis pertains to the full set of taxa modelled in the SSDM (n=104).

The median taxon tolerance values (TT_{crit}) for the least tolerant taxa are shown in Figure 4-4. The five least tolerant taxa for each stressor predominantly belonged to EPT orders. However, many non-EPT groups were also well represented among the least tolerant. Moreover, there was substantial uncertainty in TT_{crit} for each taxon, with upper and lower confidence intervals often spanning the entire stressor domain[7] (Appendix H).

---

[7] Note that the width of ST_{crit} confidence intervals in Figure 4-4 include uncertainty caused by the different multi-stressor scenarios (the difference between optimistic-pessimistic lines in Figure 4-2). This uncertainty is uncertainty about the state of other stressors in the environment and is fully factored into the estimates of CT_{crit} in Figure 4-3 and should not, therefore, diminish confidence in those estimates. In practical terms, this means uncertainty in the state of multiple stressors makes the precise order or species extirpation resulting from any single stressor highly uncertain.

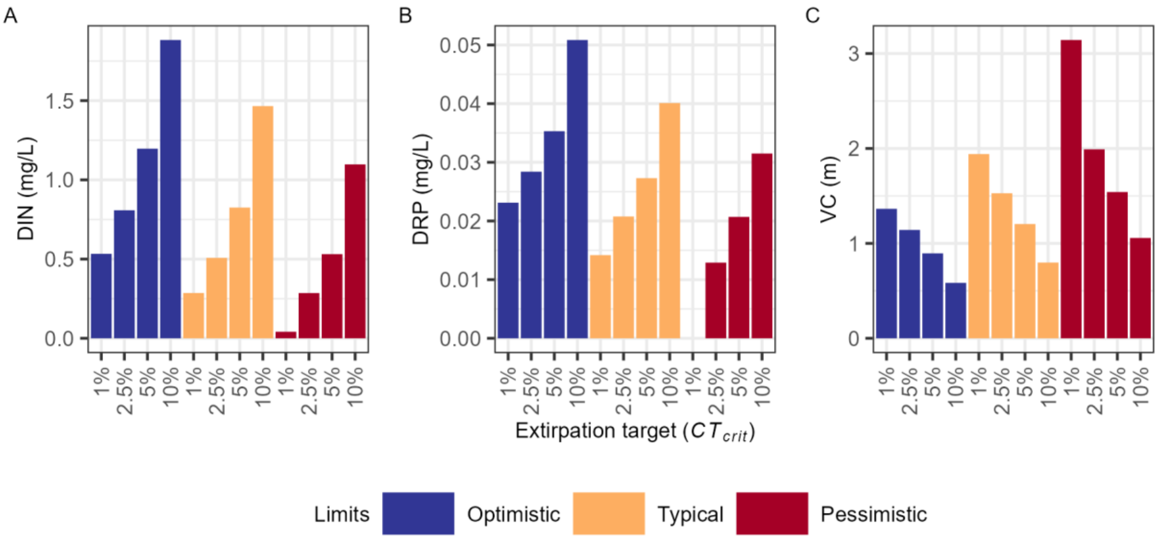


Figure 4-3: Thresholds for a) DIN, b) DRP and c) VC required to limit community extirpation to 1%, 2.5%, 5%, or 10% ($CT_{X,crit}$), for optimistic (blue), pessimistic (red) and typical (orange) thresholds. The typical, pessimistic, and optimistic multi-stressor scenarios show thresholds for each stressor while remaining stressors vary about their median, worst or least stressful conditions, respectively. Thresholds are repeated in table form in Appendix G**.**

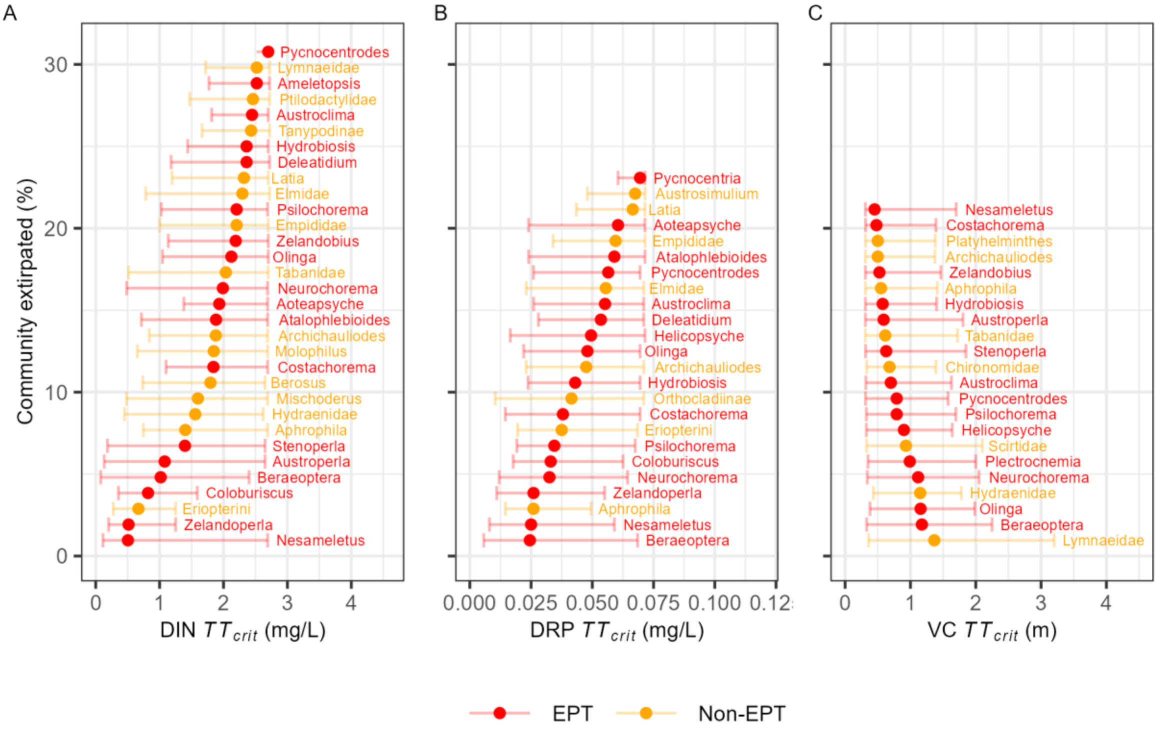


Figure 4-4: Median taxon tolerance values (TT_{crit}) for a) DIN, b) DRP, and c) VC resulting from the optimistic-pessimistic BEA Monte-Carlo simulation. TT_{crit} indicates the value of each stressor above which the taxon is likely extirpated. Taxa whose median TT_{crit} values were beyond the domain of the data (i.e., +Inf for

DIN and DRP or -Inf for VC) are not shown. Median and confidence intervals for TT_{crit} values for all taxa and stressors are shown as points and error bars and are available in Table H-1.

4.4 Conclusions

Taxon stressor sensitivity modelled by the SSDM was ecologically important:

- Individual stressors had significant and strong effects on community extirpation. Depending on the stressor, between 20-30% of the modelled community was extirpated as stressors worsened.
- Relative to the effects of DRP and VC, DIN had the strongest detrimental effects on community extirpation.
- Extirpation was strongly influenced by multi-stressor effects. The community extirpated by each single stressor increased approximately 10-fold when additional multi-stressor effects were taken into account.

Implications and outcomes:

- In order to achieve a specific invertebrate community target, thresholds placed on one water quality stressor may have to be much more stringent when/where the values of other water quality stressors are high.
- Regional plans that aim to reduce and/or limit multiple stressors in concert will likely result in better environmental outcomes than plans focused on reduction of single stressors in isolation.

5 Macroinvertebrate index model fitting and evaluation

5.1 Summary

The objective of this section was to fit and evaluate the predictive accuracy, sensitivity, and assumptions of macroinvertebrate index models (MIMs).

Model fitting, accuracy, and sensitivity of MIMs fitted to each index (MCI, QMCI and ASPM) broadly followed the same procedures as those of the SSDM (Section 3). We also used the SSDM extirpation simulations (Section 4) to evaluate index assumptions regarding taxon stressor tolerance.

Conclusions from this section were:

- No major violations of statistical assumptions were observed for the ASPM MIM. While slight departures from normality were observed at extreme values of MCI and QMCI, the fitted stressor effects of the MCI and QMCI MIMs were robust to these.
- Cross-validated predictive accuracy of MIMs was low for all indices.
- Macroinvertebrate indices were insensitive to DIN, DRP and VC. While the MIMs predicted that MCI and QMCI would increase if DIN, DRP and VC stressors improved, these increases would be insufficient to lift sites beyond their current NPSFM attribute bands.
- Index assumptions regarding taxon stressor tolerance were generally not met. This likely explains why indices were insensitive to stressors.
- MCI, QMCI and ASPM were not suitable indicators of DIN, DRP or VC pollution.

5.2 Methods

5.2.1 MIM parameterisation, selection and fit

Macroinvertebrate index models were developed for MCI, QMCI and ASPM. Indices were calculated as described in Section 2. The approach used to parameterise and select MIMs was very similar to that used for each taxon in the SSDM with some adjustments as follows.

The gamm4 package (Wood and Scheipl, 2017) was used to fit MCI and QMCI MIMs using a Gaussian distribution. The ASPM index, however, is best described as a continuous probability beta distribution (i.e., $0<\text{ASPM}<1$). The gamm4 package does not support this distribution family. We therefore used the mgcv package (Wood 2017) to fit ASPM MIMs with a beta distribution.

All MIM GAMMs were fitted using the same smoother basis dimension parameterisations as used for each taxon in the SSDM. However, we did not test the effects of increasing K on model fit for MIMs as the effective degrees of freedom of final models was consistently <2. Consequently $K = 6$ was considered adequate to characterise the non-linearity of the MIMs.

The statistical assumptions of MIMs are broadly the same as those for the SSDM (Sections 3.2.2). The evaluation of those assumptions therefore followed similar procedures, but with adjustments to accommodate the different distributions of the indices. Specifically, we evaluated the assumption that MCI and QMCI were normally distributed using standardised residuals. The assumption that ASPM was beta distributed was assessed using quantile residuals following Hartig (2022).

5.2.2 MIM predictive accuracy and sensitivity

We used 10-fold cross-validation to determine predictive accuracy of each MIM. Nash-Sutcliffe Efficiency (NSE, Nash and Sutcliffe 1970) was the statistic used to determine predictive accuracy during cross-validation. The NSE statistic provides a measure of predictive performance by indicating how closely a plot of observed versus predicted values lies to the 1:1 line (i.e., the degree to which two sets of values coincide). NSE values range from $-\infty$ to 1. An NSE of 1 corresponds to a perfect match between predictions and the observed data, an NSE of 0 indicates that the model predictions are as accurate as the mean of the observed data; and an NSE less than 0 indicates that the observed mean is a better predictor than the model.

We also calculated the coefficient of determination (R^2) of each final model. The coefficient of determination was calculated while iteratively excluding each random effect from model predictions, to assess the variance attributed to each effect.

Sensitivity of each macroinvertebrate index was estimated for each stressor following the same procedures as for PO of taxa in the SSDM, but instead using the predicted values of each index.

5.2.3 Evaluating the taxon-specific stressor tolerance assumptions of macroinvertebrate indices

Values of MCI, QMCI and ASPM are essentially based on weighted summations of taxon tolerance scores—the a_i-values of Equations 1 and 2. These tolerance scores are based on expert opinion, and are not estimated from data. When these metrics are used as indicators of specific water quality stressors the following assumptions are made:

1. taxon tolerance scores derived from expert opinion positively correlate with data-based estimates of taxon-specific tolerances (derived from the BEA, in our case); and
2. taxon-specific tolerances are correlated across multiple stressors or, put another way, one set of taxon-specific tolerances is indicative of tolerances to all stressors we are applying the community indices to.

The first assumption was evaluated by correlating the median taxon tolerance (TT_{crit}) values simulated by the BEA (Table H-1) against the hard-bottom tolerance scores used by each index. Assumption 2 was assessed by analysing the co-variance in TT_{crit} values among stressors.

5.3 Results

5.3.1 MIM parameterisation, selection and fit

The final model for MCI and QMCI MIMs included all three stressors as additive effects but was the NULL model for ASPM (Table 5-1). Consequently, there was support for stressor effects on MCI and QMCI but not ASPM.

Residual QQ plots for both the MCI (Figure I-1) and QMCI (Figure I-2) MIM showed departures from normality at the tails of the residual distribution. This was probably caused by the fact that both indices have upper and lower bounds. That is, MCI and QMCI are bounded between 0-200 and 0-10, respectively, and therefore they do not follow a normal distribution at extreme ends of their distributions. However, the residuals were homogeneous with respect to each stressor (residuals versus stressor plots in Figure I-1–Figure I-2), suggesting the stressor effects were unbiased and robust to this departure from normality. The QQ plot for ASPM was approximately linear and residuals were homogeneous with respect to each stressor suggesting the assumption that ASPM was beta distributed was correct (Figure I-3).

Table 5-1: Model selection and performance results for the MIMs. The final model is that which minimised AIC among candidate models for each index. Model Akaike weight (AIC_w), total effective degrees of freedom (EDF), cross-validated Nash-Sutcliffe's Efficiency (CV NSE) and parameter-specific coefficient of determination (R^2) are shown.

Metric	Final model	AIC	AIC_w	Total EDF	CV NSE	R^2 stressors	R^2 Site	R^2 CZ/Year	R^2 CZ
MCI	DIN+DRP+VC	21763.76	0.53	4	0.04	0.12	0.65	0.03	0.04
QMCI	DIN+DRP+VC	8852.3	0.56	3.5	0.02	0.07	0.55	0.05	0.04
ASPM	NULL	-5664.67	0.48	0	-0.1	-	-	-	-

5.3.2 MIM predictive accuracy

Model predictive performance for all MIMs was very low, with NSE reaching only as high as 0.04 (Table 5-1). The R^2 of the fixed effects of the final model was also very low, with a substantially larger amount of variation attributed to the random effects, particularly site (Table 5-1).

5.3.3 MIM stressor sensitivity

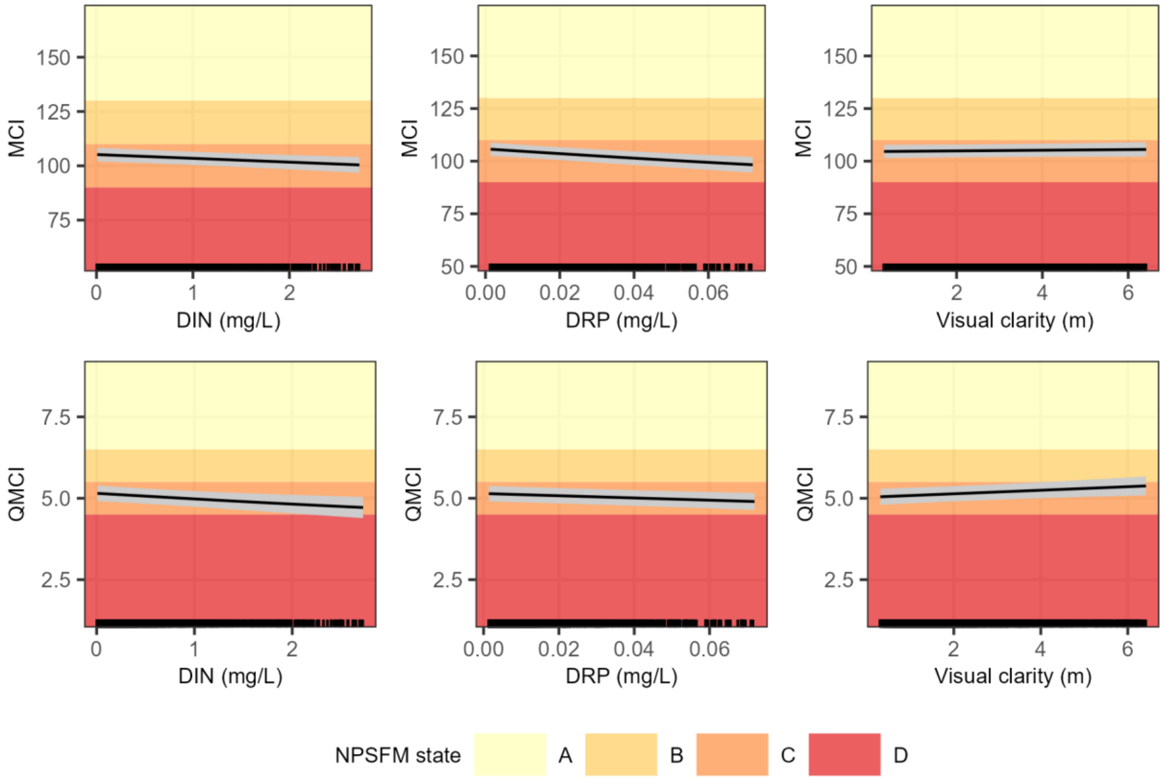


Figure 5-1: Partial plots of MCI (top row), QMCI (bottom row) MIM predictions to each stressor relative to NPSFM state bands (colour).

Unlike the SSDM, the sensitivity of macroinvertebrate indices to stressors was very low. In absolute units, the sensitivity of MCI and QMCI to each stressor was insufficient to shift macroinvertebrate attribute states beyond a single NPSFM band (Figure 5-1). Given the data and our modelling approach, the NPSFM macroinvertebrate attributes appear to be very poor indicators of DIN, DRP and VC impacts.

5.3.4 Evaluating macroinvertebrate index taxon tolerance assumptions

Support for the first assumption—that index tolerance scores positively correlate with data-based stressor tolerance—was very weak (Figure 5-2). The correlations between MCI tolerance and stressor tolerance modelled by the SSDM were weak for all three stressors (Spearman's Rho coefficient = 0.36, 0.40 and 0.16 for DIN, DRP and VC tolerance, respectively).

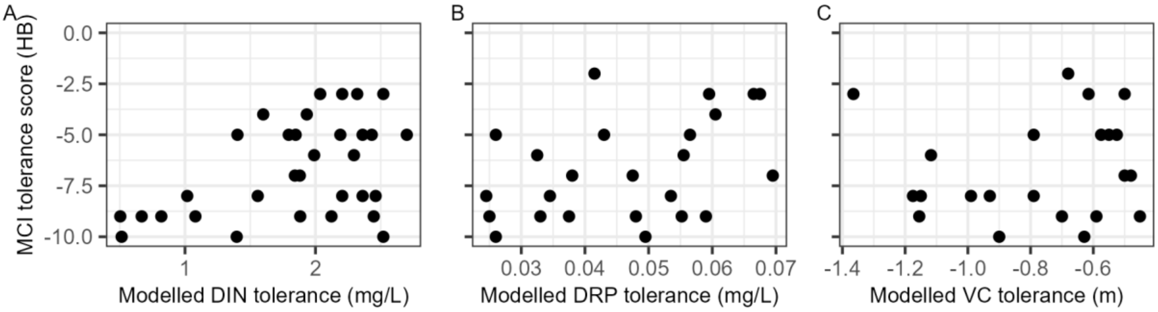


Figure 5-2: Covariance between MCI hard-bottom (HB) tolerance scores (y-axis) and modelled a) DIN, b) DRP, and c) VC, tolerance. Modelled stressor tolerances are median TT_{crit} values as modelled by the SSDM from Table H-1, excluding any taxa whose tolerances were beyond the stressor domains (i.e., Inf/-Inf). Some axes were inversed to ensure the least tolerant taxa were on the bottom left of each panel.

Support for the second assumption—that taxon tolerances are the same for all stressors—was also very weak (Figure 5-3). While simulated DIN and DRP tolerance was weakly positively correlated (Spearman's Rho = 0.67) (Figure 5-3a), there were no correlations between either nutrient stressor and VC (DIN – VC Rho = 0.30, DRP – VC Rho = 0.32) (Figure 5-3ab). Thus, taxon tolerance was typically different for each stressor.

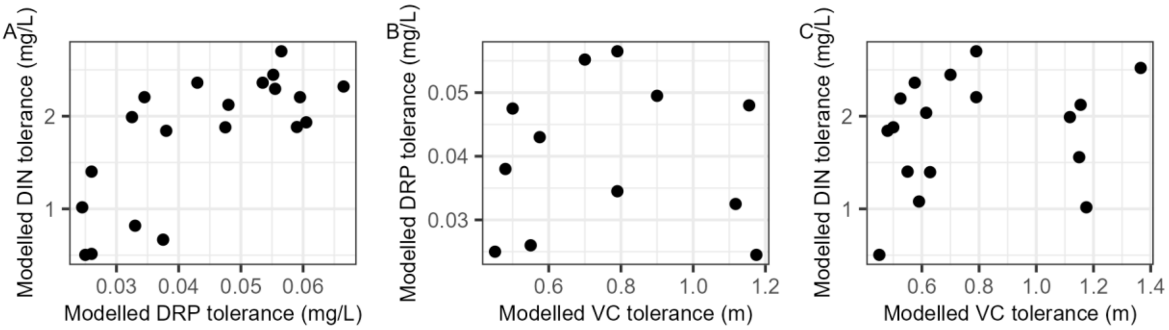


Figure 5-3: Covariance between modelled tolerance of a) DIN and DRP, b) DRP and VC, and c) DIN and VC. Modelled stressor tolerances are median TT_{crit} values as modelled by the SSDM from Table H-1, excluding any taxa whose tolerances were beyond the stressor domains (i.e., Inf/-Inf).

5.4 Conclusions

- No major violations of statistical assumptions were observed for the ASPM MIM. While slight departures from normality were observed at extreme values of MCI and QMCI, the fitted stressor effects of the MCI and QMCI MIMs were robust to these.
- Cross-validated predictive accuracy of MIMs was low for all indices.
- Macroinvertebrate indices were insensitive to DIN, DRP and VC. This means it would not be possible to achieve NPSFM macroinvertebrate target states by reducing DIN, DRP or VC as specified by national policy.
- Index assumptions regarding taxon stressor tolerance were generally not met. This likely explains why indices were insensitive to stressors.
- MCI, QMCI and ASPM were not suitable indicators of DIN, DRP or VC pollution.

6 Discussion

6.1 Summary

This report is the second in a series specifically framed around offering information on ecological changes anticipated from alternative pollutant regimes modelled by the FWMT (Healthy Waters, Auckland Council). The earlier NIWA-FWMT report had prioritised the development of macroinvertebrate response models as the most suitable means of conveying broader ecosystem health information from changes in contaminant regime, including the exploration of index-based models (MIMs) and development of taxon-specific models (SSDM) prior to development of community extirpation models (BEA).

The objectives of this report were to 1) evaluate MIMs and SSDMs as approaches to model macroinvertebrate responses to DIN, DRP and VC, and 2) develop lookup tables quantifying those responses that can be used with FWMT output.

The models were evaluated using multiple criteria:

- Model fit – did the models meet the statistical assumptions underlying them? Was there any bias in estimated stressor effects?
- Uncertainty – how uncertain were model predictions?
- Sensitivity – how sensitive are model responses to each stressor?
- Ecological plausibility – how well do models align with our ecological understanding?

Both MIMs and the SSDM were difficult to distinguish based on Criterion 1. Residual diagnostics of both model approaches indicated they were generally well-fitted, and there was little bias in their predicted stressor effects (see Section 3.2.2 for the SSDM and Section 5.2.1 for MIMs).

The two approaches were also difficult to distinguish based on Criterion 2. The accuracy of model predictions made on novel data was low for both the SSDM (Figure 3-2a) and the MIMs (Table 5-1). Irrespective of the response variable, the variance explained by stressor terms of best models was low (<10%). The low amount of variance explained by the stressors is consistent with previous efforts to model responses of New Zealand macroinvertebrate to nutrients and suspended sediment (Clapcott and Goodwin 2014, Wagenhoff et al. 2017). There is clearly a very high degree of uncertainty associated with predicting macroinvertebrate communities as a function of nutrients and suspended sediment.

There were clear differences between approaches with respect to the sensitivity of models to stressors and the ecological plausibility of model fits (Criteria 3-4). The sensitivity of PO of taxa to each stressor predicted by the SSDM was often very high (Figure 3-3). The BEA—which has as its foundation the SSDM—indicated this sensitivity was high and could lead to 20-30% of the macroinvertebrate community being extirpated by each stressor (Figure 4-2). The SSDM and BEA predicted responses to stressors that were biologically plausible, and generally consistent with contemporary models (Clapcott et al. 2017, Wagenhoff et al. 2017). Consequently, nutrient and VC criteria to meet various macroinvertebrate community targets were estimated and presented in Section 4.3 using the BEA derived from the SSDM.

In contrast, the MIMs were insensitive to stressors and ecologically implausible – that is the assumptions they made regarding taxon tolerances to multiple stressors were not supported. Given the data and modelling approach, MCI and QMCI changed very little as a function of variation in DIN, DRP and VC (Figure 5-1). ASPM was completely insensitive to each stressor. The insensitivity of macroinvertebrate attributes was likely a result of the taxon-specific tolerance scores derived from expert opinion being poorly related to observed taxon-specific tolerances derived from data (our SSDM; Figure 5-2 - Figure 5-3).

Models that directly related NPSFM attributes to water quality stressors (our MIMs) clearly have high relevance to the NPSFM. However, despite this relevance, we found no evidence that they are effective for understanding or managing the effects of nutrients or visual clarity stressors on instream values (macroinvertebrates in our case). By contrast, we presented approaches based on an SSDM that, while arguably being less directly relevant to central policy, should be more effective for managing multiple water quality stressors to meet macroinvertebrate ecological objectives.

6.2 Limitations, assumptions, and application of the SSDM

Our multilevel modelling framework allows us to partition spatial and temporal variance in macroinvertebrate occurrence into two broad classes of effects: the ‘fixed’ effects of the stressors of interest, and the ‘random’ effects of other sources of spatial and temporal variation that are not the focus of our investigation. Our fixed-effect terms were parameterised to capture the effects of stressors whose dynamics are outputs of Auckland Council’s FWMT.

Our models were relatively simple, focusing on the effects of DIN, DRP and VC fixed effects. Variance attributed to processes varying among sites (e.g., habitat, elevation) and climate were partitioned into random effects. The SSDM therefore assumed that stressor effects were independent of those processes and do not vary simply due to changes in spatial location or time.

A consequence of this assumption is that SSDM predictions to any site were highly uncertain. The stressor fixed effects explained only a small amount of variance relative to the random effects, particularly site, which cannot be used for inference.

Whether such uncertainty is acceptable for decision making depends on the intended application of the model (MacNally et al. 2017, White et al. 2023). Additional parameters could be included in the SSDM to help reduce this uncertainty. However, many of the processes associated with the random effects are likely not simulated by the FWMT (e.g., source-sink dynamics, proximity to colonists), or are static (e.g., elevation). The value of adding these variables as fixed effects is therefore questionable, given our objectives.

Ultimately, it is most important that we are confident about the effects we can simulate with the FWMT. By keeping models simple, we ensured that:

- The risk of confounding stressor effects with other covarying processes was minimised (Dormann et al. 2013).
- The SSDM was not over- or under-fitted, thereby making the stressor effects more generally applicable (Merow et al. 2014).

The resulting stressor effects estimated by the SSDM were statistically robust, biologically consistent, and ecologically important. The model was not accurate at predicting exactly which taxa will be present at a given site, given all processes operating at that site. It is unrealistic to expect this from correlative statistical models. However, the model should be robust for evaluating the behaviour of macroinvertebrate community dynamics to nutrient and sediment scenarios, assuming other drivers are constant.

The BEA presented in Section 5 is one such application. The analysis specifies stressor thresholds required to achieve macroinvertebrate extirpation targets under different multi-stressor scenarios. These thresholds can be compared to nutrient and sediment forecasts made by the FWMT to assess how each stressor may be limiting taxon presence. The thresholds do not imply that the taxa predicted to be present will be truly observed at any given site. Other unmodelled processes are likely to additionally limit which taxa will be present at a site. However, the analysis provides a tool for managers to diagnose how nutrients and sediment may be limiting macroinvertebrate community composition independently of other factors.

6.3 Next steps

This study showed that the SSDM was a good foundation for modelling the response of macroinvertebrate composition to nutrient and sediment scenarios. The next steps should continue to develop application of the SSDM to Auckland Council's specific requirements, particularly simulation of continuous community responses to FWMT outputs. To do so there are several factors that need to be considered:

1. Should SSDM forecasts consider only a subset of taxa important to the Auckland region?
2. What alternative approaches are suitable to aggregate responses of individual taxa from the SSDM into community responses?
3. How should we calibrate FWMT TSS forecasts into VC for input into the SSDM?
4. Should additional FWMT parameters should be added to the SSDM as fixed effects?

These points are expanded below.

1. All modelling activities in this report were based on national macroinvertebrate datasets. However, less than 70% of those taxa were sensitive to the stressors. Because some taxa may be restricted to certain regions, not all these taxa are necessarily relevant to the Auckland region. We therefore recommend a workshop with macroinvertebrate experts to determine if models should only consider certain taxa, and which ones are important. This workshop might also identify key taxa, that are indicators of biological, functional, or cultural significance in Auckland for forecasting. Sub-setting to such taxa may make SSDM forecasts more regionally relevant. This wouldn't require model refits, but rather selection of taxa from the existing SSDM from which to draw inferences.
2. In this report, the SSDM was used via BEA to quantify water quality thresholds that can be used to evaluate FWMT output. Alternative algorithms could be applied to the existing SSDMs to model different aspects of the macroinvertebrate community, such as taxon turnover (Ellis et al. 2012), or richness (Schipper et al. 2014). These approaches are better suited to capture community responses to continuous observations of nutrients and sediment.
3. The FWMT outputs total suspended solids (TSS), not VC. To forecast the effects of TSS using the SSDM, we need to develop functions that convert TSS to VC. Recent modelling work

recommended against this due to what they called “high” residual error in VC-TSS linear regressions (RMSE: 22-32 cm) (Montgomery et al. 2022). However, the magnitude of statistical uncertainty that may be deemed tractable (or problematic) very much depends on how model outputs are to be used (MacNally et al. 2017, White et al. 2023). The uncertainty in VC-TSS calibrations, while not acceptable for the objectives considered by Montgomery et al. (2022), may be acceptable for macroinvertebrate modelling, where:

- We require relationships between annual medians (not the daily observations used by Montgomery et al. 2022, which tend to inflate RMSE relative to relationships at coarser resolutions); and
- Where the statistical uncertainty in the relationship between TSS and VC is actually small, relative to the statistical uncertainty about the relationships between stressors and taxon responses in the SSDM.

Further, one could suggest that uncertainty about the relationship between TSS and VC could be further reduced through more advanced partitioning of error sources using multilevel modelling—like the models used here as inputs to the SSDM. An analysis using multilevel models may help us to better reduce uncertainty concerning the relationship between VC and TSS, as we could (a) account for error induced by temporal autocorrelation, and (b) borrow strength across sampling units in space and time to better isolate the ‘fixed-effects’ between VC and TSS. Simulations could also be used to assess the significance of it to macroinvertebrate forecasts with the SSDM.

4. The SSDM considered only nutrients and sediment as fixed effects that could be simulated by the FWMT. The FWMT also simulates river flow. Inclusion of flow as a fixed effect in the SSDM would be possible via forecasts produced by NIWAs national flow models (TopNet). However, this would require investment in gathering/tidying hindcasted hydrology timeseries for the national dataset and SSDM refit/re-evaluation.

There is also a risk that the indirect and direct effects of flow, sediment and nutrients become confounded in the translation of FWMT output to the SSDM. Flow likely has direct effects on macroinvertebrate presence, but also indirect effects via its influence on nutrients and sediment dynamics. While it may be possible to partition the indirect and direct pathways of flow, nutrients, and sediment in the FWMT, our ability to do so with the SSDM is more limited, owing to the correlational nature of the data. Failure to properly translate the indirect and direct pathways of flow between the FWMT and the SSDM may bias the influence of each process in simulations of the SSDM using FWMT output.

This risk needs to be considered given the relative importance of flow and stressor management in the Auckland region. If stressor management is a more pressing management concern than flow, our approach of partitioning flow into random effects via climate zones, may be more robust than attempting to model flow as a fixed effect.

7 Acknowledgements

We would like to thank Tom Stephens and Theodore Kpodonu, Anika Kuczynski and Elizabeth Graham for many helpful and insightful discussions during development of this report. We are also very grateful to John-Mark Wooley for providing the climate zone geo-spatial layers used to partition data in this report. Thank you also to Mike Bargh, Brian Smith, Juliet Milne and Richard Measures for help with the NRQWN data.

8 References

Akaike, H. (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19(6): 716–723.

Auckland Council (2021a) Freshwater management tool: report 1. Baseline Inputs. Prepared by the Auckland Council Healthy Waters Department, Paradigm Environmental, and Morphum Environmental Ltd. for Auckland Council.

Auckland Council (2021b) Freshwater management tool: report 2. Baseline configuration and performance. Prepared by the Auckland Council Healthy Waters Department, Paradigm Environmental, and Morphum Environmental Ltd. for Auckland Council.

Auckland Council (2021c) Freshwater management tool: report 3. Current State Assessment (Rivers). Prepared by the Auckland Council Healthy Waters Department, Paradigm Environmental, and Morphum Environmental Ltd. for Auckland Council.

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W. (2012) Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution, 3(2): 327-338.

Burnham, K. P., Anderson, D. R. (2002) Model selection and multimodel inference: A practical information-theoretic approach. Springer, New York.

Clapcott, J., Wagenhoff, A., Neale, M., Storey, R., Smith, B., Death, R., Harding, J., Matthaei, C., Quinn, J., Collier, K., Atalah, J., Goodwin, E., Rabel, H., Mackman, J., Young, R. (2017) Macroinvertebrate metrics for the National Policy Statement for Freshwater Management. Prepared for the Ministry for the Environment. Cawthron Report No. 3073. 139 p. plus appendices.

Collier, K.J. (2008) Average score per metric: An alternative metric aggregation method for assessing wadeable stream health. New Zealand Journal of Marine and Freshwater Research, 42(4): 367-378.

Cormier, S.M., Suter, G.W. (2013) A method for deriving water-quality benchmarks using field data. Environmental Toxicology and Chemistry, 32(2): 255-262.

Depree, C., Clapcott, J., Booker, D.J., Franklin, P., Hickey, C., Matheson, F., Shelley, J., Unwin, M., Goodwin, E., Mackman, J., Rabel, H., Wagenhoff, A. (2017) Development of ecosystem health thresholds for suspended and deposited sediment in New Zealand rivers and streams. NIWA client report, 2017076HN: 315.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S. (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography, 36(1): 27-46.

Dunn P.K., Smyth, G.K. (1996) Randomized quantile residuals. Journal of Computational and Graphical Statistics, 5:1-10.

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

Franklin, P., Stoffels, R., Clapcott, J., Booker, D.J., Wagenhoff, A., Hickey, C. (2019) Deriving potential fine sediment attribute thresholds for the National Objectives Framework. NIWA client report, 2019039HN: 290.

Griffith, M.B., Zheng, L., Cormier, S.M. (2018) Using extirpation to evaluate the ionic tolerance of freshwater fish. Environmental Toxicology and Chemistry, 37(3):871-833.

Hartig, F. (2022) DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. R package version 0.4.6. http://florianhartig.github.io/DHARMa

Helsel, D.R. (2012) Statistics for Censored Environmental Data Using Minitab and R. Second edition. Wiley.

Kidson, J.W., Renwick, J.A. (2001) Patterns of convection in the tropical pacific and their influence on New Zealand weather. International Journal of Climatology, 22: 151-174.

Landwehr, J.M., Pregibon, D., Shoemaker, A.C. (1984) Graphical methods for assessing logistic regression models. Journal of the American Statistical Association, 79(385):61-71.

Lui, C., Berry, P.M., Dawson, T.P., Pearson, R.G. (2005) Selecting thresholds of occurrence in the prediction of species distributions. Ecography, 28(3): 385-393.

MacNally, R., Duncan, R.P., Thomson, J.R., Yen, J.D.L. (2017) Model selection using information criteria, but is the “best” model any good? Journal of Applied Ecology, 55:1441-1444.

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: 639-649.

Merow, C., Smith, M.J., Edwards Jr, T.C., Guisan, A., McMahon, S.M., Normand, S., Thuiller, W., Wüest, R.O., Zimmermann, N.E., Elith, J. (2014) What do we gain from simplicity versus complexity in species distribution models? Ecography, 37(12): 1267-1281.

Ministry for the Environment (2020) National Policy Statement for Freshwater Management 2020. Wellington: Ministry for the Environment.

Ministry for the Environment (2022) Setting instream nutrient concentration thresholds for nutrient-affected attributes in rivers: Guidance on implementing Clause 3.13 of the NPS-FM. Wellington: Ministry for the Environment.

Montgomery, J., Semadeni-Davies, A., Haddadchi, A. (2022) Relating suspended sediment concentration to visual clarity for rivers in Auckland region: Stage 1: review of available datasets and development of relationships using existing data. NIWA client report, 2022173CH: 49.

Mullen, A.B. (1995) On the linearity and stability of Southern Oscillation-climate relationships for New Zealand. International Journal of Climatology, 15: 1365-1386.

Nash, J.E., Sutcliffe, J.V. (1970) River flow forecasting through conceptual models: Part I — A Spatial modelling of river water-quality state 47 discussion of principles. Journal of Hydrology, 10: 282–290.

NEMS (2022) National Environmental Monitoring Standards - Macroinvertebrates: Collection and Processing of Macroinvertebrate Samples from Rivers and Streams.

Olsen, L. R. (2017) groupdata2: Creating groups from data. R package version 1.0.0. https://CRAN.R-project.org/package=groupdata2

Orr, J.A., Vinebrooke, R.D., Jackson, M.C., Kroeker, K.J., Kordas, R.L., Mantyka-Pringle, C., Van den Brink, P.J., De Laender, F., Stoks, R., Holmstrup, M., Matthaei, C.D., Monk, W.A., Penk, M.R., Leuzinger, S., Schäfer, R.B., Piggott, J.J. (2020) Towards a unified study of multiple stressors: divisions and common goals across research disciplines. Proceedings of the Royal Society B, 287: 20200421.

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): e49873.

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

Pinheiro, J. C., Bates, D. M. (2000) Mixed-effects models in S and S-plus. Springer Verlag New York LLC.

Roark, S.A., Wolf, C.F., DeJong, G.D., Gensemer, R.W., Canton, S.P. (2013) Influences of subsampling and modelling assumptions on the US Environmental Protection Agency field-based benchmark for conductivity. Integrated Environmental Assessment and Management, 9(3): 522-534.

Shackleton, M., Holland, A., Stitz, L., McInerney, P. (2019) Macroinvertebrate Responses to Conductivity in Different Bioregions of Victoria, Australia. Environmental Toxicology and Chemistry, 38(6): 1334-1342.

Smith, D.G., McBride, G.B. (1990) New Zealand's National Water Quality Monitoring Network – Design and First Year's Operation. Water Resources Bulletin, 26:767-775.

Stark, J.D. (1985) A macroinvertebrate community index of water quality for stony streams. Water & Soil Miscellaneous Publication 87: 53 p.

Stark, J.D., Boothroyd, I..KG., Harding, J.S., Maxted, J.R., Scarsbrook, M.R. (2001) Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Working Group Report No.1. Prepared for the Ministry for the Environment. Sustainable Management Fund Project No. 5103. 57 p.

Stark, J.D., Maxted, J.R. (2007) A user guide for the macroinvertebrate community index. Cawthron Report, Prepared for the Ministry for the Environment.

Wagenhoff, A., Clapcott, J., Lau, K.E., Lewis, G.D., Young, R.G. (2017) Identifying congruence in stream assemblage thresholds in response to nutrient and sediment gradients for limit setting. Ecological Applications, 27(2): 469-484.

White, R.S.A, Graham, E., Kuczynski, A., Kilroy, C., Stoffels, R. (2022) Prospects for ecological response modelling using FWMT output. NIWA client report, 2022235CH: 67.

White, R.S.A., Stoffels, R.J., Whitehead, A.L. (2023) Assigning trend-based conservation status despite high uncertainty. Conservation Biology, 37: c14084.

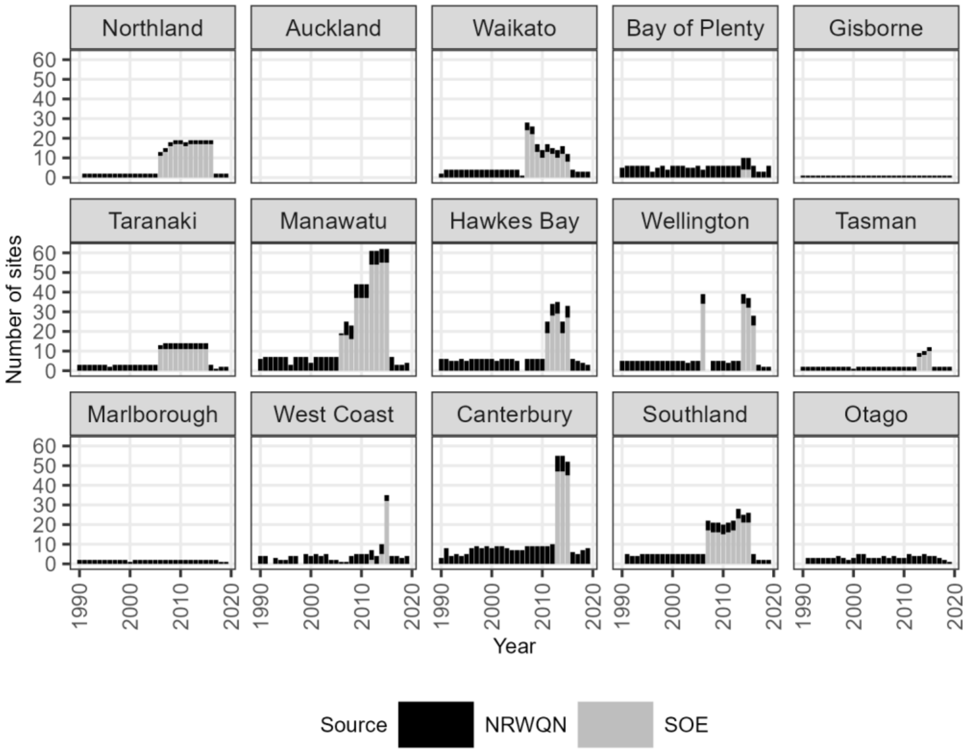
Wood, S. N. (2017) Generalised additive models: An introduction with R (2nd ed.). Chapman and Hall/CRC Texts In Statistical Science.

Wood, S.N., Scheipl, F. (2017) gamm4: Generalised additive mixed models using mgcv and lme4. R package version 0.2-5.

Zhao, Q., Jia, X., Xia, R., Lin, J., Zhang, Y. (2016) A field-based method to derive macroinvertebrate benchmark for specific conductivity adapted for small data sets and demonstrated in the Hun-Tai River Basin, Northeast China. Environmental Pollution, 216: 902-910.

Zuur, A., Ieno, E. N., Walker, N., Savaliev, A. A., Smith, G. M. (2009) Mixed effects models and extensions in ecology with R. Springer.

Appendix A Spatial and temporal distribution of sites



f

Figure A-1: Distribution of sites among regions, years and sources.

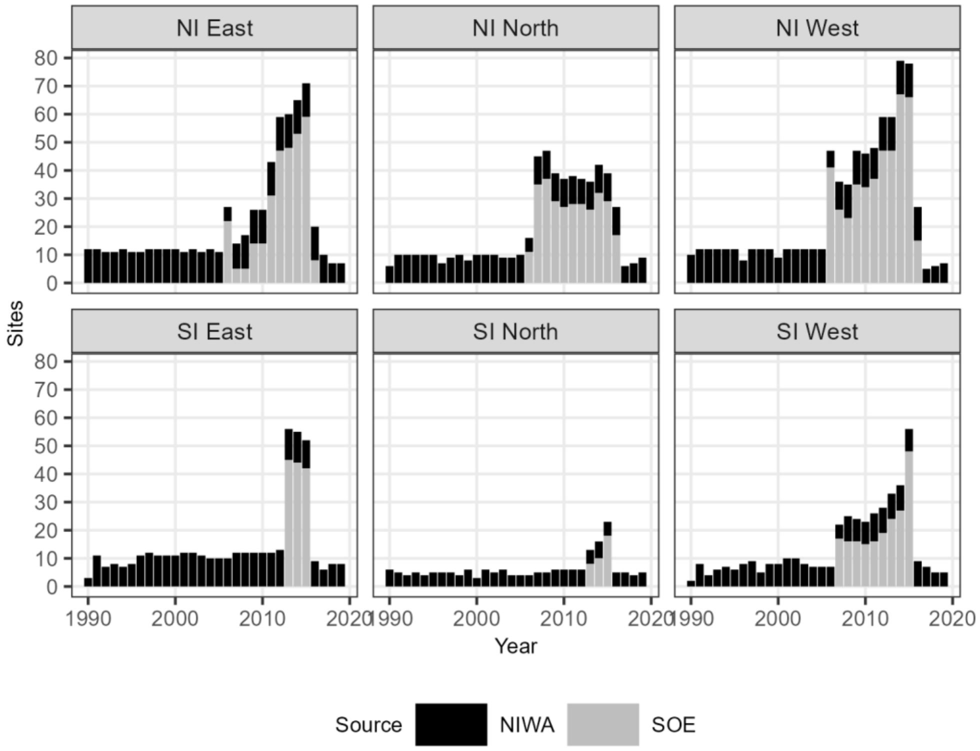


Figure A-2: Distribution of sites among climate zones, years, and sources.

Appendix B Collinearity among stressor variables

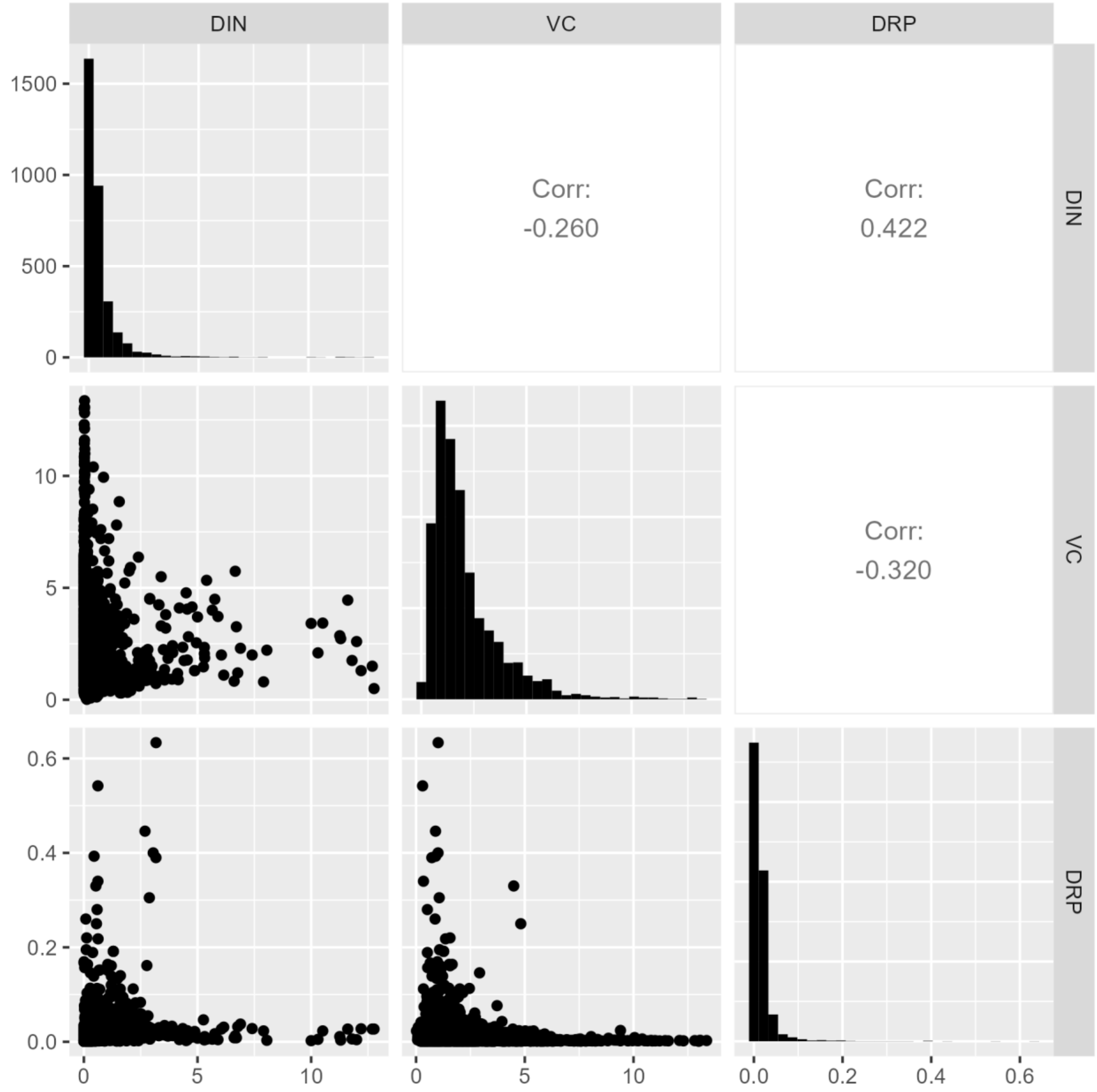


Figure B-1: Collinearity among stressor variables. Spearman correlation coefficients are shown on the upper panels. Relationships between stressors are shown on the lower panels. Frequency histograms of each stressor are shown on the diagonal panels.

Appendix C Effects of increasing the basis dimension of GAMM smoothers on model selection for the SSDM.

We assessed the effects of increasing the basis dimension on GAMM smoothers on model selection for five macroinvertebrate species representing a range of rarity, life-histories and feeding strategies. We applied the same model selection procedure to these taxa, described in Section 3.2.1, after incrementally setting *K* at 6, 12 and 24. For each setting of *K*, the basis dimension for univariate smoothers was set to 1*K*, while it was set to 2*K* and 4*K* for bivariate and multivariate smoothers, respectively. The table below shows that increasing *K* had no effect on the final model selected for each taxon. There were no consistent changes in AIC (ΔAIC) with increasing the smoother basis dimension*,* with both increases in decreases in AIC observed as *K* increased, among taxa. There was also no overwhelming support for any particular basis dimension setting as indicated by Akaike Weights (wAIC) being approximately equal across *K* settings within each taxon. This implied that *K* = 6 was sufficient to capture the complexity of non-linearity and parameterisation of stressor effects on taxon PO.

Taxon	AIC	Final model	K	ΔAIC	wAIC
Austrosimulium	3369.56	DRP	6	0	0.33
	3369.56	DRP	12	0	0.33
	3369.56	DRP	24	0	0.33
Deleatidium	1431.31	DIN:DRP	6	2.78	0.12
	1428.93	DIN:DRP	12	0.39	0.4
	1428.53	DIN:DRP	24	0	0.48
Hydrobiosis	2233.68	DIN:DRP:VC	6	0	0.33
	2233.68	DIN:DRP:VC	12	0	0.33
	2233.68	DIN:DRP:VC	24	0.01	0.33
Pycnocentria	3139.47	DIN+DRP	6	0.01	0.34
	3139.46	DIN+DRP	12	0	0.35
	3139.69	DIN+DRP	24	0.23	0.31
Zelandoperla	2007.43	DIN:DRP	6	0.02	0.33
	2007.41	DIN:DRP	12	0	0.34
	2007.42	DIN:DRP	24	0.01	0.33

Appendix D SSDM Model selection results

Table D-1: Model selection, predictive accuracy and sensitivity results for models of individual taxa of the SSDM. The final model is that which minimised AIC among candidate models for each metric. AIC_w is the Akaike weight of the final model, total EDF is the total effective degrees of freedom from all fixed effect smoothers in the final model, AUC_{cv} is the mean model AUC from the 10-fold cross-validation procedure. Absolute sensitivty (PO range) and relative sensitivity (PO range/maximum PO) from model partial predictions across each stressor domain are also shown.

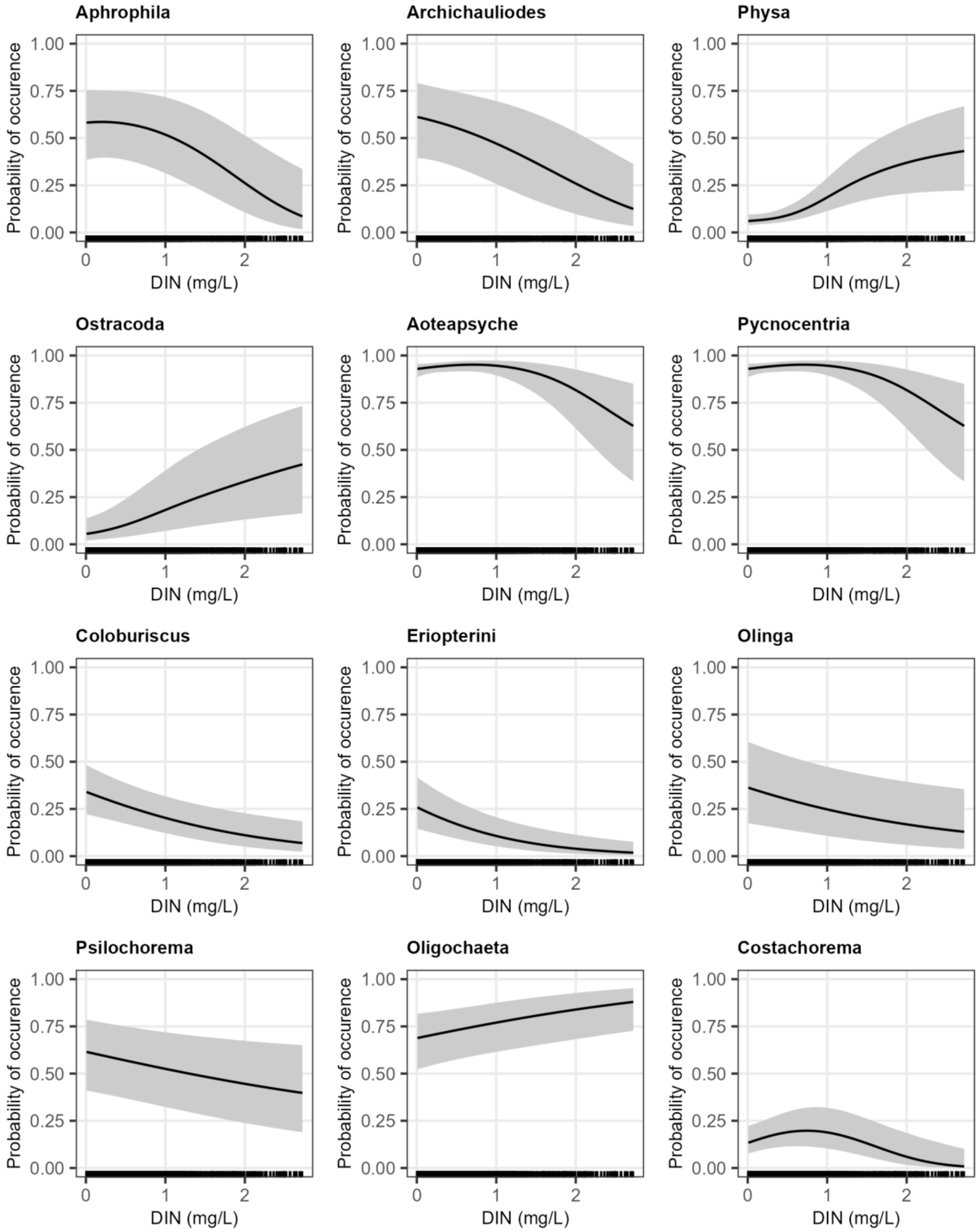
						DIN Sensitivity		DRP Sensitivity		VC Sensitivity	
Taxon	**Final model**	**AIC**	**AIC_w**	**Total EDF**	**AUC_{CV}**	**Absolute**	**Relative**	**Absolute**	**Relative**	**Absolute**	**Relative**
Acanthophlebia	NULL	513	0.52	0	0.5	<0.01	0	<0.01	0	<0.01	0
Acarina	NULL	2645	0.53	0	0.5	<0.01	0	<0.01	0	<0.01	0
Acroperla	NULL	669	0.49	0	0.5	<0.01	0	<0.01	0	<0.01	0
Ameletopsis	DIN	504	0.27	1	0.68	<0.01	0.98	<0.01	0	<0.01	0
Amphipoda	DIN:DRP:VC	1746	0.38	9	0.55	0.01	0.4	0.02	0.56	0.01	0.63
Aoteapsyche	DIN:DRP	1894	0.69	9.11	0.72	0.32	0.34	0.4	0.41	<0.01	0
Aphrophila	DIN:DRP:VC	2780	1	9	0.62	0.5	0.85	0.49	0.76	0.33	0.41
Archichauliodes	DIN:DRP:VC	2784	0.73	9	0.61	0.49	0.8	0.33	0.53	0.29	0.36
Atalophlebioides	DIN:DRP	535	0.26	2	0.66	<0.01	0.78	<0.01	0.8	<0.01	0
Austroclima	DIN:DRP:VC	2980	0.41	9	0.53	0.06	0.28	0.11	0.5	0.15	0.51
Austroperla	DIN:VC	1079	0.46	2.29	0.73	<0.01	1	<0.01	0	<0.01	0.69
Austrosimulium	DRP	3370	0.2	1	0.57	<0.01	0	0.1	0.18	<0.01	0
Beraeoptera	DIN+DRP+VC	1864	0.39	3.15	0.69	0.03	0.9	0.02	0.79	0.06	0.78
Berosus	DIN	901	0.35	1	0.49	<0.01	0.83	<0.01	0	<0.01	0
Ceratopogonidae	NULL	986	0.32	0	0.5	<0.01	0	<0.01	0	<0.01	0
Chironomidae	DIN:DRP:VC	3472	0.4	9	0.54	0.18	0.27	0.01	0.02	0.19	0.31
Chironomus	DRP:VC	644	0.26	2	0.69	<0.01	0	<0.01	0.45	<0.01	0.84
Cladocera	DRP	476	0.47	1.8	0.72	<0.01	0	<0.01	0.97	<0.01	0
Collembola	DRP	556	0.35	1.69	0.68	<0.01	0	<0.01	0.95	<0.01	0
Coloburiscus	DIN:DRP	2617	0.69	2	0.56	0.27	0.8	0.23	0.66	<0.01	0
Confluens	NULL	726	0.34	0	0.5	<0.01	0	<0.01	0	<0.01	0
Copepoda	DIN:DRP	579	0.29	2.51	0.8	<0.01	0.84	<0.01	0.65	<0.01	0
Costachorema	DIN:DRP:VC	2743	1	9	0.58	0.19	0.96	0.18	0.89	0.12	0.46
Deleatidium	DIN:DRP	1431	0.72	7.36	0.73	0.06	0.06	0.12	0.12	<0.01	0
Dytiscidae	NULL	380	0.51	0	0.5	<0.01	0	<0.01	0	<0.01	0
Elmidae	DIN:DRP	1556	0.72	8.2	0.71	0.07	0.07	0.12	0.12	<0.01	0
Empididae	DIN:DRP	2231	0.24	2	0.54	0.03	0.38	0.03	0.37	<0.01	0
Ephydridae	DRP	818	0.53	1.85	0.55	<0.01	0	0.03	0.74	<0.01	0
Eriopterini	DIN:DRP	2618	0.63	2	0.65	0.24	0.93	0.16	0.65	<0.01	0
Ferrissia	NULL	772	0.48	0	0.5	<0.01	0	<0.01	0	<0.01	0
Gyraulus	DRP	990	0.3	1	0.57	<0.01	0	<0.01	0.46	<0.01	0

						DIN Sensitivity		DRP Sensitivity		VC Sensitivity	
Taxon	**Final model**	**AIC**	**AIC$_w$**	**Total EDF**	**AUC$_{CV}$**	**Absolute**	**Relative**	**Absolute**	**Relative**	**Absolute**	**Relative**
Helicopsyche	DRP:VC	1701	0.36	5.05	0.62	<0.01	0	0.01	0.89	0.01	0.62
Hexatomini	NULL	654	0.56	0	0.5	<0.01	0	<0.01	0	<0.01	0
Hirudinea	NULL	925	0.42	0	0.5	<0.01	0	<0.01	0	<0.01	0
Hudsonema	DIN	2349	0.6	3.57	0.6	0.16	0.69	<0.01	0	<0.01	0
Hydra	DIN	438	0.35	1.82	0.73	<0.01	0.99	<0.01	0	<0.01	0
Hydraenidae	DIN:VC	1716	0.33	6.63	0.61	0.04	0.87	<0.01	0	0.06	0.79
Hydrobiosella	NULL	563	0.53	0	0.5	<0.01	0	<0.01	0	<0.01	0
Hydrobiosis	DIN:DRP:VC	2234	0.43	9	0.64	0.08	0.08	0.14	0.16	0.1	0.1
Hydrochorema	NULL	354	0.29	0	0.5	<0.01	0	<0.01	0	<0.01	0
Hydrophilidae	NULL	1163	0.54	0	0.5	<0.01	0	<0.01	0	<0.01	0
Hygraula	NULL	594	0.36	0	0.5	<0.01	0	<0.01	0	<0.01	0
Ichthybotus	NULL	407	0.54	0	0.5	<0.01	0	<0.01	0	<0.01	0
Isopoda	NULL	296	0.56	0	0.5	<0.01	0	<0.01	0	<0.01	0
Latia	DIN:DRP	1260	0.55	2	0.51	<0.01	0.91	<0.01	0.74	<0.01	0
Lymnaeidae	DIN:DRP:VC	594	0.21	9	0.59	<0.01	0.72	<0.01	0.81	<0.01	0.92
Maoridiamesa	DRP	169	0.39	2.15	0.32	<0.01	0	<0.01	0.84	<0.01	0
Mauiulus	NULL	1047	0.47	0	0.5	<0.01	0	<0.01	0	<0.01	0
Megaleptoperla	NULL	1266	0.3	0	0.5	<0.01	0	<0.01	0	<0.01	0
Microvelia	NULL	734	0.24	0	0.5	<0.01	0	<0.01	0	<0.01	0
Mischoderus	DIN	1635	0.35	1	0.51	0.03	0.68	<0.01	0	<0.01	0
Molophilus	DIN:VC	1401	0.29	2	0.65	0.03	0.8	<0.01	0	0.03	0.7
Muscidae	DIN:DRP	2632	0.39	5.52	0.55	0.14	0.54	0.03	0.19	<0.01	0
Nematoda	DIN:DRP	1411	0.42	7.08	0.65	0.05	0.72	0.05	0.73	<0.01	0
Nematomorpha	NULL	449	0.33	0	0.5	<0.01	0	<0.01	0	<0.01	0
Nemertea	DRP:VC	1126	0.29	2	0.58	<0.01	0	<0.01	0.29	0.01	0.81
Neozephlebia	NULL	915	0.55	0	0.5	<0.01	0	<0.01	0	<0.01	0
Nesameletus	DIN:DRP:VC	2241	0.97	9	0.66	0.15	0.98	0.15	0.98	0.25	0.72
Neurochorema	DIN:DRP:VC	2819	0.82	9	0.63	0.14	0.82	0.16	0.85	0.24	0.7
Nothodixa	NULL	154	0.35	0	0.5	<0.01	0	<0.01	0	<0.01	0
Oecetis	NULL	415	0.28	0	0.5	<0.01	0	<0.01	0	<0.01	0
Oeconesidae	NULL	319	0.38	0	0.5	<0.01	0	<0.01	0	<0.01	0
Oligochaeta	DIN:DRP:VC	3307	0.57	9	0.57	0.19	0.22	0.2	0.23	0.12	0.16
Olinga	DIN:DRP:VC	2595	0.32	9	0.68	0.23	0.64	0.19	0.52	0.54	0.75
Orthocladiinae	DRP	1810	0.28	2.72	0.58	<0.01	0	0.06	0.07	<0.01	0
Orthopsyche	NULL	649	0.43	0	0.5	<0.01	0	<0.01	0	<0.01	0
Ostracoda	DIN:DRP	2241	0.7	7.5	0.66	0.37	0.87	0.27	0.84	<0.01	0
Oxyethira	DIN:VC	3547	0.42	2	0.55	0.16	0.27	<0.01	0	0.14	0.3
Paracalliope	DIN+DRP	1283	0.74	5.99	0.79	0.08	0.9	0.05	0.98	<0.01	0

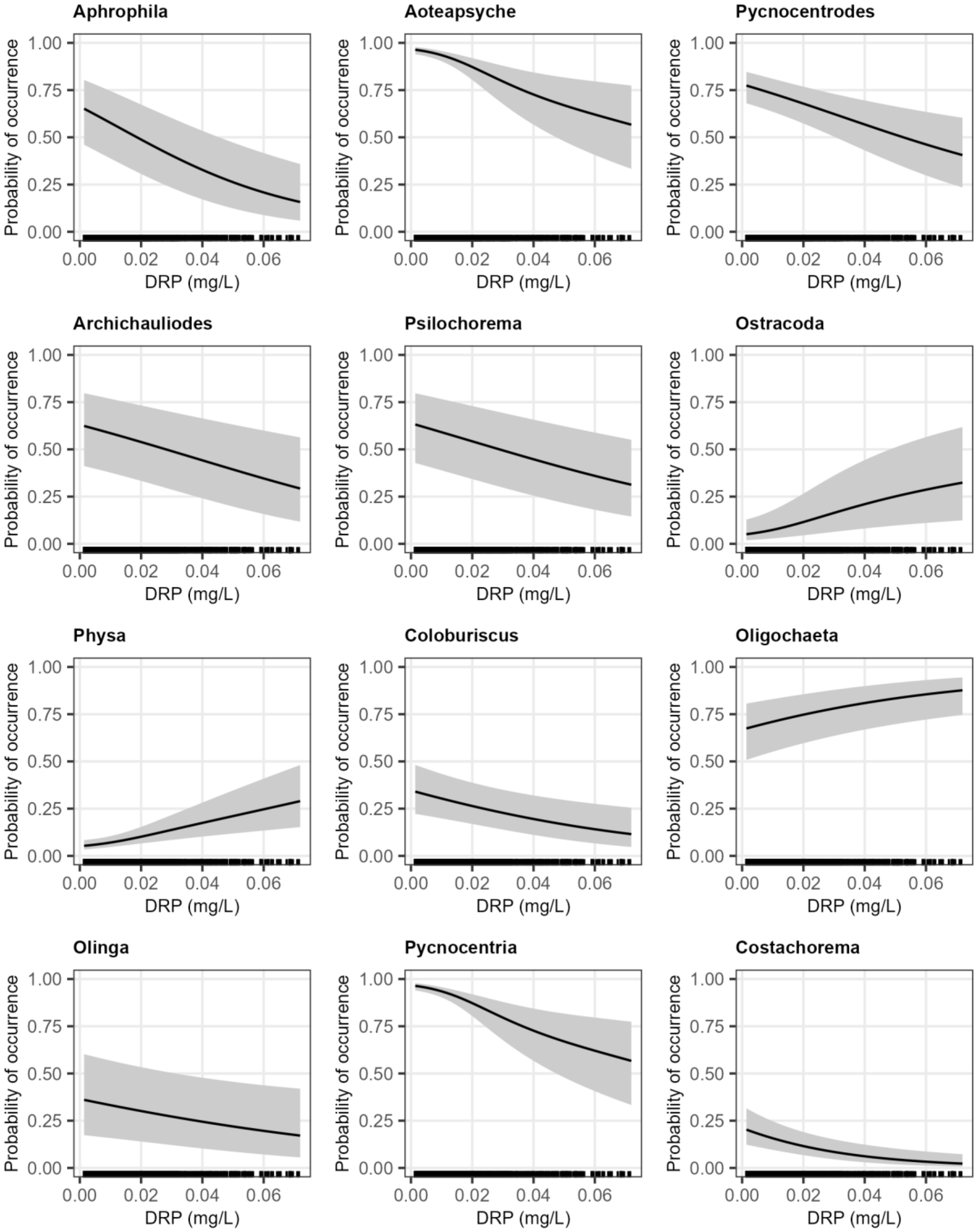
						DIN Sensitivity		DRP Sensitivity		VC Sensitivity	
Taxon	**Final model**	**AIC**	**AIC$_w$**	**Total EDF**	**AUC$_{CV}$**	**Absolute**	**Relative**	**Absolute**	**Relative**	**Absolute**	**Relative**
Paradixa	DIN:VC	357	0.26	2.46	0.8	<0.01	0.93	<0.01	0	<0.01	0.98
Paraleptamphopus	NULL	272	0.41	0	0.5	<0.01	0	<0.01	0	<0.01	0
Paralimnophila	NULL	985	0.25	0	0.5	<0.01	0	<0.01	0	<0.01	0
Paranephrops	NULL	301	0.42	0	0.5	<0.01	0	<0.01	0	<0.01	0
Paratya	NULL	729	0.3	0	0.5	<0.01	0	<0.01	0	<0.01	0
Paroxyethira	DIN:DRP:VC	904	0.63	9	0.7	<0.01	0.3	0.06	0.92	<0.01	0.58
Philorheithrus	NULL	211	0.47	0	0.5	<0.01	0	<0.01	0	<0.01	0
Physa	DIN:DRP	2266	0.46	7.69	0.69	0.37	0.86	0.24	0.82	<0.01	0
Platyhelminthes	DIN:DRP:VC	2537	0.83	9	0.62	0.13	0.59	0.15	0.63	0.03	0.28
Plectrocnemia	VC	550	0.3	1.77	0.72	<0.01	0	<0.01	0	<0.01	0.93
Polychaeta	NULL	252	0.5	0	0.5	<0.01	0	<0.01	0	<0.01	0
Polyplectropus	NULL	668	0.4	0	0.5	<0.01	0	<0.01	0	<0.01	0
Potamopyrgus	DIN:DRP	2305	0.94	7.17	0.69	0.12	0.12	0.1	0.11	<0.01	0
Psilochorema	DIN:DRP:VC	2941	0.54	9	0.66	0.22	0.35	0.32	0.5	0.4	0.47
Psychodidae	DRP	438	0.21	1	0.66	<0.01	0	<0.01	0.52	<0.01	0
Ptilodactylidae	DIN	695	0.25	1	0.61	<0.01	0.97	<0.01	0	<0.01	0
Pycnocentria	DIN+DRP	3139	0.42	3.24	0.57	0.3	0.49	0.18	0.47	<0.01	0
Pycnocentrodes	DIN:DRP:VC	3014	0.97	9	0.6	0.05	0.06	0.37	0.48	0.25	0.28
Rallidens	NULL	514	0.52	0	0.5	<0.01	0	<0.01	0	<0.01	0
Scirtidae	VC	473	0.26	1.31	0.66	<0.01	0	<0.01	0	<0.01	0.87
Sigara	DIN:DRP:VC	1050	0.29	9	0.62	0.01	0.6	0.03	0.84	<0.01	0.48
Sphaeriidae	DIN	1227	0.39	3.21	0.71	0.06	0.95	<0.01	0	<0.01	0
Staphylinidae	NULL	495	0.34	0	0.5	<0.01	0	<0.01	0	<0.01	0
Stenoperla	DIN:VC	1371	0.34	2	0.75	<0.01	0.98	<0.01	0	<0.01	0.7
Stictocladius	NULL	217	0.5	0	0.5	<0.01	0	<0.01	0	<0.01	0
Stratiomyidae	DRP	300	0.33	1	0.57	<0.01	0	<0.01	0.39	<0.01	0
Tabanidae	DIN:VC	1006	0.26	2	0.6	0.01	0.78	<0.01	0	0.01	0.54
Tanypodinae	DIN	3326	0.22	1	0.53	0.08	0.26	<0.01	0	<0.01	0
Tanytarsini	DIN	2733	0.37	2.16	0.5	0.12	0.39	<0.01	0	<0.01	0
Tanytarsus	NULL	496	0.55	0	0.5	<0.01	0	<0.01	0	<0.01	0
Triplectides	DIN:DRP	1291	0.27	7.25	0.66	0.05	0.77	0.05	0.79	<0.01	0
Xanthocnemis	NULL	586	0.34	0	0.5	<0.01	0	<0.01	0	<0.01	0
Zelandobius	DIN:VC	2299	0.21	2.08	0.54	0.04	0.46	<0.01	0	0.05	0.39
Zelandoperla	DIN:DRP	2007	0.67	2	0.66	0.07	0.98	0.07	0.92	<0.01	0
Zephlebia	DIN:DRP:VC	2219	0.92	9	0.61	0.06	0.44	0.04	0.36	0.08	0.92

Appendix E Partial plots of the taxa with the highest absolute sensitivity to each stressor

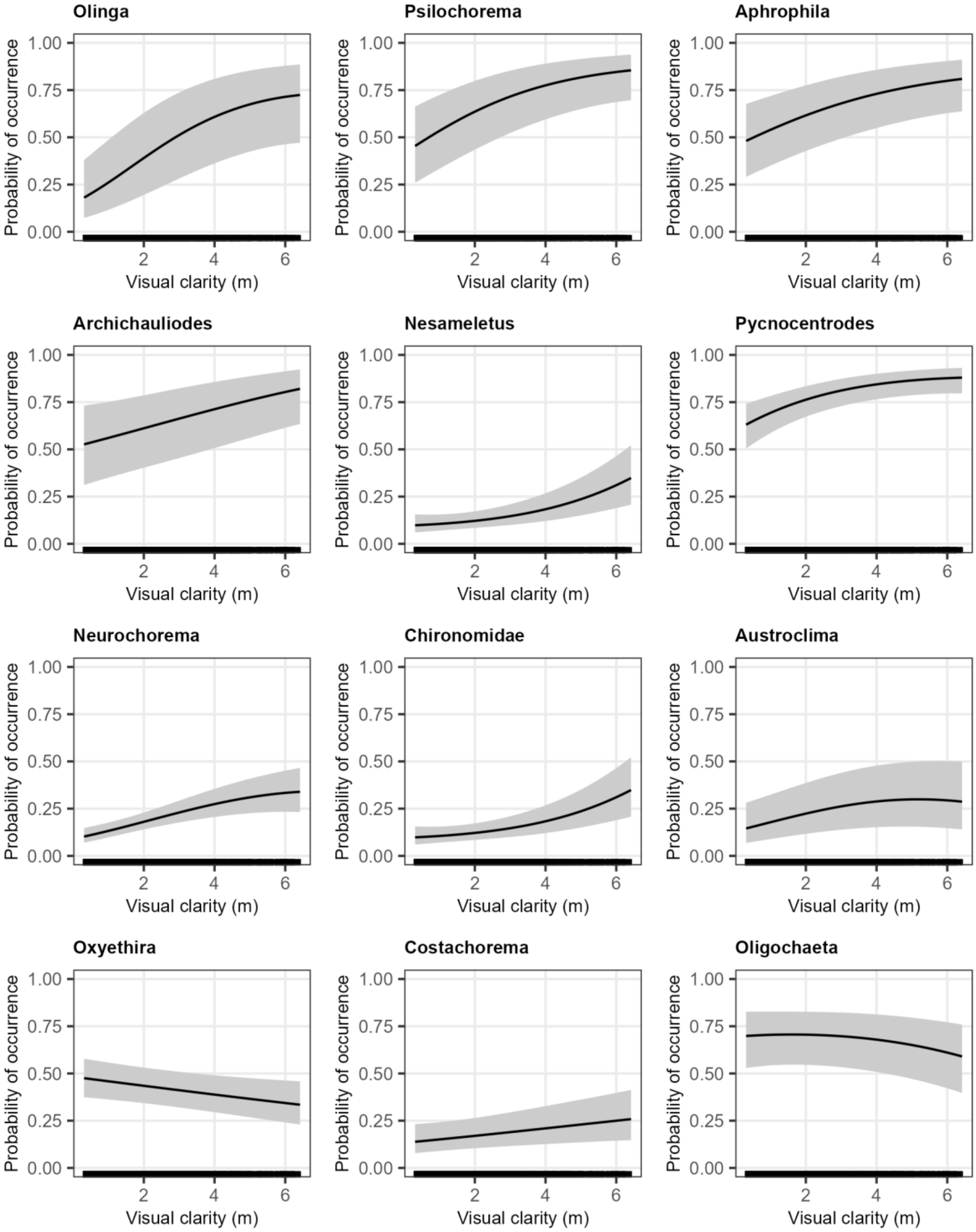
a) DIN



b) DRP


Aphrophila
Aoteapsyche
Pycnocentrodes
Archichauliodes
Psilochorema
Ostracoda
Physa
Coloburiscus
Oligochaeta
Olinga
Pycnocentria
Costachorema
Probability of occurrence
0.00
0.25
0.50
0.75
1.00
DRP (mg/L)
0.00
0.02
0.04
0.06

c) VC


Olinga
Psilochorema
Aphrophila
Archichauliodes
Nesameletus
Pycnocentrodes
Neurochorema
Chironomidae
Austroclima
Oxyethira
Costachorema
Oligochaeta
Probability of occurrence
0.00
0.25
0.50
0.75
1.00
Visual clarity (m)
2
4
6

Appendix F Sensitivity of $CT_{X,crit}$ to different multi-stressor conditions

Table F-1: Critical values of VC required to limit macroinvertebrate community extirpation to ≤1% ($CT_{1,crit}$), ≤2.5% ($CT_{2.5,crit}$), ≤5% ($CT_{5,crit}$) or ≤10% ($CT_{10,crit}$), under different conditions of DIN and DRP. All critical values are the medians estimated from 1000 realisations of the BEA algorithm for a particular set of DIN and DRP conditions. All values are equivalent to 12 monthly annual median exposures. Inf values indicate critical values that are beyond the domain of the data used. This means it is not possible to improve (increase) VC sufficiently to limit extirpation to that level given the values of DIN and DRP.

DIN:DRP conditions		Critical VC values required to limit macroinvertebrate extirpation			
DIN (mg/L)	DRP (mg/L)	VC $CT_{1,crit}$ (m)	VC $CT_{2.5,crit}$ (m)	VC $CT_{5,crit}$ (m)	VC $CT_{10,crit}$ (m)
0.03	0.006	3.4	1.8	1.1	0.7
0.03	0.01	3	1.8	1.2	0.8
0.03	0.018	2.4	1.9	1.5	1
0.03	0.054	4.5	3.5	2.8	1.5
0.24	0.006	2.9	1.9	1.2	0.8
0.24	0.01	2.7	1.8	1.4	0.9
0.24	0.018	2.7	2.2	1.7	1.1
0.24	0.054	4.2	3.5	2.8	1.7
1.3	0.006	6.2	4.1	2.7	1.5
1.3	0.01	5.6	4.1	2.8	1.5
1.3	0.018	5.2	4.1	2.8	1.6
1.3	0.054	6.1	4.5	3.1	2.3
2.2	0.006	Inf	5	3.5	1.7
2.2	0.01	Inf	5	3.5	1.7
2.2	0.018	6.3	4.9	3.6	1.9
2.2	0.054	Inf	4.9	3.3	2.2

Table F-2: Critical values of DIN required to limit macroinvertebrate community extirpation to ≤1% ($CT_{1,crit}$), ≤2.5% ($CT_{2.5,crit}$), ≤5% ($CT_{5,crit}$) or ≤10% ($CT_{10,crit}$), under different conditions of VC and DRP. All critical values are the medians estimated from 1000 realisations of the BEA algorithm for a particular pair of VC and DRP conditions. All values are equivalent to 12 monthly annual median exposures. -Inf values indicate critical values that are beyond the domain of the data used. This means it is not possible to improve (decrease) DIN sufficiently to limit extirpation to that level given the values of VC and DRP.

VC:DRP constants		Critical DIN values required to limit macroinvertebrate extirpation			
VC (m)	DRP (mg/L)	DIN $CT_{1,crit}$ (mg/L)	DIN $CT_{2.5,crit}$ (mg/L)	DIN $CT_{5,crit}$ (mg/L)	DIN $CT_{10,crit}$ (mg/L)
0.6	0.006	-Inf	-Inf	0.36	1.09
1.4	0.006	0.13	0.36	0.68	1.34
2.2	0.006	0.29	0.51	0.85	1.49
3	0.006	0.29	0.54	0.95	1.63
0.6	0.01	-Inf	-Inf	0.24	0.99
1.4	0.01	0.02	0.24	0.57	1.22
2.2	0.01	0.23	0.42	0.75	1.42
3	0.01	0.24	0.48	0.85	1.57
0.6	0.018	-Inf	-Inf	0.07	0.74
1.4	0.018	-Inf	-Inf	0.38	1.08
2.2	0.018	0.02	0.25	0.57	1.31
3	0.018	0.18	0.37	0.64	1.47
0.6	0.054	-Inf	-Inf	-Inf	0.24
1.4	0.054	-Inf	-Inf	-Inf	0.74
2.2	0.054	-Inf	-Inf	0.22	1.25
3	0.054	-Inf	-Inf	0.37	1.49

Table F-3: Critical values of DRP required to limit macroinvertebrate community extirpation to ≤1% ($CT_{1,crit}$), ≤2.5% ($CT_{2.5,crit}$), ≤5% ($CT_{5,crit}$) or ≤10% ($CT_{10,crit}$), under different conditions of VC and DRP. All critical values are the medians estimated from 1000 realisations of the BEA algorithm for a particular pair of VC and DIN conditions. All values are equivalent to 12 monthly annual median exposures. -Inf values indicate critical values that are beyond the domain of the data used. This means it is not possible to improve (decrease) DRP sufficiently to limit extirpation to that level given the values of VC and DIN.

VC:DIN constants		Critical DRP values required to limit macroinvertebrate extirpation			
VC (m)	DIN (mg/L)	DRP $CT_{1,crit}$ (mg/L)	DRP $CT_{2.5,crit}$ (mg/L)	DRP $CT_{5,crit}$ (mg/L)	DRP $CT_{10,crit}$ (mg/L)
0.6	0.03	0.002	0.007	0.012	0.031
1.4	0.03	0.009	0.013	0.019	0.041
2.2	0.03	0.013	0.02	0.033	0.051
3	0.03	0.014	0.025	0.041	0.057
0.6	0.24	-Inf	0.004	0.01	0.026
1.4	0.24	0.006	0.012	0.019	0.036
2.2	0.24	0.013	0.02	0.029	0.048
3	0.24	0.016	0.026	0.038	0.055
0.6	1.3	-Inf	-Inf	-Inf	0.012
1.4	1.3	-Inf	-Inf	-Inf	0.027
2.2	1.3	-Inf	-Inf	0.008	0.048
3	1.3	-Inf	-Inf	0.018	0.058
0.6	2.2	-Inf	-Inf	-Inf	0.009
1.4	2.2	-Inf	-Inf	-Inf	0.031
2.2	2.2	-Inf	-Inf	-Inf	0.049
3	2.2	-Inf	-Inf	-Inf	0.055

Appendix G Tables of critical values of stressors required to meet macroinvertebrate targets for optimistic-pessimistic thresholds

Table G-1: Critical values of DIN required to limit macroinvertebrate community extirpation to ≤1% ($CT_{1,crit}$), ≤2.5% ($CT_{2.5,crit}$), ≤5% ($CT_{5,crit}$) or ≤10% ($CT_{10,crit}$), for optimistic, median and pessimistic thresholds. Optimistic $CT_{X,crit}$ values assume that VC and DRP are currently at levels causing minimal to no extirpation of any taxon. Pessimistic $CT_{X,crit}$ values assume VC and DRP are currently at levels causing high community extirpation. Median $CT_{X,crit}$ values assume VC and DRP are currently at levels cause moderate community extirpation. -Inf values indicate critical values that are beyond the domain of the data used. This means it is not possible to improve (decrease) DIN sufficiently to limit extirpation to that level for a given scenario.

	Critical DIN values required to limit macroinvertebrate extirpation			
Thresholds	**DIN $CT_{1,crit}$ (mg/L)**	**DIN $CT_{2.5,crit}$ (mg/L)**	**DIN $CT_{5,crit}$ (mg/L)**	**DIN $CT_{10,crit}$ (mg/L)**
Optimistic	0.53	0.81	1.2	1.88
Median	0.29	0.51	0.83	1.47
Pessimistic	0.04	0.29	0.53	1.1

Table G-2: Critical values of DRP required to limit macroinvertebrate community extirpation to ≤1% ($CT_{1,crit}$), ≤2.5% ($CT_{2.5,crit}$), ≤5% ($CT_{5,crit}$) or ≤10% ($CT_{10,crit}$), for optimistic, median and pessimistic thresholds of VC and DIN. Optimistic $CT_{X,crit}$ values assume that VC and DIN are currently at levels causing minimal to no extirpation of any taxon. Pessimistic $CT_{X,crit}$ values assume VC and DIN are currently at levels causing high community extirpation. Median $CT_{X,crit}$ values assume VC and DIN are currently at levels cause moderate community extirpation. -Inf values indicate critical values that are beyond the domain of the data used. This means it is not possible to improve (decrease) DRP sufficiently to limit extirpation to that level for a given scenario.

	Critical DRP values required to limit macroinvertebrate extirpation			
Thresholds	**DRP $CT_{1,crit}$ (mg/L)**	**DRP $CT_{2.5,crit}$ (mg/L)**	**DRP $CT_{5,crit}$ (mg/L)**	**DRP $CT_{10,crit}$ (mg/L)**
Optimistic	0.023	0.028	0.035	0.051
Median	0.014	0.021	0.027	0.04
Pessimistic	-Inf	0.013	0.021	0.031

Table G-3: Critical values of VC required to limit macroinvertebrate community extirpation to ≤1% ($CT_{1,crit}$), ≤2.5% ($CT_{2.5,crit}$), ≤5% ($CT_{5,crit}$) or ≤10% ($CT_{10,crit}$), for optimistic, median and pessimistic thresholds. Optimistic $CT_{X,crit}$ values assume that DRP and DIN are currently at levels causing minimal to no extirpation of any taxa. Pessimistic $CT_{X,crit}$ values assume DRP and DIN are currently at levels causing high community extirpation. Median $CT_{X,crit}$ values assume DRP and DIN are currently at levels causing moderate community extirpation. -Inf values indicate critical values that are beyond the domain of the data used. This means it is not possible to improve (increase) VC sufficiently to limit extirpation to that level for a given scenario.

	Critical VC values required to limit macroinvertebrate extirpation			
Thresholds	**VC $CT_{1,crit}$ (m)**	**VC $CT_{2.5,crit}$ (m)**	**VC $CT_{5,crit}$ (m)**	**VC $CT_{10,crit}$ (m)**
Optimistic	1.4	1.1	0.9	0.6
Median	1.9	1.5	1.2	0.8
Pessimistic	3.2	2	1.5	1.1

Appendix H Taxon tolerance thresholds (TT_{crit}) to each stressor

Table H-1: Taxon tolerance thresholds (TT_{crit}) to each stressor. Values within cells of the first 9 columns show the median (Med), lower (L95), and upper (U95) 95% confidence interval of TT_{crit} (the value of a stressor at which the taxon is likely extirpated) estimated from the optimistic-pessimistic Monte-Carlo simulations. Taxa are ordered by their mean rank tolerance among all three stressors (final three columns). Tolerance values beyond the domain of the stressor are denoted as Inf or -Inf for nutrients and VC, respectively. This implies that the taxa are tolerant of higher nutrients, or lower VC than that simulated.

	DIN TT_{crit} (mg/L)			**DRP TT_{crit} (mg/L)**			**VC TT_{crit} (m)**			**Rank tolerance**		
Taxon	**Med**	**L95**	**U95**	**Med**	**L95**	**U95**	**Med**	**L95**	**U95**	**DIN**	**DRP**	**VC**
Beraeoptera	1.02	0.08	2.4	0.024	0.006	0.068	1.17	0.33	2.25	5	1	2
Nesameletus	0.5	0.11	2.69	0.025	0.008	0.059	0.45	0.31	1.7	1	2	22
Neurochorema	1.99	0.49	2.69	0.032	0.012	0.064	1.12	0.34	2.05	17	5	5
Aphrophila	1.4	0.74	2.69	0.026	0.015	0.05	0.55	0.31	1.41	8	4	17
Olinga	2.12	1.04	2.7	0.048	0.022	0.07	1.15	0.38	1.99	19	13	4
Psilochorema	2.2	1.03	2.69	0.034	0.019	0.068	0.79	0.32	1.7	22	7	10
Costachorema	1.84	1.1	2.69	0.038	0.015	0.07	0.48	0.31	1.39	12	9	21
Archichauliodes	1.88	0.84	2.69	0.048	0.023	0.071	0.5	0.31	1.38	15	13	20
Hydrobiosis	2.36	1.44	2.7	0.043	0.024	0.07	0.58	0.31	1.4	26	11	16
Austroclima	2.45	1.81	2.7	0.055	0.026	0.071	0.7	0.32	1.63	28	16	11
Pycnocentrodes	2.7	2.52	Inf	0.056	0.026	0.07	0.79	0.31	1.58	32	18	10
Zelandoperla	0.51	0.2	1.25	0.026	0.011	0.055	-Inf	-Inf	-Inf	2	4	104
Coloburiscus	0.82	0.36	1.59	0.033	0.018	0.063	-Inf	-Inf	-Inf	4	6	104
Eriopterini	0.67	0.27	1.25	0.038	0.02	0.068	-Inf	-Inf	-Inf	3	9	104
Hydraenidae	1.56	0.45	2.62	Inf	Inf	Inf	1.15	0.43	1.78	9	104	4
Austroperla	1.08	0.13	2.65	Inf	Inf	Inf	0.59	0.31	1.81	6	104	15
Stenoperla	1.4	0.18	2.65	Inf	Inf	Inf	0.63	0.31	1.85	8	104	13
Helicopsyche	Inf	Inf	Inf	0.05	0.016	0.072	0.9	0.32	1.64	104	14	8
Lymnaeidae	2.52	1.72	Inf	Inf	0.068	Inf	1.36	0.36	3.2	31	104	1
Tabanidae	2.04	0.51	2.7	Inf	Inf	Inf	0.62	0.31	1.72	18	104	14
Atalophlebioides	1.88	0.72	2.69	0.059	0.024	0.072	-Inf	-Inf	-Inf	15	19	104
Aoteapsyche	1.93	1.38	2.69	0.06	0.024	0.072	-Inf	-Inf	-Inf	16	21	104
Zelandobius	2.19	1.14	2.7	Inf	Inf	Inf	0.52	0.31	1.47	20	104	18
Deleatidium	2.36	1.18	2.72	0.054	0.028	0.071	-Inf	-Inf	-Inf	26	15	104
Elmidae	2.3	0.78	2.72	0.056	0.023	0.071	-Inf	-Inf	-Inf	23	18	104
Empididae	2.2	1	2.7	0.06	0.034	0.072	-Inf	-Inf	-Inf	22	21	104
Latia	2.32	1.2	2.7	0.066	0.043	0.072	-Inf	-Inf	-Inf	24	22	104
Plectrocnemia	Inf	Inf	Inf	Inf	Inf	Inf	0.99	0.35	2	104	104	6
Scirtidae	Inf	Inf	Inf	Inf	Inf	Inf	0.93	0.33	2.1	104	104	7
Mischoderus	1.6	0.48	2.69	Inf	Inf	Inf	-Inf	-Inf	-Inf	10	104	104
Orthocladiinae	Inf	Inf	Inf	0.042	0.01	0.071	-Inf	-Inf	-Inf	104	10	104

	DIN TT_{crit} (mg/L)			DRP TT_{crit} (mg/L)			VC TT_{crit} (m)			Rank tolerance		
Taxon	**Med**	**L95**	**U95**	**Med**	**L95**	**U95**	**Med**	**L95**	**U95**	**DIN**	**DRP**	**VC**
Berosus	1.79	0.73	2.65	Inf	Inf	Inf	-Inf	-Inf	-Inf	11	104	104
Chironomidae	Inf	Inf	Inf	Inf	Inf	Inf	0.68	0.33	1.39	104	104	12
Molophilus	1.85	0.65	2.69	Inf	Inf	Inf	-Inf	-Inf	-Inf	13	104	104
Platyhelminthes	Inf	Inf	Inf	Inf	Inf	Inf	0.5	-Inf	1.38	104	104	20
Austrosimulium	Inf	Inf	Inf	0.068	0.048	0.072	-Inf	-Inf	-Inf	104	23	104
Pycnocentria	Inf	Inf	Inf	0.07	0.06	Inf	-Inf	-Inf	-Inf	104	24	104
Tanypodinae	2.43	1.66	2.72	Inf	Inf	Inf	-Inf	-Inf	-Inf	27	104	104
Ptilodactylidae	2.46	1.47	2.72	Inf	Inf	Inf	-Inf	-Inf	-Inf	29	104	104
Ameletopsis	2.52	1.77	2.72	Inf	Inf	Inf	-Inf	-Inf	-Inf	31	104	104
Acanthophlebia	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Acarina	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Acroperla	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Amphipoda	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Ceratopogonidae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Chironomus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Cladocera	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Collembola	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Confluens	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Copepoda	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Dytiscidae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Ephydridae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Ferrissia	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Gyraulus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Hexatomini	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Hirudinea	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Hudsonema	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Hydra	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Hydrobiosella	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Hydrochorema	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Hydrophilidae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Hygraula	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Ichthybotus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Isopoda	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Maoridiamesa	Inf	Inf	Inf	Inf	0.068	Inf	-Inf	-Inf	-Inf	104	104	104
Mauiulus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Megaleptoperla	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Microvelia	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Muscidae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104

	DIN TT_{crit} (mg/L)			DRP TT_{crit} (mg/L)			VC TT_{crit} (m)			Rank tolerance		
Taxon	**Med**	**L95**	**U95**	**Med**	**L95**	**U95**	**Med**	**L95**	**U95**	**DIN**	**DRP**	**VC**
Nematoda	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Nematomorpha	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Nemertea	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Neozephlebia	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Nothodixa	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Oecetis	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Oeconesidae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Oligochaeta	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Orthopsyche	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Ostracoda	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Oxyethira	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Paracalliope	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Paradixa	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Paraleptamphopus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Paralimnophila	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Paranephrops	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Paratya	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Paroxyethira	Inf	2.5	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Philorheithrus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Physa	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Polychaeta	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Polyplectropus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Potamopyrgus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Psychodidae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Rallidens	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Sigara	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Sphaeriidae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Staphylinidae	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Stictocladius	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Stratiomyidae	Inf	Inf	Inf	Inf	0.069	Inf	-Inf	-Inf	-Inf	104	104	104
Tanytarsini	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Tanytarsus	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Triplectides	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Xanthocnemis	Inf	Inf	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104
Zephlebia	Inf	2.62	Inf	Inf	Inf	Inf	-Inf	-Inf	-Inf	104	104	104

Appendix I MIM residual diagnostic plots

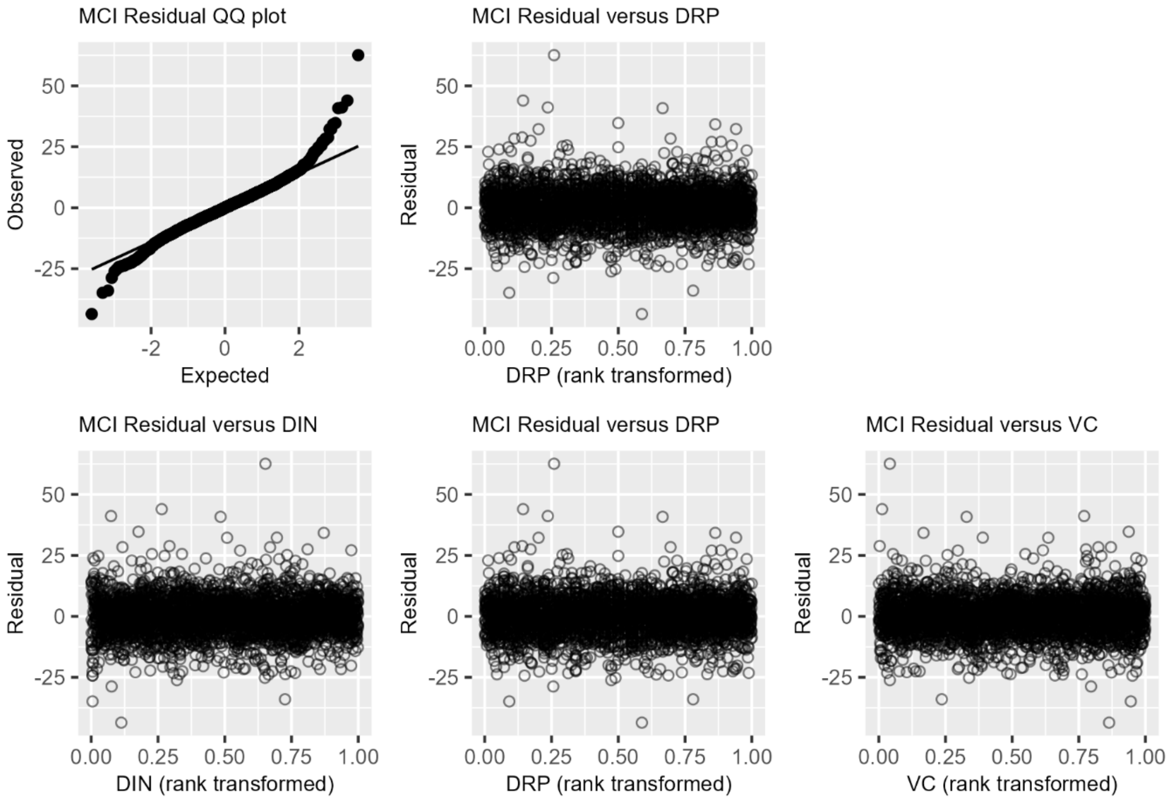


Figure I-1: Residual diagnostic plots for the final MCI MIM.

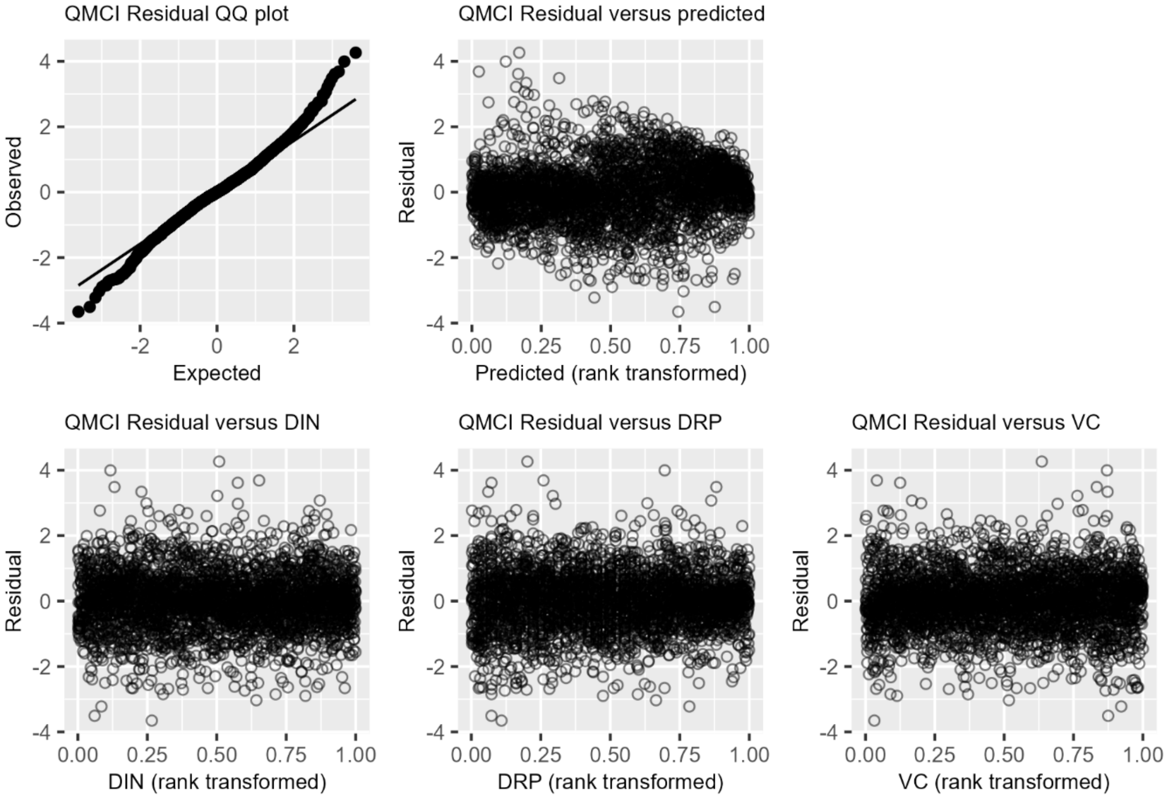


Figure I-2: **Residual diagnostic plots for the final QMCI MIM.**

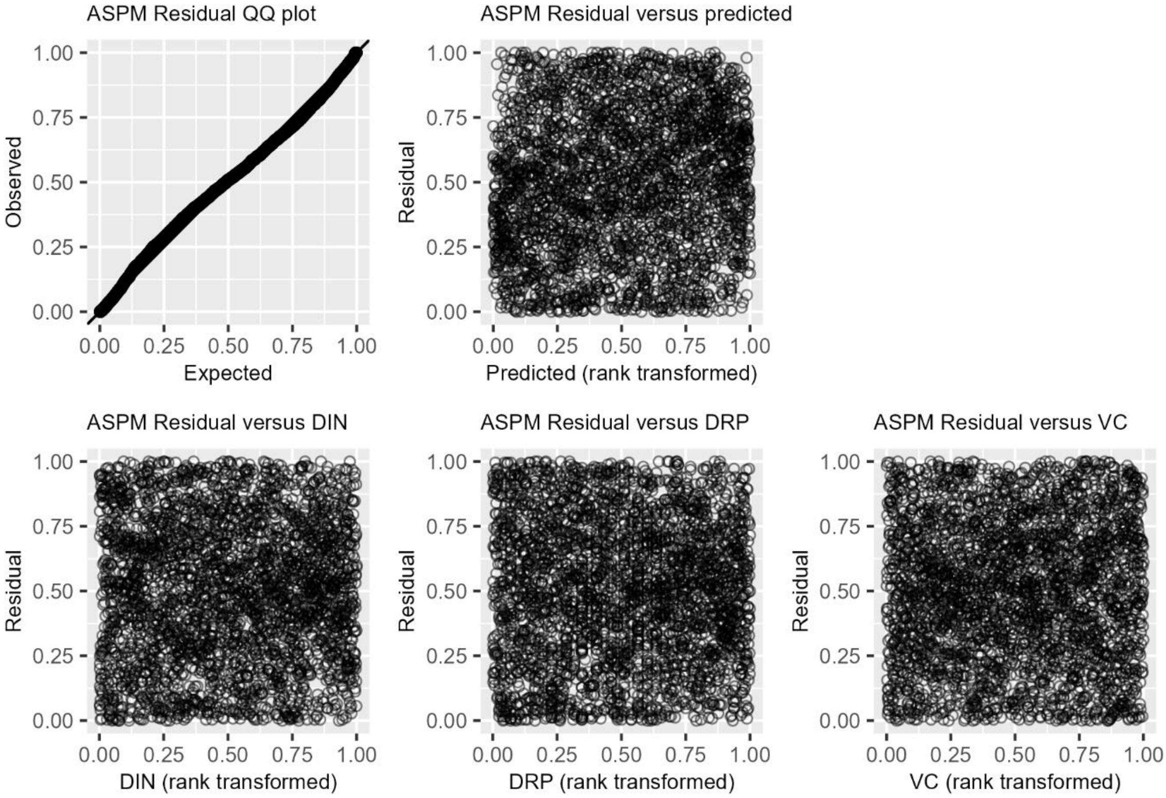


Figure I-3: **Residual diagnostic plots for the final ASPM MIM.**