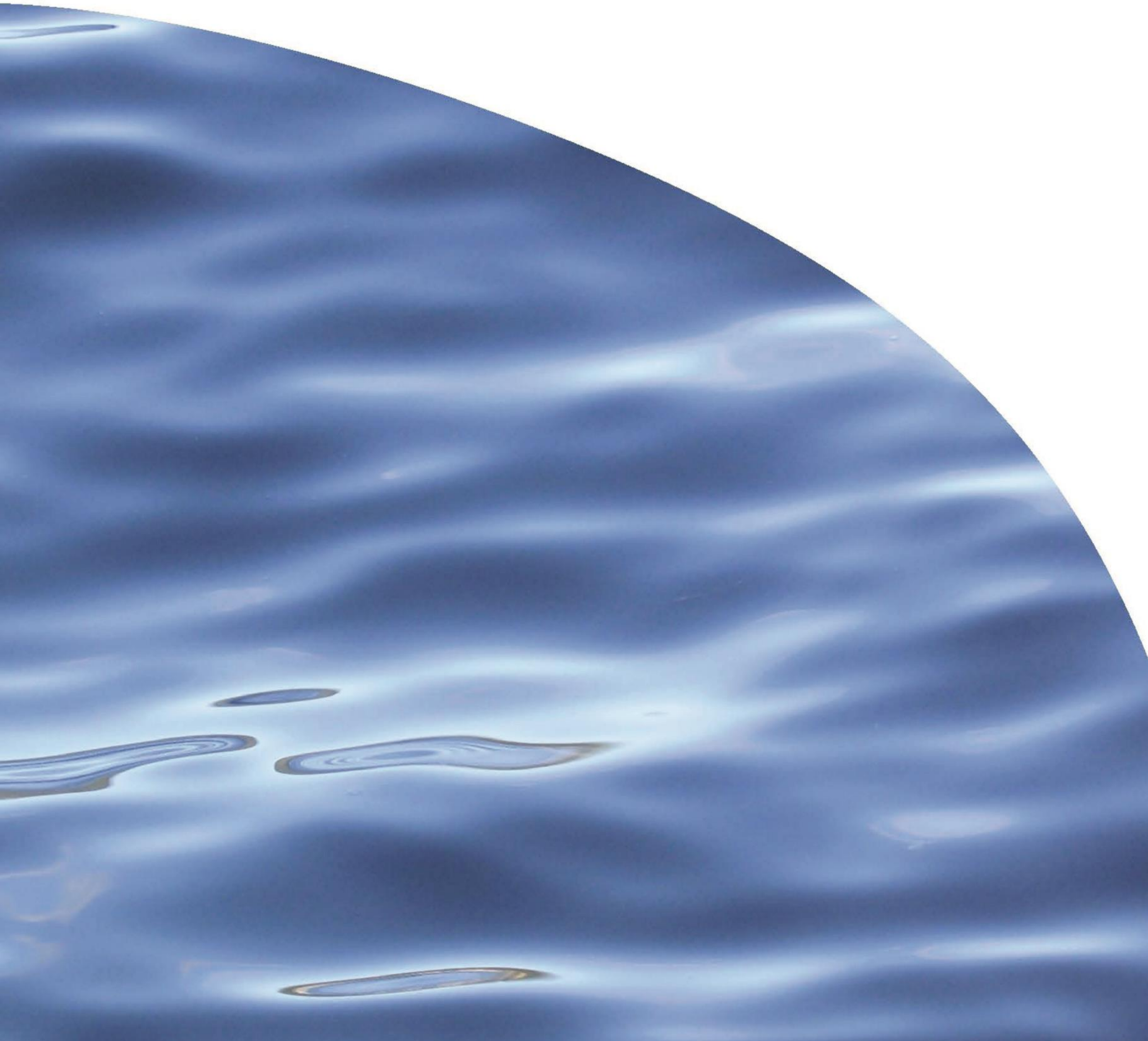
REPORT NO. 3197

DEVELOPMENT AND VALIDATION OF STRESSOR-SPECIFIC MACROINVERTEBRATE METRICS



DEVELOPMENT AND VALIDATION OF STRESSOR-SPECIFIC MACROINVERTEBRATE METRICS

ANNIKA WAGENHOFF[1], ERIC GOODWIN[1], JAVIER ATALAH[1], BRIAN SMITH[2], JON HARDING[3]

[1] Cawthron Institute
[2] NIWA
[3] University of Canterbury

Prepared for Ministry for the Environment

CAWTHRON INSTITUTE
98 Halifax Street East, Nelson 7010 | Private Bag 2, Nelson 7042 | New Zealand
Ph. +64 3 548 2319 | Fax. +64 3 546 9464
www.cawthron.org.nz

REVIEWED BY:
Joanne Clapcott

APPROVED FOR RELEASE BY:
Roger Young

ISSUE DATE: 20 December 2018

RECOMMENDED CITATION: Wagenhoff A, Goodwin E, Atalah J, Smith B, Harding J 2018. Development and validation of stressor-specific macroinvertebrate metrics. Prepared for Ministry for the Environment. Cawthron Report No. 3197. 83 p. plus appendices.

EXECUTIVE SUMMARY

In 2016, Ministry for the Environment (MfE) commissioned a project aimed at investigating macroinvertebrate metrics for the NPS-FM (Clapcott et al. 2017b). As part of this project, significant effort was spent on proof of concept for development of stressor-specific macroinvertebrate metrics for the two major stressors in stream ecosystems: increased loads of nutrients (nitrogen and phosphorus) and fine sediment. Stressor-specific metrics are developed predominantly to respond to a single stressor. In resource management, such metrics are useful for detection of the ecological effects of a single stressor and for diagnosis of the dominant stressor in multiple stressor environments. The effects of nutrient enrichment were attributed to algal proliferation while effects of increased fine sediments were attributed to the deposition on the streambed.

This current project built on previous work and further developed and validated these metrics with the overarching aims to:

1. provide a list of indicator taxa along with indicator values for each stressor
2. provide a list of the best-performing metrics, from among a range of candidate metrics, i.e. formulae for calculation of metric site scores from a macroinvertebrate sample and the information provided in the indicator taxa tables
3. validate the metrics by testing their response to the focal stressor gradients and testing their diagnostic power
4. compare metric response to that of the existing metrics commonly used as indicators of ecosystem health—the MCI and EPT metrics—and to that of macroinvertebrate traits, the latter which have previously been suggested to potentially serve as stressor-specific metrics.

To achieve these aims, the following main tasks were performed.

Assigning tolerance scores to indicator taxa
Indicator taxa were identified and indicator values assigned from output of a gradient forest analysis performed on relative abundance macroinvertebrate data (MCI-level taxonomy) using a dataset consisting of State of the Environment data and research data. The gradient forest analysis builds random forest models for each taxon which included the focal stressor attributes, % sediment cover and chlorophyll-*a*, but also dissolved nutrient concentrations and a set of natural environmental predictors. The fitted functions shown in partial dependence plots for sediment cover and chlorophyll-*a* were used to identify taxa that have a consistent negative or positive response shape and therefore were called decreasers and increasers, respectively. Decreasers are sensitive taxa that, with increasing stressor intensity, proportionally decrease within the macroinvertebrate community, while increasers are tolerant taxa that proportionally increase. However, as a second step, expert knowledge was used to assess whether the response curves were ecologically meaningful before taxa

made it into the list of indicator taxa. For indicator taxa, the species turnover functions were used to assign indicator values ranging from 1 to 10 (10 for the most sensitive taxa and 1 for the most tolerant taxa) and values ranging from 1 to 3 representing the relative strength of response.

In total, 49 indicator taxa were identified from their response to chlorophyll-*a* or % sediment cover, out of which there were 32 for periphyton-specific metrics (26 decreasers, 6 increasers) and 37 indicator taxa for sediment-specific metrics (31 decreasers, 6 increasers). The indicator taxa covered a broad range of taxonomic groups; genera within the EPT insect orders were well represented. Out of the 49 indicator taxa, 29 responded to only a single stressor, either chlorophyll-*a* or sediment, and two taxa responded to both stressors but were decreasers for one and increasers across the other stressor gradient. These 31 'discriminative' taxa have the potential to provide diagnostic power of the resulting stressor-specific metrics. In addition to these 31 discriminative taxa, there were further taxa that were either decreasers or increasers with respect to both stressors but differed in their degree of sensitivity or tolerance. Hence they also provide another level of discrimination between the effects of the two stressors.

Calculating metrics

Twenty-one different ways of calculating a metric score were applied using taxa tolerance scores along with information on the relative strength of response of each taxon to chlorophyll-*a* and sediment. Metrics included those that only took into account decreasers or increasers (e.g. the metric Number of decreasers), those that used the indicator values that had been assigned to both decreasers and increasers (e.g. Sediment MCI, Nutrient MCI) with and without a weighting factor of the strength of response, among other metrics.

Validation of the stressor-specific metrics

Firstly, single-stressor analysis on an independent field observational dataset using simple linear regression models was performed to reduce the candidate set of 21 stressor-specific metrics based on their relationship with the respective stressor. Eight metrics were retained based on a ranking of metrics, which was relatively consistent across sediment and periphyton-specific metrics. The stressor-specific MCI and SQMCI were consistently among the best performing metrics with regard to responding to the focal stressor 'gradient' while those based on information of only decreasers or increasers performed less well. This suggests that inclusion of both decreasers and increasers and the level of sensitivity/tolerance quantified as indicator values in metric calculation significantly improved performance of the metrics.

Secondly, the metrics in the reduced candidate set were tested for their diagnostic power using multiple linear regression analysis and hierarchical variance partitioning. These analyses were performed on the dataset used to develop the metrics and on independent experimental data from a mesocosm study to provide multiple lines of evidence about the usefulness of these metrics. Multiple-stressor analysis on the national field observational dataset (used to develop the metrics) showed that several of the eight metrics responded to the focal stressor gradient only. This suggests that the stressor-specific metrics have diagnostic power. The experimental data were only suitable to test the effects of increasing sedimentation as nutrient enrichment in this experiment did not translate into different levels of periphyton biomass. Sediment level was an important predictor of some of the sediment-specific metrics, but periphyton-specific metrics were also influenced by sediment level, suggesting that many of the indicator taxa which made up metric scores were sensitive to both stressors. Nevertheless, Sediment MCI and Periphyton MCI were not correlated in the experimental dataset, suggesting that if used together they can distinguish between the stressor effects if one stressor (sediment in this case) has a strong effect but not the other (nutrients in this case).

Comparison of stressor-specific and existing metrics and traits

The same multiple-stressor analyses on metrics MCI, SQMCI, EPT richness, %EPT richness and %EPT abundance showed that the majority of these responded to both stressors, confirming their sensitivity to multiple stressors and providing evidence for the usefulness of stressor-specific metrics as complementary tools. Analyses of taxa traits showed that there were few trait modalities that responded to algal proliferation only, while there were several that responded to sedimentation only. This suggests that traits may provide additional useful tools as stressor-specific metrics.

The recently updated macroinvertebrate trait database (Phillips & Smith 2018) is publicly available on the web, however calculation of the trait metrics is not as simple as calculation of the stressor-specific metrics. This is mainly due to trait information being presented mainly at the species level while macroinvertebrate taxa often are identified at the genus or even higher taxonomic levels. Instructions for how to assign trait affinity scores to higher taxonomic levels or provision of nationally-applicable affinity scores for higher levels, in addition to instructions for how to calculate trait metrics, would increase uptake of traits in resource management.

Recommendation of the best stressor-specific metrics for use in resource management

Overall, our analyses suggest that the new stressor-specific metrics have diagnostic power if periphyton and sediment-specific metrics are used in concert. We recommend calculation of metric scores for the following eight types of stressor-specific metrics:

- Sediment or Periphyton MCI
- Sediment or Periphyton SQMCI
- Number of sensitive taxa
- % Sensitive taxa
- % Sensitive individuals

- Number of tolerant taxa
- % Tolerant taxa
- % Tolerant individuals.

Among these, the stressor-specific MCI and SQMCI are likely to be the most reliable in detecting and disentangling the effects of sedimentation and nutrient enrichment via eutrophication.

Calculation is based on a macroinvertebrate sample collected semi-quantitatively at a stream site and processed to determine coded abundances. Sediment and Periphyton MCI, Number of sensitive/tolerant taxa and % Sensitive/tolerant metrics, however, only require presence-absence data.

These metrics are ready for use by regional councils to diagnose the causes of degraded ecosystem health and assess policy effectiveness. For example, these metrics may be applied on a site-specific scale to

1. track the likely effects of these two stressors at monitoring sites over time
2. detect early signs of ecological degradation
3. diagnose whether nutrient enrichment or sedimentation or both are the likely leading causes of degraded stream ecosystem health
4. estimate the ecological effects of a significant disturbance event
5. measure restoration success.

These metrics may also be applied on a regional scale to:

6. track the likely effects of these two stressors across the region
7. diagnose whether nutrient enrichment or sedimentation or both are the likely leading causes of degraded stream ecosystem health in the region
8. test which land use practices have the largest effects on stream ecosystem health, and why
9. test policy effectiveness on stream ecosystem health, specifically with respect to measures that control or reduce the inputs of nutrients and fine sediments
10. define thresholds for nutrient concentrations, periphyton and sediment aimed at protection of ecosystem health, or thresholds for the metrics themselves to use within an ecosystem health assessment framework.

It is important to note these metrics, if used for diagnostic purposes, have to be used in conjunction as sediment-specific and periphyton-specific metrics are also influenced by increasing chlorophyll-*a* and sediment cover, respectively, due to considerable overlap of indicator taxa with similar response direction to both stressor gradients. Hence, for example, a declining trend of the Sediment MCI at a site is not necessarily indicative of increasing sedimentation but sedimentation is more likely the reason for impairment if the Periphyton MCI does not show such a strong declining trend. Also, the absolute metric scores should not be compared among sediment-specific and periphyton-specific metrics. Hence, in order to compare the magnitude of impairment attributable to each stressor, in most cases it will be necessary to determine what the reference metric score should be and quantify impact in relation to reference.

Furthermore, these metrics are also likely to be affected by other stressors and in particular to changes in flow. For example, water abstraction increases sediment deposition and algal proliferation. In a separate report, for the same overarching project, the New Zealand specific Lotic-invertebrate Index for Flow Evaluation (LIFENZ) has been tested and confirmed to be a hydrologically-sensitive macroinvertebrate metric (Greenwood 2018). Use of the sediment-specific and periphyton-specific metrics in conjunction with the LIFENZ may help in disentangling the effects of sedimentation, nutrient enrichment and flow reduction. However, this has not been tested to date.

Finally, these metrics, as all macroinvertebrate metrics, are influenced by natural environmental gradients, which is a result of the species being adapted to different environmental conditions. Hence, metric scores observed at reference sites will differ according to the specific natural environmental conditions (often called ‘stream type’) and equally metric scores at impact sites cannot necessarily be compared among sites. The influence of natural environmental gradients will introduce variability in regional analyses where a time-for-space approach is being used but broad patterns may still be detectable. However, in order to address the questions listed above, and as already mentioned, in most cases it will be necessary to determine what the reference metric score should be and quantify impact in relation to reference.

Application of the metrics will demonstrate their usefulness in practice and provide an opportunity to further improve the metrics. Scope for improvement of the metrics is mainly through identifying further indicator taxa. More robust stressor data linked to macroinvertebrate data will provide new, high-quality datasets for future analyses. Identification of further indicator taxa and more accurate assignment of indicator values for sediment-specific metrics would most likely benefit from datasets where annual medians of sediment cover linked to each macroinvertebrate sample are calculated from monthly samples, and also from the application of standardised deposited sediment methods. The difficulty of reliably identifying indicator taxa for periphyton-specific metrics, on the other hand, is likely related predominately to the multiple pathways via which algal proliferation can affect taxa. However, possibly linking an upper percentile of monthly chlorophyll-*a* samples to each macroinvertebrate sample rather than an annual median may improve the likelihood of identifying indicator taxa.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION

1.1. Background

Benthic macroinvertebrates have been used worldwide as biological indicators in stream ecosystems for decades. In most cases, biotic indices or metrics are based on the composition of the macroinvertebrate community and combine the information on the presence-absence or abundance of multiple taxa with known sensitivities to human impacts into a single number.

1.1.1. Existing macroinvertebrate metrics

In the early days of using macroinvertebrate as indicators of human impacts, metrics were developed to indicate water quality and particularly targeted organic pollution and nutrient enrichment. In New Zealand, the use of biotic indices started in the mid-1980s with development of the New Zealand-specific Macroinvertebrate Community Index (MCI) by Stark (1985). The MCI is calculated from tolerance values that Stark assigned to about 180 macroinvertebrate taxa. Tolerance value assignment for most taxa was based on their relative abundance within streams that were categorised into one of three pollution classes that were selected to represent different degrees of organic and nutrient pollution, but tolerance values assigned to other taxa were purely based on expert knowledge.

The MCI was intended to be used as an index of water quality in stony streams. However, as the MCI was sensitive to other stressors (e.g. deposited fine sediment) and as macroinvertebrates play an important role in stream ecosystems, regional councils started to use the MCI as an indicator of overall ecosystem health in all stream types. Similarly, taxa of the insect orders Ephemeroptera (mayflies), Plecoptera (stoneflies) and Trichoptera (caddis flies) are known to be sensitive to a range of stressors, hence the so-called EPT metrics are calculated from their presence-absence or abundance, either in absolute numbers or in relation to the other likely less-sensitive taxa in the sample, and used worldwide as indicators of stream ecosystem health.

While the strength of the MCI and EPT metrics lies in the sensitivity to multiple stressors, their weakness necessarily lies in their inability to diagnose the cause of degraded stream health, which could be due to one or more stressors. Diagnosis is a crucial step toward implementation of effective management actions—such as setting rules and regulations for resource use or deciding on practical rehabilitation measures. The recent 2017 amendments to the National Policy Statement for Freshwater Management (NPS-FM, MfE 2014) require regional councils to use the MCI as a monitoring tool. Furthermore, if the MCI score falls below 80 or shows a declining trend, councils are required to establish methods to respond and ensure that methods investigate the causes of declining trends, seek to halt declining trends, or seek to improve the MCI score if it is below 80.

Analysis of the state and trends of water and habitat quality parameters relevant to the health of macroinvertebrate communities is one approach to investigating causes of degraded stream health. However, these data are not always available at respective monitoring sites. Water quality monitoring is costly due to laboratory analyses, and both water and habitat quality investigations are costly due to the requirement for frequent sampling (usually monthly) to capture overall conditions leading to ecological effects. In addition, changes to the flow regime may have also caused degradation in macroinvertebrate communities but flow monitoring is also very costly.

1.1.2. Stressor-specific metrics

An alternative and cost-effective way for investigation of causes of stream degradation would be the use of stressor-specific macroinvertebrate metrics calculated from the data that are already at hand at focal monitoring sites. Or, if water managers wanted to investigate causes at a regional scale or for a specific catchment at a higher resolution, one-off or repeated annual macroinvertebrate sampling may help in discerning the most common cause of degraded stream health.

In addition to the use of stressor-specific metrics as diagnostic tools, the strength of the relationship of these metrics with the respective stressor makes them useful for identifying thresholds that help with definition of instream objectives (targets) for manageable stressors, e.g. development of a sediment or nutrient attribute for the National Objectives Framework (NOF) within the NPS-FM, to manage streams for ecosystem health. The sediment-specific macroinvertebrate metrics that have been developed in the previous MfE project have already been used to identify thresholds to help development of a deposited fine sediment NOF attribute (Depree et al*.* 2018).

Stressor-specific macroinvertebrate metrics are developed to respond predominantly to a single stressor. Development of such metrics involves identification of indicator taxa for that specific stressor and may also involve assignment of indicator values to them. However, there are broadly five potential limitations to the ability to develop such indices. First, if two or more stressors have the same mode of action in affecting macroinvertebrate species it is impossible to tease apart their individual effects. Secondly, there must be species that are sensitive to one stressor but tolerant to the other, these are discriminative taxa. Thirdly, for broad-scale use of metrics, in this case across New Zealand, indicator taxa will have to be widespread. Fourthly, there may be limitations related to properties of datasets such as high correlation between the focal stressors which prevents teasing apart their individual effects. Finally, presence of strong interactive effects between focal stressors prevents teasing apart their individual effects using statistical models that do not account for interactions.

Stressor-specific macroinvertebrate metrics have been developed overseas using various approaches to indicate the effects of deposited fine sediment (Zweig & Rabeni 2001; Bryce et al*.* 2010; Relyea et al*.* 2012; Extence et al*.* 2013; Murphy et al*.* 2015; Hubler et al*.* 2016), of phosphorus or nitrate (Smith et al*.* 2007) or eutrophication (Haase & Nolte 2008) and a macroinvertebrate index has been developed to indicate the impact of flow regimes (Extence et al*.* 1999). There are also examples of stressor-specific macroinvertebrate indices developed in New Zealand. The Acid Mine Drainage Index (AMDI) was developed to assess chemical pollution from coal mining impacts taking into account the combined effects of increased metals and conductivity and lowered pH (Gray & Harding 2012) and the LIFENZ (Greenwood et al*.* 2016) is a hydrologically-sensitive macroinvertebrate index developed based on the British example, the Lotic Index for Flow Evaluation (Extence et al. 1999). These two New Zealand-specific indices have not found wide application in resource management, possibly because little work has been done validating these indices and demonstrating their usefulness with practical examples.

Recently, the Ministry for the Environment (MfE) commissioned a large project aimed at investigating macroinvertebrate metrics for the NPS-FM (MfE contract no. 21630, Clapcott et al*.* 2017a, 2017b). As part of this project, significant effort was spent on proof of concept for development of nutrient-specific and sediment-specific metrics as increased loads of nutrients (nitrogen and phosphorus) and fine sediment are major stressors in stream ecosystems.

We define a *stressor* as an attribute that, as a consequence of human activity, has exceeded its normal range of variation and affects macroinvertebrate taxa and communities. As we expected nutrient enrichment to affect macroinvertebrate communities mainly via algal proliferation, which can be measured as chlorophyll-*a*, we refer to nutrients and chlorophyll-*a* as stressors. Equally, as we expect increased sediment loads to have major effects on macroinvertebrate communities via deposition, which can be measured as % sediment cover on the stream bed, we refer to fine sediment and % sediment cover as stressors.

Periphyton-specific metrics were developed to respond to algal proliferation while sediment-specific metrics were developed to respond to increased fine sediment deposition. The statistical approach to identifying indicator taxa and assigning indicator values to them was based on models that quantify the stressor-response curves of individual taxa and species turnover functions taking into account the effects of multiple stressors. Model output was derived from ‘gradient forest’ analysis, developed by Ellis et al. (2012) to identify community thresholds across stressor gradients. Use of the gradient forest analysis for stressor-specific metric development, however, is novel. Expert knowledge was used to assess whether the response curves were ecologically sound. Overall, the results suggested that these metrics may be able to tease apart the individual effects of the two stressors, however, this has not been tested in detail. As part of a larger project (MfE contract no. 18502), MfE commissioned further work to develop and validate these stressor-specific metrics (presented in this report) and to also further investigate the flow-specific LIFENZ (Greenwood 2018).

1.2. Project scope and aims

The project reported here builds on previous work commissioned by MfE (Clapcott et al. 2017b) that provided proof of concept for development of stressor-specific macroinvertebrate metrics for nutrient enrichment and fine sediment deposition. The overarching aims of this project are to

1. provide a list of indicator taxa along with indicator values for each stressor
2. provide a list of the best-performing metrics, from a range of candidate metrics for use in freshwater management, i.e. formulae for calculation of metric site scores from a macroinvertebrate sample and the information provided in the indicator taxa tables
3. validate the metrics by testing their response to the focal stressor gradients and testing their diagnostic power (which will contribute to deciding upon the best performing metrics)
4. compare metric response to that of the existing metrics, the MCI and EPT metrics, and to that of macroinvertebrate traits, the latter which have been suggested to potentially serve as stressor-specific metrics.

To achieve these aims, we performed the following tasks:

- gradient forest analysis on a larger dataset by adding State of the Environment (SoE) data to the previously used research dataset (Section 2.1.2)
- gradient forest analysis on macroinvertebrate presence-absence data in addition to analysis on relative abundance data (Section 2.1.3)
- seek expert opinion from a second macroinvertebrate ecologist to assess whether the empirically-derived response curves are ecologically sound (Section 2.1.4)
- add more metrics to the previously suggested list of ten candidate metrics, including metrics that use the strength of response of the indicator taxa to the focal stressor as a weighting factor in metric score calculation (Section 2.1.7)
- investigate the relationships of the metrics and other existing metrics as well as macroinvertebrate traits with the two focal stressor gradients using (1) an independent field observational dataset (i.e. not used in metric development), (2) the training data (i.e. data used in metric development), and (3) an experimental dataset. (Section 2.2)

While this project builds on previous work, we present a level of detail so that this is a stand-alone report but refer to previous reports (Clapcott et al. 2017a, 2017b) for those who require even more detail, for example on dataset compilation. This project is part of a much larger MfE project (MfE contract no. 18502 – Water Quality and Flows – State and Trends) composed of multiple individual projects. Among these, investigation of the flow-sensitive macroinvertebrate index LIFENZ is closely related to our project but presented in a separate report (Greenwood 2018).

2. METHODS

2.1. Development of stressor-specific metrics

2.1.1. Overview and rationale for using a multi-stressor analysis

There are three main steps to metric development: (1) identification of sensitive or tolerant taxa (indicator taxa) and/or assignment of indicator values to taxa, (2) combining the information on indicator taxa and indicator values into a metric score, and (3) validating the metric, i.e. testing whether the metrics respond to the stressor that they were developed for and whether they are able to discriminate between the different stressors.

In the literature, there is a vast range of approaches to identification of indicator taxa and assignment of indicator values (step 1). Two broad distinctions are whether these are determined empirically (from data) or based on expert knowledge. During the previous MfE project, we tasked multiple New Zealand experts on benthic macroinvertebrates to assign indicator values on a four-point scale to taxa based on their expected response to either nutrient enrichment via algal proliferation or to sedimentation. There was little distinction between the indicator values assigned to the taxa for either nutrient enrichment or sedimentation which makes it impossible to develop stressor-specific metrics. Furthermore, for several taxa there was strong disagreement with respect to indicator value assignment among experts. We concluded that it may be difficult for experts to disentangle the effects of these two stressors as their experience is drawn from observations of streams where these two stressors commonly co-occur. Secondly, there are multiple mechanisms for each of the two stressors via which they could affect individual taxa. For example, the mechanisms could be more directly through clogging of the gills by fine sediments or more indirectly through degradation of the food sources or through influencing biological interactions such as competition or predation. Altogether, this complexity makes it difficult to predict taxon response based on knowledge of their traits.

Consequently, we decided to use an empirical approach to identification of indicator taxa and assignment of indicator values along with an approach that takes into account the effects of multiple stressors. We then used expert knowledge to assess whether the response curves are ecologically sound. We used gradient forest (GF) analysis (Ellis et al. 2012), which is a relatively new statistical approach developed for calculating the importance of environmental drivers and identifying assemblage thresholds using regression tree-based random forest (RF) models for individual taxa . For example, this approach has recently been used to identify congruence in stream assemblage thresholds among macroinvertebrate, periphyton and bacterial assemblages in response to nutrient and sediment gradients (Wagenhoff et al. 2017a). Here, however, we adopted the analytical approach to identify indicator taxa, assign indicator values and quantify the strength of the response. To our knowledge, GF has never been applied for this purpose.

The strength of flexible models, such as RF, is the ability to model complex, nonlinear response shapes and the inclusion of multiple predictors that strengthen the evidence of correlative association between stressors and taxon responses. Furthermore, potential complex interactions among stressors and also natural environmental variables are automatically handled (Cutler et al*.* 2007). These characteristics are promising for disentangling the effects of multiple stressors (here sediment and nutrients/periphyton biomass) and natural environment gradients typically prevalent in broad spatial scale datasets. However, the higher the correlations between drivers within a dataset the less able we are to disentangle the effects of multiple drivers. While RF models are capable of dealing with correlated predictors well, it is unknown at what degree of correlation the interpretation of the results (i.e. assignment of indicator values) becomes less reliable. Validation of the metrics using independent spatial datasets that also contain minimum correlation between stressors, or using experimental datasets, would be best to test the validity of stressor-specific indicator values and hence metrics.

2.1.2. Dataset

The macroinvertebrate dataset used for stressor-specific metric development was compiled during previous MfE projects and consisted of research and State of the Environment (SoE) data from sites spread across New Zealand (Figure 1). The samples (N = 501) included measurements of deposited fine sediment assessed instream, periphyton chlorophyll-*a* as well as of dissolved inorganic nitrogen (DIN) and dissolved reactive phosphorus (DRP) alongside macroinvertebrate data. The requirement of these stressor data for our multiple-stressor analysis mainly contributed to a sample size that was considerably smaller than that of the original large dataset.

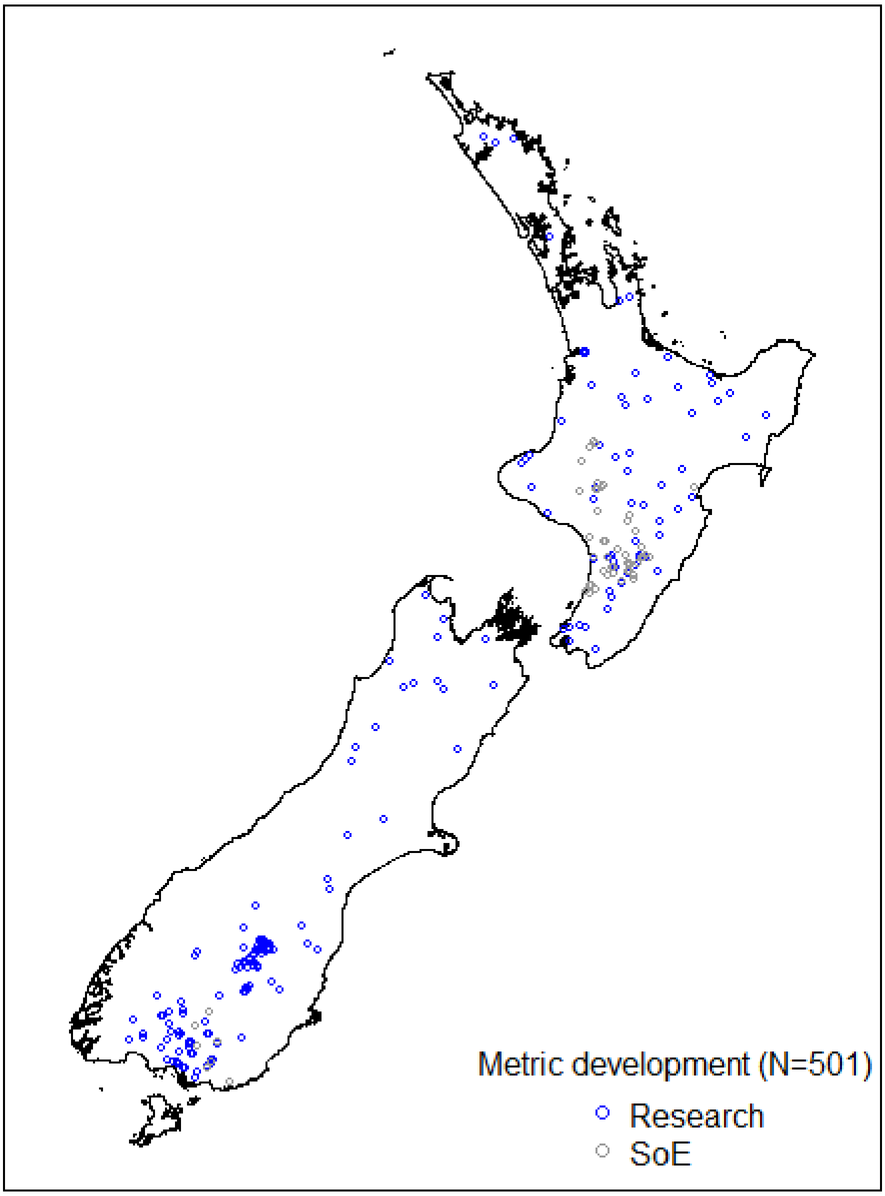


Figure 1. Map of New Zealand showing the spatial spread of the 501 samples points used for stressor-specific metric development.

Of the 501 samples, the majority (N = 297) were selected from a research dataset that was specifically compiled for stressor-specific metric development (MfE contract no. 21630) which is described in detail in Clapcott et al. (2017a). Seven individual research studies were included: five published field surveys (Quinn & Hickey 1990; Wagenhoff et al. 2011; Lange et al. 2014; Macher et al. 2016; Wagenhoff et al. 2017b) and unpublished data from a MSc thesis (Holmes 2008, only from control sites) and from the Whatawhata research station provided by Elizabeth Graham and John Quinn. A maximum number of five samples from a single stream location were included that were taken usually two or more years apart from each other. Selected samples from the research dataset were collected during the period from 1987 to 2012.

The remaining samples (N = 204) were selected from an SoE and National River Water Quality Network (NRWQN) macroinvertebrate dataset originally compiled by Martin Unwin and matched with stressor data for two previous projects (MfE contracts 21511 and 21630) described in depth in Clapcott et al. (2017b). Briefly, due to potentially significant annual variation of sediment, nutrient and periphyton conditions at a single site, a median value had been calculated from all available stressor data collected within the same month and the 12 months prior to macroinvertebrate sampling. However, monthly (or more) observations were not always available. For deposited sediment in particular, often only one or two observations were available and hence used as representing the 12-month time period. On the other hand, for chlorophyll-*a*, DIN and DRP, the medians were predominately calculated from monthly samples. The vast majority of samples that we selected for metric development were collected by Horizons Regional Council with only a few samples collected by Hawke's Bay Regional Council and Environment Southland. The data were collected during the period from 2010 to 2015, hence a maximum number of six samples from each of 59 SoE monitoring sites were included.

Generally, macroinvertebrates were sampled quantitatively (with a Surber sampler) or semi-quantitatively (with a kicknet) in the field and processed in the lab using standard protocols (Stark et al. 2001). The taxonomic resolution originally provided appeared to be mostly to the 'most practical level', being a level that is feasible with use of a stereomicroscope and standard keys such as Winterbourn et al. (2006). For EPT taxa, this meant to at least genus but often to species level, and overall to at least the level required to calculate the MCI (mostly genus level for EPT taxa but higher levels for other taxa). We brought the taxonomic resolution for all samples to that required to calculate the MCI (hard-bottom or soft-bottom versions (Stark & Maxted 2007a)). This decision was made for two reasons. Firstly, use of the same taxon list for MCI and stressor-specific metrics makes calculation of site scores simple, practicality and comprehension being important determinants for whether these new metrics will be taken up by management agencies. Secondly, councils have been collecting macroinvertebrate data routinely for over 20 years with data processed to mainly MCI-level taxonomic resolution, hence, site scores can be calculated retrospectively for long-term datasets to answer various applied research questions. The macroinvertebrate datasets were checked for use of old and new taxon names and adjusted accordingly. Overall, in our dataset 131 out of the 191 taxa on the MCI taxon list (hard-bottom or soft-bottom versions, (Stark & Maxted 2007a)) were represented.

The protocols for measuring the stressor variables were also generally consistent across the samples. Deposited fine sediment was always assessed instream and predominantly as a visual estimate of the percentage cover of fine sediment (< 2 mm) on the streambed (mean of several observations taken at each sampling site). Only on some occasions was the deposited fine sediment estimate derived from a Wolman pebble count. Both of these methods for deposited fine sediment assessment are described in the standard protocols (Clapcott et al*.* 2011). Periphyton chlorophyll-*a* estimates were derived from scrapings and lab methods applied similar to those described in standard protocols (Biggs & Kilroy 2000). Finally, DIN and DRP were determined from water grab samples using standard lab methods similar to those described by APHA (2005). The dataset contains broad gradients in these stressor variables, for example, deposited sediment ranged from 0 to 100% and periphyton chlorophyll-*a* from 0 to 374 mg/m^2 (Table 1).

For each macroinvertebrate sampling site, we retrieved information on elevation, position in the stream network, catchment land use, surface water abstraction pressure and relevant environmental variables from existing databases via the NZReach ID within the national river network (REC v1, Snelder & Biggs 2002). Environmental variables describe catchment geology and morphology, segment slope, bed substrate, flow, temperature and shading and have been previously shown to influence macroinvertebrate communities (e.g. Clapcott et al. 2016; Wagenhoff et al. 2017a, b). These environmental variables as well as an estimate of surface water abstraction pressure were used as predictors in the models alongside the focal stressor variables to account for their potential effects on macroinvertebrate taxa. Summary statistics for all variables as well as sources of the databases and Spearman rank correlations among these variables can be found in Table 1 and Figure 2, respectively.

Table 1. Summary statistics and description of variables describing characteristics of the macroinvertebrate sampling sites including segment-scale and upstream/catchment-scale variables (denoted with 'Seg' and 'US', respectively), see * below for data sources.

Variable	**Description**	**Mean**	**Median**	**Min**	**Max**
Order	Stream order	4.5	5.0	1	7
Elevation	Altitude of the stream segment (m)	198	130	10	978
Land use					
T1NativeVeg	Percentage catchment land cover in native vegetation including forest and scrubland	30	20	0	100
T2PastoralHeavy	Percentage catchment land cover in heavy pastoral land uses including exotic grassland, short rotation cropland, orchards and vineyards	46	49	0	100
Stressors					
SEDCOVER	% sediment cover (< 2 mm) estimated instream	18	11	0	100
CHLA	Periphyton biomass measured as chlorophyll-*a* from rock scrapings (mg/m^2)	34	14	0	374
DIN	Dissolved inorganic nitrogen (mg/L)	0.446	0.150	0.005	11.6
DRP	Dissolved reactive phosphorus (mg/L)	0.017	0.011	0.001	0.404
StAccMaxRate	Total upstream accumulated maximum (consented) rate divided by modelled median flow (maxrateToQ50 in the database); greater than 1 means that more than the median flow is consented	0.045	0.001	0	3.9
Environmental predictors					
MALF	Mean annual low flow for a segment (m^3/s)	3.9	0.8	0.0002	45.2
FRE3	Annual frequency of flood events > 3x median annual flow for a segment	12.9	13.1	1.8	37.4
SegFlowStability	Ratio of mean annual low flow / mean annual flow	0.17	0.15	0.04	0.49
SegSumT	Summer air temperature (°C)	16.3	16.8	12.0	19.6
SegTSeas	Seasonal air temperature range (°C)	0.2	0.7	-4.1	2.6
SegShade	Riparian shade (proportion)	0.16	0.06	0	0.80
SegSlope	Slope (°)	0.6	0.2	0	12.3
SegSubstrate	Weighted average of proportional cover of bed substrate using categories of 1 – mud, 2 – sand, 3 – fine gravel, 4 – coarse gravel, 5 – cobble, 6 – boulder, 7 – bedrock	3.8	4.0	1.6	5.6
USSlope	Average slope in the catchment (°)	13.0	13.1	0.3	29.5
USRainDays	Days per month with rainfall greater than 25 mm	10.8	10.1	1.2	79.4

Variable	Description	Mean	Median	Min	Max
USCalcium	Average calcium concentration of underlying rocks, 1 = very low to 4 = very high	1.5	1.5	1.0	2.9
USPhosphorus	Average phosphorus concentration of underlying rocks, 1 = very low to 5 = very high	2.2	2.0	1.0	5.0
USHardness	Average hardness of underlying rocks, 1 = very low to 5 = very high	2.9	2.9	1.0	5.0

* Stream order and elevation were retrieved from the River Environment Classification database (Snelder & Biggs 2002); catchment land use was calculated from Land Cover Data Base layer v3 (https://lris.scinfo.org.nz/layer/304-lcdb-v30-deprecated/); catchment and segment-scale predictors were retrieved from the Freshwater Ecosystems New Zealand database (Leathwick et al. 2011); flow statistics MALF and FRE3 (details in Booker 2013; Booker & Woods 2014) and StAccMaxRate (details in Booker et al. 2016) were downloaded from the Ministry for the Environment website (https://data.mfe.govt.nz/table/2536-natural-river-flow-statistics-predicted-for-all-river-reaches/; accessed 23 August 2016 and https://data.mfe.govt.nz/table/3614-accumulated-freshwater-takes-201314/; accessed 15 June 2017, respectively).

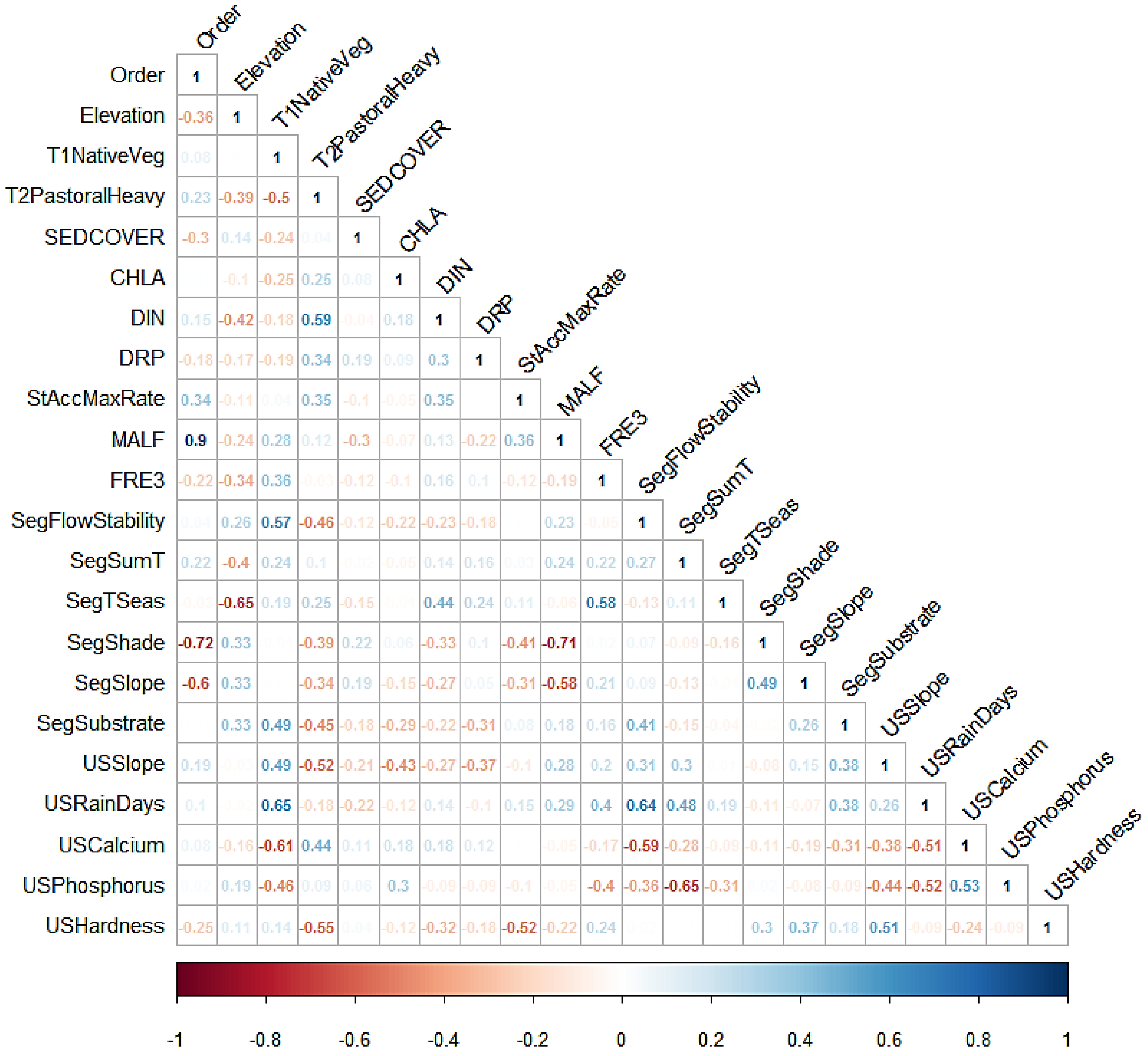


Figure 2. Spearman rank correlations among stressors and land use and other environment predictor variables. The level of shading is reflective of correlation strength, see scale bar at the bottom of the figure.

2.1.3. Description of the gradient forest analysis

The GF analysis is described in detail by Ellis et al. (2012). It is performed in the freely-available statistical programme R (R Core Team 2016, here in R version 3.3.2) using two R packages provided by the same authors. A shorter description of the computational method of the GF analysis available in Wagenhoff et al. (2017a) is given below.

First, R package ‘extendedForest’ builds RF models for each taxon consisting of 500 regression trees. Each regression tree is fitted to a bootstrap sample of the observations and partitioning of the data is performed by the best split (minimising error variance) tested on a random subsample of the predictors (Cutler et al. 2007). Each split is associated with an importance value reflecting the degree of change in abundance. The RF model predictions are averages of the predictions of each tree. The goodness-of-fit measure R^2 (pseudo R^2) is the proportion of the variance explained by the RF model and derived through cross-validation, estimated from the error variance of the out-of-bag sample process (Ellis et al. 2012; Wagenhoff et al. 2017a). The package extendedForest calculates an improved, more robust measure of predictor importance within each RF model by taking into account correlation between predictors. We used a default correlation threshold of 0.5. From the RF model, partial dependence plots can be produced that are the fitted functions for each predictor gradient showing the marginal effect of each predictor on the ecological response when all other predictors are held at constant values—in this study, at their mean values.

Secondly, the R package ‘gradientForest’ uses information only of those taxon-specific RF models with $R^2 > 0$. It computes species turnover functions using information on the predictor splits and associated importance within all trees in the forest by accumulating the split importance values across each environmental gradient. As the distribution of splits is biased toward values that are more densely sampled, split importance is standardised by density. Furthermore, turnover is normalised to sum to the amount of variation explained by the predictor (R^2_{fp})[1]. In essence, species turnover functions depict change in abundance that is normalised so that total turnover over the range of the predictor p is R^2_{fp}. The GF output also provides the importance of each predictor for overall compositional turnover, incorporating all species, which is calculated by averaging across species (Ellis et al. 2012).

Two GF analyses were undertaken, one using macroinvertebrate presence-absence data and one using relative abundance data after log-transformation to approximate normal error distributions, using the formula $\ln(x + \min(x > 0))$ where x is the relative abundance expressed as a proportion. Comparison of the partial dependence plots from GF analysis on the log-transformed relative abundances and on presence-absence of the taxa suggested that response shapes across the focal stressor gradients were less informative for the models based on presence-absence data, and hence we preceded with the analysis of the relative abundance data only.

Analysis was restricted to taxa that had at least 10 non-zero values, found in at least 10 out of the 501 samples. Out of 131 taxa (MCI-level taxonomy) in the dataset, 87 taxa entered the analysis while 44 taxa were too infrequent. The GF approach first calculates RF models for each of these taxa. In total 18 predictors were included in the models, the four focal stressor variables, water abstraction pressure and 13 environmental predictors (Table 1). For all taxa with an RF model that had an $R^2 > 0$,

---

[1] R^2_{fp} is the portion of the random forest model R^2 that has been attributed to the predictor according to its conditional importance; subscript ‘f’ for ‘forest’ specifies that the goodness of fit R^2 is for a random forest model consisting of a forest of regression trees (in Ellis et al. (2012) it is specified as R^2_f but we use R^2 for simplicity) while subscript ‘p’ for ‘predictor’ specifies that it is the portion that is attributed to a predictor.

i.e. for which our 18 predictors had at least some explanatory power, species turnover functions were calculated.

2.1.4. Identification of 'decreasers' and 'increasers' in response to a stressor gradient

The assumption is that taxa decreasing across the stressor gradient ('decreasers') are sensitive taxa while taxa increasing across the stressor gradient ('increasers') are taxa tolerant to the stressor, or even favouring the stressor. Both decreasers and increasers are indicator taxa for the effects of the stressors on macroinvertebrate communities. Species turnover functions do not depict whether a taxon decreased or increased across the stressor gradient. Hence, decreasers and increasers were identified using both the modelled response shapes as well as expert opinion, as follows.

First, the partial dependence plots of the RF models (with overlay of data points) were visually investigated, and taxa with response shapes that are either predominantly negative or positive were named decreasers or increasers, respectively. Taxa that showed neither of these patterns were classified as 'unclear' (examples shown in the Results), which may be a result of taxa being unresponsive to the stressor gradient or due to insufficient data. Random forest models are good predictive models but because they are not parametric models, it is not possible to test for statistical significance of the parameter estimates (i.e. response shape). Nevertheless, a clear positive or negative response shape provides some evidence for taxa increasing or decreasing to a stressor gradient, respectively. On the other hand, the decision of whether small, directional variations in the response shape are meaningful or due to unusual data points are somewhat subjective. Hence, as a second step, two macroinvertebrate experts (Jon Harding from the University of Canterbury and Brian Smith from NIWA) were presented with the partial dependence plots and initial assignment due to model output and they assessed independently whether the response shapes make ecological sense. Final assignment was made based on both analytical output and expert opinion about taxa response to the stressor gradients.

2.1.5. Indicator value assignment

Indicator values were assigned to rank the sensitivity or tolerance of the decreasers and increasers, respectively. Ranking was performed by identifying the value in the stressor variable where the species turnover function reaches 25%[2] of the total turnover for each % sediment cover and chlorophyll-*a*, which we now call the 'threshold' (example shown in the Results). Ranking was done by splitting the range from the lowest to the highest threshold (within each of the decreaser and increaser groups) into 5 equal bins and then assigning indicator values to them as follows. Indicator values for decreasers were assigned to range from 10 to 6 (i.e. 10, 9, 8, 7 or 6) with values of 10 being assigned to the most sensitive taxa that are associated with low stressor values (i.e. those with the lowest 'thresholds'), and values of 6 being

---

[2] In the exploratory stage of metric development, we had used 50% of the maximum cumulative importance to assign the thresholds and it appeared that tolerance value assignment was not much influenced.

assigned to sensitive taxa that can tolerate a certain degree of sediment or periphyton biomass before being negatively affected. Indicator values for increasers were assigned to a range from 1 to 5 (i.e. 1, 2, 3, 4 or 5) with values of 1 being assigned to the taxa that are most associated with high stressor values (i.e. those with the highest 'thresholds'), and values of 5 being assigned to tolerant taxa that already increase at lower amounts of sediment or periphyton biomass, therefore being less strongly associated with only high stressor values as those with values of 1.

We split the range from lowest to the highest threshold into 5 equal-sized bins in two different ways, based (1) on the raw stressor gradient, and (2) on the natural log-scaled stressor gradient (i.e. this resulted in two alternative indicator values for each taxon). As taxa often respond in a log-linear fashion to stressors, in particular to nutrient enrichment, the latter assignment would theoretically be able to better differentiate sensitivity between the most sensitive taxa.

A 1-10 scale was chosen for tolerance assignment as managers in New Zealand are familiar with this scale, which has been used for indicator value assignment for the most commonly-used macroinvertebrate metric to indicate stream health, the MCI. However, the approaches to indicator value assignment between the MCI and our approach fundamentally differ in that our approach uses stressor-response shapes to identify indicator taxa (decreasers and increasers) whereas the approach adopted for the MCI is simply based on association of a taxon with an *a priori*-selected pollution gradient consisting of three pollution classes (Stark 1985). In contrast to the MCI approach, taxa that did not show a clear positive or negative response shape were not assigned indicator values with our described approach, hence the list of indicator taxa is necessarily smaller. As one could argue that taxa with no clear response to the stressor gradients are somewhat tolerant, we subsequently assigned these taxa an indicator value of 5 and calculated metrics with and without their inclusion.

2.1.6. Weighting factor

Two taxa may be assigned the same indicator value based on similarity in the stressor value where 25% turnover occurs, but the strength of the response to the stressor gradient greatly differs. This difference can be captured in metric calculation by adding a weighting factor to each indicator taxon. For example, the presence of a taxon with an indicator value of 10 and a strong response to the stressor will lead to a relatively higher metric score than presence of a taxon with an indicator value of 10 but a weaker response to the stressor gradient. We calculated weighting factors for the taxa based on their total turnover over the range of the predictor (maximum cumulative importance as seen in the species turnover functions), which is equivalent to R^2_{fp} (see Section 2.1.3 and example shown in the Results). We assigned a weighting factor of either 1, 2 or 3 which was done by splitting the range from the lowest to the highest R^2_{fp} within each the decreaser and increaser groups into 3 equal bins. A weighting factor of 3 was assigned to those taxa with the highest R^2_{fp}, in other words taxa that had shown the greatest turnover across the predictor gradient.

2.1.7. Metric calculation

A community-level metric can be constructed in different ways from information on indicator taxa. The versions we constructed distinguish: (1) if only decreasers or increasers contribute to metric calculation or both, (2) whether decreasers (or increasers) contribute equally or whether we distinguish according to their degree of sensitivity (or tolerance) by incorporating indicator values, (3) whether their strength of response is captured by incorporating weighting factors, (4) whether information on presence-absence or on relative abundance is used, (5) whether metrics include the presence or relative abundance of non-scoring taxa, and (6) whether taxa that showed a non-response in our analyses were included. Overall, this resulted in 21 metrics for each stressor (Table 2). Fourteen of the metrics are calculated based on indicator values in a similar way to the MCI and the SQMCI which are two existing macroinvertebrate metrics developed by Stark (1985, 1998). The MCI is calculated based on presence-absence data as follows

$$\mathrm{MCI} = \frac{\sum_{i=1}^{i=S} a_i}{S} \text{ x } 20$$

where S = the total number of scoring taxa in the sample, and a_i = the tolerance value for the i^{th} taxon. The scaling factor of 20 has been added to distinguish between MCI site scores and site scores of the MCI variants, taking into account taxon counts (e.g. SQMCI).

The SQMCI is calculated typically based on coded abundance data (assigned to the R = Rare, C = Common, A = Abundant, VA =Very Abundant and VVA = Very Very Abundant classes) from semi-quantitative samples (e.g. kicknet) as follows

$$\mathrm{SQMCI} = \sum_{i=1}^{i=S} \frac{(n_i \text{ x } a_i)}{N}$$

where S = the total number of scoring taxa in the sample, n_i = the coded abundance for the i^{th} scoring taxon, a_i = the tolerance value for the i^{th} taxon and N = the total of the coded abundances of the scoring taxa for the entire sample. The coded abundance is the lower boundary of the respective abundance classes assigned to each taxon based on the counts: R (1-4), C (5-19), A (20-99), VA (100-499) and VVA (500+). If macroinvertebrates are counted rather than assigned to abundance classes, then abundances (counts) are used instead of coded abundances in the above formula.

Table 2. List of the 21 different metrics for each of the two stressors. Note that the code of the metrics is preceded by either ‘sed’ or ‘chl’ to specify which stressor it is developed for. The abbreviations ‘MCI’ and ‘SQMCI’ refer to formulas used to calculate existing metrics developed by Stark (1985) and Stark & Maxted (2007b).

Code	Name	Description
_MCI	Sediment or Periphyton MCI	MCI formula (raw-scale indicator value assignment)
_MCI_log	Sediment or Periphyton MCI (log)	MCI formula (log-scale indicator value assignment)
_MCI_weighted	Sediment or Periphyton MCI weighted	MCI formula plus weighting factor (raw-scale indicator value assignment)
_MCI_log_weighted	Sediment or Periphyton MCI weighted (log)	MCI formula plus weighting factor (log-scale indicator value assignment)
_MCI_plusnonind	Sediment or Periphyton MCI plus non-responsive taxa	MCI formula including non-responsive taxa (raw-scale indicator value assignment)
_MCI_log_plusnonind	Sediment or Periphyton MCI plus non-responsive taxa (log)	MCI formula including non-responsive taxa (log-scale indicator value assignment)
_SQMCI	Sediment or Periphyton SQMCI	SQMCI formula (raw-scale indicator value assignment)
_SQMCI_log	Sediment or Periphyton SQMCI (log)	SQMCI formula (log-scale indicator value assignment)
_SQMCI_weighted	Sediment or Periphyton SQMCI weighted	SQMCI formula plus weighting factor (raw-scale indicator value assignment)
_SQMCI_log_weighted	Sediment or Periphyton SQMCI weighted (log)	SQMCI formula plus weighting factor (log-scale indicator value assignment)
_MCI_decreasers	Sediment or Periphyton MCI decreasers only	MCI formula but only including decreasers (raw-scale indicator value assignment)
_MCI_log_decreasers	Sediment or Periphyton MCI decreasers only (log)	MCI formula but only including decreasers (log-scale indicator value assignment)
_MCI_increasers	Sediment or Periphyton MCI increasers only	MCI formula but only including increasers (raw-scale indicator value assignment)
_MCI_log_increasers	Sediment or Periphyton MCI increasers only (log)	MCI formula but only including increasers (log-scale indicator value assignment)
_richness_decreaser	Number of sensitive taxa	Number of decreaser taxa
_pct_richness_decreaser	% Sensitive taxa	Number of decreaser taxa divided by total number of taxa
_pct_abund_decreaser	% Sensitive individuals	Number of decreaser individuals divided by total number of individuals
_richness_increaser	Number of tolerant taxa	Number of increaser taxa
_pct_richness_increaser	% Tolerant taxa	Number of increaser taxa divided by total number of taxa
_pct_abund_increaser	% Tolerant individuals	Number of increaser individuals divided by total number of individuals
_richness_ratio_Decr_Incr	Ratio of sensitive to tolerant taxa	Number of decreaser taxa divided by number of increaser taxa

2.2. Validation of stressor-specific metrics

2.2.1. Data analysis

Overview

Validation was performed on three different datasets using different statistical methods. Datasets were composed either of field observational data or from a controlled mesocosm experiment. Statistical approaches encompassed single-stressor and multiple-stressor analyses.

First, we used a field observational dataset composed of samples from the large research and SoE dataset that had not been used in metric development (therefore independent data) and applied a single stressor analysis. This analysis was used to reduce the candidate set of the 21 stressor-specific metrics shown in Table 2 (for each sediment and periphyton) based on their relationship with the respective stressor. Secondly, the metrics in the reduced candidate set were tested in multiple-stressor analyses aimed at testing their diagnostic power. We ran multiple-stressor analyses on the dataset used to develop the metrics as well as on an independent experimental dataset to provide multiple lines of evidence about the usefulness of these metrics. Existing metrics and macroinvertebrate traits were included in analyses (see Sections 2.2.2 and 2.2.3 for descriptions of how metrics were calculated). The combined results informed our recommendation on the best metrics for use in resource management.

Single-stressor analysis on independent field observational data

Simple linear regression analyses were performed on field observational data not used for metric development to reduce the candidate set metrics based on their relationship with the respective stressor. For sediment, there were 634 samples which were collated from a total of 15 individual research studies including ten published studies (Quinn & Hickey 1990; Collier & Smith 2005; Matthaei et al*.* 2006; Townsend et al*.* 2008; Magbanua et al*.* 2010; Reid et al*.* 2010; Wagenhoff et al. 2011; Burdon et al*.* 2013; Harding & Jellyman 2015; Wagenhoff et al. 2017b) and five unpublished studies (a MSc thesis (Eivers 2006), Whatawhata research station data provided by Elizabeth Graham and John Quinn, a technical report (Storey et al. 2009) as well as two targeted field studies conducted in 2017 for a previous MfE sediment project). For chlorophyll-*a* there were 322 samples available from three research studies (Reid et al. 2010; Eivers 2006; and the Whatawhata data). However, because most studies sampled same sites repeatedly, the spatial spread in the data was relatively scarce. Hence, additional data were obtained from the large SoE dataset to increase the validation dataset to 1986 sample points. The spatial spread of these two datasets can be seen in Figure 3, and summary statistics of % sediment cover, chlorophyll-*a* and other characteristics of the sampling sites in Table 3.

To improve linear fit, chlorophyll-*a* was log-transformed, using the natural logarithm and adding a small constant (0.5 * min(x > 0)). The adjusted coefficient of determination values of the regression models (R^2, the proportion of variation explained by the model) were used to compare the 21 candidate sediment and periphyton-specific metrics. We compared metrics calculated based on (1) raw-scale vs. log-scale indicator value assignment, (2) indicator taxa only vs. inclusion of non-indicator taxa which were assigned an indicator value of 5, (3) inclusion of indicator values for decreasers and increasers vs. for decreasers or increasers only, (4) inclusion of a weighting factor vs. omitting use of a weighting factor, and (5) presence-absence (MCI-style calculation) vs. relative abundances (SQMCI-style calculation).

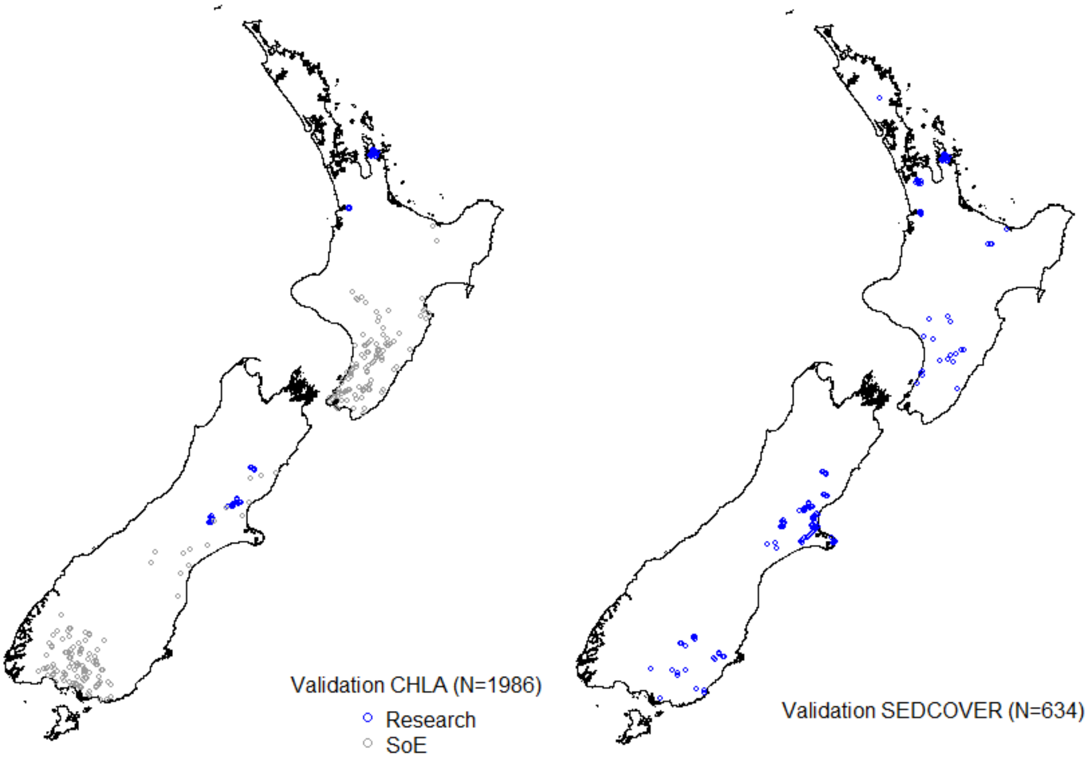


Figure 3. Map of New Zealand showing the spatial spread of the 1986 and 634 sample points used for single-stressor validation analysis of periphyton-specific metrics (left) and sediment-specific metrics (right), respectively. The samples were selected from a larger research (blue dots) or State of the Environment (SoE) dataset (grey dots) representing field observations that had not been used in metric development (independent dataset).

Table 3. Summary statistics of the stressors, catchment land use, stream order and elevation of the macroinvertebrate sampling sites within the datasets used for single-stressor validation analysis of sediment-specific metrics (validation SEDCOVER) and periphyton-specific metrics (validation CHLA). See Table 1 for description of the variables.

	Validation SEDCOVER					**Validation CHLA**				
Variable	**Mean**	**Median**	**Min**	**Max**	**Count**	**Mean**	**Median**	**Min**	**Max**	**Count**
Order	2.2	2.0	1	7	592	4.2	4.0	1	7	1977
Elevation	158	81	3	1094	592	141	100	3	832	1977
Land use										
T1NativeVeg	29	10	0	100	592	36	26	0	100	1977
T2PastoralHeavy	23	2	0	100	592	32	22	0	100	1977
Stressors										
SEDCOVER	27	19	0	100	634	15	13	0	70	399
CHLA	86	50	0	947	321	45	20	0	947	1986
DIN	0.698	0.228	0.001	8.8	225	0.488	0.216	0.002	5.0	627
DRP	0.029	0.021	0.002	0.238	223	0.022	0.015	0.002	0.192	455

Multiple-stressor analysis on field observational data

Multiple-stressor analysis was performed to test the diagnostic power of the stressor-specific metrics. The analysis was performed on the dataset used for metric development (N = 501). While this dataset is not completely independent as it was used to inform identification of indicator taxa, here it is used to validate and compare with each other the different stressor-specific metrics that aggregate knowledge of indicator taxa in different ways. For existing macroinvertebrate metrics and traits, however, the dataset is completely independent.

Multiple linear regression models were used to describe the relationship between 80 macroinvertebrate metrics (response variables) and 18 predictors including the stressors and environmental variables (see Section 2.1.2). Tested metrics included stressor-specific metrics selected from the single-stressor analysis (there were eight per stressor) as well as five existing metrics (MCI, SQMCI, EPT richness, %EPT richness, %EPT abundance) and 59 'trait metrics' (trait modalities). Initial data exploration was conducted following the protocol suggested by Zuur et al. (2010) to check for outliers and the normality of variable distributions. All response variables and predictor variables were transformed to improve normality using the Yeo-Johnson and Box-Cox transformations, respectively. After transformation, all variables were centred and scaled to allow direct comparison of regression coefficients and inference about relative effects sizes among stressors and metrics. Collinearity among predictor variables was checked using a variance inflation factor (VIF, Zuur et al. 2010), with all 18 predictors having VIFs of < 5 and hence being included in the models. Linear models were fitted with the predictor variables as fixed effects and final predictor variables were selected, for each metric, using a backwards procedure based on the generalised Schwarz's Bayesian information criterion (BIC). Final model diagnostics were checked by plotting residuals versus fitted and normal Q-Q plots of the residuals (results not shown).

Hierarchical partitioning of R^2 values for each metric was used to determine the proportion of variance explained independently by each predictor variable (Chevan & Sutherland 1991; MacNally 2000). This method allows identification of variables whose independent correlation with the dependent variable is strong, in contrast to variables that have little independent effect but have a high correlation with the dependent variable resulting from joint correlation with other independent variables.

Multiple-stressor analysis on experimental data

A second multiple-stressor analysis was performed to test the diagnostic power of stressor-specific metrics, but this time using an experimental dataset derived from a stream-side mesocosm study. Analysis of field observational data and experimental data provides complementary information, as both approaches have strengths and weaknesses. Analysis of the field observational data, using a time-for-space approach, has a high degree of realism. But correlation among stressors and natural environmental variables (measured and unmeasured) can bias estimates of the stressor effects and hence impede inference of cause-effect relationships. Experimental data, on the other hand, provide a high level of control allowing to establish cause-effect relationships but with lesser degree of realism associated with spatial and temporal limitations.

The mesocosm experiment is described in detail in Wagenhoff et al. (2012), and briefly summarised here. The set-up was situated next to the 5th-order Kauru River in North Otago which supplied relatively low-nutrient water, pumped from the river, into 128 circular flow-through mesocosms (25 cm diameter). Organisms arrived with the water and a colonisation period of two weeks prior to experimental manipulations allowed establishment of periphyton. To boost the macroinvertebrate community, macroinvertebrates were collected from the river and seeded into the mesocosms via the header tanks a day prior to the experimental period.

The full-factorial experimental design consisted of eight levels of deposited sediment crossed with eight levels of nutrient concentrations and two replicates of each treatment combination. The achieved average sediment cover levels ranged from 0 to 100% (equivalent to 0 to 1020 g per mesocosm), while nutrient concentrations ranged from 0.036 mg DIN/L and 0.0014 mg DRP/L to 6.90 mg DIN/L and 0.446 mg DRP/L. Levels were chosen so that there was an exponential increase across the stressor gradients. Fine sediment was weighed out and placed directly into the mesocosms, nutrient concentrations were increased by dripping highly-concentrated nutrient solutions into header tanks (each header tank fed 16 mesocosms). The experiment ran for three weeks, in autumn, after which macroinvertebrates were processed using standard protocols (Winterbourn et al. 2006). At the end of the experiment periphyton was sampled by scraping a surface area of 11.5 cm^2 from each mesocosm to determine chlorophyll-*a* using spectrophotometry.

Multiple linear regressions were performed on each of the 80 macroinvertebrate metrics as described for the field observational dataset, however there were some differences. Predictor variables were in a nominal scale (the eight levels of the nutrient and sediment treatments) and these were not transformed. An interaction term between sediment and nutrient was included in the linear models in addition to the main effects as previous analyses showed interactive effects on some macroinvertebrate metrics (Wagenhoff et al. 2012). Additionally, we did not use model selection and instead presented regression coefficients for all three predictor terms regardless of statistical significance. As the experimental design was full-factorial and hence sediment and nutrients were not correlated, regression coefficients were used to directly compare the relative importance of both stressors without the need of hierarchical variance partitioning.

2.2.2. Calculation of existing macroinvertebrate metrics

The MCI and SQMCI (Stark & Maxted 2007a), were calculated using hard-bottom indicator values provided in Appendix 1. Note that in several cases samples were collected quantitatively (i.e. with a Surber sampler), hence technically, the resulting metric would be called QMCI (Quantitative MCI), but these samples were combined with those collected semi-quantitatively.

Three EPT metrics were calculated excluding the pollution-tolerant taxa within the trichopteran family, Hydroptilidae (*Oxyethira* and *Paroxythira*): (1) EPT richness, which is the number of EPT taxa in a sample, (2) %EPT richness, which is the number of EPT taxa divided by total taxon richness, and (3) %EPT abundance, which is the number of EPT individuals divided by total number of individuals. The EPT metrics were calculated based on the highest taxonomic resolution available for each taxon, which was often to species level.

2.2.3. Macroinvertebrate trait calculation

The trait composition of the macroinvertebrate community is derived by combining its taxonomic composition with the trait ‘profile’ of each taxon present. The trait profile of a taxon can be represented by its affinity to different modalities (or categories) within each trait (see examples for modalities and affinity scores in Table 4 and Table 5, respectively). Affinities in the New Zealand traits database are coded using integers from 0 to 3, representing the strength of affinity for each modality. For a variety of reasons taxa can have affinity scores > 0 for more than one modality. For example, in terms of feeding, this can reflect: (1) a taxon changes feeding preferences during different stages of maturity, (2) a taxon shows different feeding behaviour in different environments, and/or (3) a taxon encompasses species that differ in their feeding preferences.

Table 4. Modalities (categories) within three example traits.

Trait	**Modality**					
	1	**2**	**3**	**4**	**5**	**6**
Body size	≤ 5 mm	5–10 mm	10–20 mm	> 20–40 mm	> 40 mm	
Feeding	Shredder	Scraper	Deposit-feeder	Filter-feeder	Predator	Algal piercer
Attachment to substrate	Swimmer	Crawler	Burrower	Attached		

Table 5. Affinity scores of three example taxa for the feeding trait based on modalities defined in Table 4.

Taxon	**Modality**					
	Shredder	**Scraper**	**Deposit-feeder**	**Filter-feeder**	**Predator**	**Algal piercer**
Taxon 1	1	3	0	0	0	0
Taxon 2	0	0	0	3	0	0
Taxon 3	0	3	0	0	2	0

The full array of affinity scores for all taxa across all trait modalities is called a traits database. Trait composition of a community is typically described in terms of trait relative abundance. Trait relative abundance for a dataset was calculated as follows. First, the dataset was arranged as a site-by-taxa matrix containing the relative abundances expressed as percentages of all taxa that occurred in the dataset. The relative abundances were log(x+1)-transformed, using the natural logarithm, to increase the relative contribution of naturally less-abundant taxa to the calculation of trait relative abundances. To simplify calculations, we converted all data to the taxonomic level required for the MCI. Secondly, a taxa-by-trait (trait modalities) matrix was prepared containing trait information, if available, for all taxa that occurred in the dataset. Trait data are not always available at the taxonomic level provided in the dataset and hence respective trait modality affinity scores have to be assigned according to best information available. In several cases, trait information was given at the species level while the taxa in our dataset typically were at genus level (MCI-level taxonomy). Here the genus was assigned an average trait score across the existing scores for the species within that genus. The affinity scores in the taxa-by-trait matrix were standardised to sum to 1 for each trait. Thirdly, matrix multiplication of the site-by-taxa and taxa-by-trait matrix result in a site-by-trait matrix. A fourth and final step is to divide the abundance of each trait modality by the sum of all modalities for that trait so that the 'trait metric' is the relative abundance of the trait modality.

The New Zealand species traits database

A traits database for New Zealand taxa was first developed for stream biomonitoring in 2004 (Phillips 2004), with several updates until 2012 (Phillips & Reid 2012a, 2012b). However, some inaccuracies remained and a large number of knowledge gaps have since been filled by inference from international literature. As part of the previous project (MfE contract no. 21630, Clapcott et al. 2017b), the New Zealand traits database was reviewed and updated using recent information and specialist knowledge of the major aquatic insect orders. Affinity scores for 16 traits with between two and five modalities each were assigned to each taxon at the lowest level of identification available. The database now represents the best specialist knowledge available for New Zealand benthic macroinvertebrate fauna. The database is maintained by NIWA, freely available for download on the web (https://www.niwa.co.nz/freshwater-and-estuaries/management-tools/aquatic-invertebrate-traits-database) and accompanied by a technical report (Phillips & Smith 2018). Table 6 presents the 16 traits and 59 trait modalities total along with the metric code we assigned to shorten the metric names.

Table 6. List of species traits modalities of 16 traits assigned to New Zealand taxa.

Trait	Modality code	Modality description	Metric code
Maximum potential size	SIZE1	≤ 5 mm	1a
	SIZE2	> 5–10 mm	1b
	SIZE3	> 10–20 mm	1c
	SIZE4	> 20–40 mm	1d
	SIZE5	> 40 mm	1e
Maximum number of descendants per reproductive cycle	DESC1	≤ 100	2a
	DESC2	> 100–1000	2b
	DESC3	> 1000–3000	2c
	DESC4	> 3000	2d
Maximum number of reproductive cycles per year	SEMI	semivoltine	3a
	UNIV	univoltine	3b
	PLURIV	plurivoltine	3c
Number of reproductive cycles per individual	CPI1	1	4a
	CPI2	≥ 2	4c
Life duration of adults	LDA1	≤ 1 day	5a
	LDA2	> 1–10 days	5b
	LDA3	> 10–30 days	5c
	LDA4	> 30–365 days	5d
	LDA5	> 365 days	5e
Reproductive technique	SINGLE	single individual	6a
	HERMA	hermaphrodism	6b
	TWO	male and female	6c
Oviposition site	SURFACE	water surface	7a
	SUBMERGED	submerged	7b
	TERRESTRIAL	terrestrial	7c
	EGGENDO	eggs endophytic	7d
Egg/egg mass	EGGFREE	free	8a
	EGGCEMENT	cemented	8b
	EGGPROTECTED	female bears eggs in/on body	8c
Dissemination potential (all stages)	DISSLOW	low (10 m)	9a
	DISSMEDIUM	medium (1 km)	9b
	DISSHIGH	high (> 1 km)	9c
Attachment to substrate of aquatic stages (excluding eggs)	SWIMMER	swimmers (water column)	10a
	CRAWLER	crawlers (epibenthic)	10b
	BURROWER	burrowers (infauna)	10c
	ATTACHED	attached	10d

Trait	Modality code	Modality description	Metric code
Body flexibility	NOFLEX	none (< 10°)	11a
	LOWFLEX	low (> 10-45°)	11b
	HIGHFLEX	high (> 45°)	11c
Body form	STREAMLINED	streamlined	12a
	FLATTENED	flattened (dorso-ventral or lateral)	12b
	CYLINDRICAL	cylindrical	12c
	SPHERICAL	spherical	12d
Feeding method	SHREDDER	shredders	13a
	SCRAPER	scrapers	13b
	DEPOSIT	deposit-feeders	13c
	FILTERFEED	filter-feeders	13d
	PREDATOR	predator	13e
	ALGALP	algal piercer	13f
Dietary preferences	SPECIALIST	strong (specialist)	14a
	MODERATESPE	moderate	14b
	GENERALIST	weak (generalist)	14c
Respiration of aquatic stages (not including eggs)	TEGUMENT	tegument	15a
	GILL	gills	15b
	PLASTRON	plastron	15c
	AERIAL	aerial	15d
Aquatic stages	ADUANDLAR	adult, larva	16a
	ADUORLAR	adult or larva	16b
	LARANDPUP	larva, pupa	16c

3. RESULTS

3.1. Development of stressor-specific metrics

3.1.1. Gradient forest outputs

Out of the 87 taxa included in the GF analysis, 73 taxa had an $R^2 > 0$, ranging from 0.02 to 0.76 (mean = 0.30, Appendix 2). The importance of each predictor for overall compositional turnover, i.e. incorporating all 73 taxa, is presented in Figure 4. SEDCOVER and CHLA ranked 10th and 14th, respectively, in terms of relative importance for the distribution of these taxa (Figure 4) suggesting that environmental drivers other than the stressors heavily influence macroinvertebrate communities (Figure 4). This is not surprising and natural environmental variables will therefore also have an influence on metrics which will have to be taken into account when comparing scores among sites (this will be discussed in Section 5). Incorporating these environmental variables as predictors in our analysis likely improved prediction of the effects that are attributed to % sediment cover and chlorophyll-*a*. The species turnover functions for SEDCOVER and CHLA are presented in Figure 5, and also individually and alongside the partial dependence plots in Appendix 3.

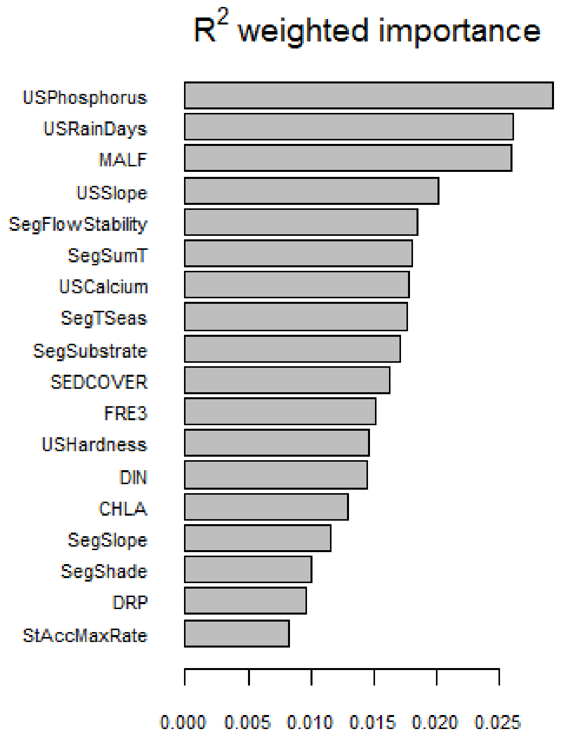


Figure 4. Overall predictor importance (in R^2 units) for species distribution (across all 73 taxa) allowing assessment of the relative importance of the environmental gradients for compositional turnover of the macroinvertebrate community.

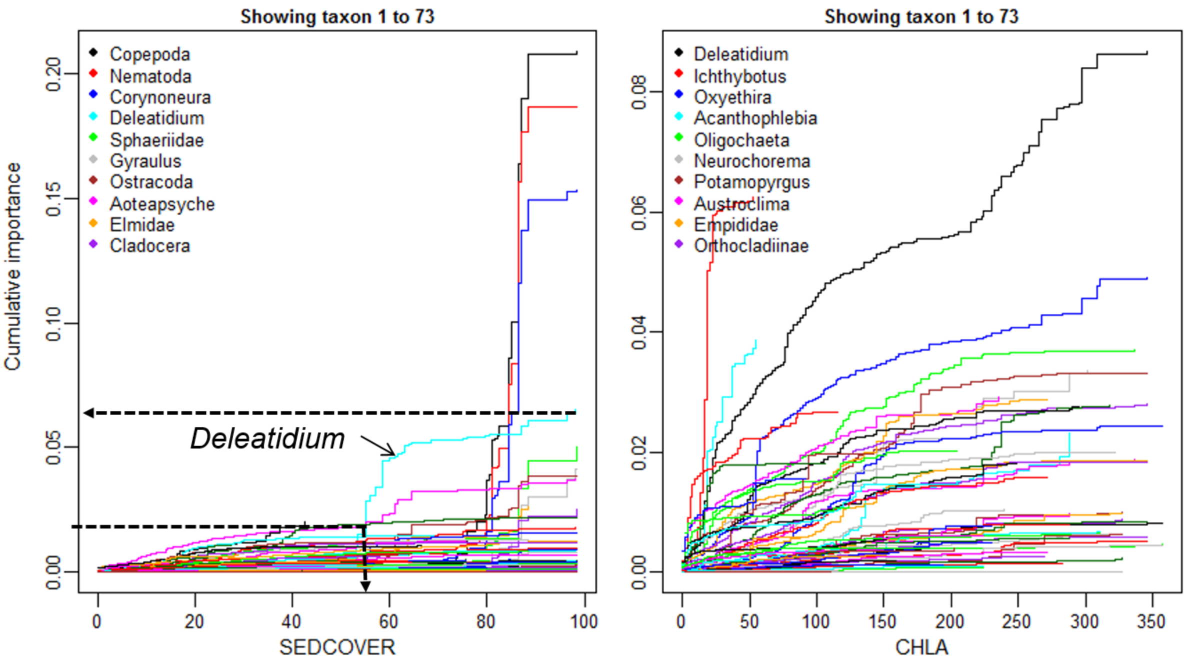


Figure 5. Species turnover functions for all 73 taxa across the % sediment cover (SEDCOVER) and chlorophyll-*a* (CHLA) gradients calculated by accumulating the split importance values that are normalised to sum to R^2_{fp} (gradient forest output), which is the total turnover. The legend lists the ten taxa with the largest total turnover. For example, total turnover of *Deleatidium* across SEDCOVER is 0.065 depicted on the y-axis (upper black arrow).

3.1.2. Decreasers and increasers

Overall, while there was some good agreement between response shapes in the partial dependence plots and expert opinion, there also were several cases where experts disagreed in that they had expected no response or even the opposite response shape. There also were some disagreements between the two experts in what response shape should be expected. We stayed on the side of caution and in many cases rejected indicator taxa as there was mixed agreement among experts concerning the response shapes.

Following consideration of both analytical output and expert opinion a total of 26 decreasers and 6 increasers were assigned in response to chlorophyll-*a*. Firstly, based on partial dependence plots for all 73 taxa (Appendix 3), 23 taxa were classified as decreasers, 20 as increasers and 3 as subsidy-stress responses[3] across the chlorophyll-*a* gradient; the remaining responses were unclear. Examples were *Deleatidium* (decreaser), Oligochaeta (increaser), *Potamopyrgus* (subsidy-stress response) and *Psilochorema* (unclear response). Sixteen out of the 23 ‘decreaser’

---

[3] A subsidy-stress response is characterised by an initial positive response direction across the stress gradient before turning into a negative response direction. Such response shapes have been found across a nutrient enrichment gradient where initial increase in nutrient concentrations can provide an initial subsidy to macroinvertebrate abundance due to increased periphyton productivity (Wagenhoff et al. 2011, 2012).

taxa were accepted as decreaser taxa and 2 were rejected after seeking expert opinion, the remaining 5 taxa were rejected because the total turnover was zero and therefore not informative. Only 6 of the 20 'increaser' taxa were accepted by the experts as increasers while the remaining were rejected, in several cases because one expert had expected negative response shapes, and for 5 taxa also because the total turnover was zero. In a single case, both experts agreed that the response shape should be negative, and the taxon was classified as a decreaser despite the positive response shape in the partial dependence plot. Two of the subsidy-stress responses were classified as decreasers after expert opinion. Finally, in 7 cases, the response shapes were not clear in the partial dependence plots, but experts agreed that the taxa should be classified as decreasers.

For % sediment cover, a total of 31 decreasers and 6 increasers were assigned. Firstly, 9 taxa were classified as decreasers, 23 as increasers and 4 as subsidy-stress responses based on partial dependence plots. After seeking expert opinion, 8 of the 9 'decreaser' taxa were accepted as decreasers and one was rejected. Only 5 of the 23 'increaser' taxa were accepted while several of the remaining were rejected partly due to inconsistencies between the experts. In four cases experts expected negative response shapes, hence these taxa were classified as decreasers despite the positive response shape in the partial dependence plot. The four taxa with subsidy-stress responses were classified as decreasers due to expert opinion. Finally, in 15 cases, the response shapes were not clear in the partial dependence plots, but experts agreed that the taxa should be classified as decreasers and in one case experts agreed it should be an increaser taxon.

The indicator taxa and model results are presented in the following section. The model results and names of the taxa that did not show a clear response shape to chlorophyll-*a* (41 taxa) or sediment cover (36 taxa) and/or were not expected to be responsive to the stressor gradient by the experts are presented in Appendix 4.

3.1.3. Indicator values and weighting factors

Indicator value assignment was done based on thresholds (presented in Appendix 2), which were defined using the species turnover functions (see Figure 5 for illustration of threshold definition for *Deleatidium* at 55% sediment cover). For chlorophyll-*a*, most of the indicator taxa were assigned an indicator value of 10 (13 taxa out of 32) with fewer taxa assigned to the other indicator values on the raw scale, whereas on the log scale the taxa were more evenly distributed (Table 7). Similarly for sediment cover, the majority of indicator taxa were assigned an indicator value of 10 (20 taxa out of 37) on the raw scale but on the log-scale these got mainly assigned an indicator value of 9 or 10. For both stressors, due to only a few increaser taxa, some of the indicator values were not assigned to any taxa, for example for sediment cover none of the taxa were assigned to indicator values of 2, 3 or 4 (Table 7).

Table 7. Indicator value assignment based on bins of chlorophyll-*a* and % sediment cover threshold values along with number of taxa per bin . The thresholds determined for each of the taxa are presented in Appendix 2.

	Chlorophyll-*a* (mg/m^2)				% Sediment cover			
Tol. value	Threshold range (raw scale)	No. of taxa	Threshold range (log scale)	No. of taxa	Threshold range (raw scale)	No. of taxa	Threshold range (log scale)	No. of taxa
Decreasers								
10	1.9–16.5	13	1.9–3.9	5	4–14.1	20	4–6.7	1
9	16.5–31.0	4	3.9–8.2	2	14.1–24.2	8	6.7–11.4	10
8	31.0–45.6	2	8.2–17.1	6	24.2–34.3	0	11.4–19.2	17
7	45.6–60.2	4	17.1–35.8	4	34.3–44.4	0	19.2–32.3	0
6	60.2–74.8	3	35.8–74.8	9	44.4–54.5	3	32.3–54.5	3
Increasers								
5	20.6–44.5	4	20.6–30.2	2	14.5–28.1	2	14.5–20.5	2
4	44.5–68.4	1	30.2–44.3	2	28.1–41.7	0	20.5–29.1	0
3	68.4–92.4	0	44.3–65.1	1	41.7–55.3	0	29.1–41.2	0
2	92.4–116.3	0	65.1–95.5	0	55.3–68.9	0	41.2–58.3	0
1	116.3–140.3	1	95.5–140.3	1	68.9–82.5	4	58.3–82.5	4

Out of the 49 indicator taxa total, 32 indicator taxa were identified for calculation of periphyton-specific metrics and 37 indicator taxa for calculation of sediment-specific metrics (Table 8). Out of these taxa, 29 responded to only a single stressor, either chlorophyll-*a* or sediment, and two taxa responded to both stressors but were decreasers for one (*Austrosimulium* for chlorophyll-*a*, *Aphrophila* for sediment) and increasers across the other stressor gradient (Table 8). These 31 'discriminative' taxa (denoted with * in Table 8) have the potential to provide diagnostic power of the resulting stressor-specific metrics. In addition to these 31 discriminative taxa, there were further taxa that were either decreasers or increasers with respect to both stressors but differed in their degree of sensitivity or tolerance and hence also provide a level of discrimination between the effects of the two stressors.

The indicator taxa for the stressor-specific metrics covered a broad range of taxonomic groups. Genera within the EPT insect orders were well represented and all taxa, except for trichopteran *Oxyethira*, were decreasers (Table 8). Further indicator taxa belong to other insect orders, mainly Coleoptera and Diptera, and one megalopteran. Finally, there were a few molluscs, a crustacean and the taxa Acarina, Nematoda and Oligochaeta (Table 8). Compared to the MCI taxon list that contains 191 taxa for both hard-bottom and soft-bottom versions combined and 182 for the hard-bottom version (MCI$_{hb}$) only, the number of indicator taxa for the stressor-specific metrics is relatively small.

Weighting factors represent the relative degree of change of a taxon across the stressor gradient within each the increaser and decreaser groups and were calculated from R^2_{fp} (presented in Appendix 2). For both stressor gradients, the majority of the taxa were assigned a weighting factor of 1 (relatively small change) and only few were assigned a weighting factor of either 2 or 3 (intermediate to relatively large change) (Table 8). Figure 5 illustrates distribution of R^2_{fp} values for SEDCOVER and CHLA although it includes all taxa, including decreasers and increasers, as well as taxa that were not identified as indicator taxa. Copepoda was rejected as an indicator taxon for SEDCOVER, hence increasers Nematoda and *Corynoneura,* showing a large turnover relative to other increasers, were assigned a weighting factor of 3, and *Deleatidium*, which was the decreaser taxon with the highest turnover was also assigned a weighting factor of 3. A weighting factor of 2 was assigned to *Aoteapsyche* and Elmidae (both decreasers across SEDCOVER). Among the indicator taxa for periphyton-specific metrics, a weighting factor of 3 was assigned to *Deleatidium* and *Ichthybotus* (both decreasers), *Oxyethira* and Oligochaeta (both increasers), and a weighting factor of 2 was assigned to *Acanthophlebia* and *Potamopyrgus* (both decreasers), *Aphrophila*, Empididae and Orthocladiinae (all increasers) (Table 8).

Table 8. Indicator value table for calculation of periphyton-specific and sediment-specific metrics for a total of 49 indicator taxa. There are two sets of indicator values for each stressor due to assignment on raw-scale or log-scale spectrums and also a weighting factor. Taxa with indicator values of 10 to 6 are decreasers (sensitive taxa) while taxa with indicator values of 5 to 1 are increasers (tolerant taxa). Discriminative taxa are marked with an asterisk (*). MCI tolerance values (hard-bottom version, Stark & Maxted 2007a) for the same 49 taxa are also given for comparison.

	Periphyton-specific metrics			Sediment-specific metrics			MCI
Taxon	Tol. value (raw)	Tol. value (log)	Weighting	Tol. value (raw)	Tol. value (log)	Weighting	Tol. value
Ephemeroptera							
Acanthophlebia	10	8	2	10	8	1	7
Ameletopsis *				10	9	1	10
Austroclima	10	9	1	10	8	1	9
Coloburiscus *				10	8	1	9
Deleatidium	9	7	3	6	6	3	8
Ichthybotus *	10	8	3				8
Nesameletus	7	6	1	10	9	1	9
Zephlebia	7	6	1	10	9	1	7
Plecoptera							
Austroperla	10	10	1	10	8	1	9
Megaleptoperla *				9	8	1	9
Stenoperla *				10	9	1	10
Zelandobius *	9	7	1				5
Zelandoperla	10	8	1	10	9	1	10
Trichoptera							
Beraeoptera	10	10	1	10	10	1	8
Confluens *				10	9	1	5
Costachorema *				10	9	1	7
Helicopsyche	7	6	1	10	8	1	10
Hydrobiosis *				6	6	1	5
Aoteapsyche	8	6	1	9	8	2	4
Orthopsyche *	10	8	1				9
Neurochorema *				10	9	1	6
Olinga *				9	8	1	9
Oxyethira *	4	3	3				2
Psilochorema *				9	8	1	8
Pycnocentria	6	6	1	9	8	1	7
Pycnocentrodes	6	6	1	9	8	1	5
Triplectides *	10	8	1				5
Coleoptera							
Elmidae *				6	6	2	6
Hydraenidae	10	9	1	10	9	1	8
Ptilodactylidae *				9	8	1	8

	Periphyton-specific metrics			Sediment-specific metrics			MCI
Taxon	Tol. value (raw)	Tol. value (log)	Weighting	Tol. value (raw)	Tol. value (log)	Weighting	Tol. value
Scirtidae	10	10	1	10	8	1	8
Diptera							
Aphrophila *	5	4	2	10	8	1	5
Austrosimulium *	8	6	1	5	5	1	3
Corynoneura *				1	1	3	2
Empididae *	5	5	2				3
Lobodiamesa *				10	9	1	5
Molophilus *	10	10	1				5
Orthocladiinae *	5	4	2				2
Paralimnophila *	10	8	1				6
Stictocladius	10	10	1	9	8	1	8
Megaloptera							
Archichauliodes *				10	8	1	7
Mollusca							
Gyraulus	1	1	1	1	1	1	3
Latia	9	7	1	10	8	1	3
Potamopyrgus *	6	6	2				4
Sphaeriidae *				1	1	1	3
Crustacea							
Paraleptamphopus *	7	6	1				5
Other							
Acarina *	9	7	1				5
Nematoda *				1	1	3	3
Oligochaeta	5	5	3	5	5	1	1

3.2. Validation of stressor-specific metrics

3.2.1. Single-stressor analysis on independent field observational data

Periphyton-specific metrics

Scatterplots, along with the simple linear regression models for each of the 21 periphyton-specific candidate metrics (Table 2) across the chlorophyll-*a* gradient ordered according to their adjusted R^2 values, are presented in Figure 6. The overall best-performing metric was chl_SQMCI with an R^2 of 0.24, however R^2 values varied widely across the metrics. Overall, metrics calculated based on indicator values reflecting the different degrees of sensitivity or tolerance of both decreasers and increasers (MCI/SQMCI-style metrics) performed better than metrics based on decreasers or increasers only that did not differentiate between the degrees of sensitivity or tolerance. The best-performing metric among the latter type was chl_pct_abund_increaser with an R^2 of 0.1 (Figure 6).

Metrics calculated with indicator values based on raw-scale assignment performed slightly better compared to the ones calculated based on log-scale assignment between all comparable versions. Equally, inclusion of the non-indicator taxa, which were assigned an indicator value of 5, did not improve model fit. MCI-style or SQMCI-style periphyton-specific metrics that were based on decreasers **and** increasers always performed much better than those MCI/SQMCI-style metrics based on either decreasers or increasers only. Inclusion in metric calculation of a weighting factor representing the strength of response of the indicator taxon to chlorophyll-*a* did not improve model performance or only slightly improved model performance between all comparable versions. Finally, SQMCI-style periphyton-specific metrics performed better than their MCI-style equivalent versions, for example the overall best-performing metric chl_SQMCI had an R^2 of 0.24 while chl_MCI had an R^2 of 0.12.

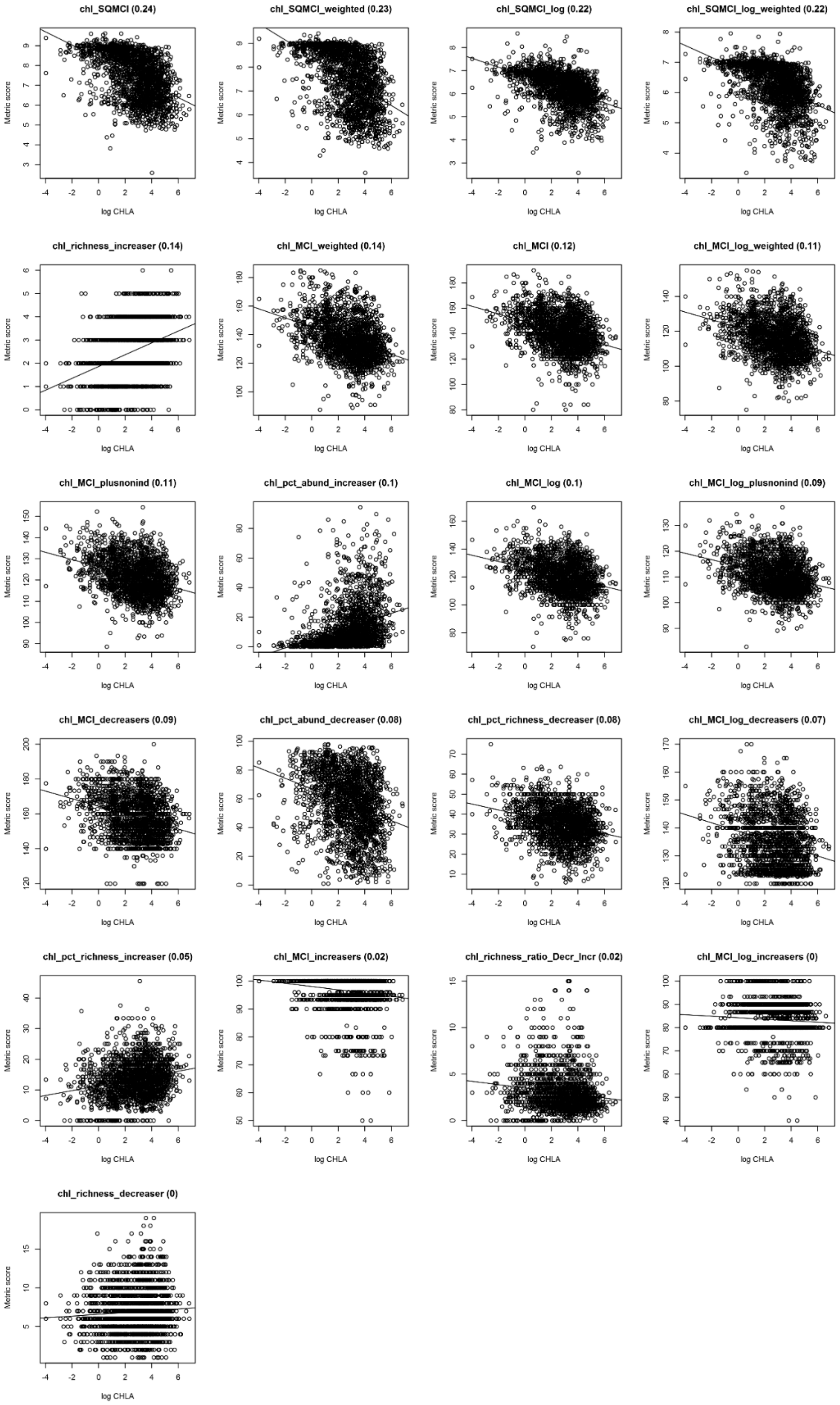


Figure 6. Scatterplots along with the simple linear regression models for each of the 21 periphyton-specific candidate metrics across the chlorophyll-*a* (CHLA in mg/m^2 and log-transformed) gradient ordered according to their adjusted R^2 values (R^2 given in parentheses).

Sediment-specific metrics

Scatterplots along with the simple linear regression models for the 21 sediment-specific candidate metrics across the sediment cover gradient ordered according to their adjusted R^2 values are presented in Figure 7. The overall best-performing metric was sed_MCI_log with an R^2 of 0.44, however R^2 values varied widely across the metrics. Overall, like periphyton-specific metrics, sediment-specific metrics calculated based on indicator values reflecting the different degrees of sensitivity or tolerance performed better than metrics based on decreasers or increasers only. The best-performing metric among the latter type was sed_pct_richness_decreaser with an R^2 of 0.32 (Figure 7).

In contrast to the periphyton-specific metrics, sediment-specific metrics calculated with indicator values based on log-scale assignment performed slightly better compared to the ones calculated based on raw-scale assignment between all comparable versions. Similar to the periphyton-specific metrics, inclusion of the non-indicator taxa did not improve model fit, and MCI-style or SQMCI-style sediment-specific metrics based on both decreasers and increasers always performed much better than those MCI/SQMCI-style metrics based on either decreasers or increasers only. Also, inclusion of a weighting factor in metric calculation representing the strength of response of the indicator taxa to sediment cover did not improve model performance for all comparable versions. Finally, in contrast to the periphyton-specific metrics, MCI-style sediment-specific metrics performed slightly better than their SQMCI-style equivalent versions, for example the overall best-performing metric sed_MCI_log had an R^2 of 0.44 while sed_SQMCI_log had an R^2 of 0.37.

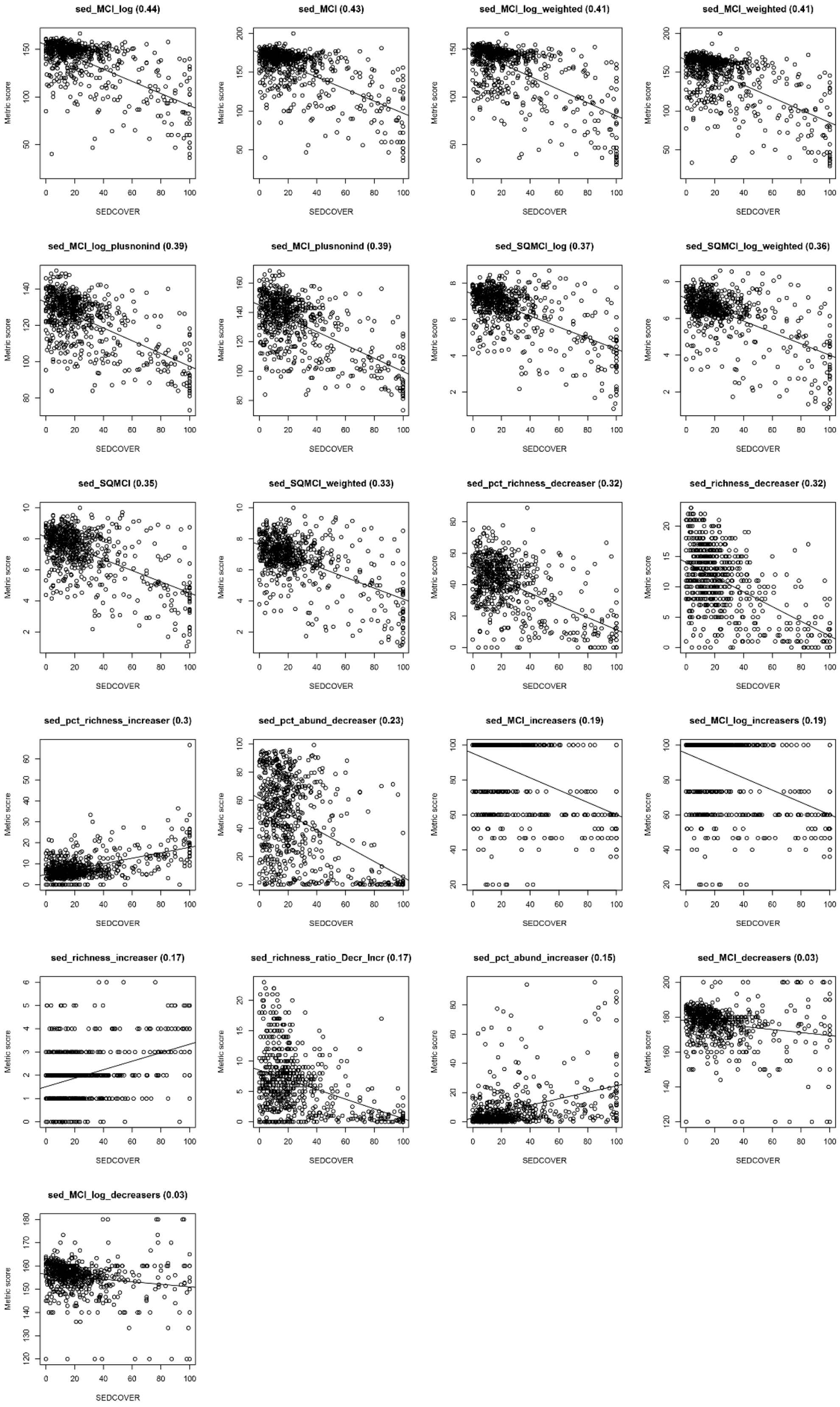


Figure 7. Scatterplots along with the simple linear regression models for each of the 21 sediment-specific candidate metrics across the % sediment cover (SEDCOVER) gradient ordered according to their adjusted R^2 values (R^2 given in parentheses).

Selection of stressor-specific metrics

We reduced the set of candidate metrics based on single-stressor model results. First, we excluded metrics calculated with indicator values that were assigned based on a log-scale range of threshold values. These did not perform as well as the ones calculated with indicator values assigned based on a raw-scale range of threshold values for periphyton-specific metrics and only slightly better for sediment-specific metrics. Assignment across the raw stressor gradient overall is simpler to explain and hence preferable.

Secondly, we excluded metrics where non-indicator taxa had received an indicator value of 5 as consistently there was no improvement in model fit. Thirdly, MCI-style or SQMCI-style sediment-specific metrics based on either decreasers or increasers only were dropped as these consistently performed worse than the MCI/SQMCI-style metrics that incorporated indicator values of both decreasers and increasers. Fourthly, we dropped metrics that included weighting factors in metric calculation, representing the strength of response of the indicator taxa to the stressor gradient, as this rarely improved model performance and a simpler metric formula is preferable.

Whether MCI-style or SQMCI-style metrics performed better differed between the two sets of stressor-specific metrics. In other words, while metrics calculated by averaging the indicator values of the taxa present (MCI-style) performed better among the sediment-specific metrics, weighting of the indicator values according to the relative abundances of the taxa (SQMCI-style) improved model performance among the periphyton-specific metrics. Hence, we decided to keep both versions to retain the flexibility of calculating at least the MCI-style stressor-specific metrics if only presence-absence data are available.

Apart from the MCI/SQMCI-style metrics, we had multiple versions of metrics combining information on either decreaser or increaser taxa without differentiating between their degree of sensitivity or tolerance (i.e. no indicator values). Despite the often low R^2 of the metrics of this type we retained them all as there was no consistent ranking in their performance across the periphyton-specific and sediment-specific metrics as well as across different datasets in exploratory analyses (data not shown). The exception was the decision to drop the ratio of decreasers vs. increasers (chl_richness_ratio_Decr_Incr, also sed_richness_ratio_Decr_Incr) as it appeared to be generally a poorly-performing metric.

Overall, eight stressor-specific metrics per stressor were further validated using multiple-stressor analyses (Table 9).

Table 9. Stressor-specific metrics (eight for each stressor) selected during single-stressor analysis for further validation using multiple-stressor analyses.

Stressor specific metric
_MCI
_SQMCI
_richness_decreaser
_pct_richness_decreaser
_pct_abund_decreaser
_richness_increaser
_pct_richness_increaser
_pct_abund_increaser

Comparison with existing metrics

Among the existing metrics, the SQMCI was the best-performing metric with respect to the chlorophyll-*a* gradient ($R^2 = 0.26$), much better than all EPT metrics (Figure 8). The SQMCI performed similarly to chl_SQMCI, which was the best periphyton-specific metric ($R^2 = 0.24$, Figure 6). With respect to the sediment cover gradient, % EPT richness was the best-performing among the existing metrics ($R^2 = 0.33$, Figure 8), however, it did not perform as well as the sediment-specific metric sed_MCI ($R^2 = 0.43$, Figure 7).

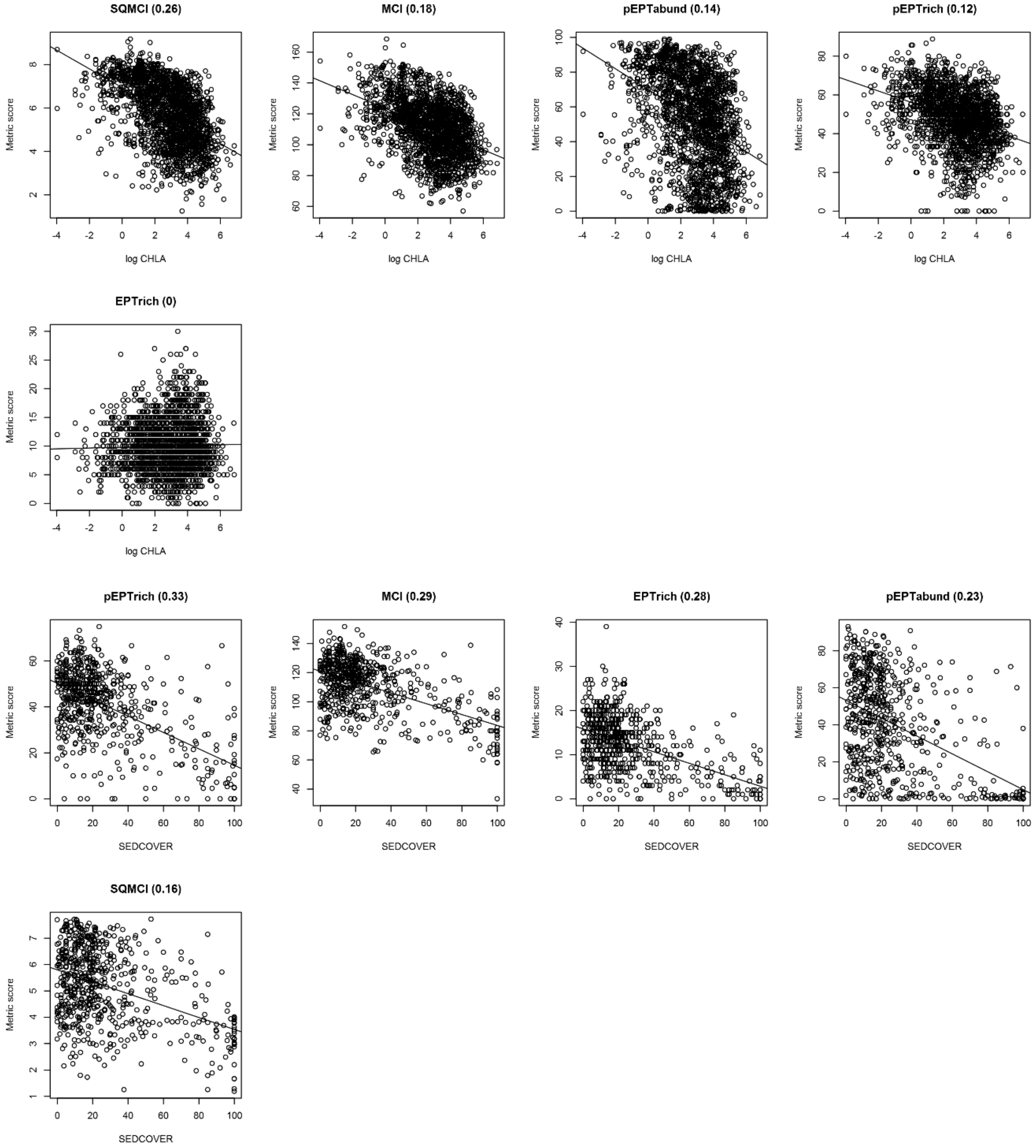


Figure 8. Scatterplots along with the simple linear regression models for existing metrics across the chlorophyll-*a* (CHLA in mg/m^2, log-transformed) and % sediment cover (SEDCOVER) gradients ordered according to their adjusted R^2 values , respectively (R^2 given in parentheses). See Section 2.2.2 for description of the metrics.

After comparing the performance of periphyton-specific metrics and existing metrics, simple regression models were built to compare scores of existing metrics and periphyton-specific metrics directly. This showed that SQMCI vs. chl_SQMCI, and MCI vs. chl_MCI were well correlated (Figure 9). Similarly, EPT richness and % EPT richness were correlated to chl_richness_decreaser and chl_pct_richness_decreaser, respectively (Figure 9). The same analyses were run for sediment-specific metrics. Correlations between EPT metric scores and the number or percentage of sediment-sensitive taxa were also strong while the relationships between MCI and sed_MCI and in particular between SQMCI and sed_SQMCI were relatively less strong (Figure 10).

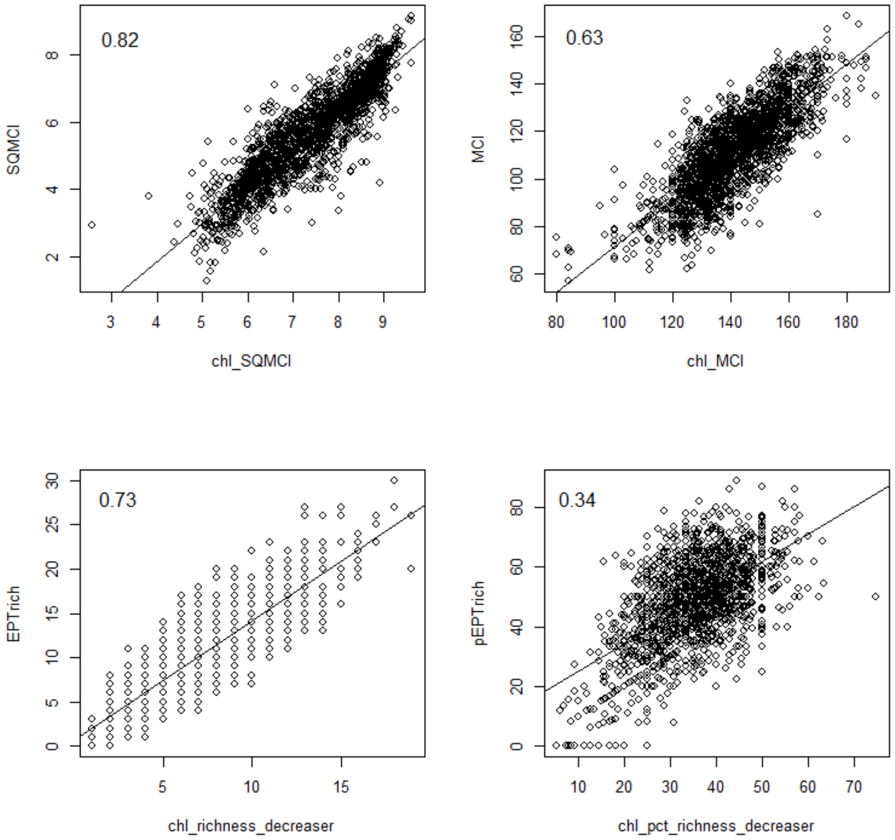


Figure 9. Scatterplots along with the simple linear regression models showing the relationship between existing metrics and comparable periphyton-specific metrics based on the field observational dataset used for single-stressor analysis. The adjusted R^2 values are given at the top left of the panels.

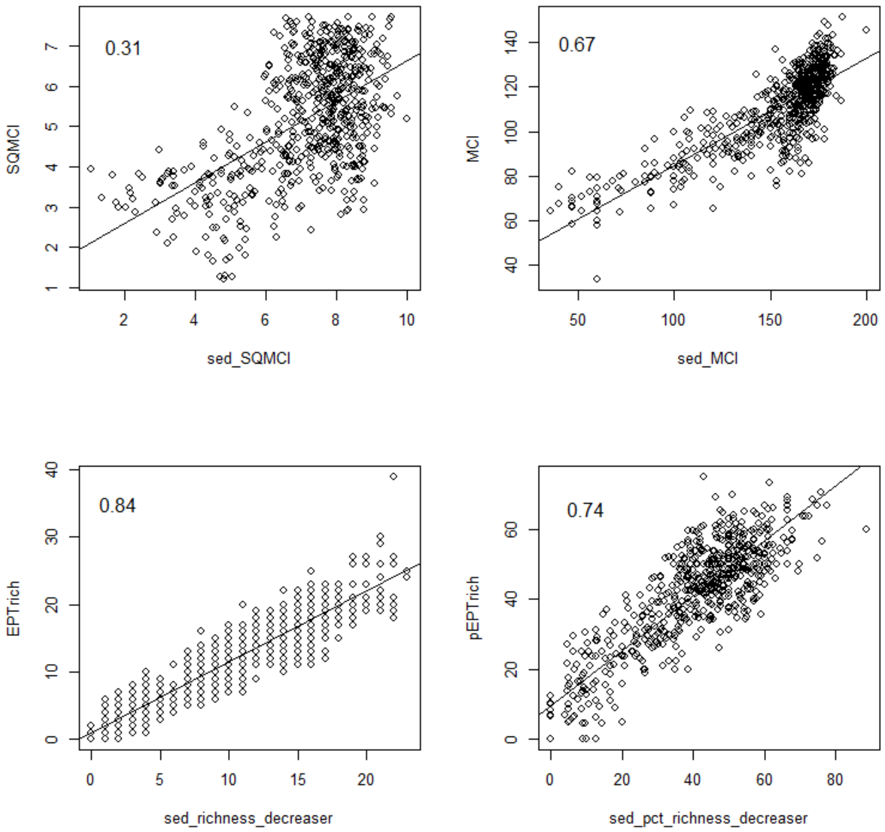


Figure 10. Scatterplots along with the simple linear regression models showing the relationship between existing metrics and comparable sediment-specific metrics based on the field observational dataset used for single-stressor analysis. The adjusted R^2 values are given at the top left of the panels.

Comparison with macroinvertebrate traits

For simplicity, only the best-performing trait metrics (a total of 25) are presented in Table 10 to compare with our stressor-specific metrics (for chlorophyll-*a* those with an $R^2 > 0.1$, for sediment cover those with $R^2 > 0.3$; see Appendix 5 and Appendix 6 for all trait models). Performance of the traits LOWFLEX (low body flexibility) and ADUORLAR (aquatic life stages: adult or larva) across CHLA were similar to that of the best-performing periphyton-specific metric chl_SQMCI (Figure 6). By contrast, none of the traits performed as well as the best-performing sediment-specific metric sed_MCI (Figure 7); however, the best three trait metrics across SEDCOVER were TWO (reproductive technique: male and female), CPI1 and CPI2 (Number of reproductive cycles per individual: 1 and ≥ 2, respectively).

Table 10. Best-performing trait metrics in the single-stressor analysis across the chlorophyll-*a* (CHLA, $R^2 > 0.1$) or % sediment cover (SEDCOVER, $R^2 > 0.3$) gradient along with adjusted R^2 values and response direction. See Table 6 for description of the traits.

Validation CHLA				**Validation SEDCOVER**			
Metric code	**Modality code**	**R^2**	**Direction**	**Metric code**	**Modality code**	**R^2**	**Direction**
11b	LOWFLEX	0.26	-	6c	TWO	0.36	-
16b	ADUORLAR	0.23	-	4a	CPI1	0.35	-
5a	LDA1	0.19	-	4c	CPI2	0.35	+
3c	PLURIV	0.15	+	3c	PLURIV	0.33	+
7b	SUBMERGED	0.14	+	16a	ADUANDLAR	0.33	+
12b	FLATTENED	0.14	-	6a	SINGLE	0.33	+
6a	SINGLE	0.13	+	5d	LDA4	0.33	+
3b	UNIV	0.12	-	7b	SUBMERGED	0.32	+
6c	TWO	0.12	-	12d	SPHERICAL	0.32	+
10b	CRAWLER	0.12	-	3b	UNIV	0.31	-
7a	SURFACE	0.12	-	2a	DESC1	0.31	+
11c	HIGHFLEX	0.12	+	7a	SURFACE	0.31	-
12c	CYLINDRICAL	0.11	+				

3.2.2. Multiple-stressor analysis on field observational data

Periphyton-specific metrics

All multiple regression models were statistically significant ($p < 0.001$). For the eight periphyton-specific metrics, the amount of variation explained (adjusted R^2) by stressors and environmental predictors ranged between 0.11 and 0.43, with chl_MCI and chl_SQMCI being the best-performing metrics ($R^2 = 0.43$ and 0.36, respectively, Figure 11). After model selection aimed at selecting the minimum adequate model for each response, all periphyton-specific metrics retained CHLA as a predictor with standardised effect sizes (coefficients) ranging from 0.11 (chl_richness_decreaser) to 0.23 (chl_pct_richness_decreaser) (Figure 11). As expected, the effects of CHLA were negative for all metrics except for those calculated based on periphyton-tolerant taxa (increasers). On the other hand, SEDCOVER was only retained in two models, chl_SQMCI and chl_pct_abund_decreaser, and effect sizes were similar to those of CHLA (Figure 11). DIN was only retained in one model, chl_pct_abund_decreaser, having a positive effect on this metric. DRP and StAccMaxRate, a measure of abstraction pressure, were not retained in any of the periphyton-specific metric models, and neither were FRE3 or USHardness. Up to five environmental predictors were retained in the models, among those USSlope (in 6 metrics), SegTSeas (in 5) and MALF (5) were the most prevalent across the eight metrics (Figure 11).

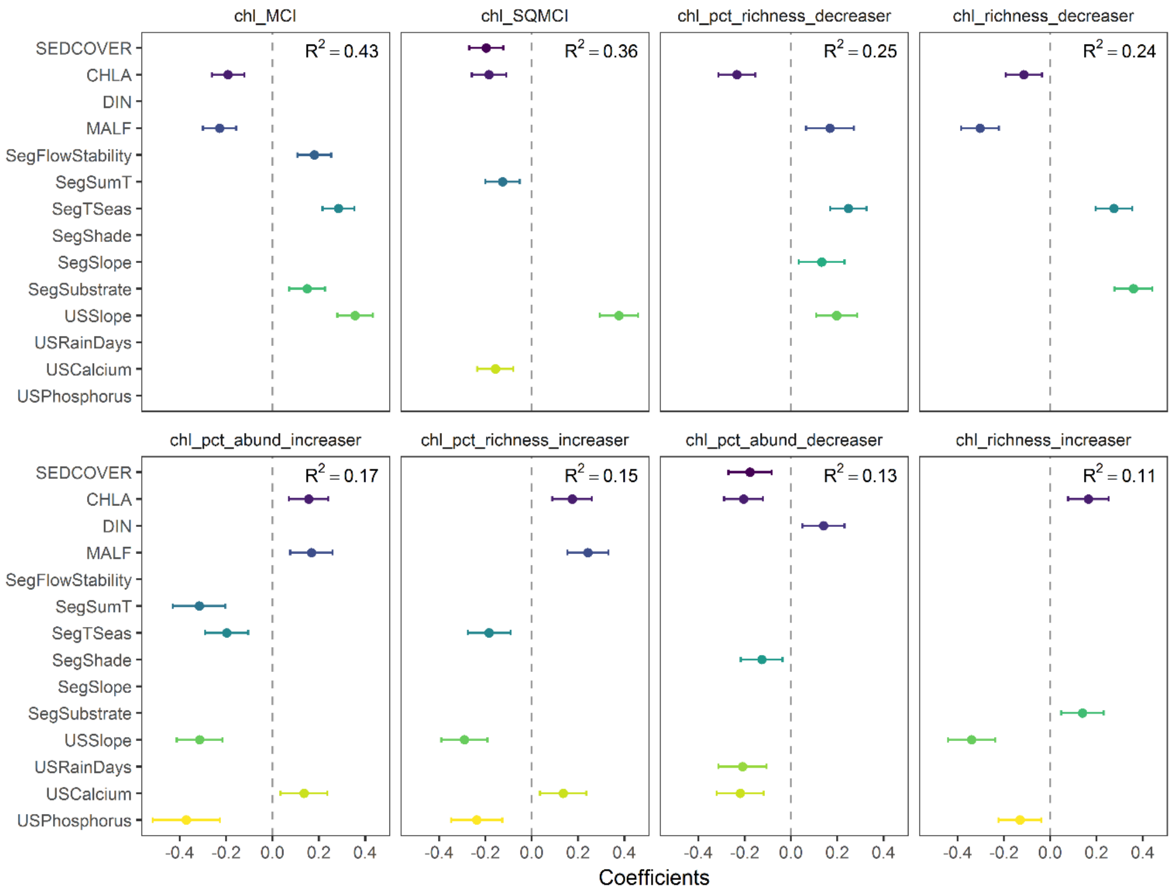


Figure 11. Regression coefficients (± 95% CI) of the predictors in the multiple linear regression models for eight periphyton-specific metrics built on field observational data. Metrics are ordered according to decreasing adjusted R^2 values. As predictors and metrics had been scaled and centred before the analyses, the regression coefficients represent standardised effect sizes which can be compared among predictors and metrics. See Table 1 and Table 2 for description of the predictors and metrics, respectively.

Hierarchical variance partitioning showed that the relative importance of CHLA among periphyton-specific metric models ranged from 10% to 30% (Figure 12), the highest was attributed to chl_pct_abund_decreaser, but SEDCOVER also had high relative importance for this metric (27%). Across all periphyton-specific metrics, CHLA explained on average 22% of the total variance (Figure 12).

Figure 12. Relative importance of predictor variables determined by hierarchical variance partitioning from multiple regression linear models for the eight periphyton-specific metrics. Metrics are ordered according to decreasing adjusted R^2 values. See Figure 11 caption for more details.

Sediment-specific metrics

All regression models were statistically significant ($p < 0.001$). For the eight sediment-specific metrics, the amount of variation explained (adjusted R^2) by stressors and environmental predictors ranged between 0.37 and 0.55, with sed_pct_richness_decreaser and sed_MCI being the two best-performing metrics ($R^2 = 0.55$ and 0.54, respectively). After model selection, all sediment-specific metrics retained SEDCOVER as a predictor with standardised effect sizes (coefficients) ranging from 0.10 (sed_pct_abund_increaser) to 0.22 (sed_pct_abund_decreaser)

(Figure 13). As expected, the effects of SEDCOVER were negative for all metrics except for metrics calculated based on sediment-tolerant taxa (increasers). CHLA was retained in a single model only, for sed_pct_abund_decreaser, having a slightly smaller effect size than SEDCOVER and also with a negative effect (Figure 13). DRP was only retained in one model, sed_richness_decreaser, having a weak positive effect on this metric. DIN and StAccMaxRate were not retained in any of the models of sediment-specific metrics, and neither were SegShade or USCalcium. Up to eight environmental predictors were retained in the models, among those SegSubstrate (in all 8 metrics), USRainDays (in 7), SegTSeas (5) and MALF (5) were the most prevalent across the eight metrics (Figure 13). SegSubstrate is the predicted natural proportional cover of bed substrate with higher values representing proportionally larger substrate sizes. The positive effect of proportionally larger substrate sizes on metrics calculated based on sediment-sensitive taxa or indicator values of all indicator taxa, and in turn, the negative effect on those calculated based on sediment-tolerant taxa (increasers) is not surprising.

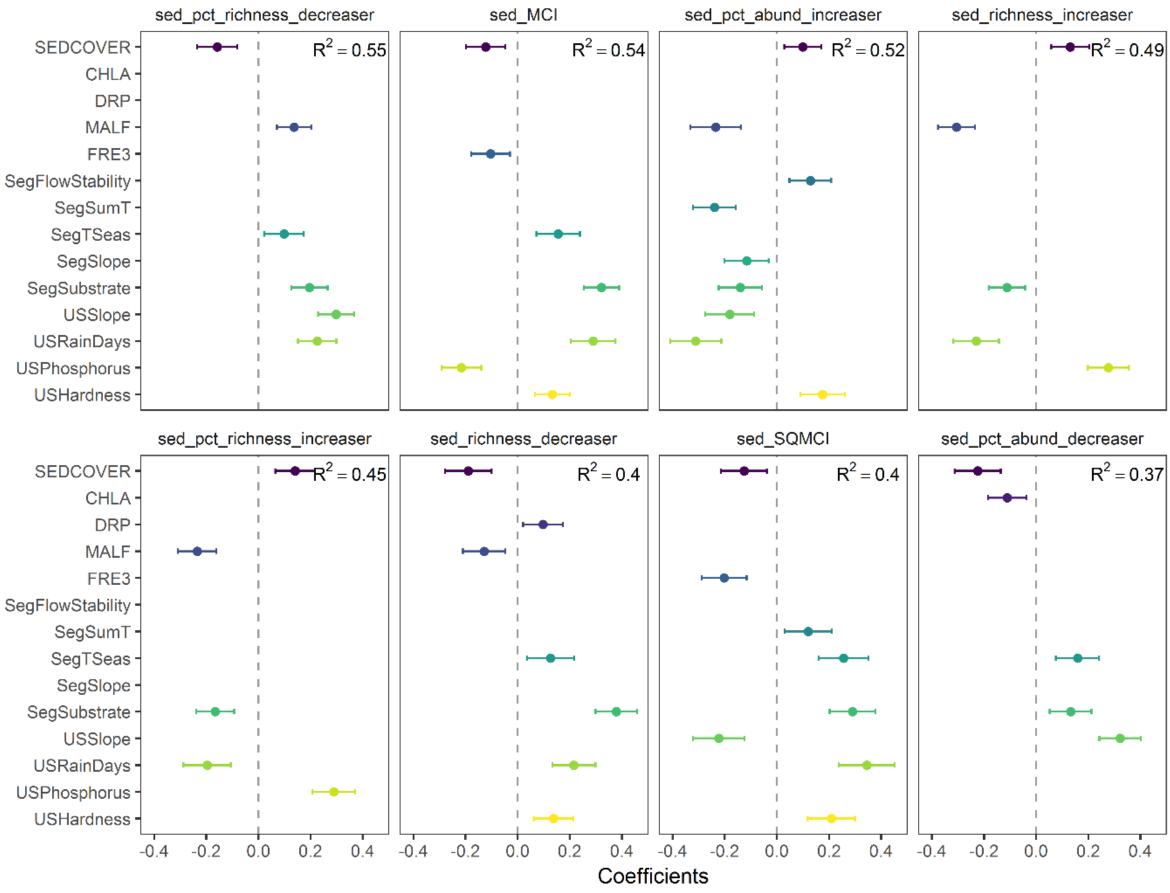


Figure 13. Regression coefficients (± 95% CI) of the predictors in the multiple linear regression models for eight sediment-specific metrics built on field observational data. Metrics are ordered according to decreasing adjusted R^2 values. See Figure 11 caption for more details.

Hierarchical partitioning of variance showed that the relative importance of SEDCOVER among the models of the sediment-specific metrics ranged from 9 to 26% (Figure 14); the highest was attributed to sed_pct_abund_decreaser, but this was the metric for which CHLA was also a predictor with a relative importance of 9%. Across all sediment-specific metrics, SEDCOVER explained on average 16% of the total variance (Figure 14).

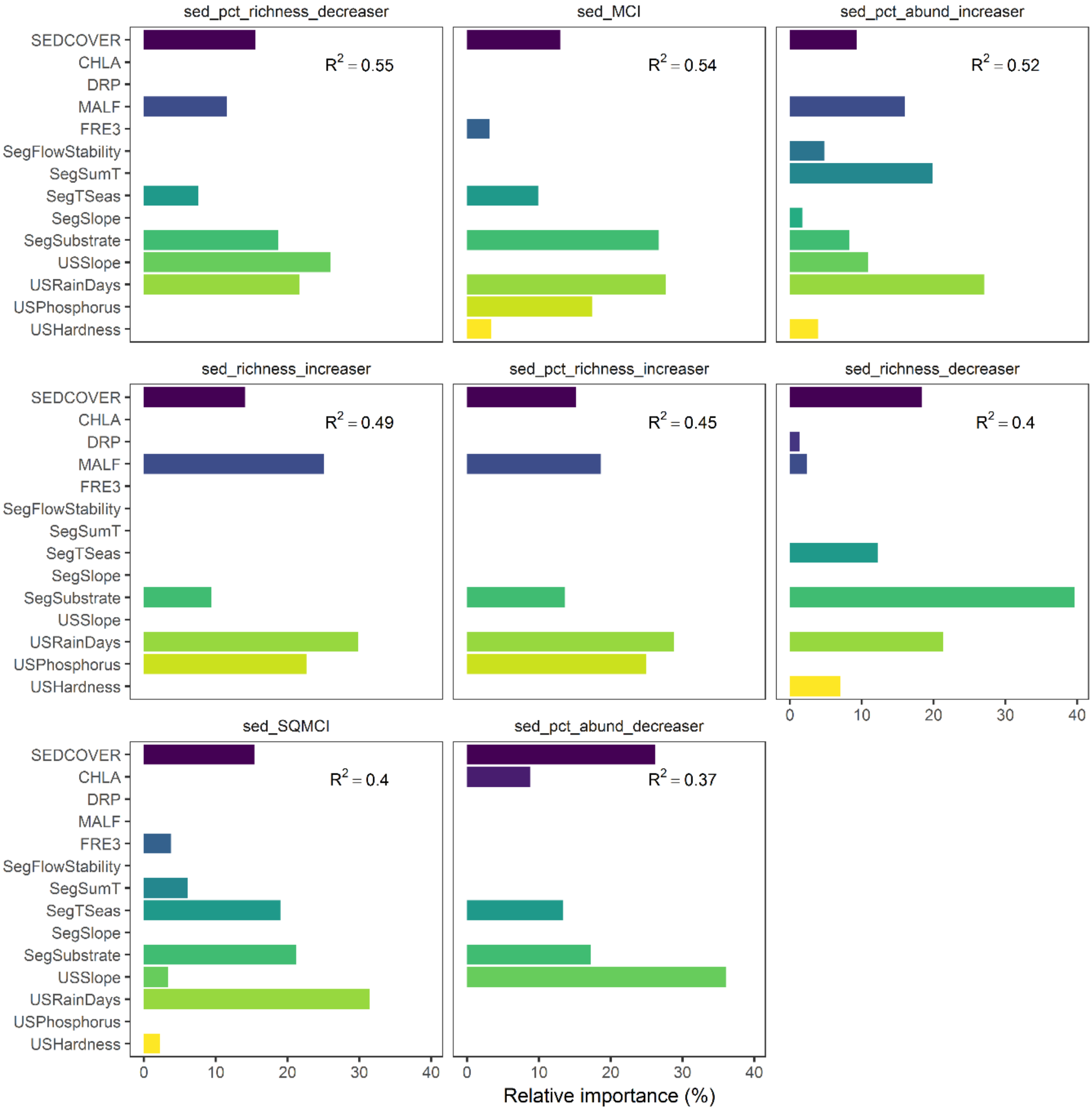


Figure 14. Relative importance of predictor variables determined by hierarchical variance partitioning from multiple linear regression models for the eight sediment-specific metrics. Metrics are ordered according to decreasing adjusted R^2 values. See Figure 11 caption for more details.

After comparing the performance of periphyton-specific metrics and sediment-specific metrics, simple regression models were built to compare scores of periphyton-specific metrics and sediment-specific metrics directly. SQMCI-style metrics were only weakly correlated but MCI-style metrics were correlated as well as the richness of periphyton-sensitive taxa with that of sediment-sensitive taxa (Figure 15). However, richness of periphyton-tolerant and sediment-tolerant taxa were only weakly correlated (Figure 15).

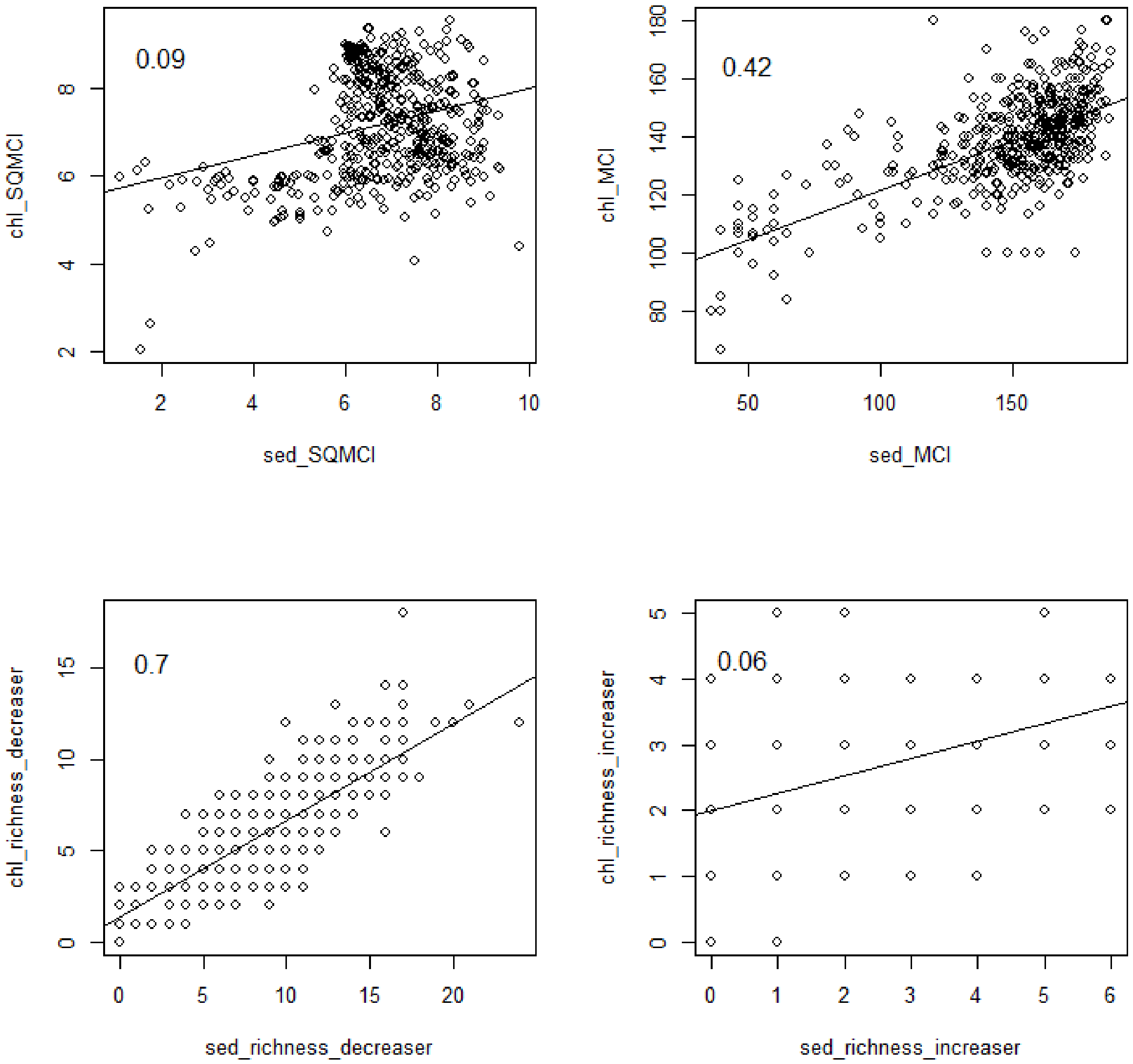


Figure 15. Scatterplots along with the simple linear regression models showing the relationship between periphyton-specific and comparable sediment-specific metrics based on the field observational dataset used for multiple-stressor analysis. The adjusted R^2 values are given at the top left of the panels.

Comparison with existing metrics

All regression models were statistically significant ($p < 0.001$). The adjusted R^2 values of the models of the existing metrics ranged between 0.3 and 0.48, with the highest amount variance explained for MCI (Figure 16). As expected, all existing metrics responded negatively to both SEDCOVER and CHLA, with the exception of EPTrich which responded to SEDCOVER only (Figure 16). Standardised effect sizes of CHLA for existing metrics ranged from 0.10 (MCI) to 0.18 (pEPTabund), which are similar to effect sizes observed for periphyton-specific metrics. Standardised effect sizes of SEDCOVER for existing metrics ranged from 0.14 (MCI) to 0.25 (SQMCI), which were slightly higher than those for sediment-specific metrics.

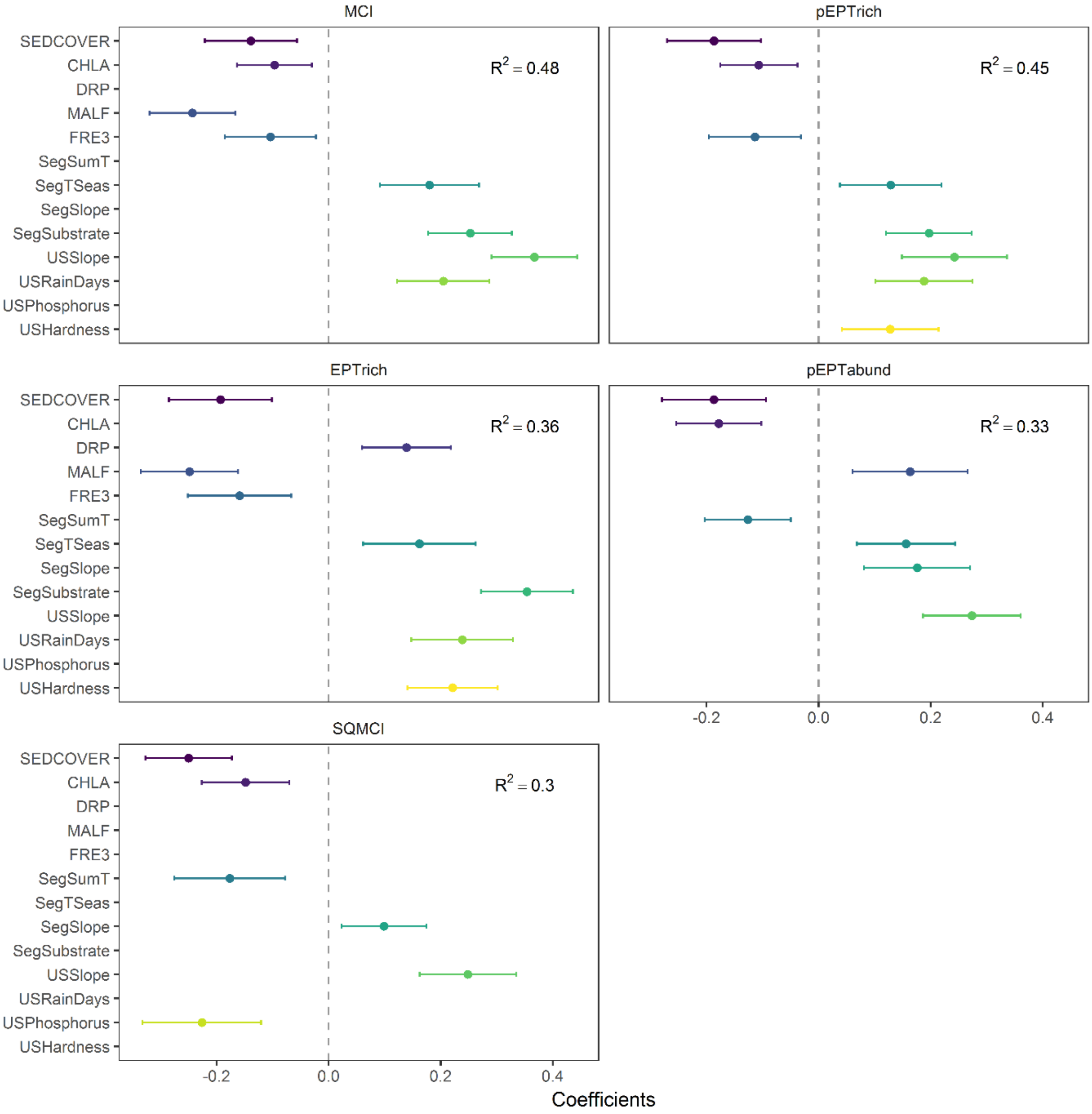


Figure 16. Regression coefficients (± 95% CI) of the predictors in the multiple linear regression models for five existing metrics built on field observational data. Metrics are ordered according to decreasing adjusted R^2 values. See Figure 11 caption for more details.

Relative importance of CHLA for existing metrics ranged from 7 to 16% (Figure 17) which was comparably lower than that for periphyton-specific metrics for which relative importance was often above 20%. Relative importance of SEDCOVER for existing metrics ranged from 11 to 28% (Figure 17), which was comparable to that for sediment-specific metrics.

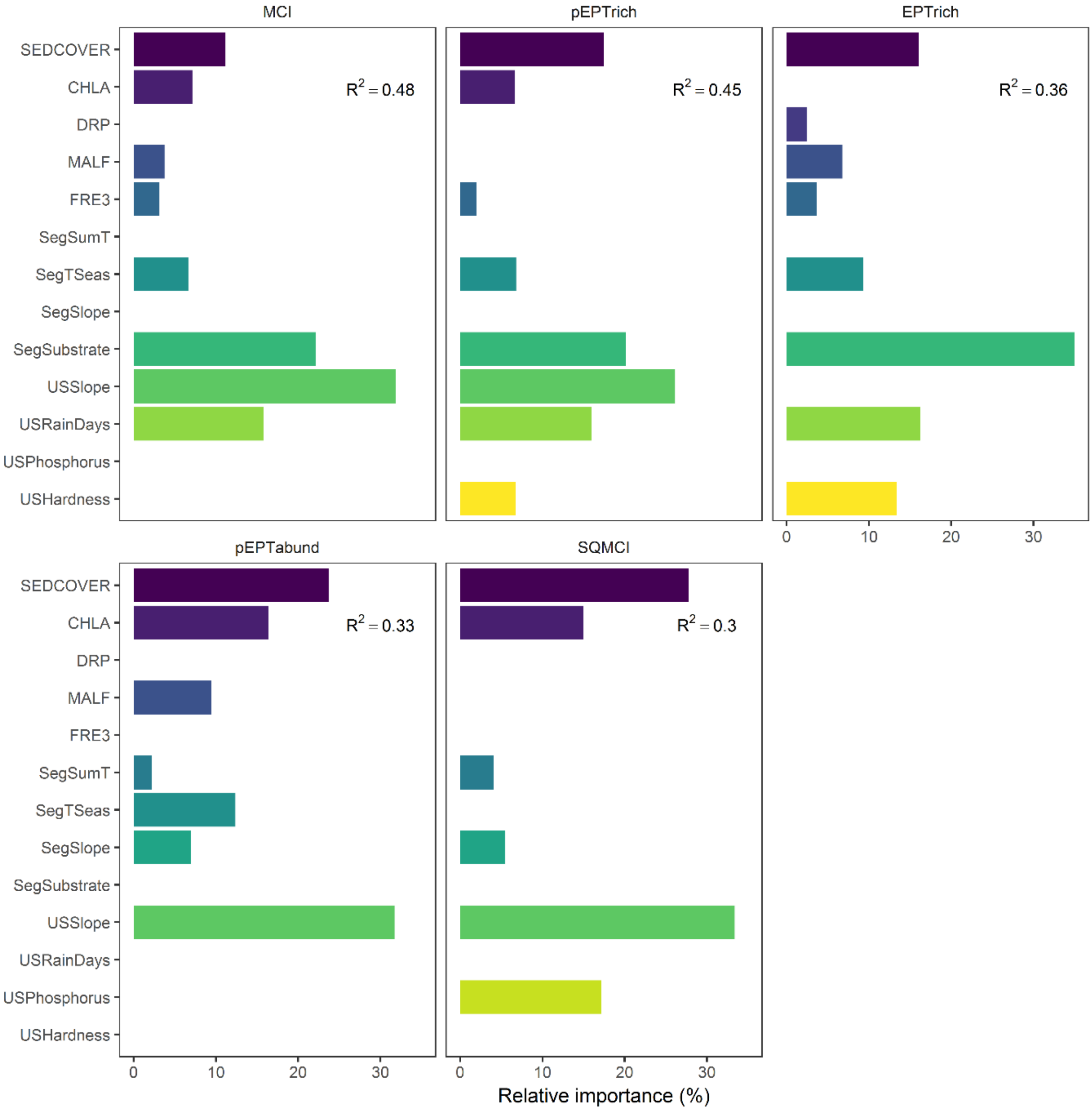


Figure 17. Relative importance of predictor variables determined by hierarchical variance partitioning from multiple linear regression models for five existing metrics. Metrics are ordered according to decreasing adjusted R^2 values. See Figure 11 caption for more details.

Comparison with macroinvertebrate traits

The focus of this analysis was on identifying trait metrics that may potentially also be useful as stressor-specific metrics in addition to our developed stressor-specific metrics. Out of the total of 59 trait modalities, CHLA and/or SEDCOVER were retained as predictors in the multiple stressor models of 36 trait metrics. All these models were statistically significant ($p < 0.001$) and the adjusted R^2 values ranged between 0.06 and 0.44. Regression coefficients, and the relative importance of all predictors retained in these 36 models along with R^2 values, are presented in Appendix 7 and Appendix 8, respectively.

For comparison, we summarised standardised effect sizes and response direction of the stressors as well as R^2 values of the new stressor-specific metrics and the 36 trait metrics (Table 11). Among the 8 trait metrics that were responsive to CHLA only, effect sizes were smaller than the better-performing periphyton-specific metrics (e.g. 0.14 for 12b-FLATTENED vs. 0.23 for chl_pct_richness_decreaser or 0.19 for chl_MCI), nevertheless all effect sizes were ≥ 0.10 (Table 11). Among the 12 trait metrics that were responsive to SEDCOVER only, on the other hand, effect sizes of 11 trait metrics were larger than that of sed_richness_decreaser (0.19) which was the sediment-specific metric with the largest effect size that only responded to SEDCOVER. The effect sizes of all SEDCOVER only trait metrics were ≥ 0.12 (Table 11). There were 2 trait metrics, 14c-GENERALIST and 7c-TERRESTRIAL, that responded to CHLA and to SEDCOVER but in opposite directions.

Table 11. Summary of stressor-specific metrics (8 each) and the trait metrics that responded to either CHLA (8) or SEDCOVER (16) or both (12) according to multiple-stressor analysis of field observational data. The metrics within each group are ordered according to their standardised effect sizes. Response direction for CHLA and SEDCOVER and model R^2 are also given.

Metric	**R^2**	**CHLA Effect size**	**Direction**	**SEDCOVER Effect size**	**Direction**
Periphyton-specific metrics					
chl_pct_richness_decreaser	0.25	0.23	-		
chl_pct_abund_decreaser	0.13	0.21	-	0.18	-
chl_MCI	0.43	0.19	-		
chl_SQMCI	0.36	0.18	-	0.20	-
chl_pct_richness_increaser	0.15	0.17	+		
chl_richness_increaser	0.11	0.16	+		
chl_pct_abund_increaser	0.17	0.16	+		
chl_richness_decreaser	0.24	0.11	-		
Sediment-specific metrics					
sed_pct_abund_decreaser	0.37	0.11	-	0.22	-
sed_richness_decreaser	0.40			0.19	-
sed_pct_richness_decreaser	0.55			0.16	-
sed_pct_richness_increaser	0.45			0.14	+
sed_richness_increaser	0.49			0.13	+
sed_SQMCI	0.40			0.13	-
sed_MCI	0.54			0.12	-
sed_pct_abund_increaser	0.52			0.10	+
Traits responsive to CHLA only					
12b-FLATTENED	0.18	0.14	-		
10d-ATTACHED	0.07	0.13	-		
5c-LDA3	0.12	0.12	-		
2b-DESC2	0.32	0.10	-		
2a-DESC1	0.38	0.11	+		
15c-PLASTRON	0.06	0.11	+		
12c-CYLINDRICAL	0.25	0.12	+		
14a-SPECIALIST	0.10	0.12	+		
Traits responsive to SEDCOVER only					
1b-SIZE2	0.21			0.29	-
2c-DESC3	0.37			0.26	-
13b-SCRAPER	0.22			0.25	-
10b-CRAWLER	0.41			0.25	-
3b-UNIV	0.38			0.23	-
14b-MODERATESP	0.18			0.23	-
13d-FILTERFEED	0.07			0.22	+
3c-PLURIV	0.40			0.22	+
8a-EGGFREE	0.28			0.22	-

Metric	R^2	CHLA Effect size	Direction	SEDCOVER Effect size	Direction
12d-SPHERICAL	0.34			0.21	+
7b-SUBMERGED	0.44			0.21	+
15a-TEGUMENT	0.20			0.18	+
3a-SEMI	0.34			0.15	-
10a-SWIMMER	0.37			0.13	+
9c-DISSHIGH	0.11			0.12	-
9a-DISSLOW	0.35			0.12	+
Traits responsive to CHLA and SEDCOVER					
5a-LDA1	0.39	0.18	-	0.28	-
15b-GILL	0.16	0.16	-	0.21	-
11b-LOWFLEX	0.39	0.15	-	0.18	-
13f-ALGALP	0.10	0.15	+	0.18	+
11c-HIGHFLEX	0.11	0.14	+	0.27	+
7a-SURFACE	0.36	0.14	-	0.20	-
10c-BURROWER	0.34	0.14	+	0.15	+
1a-SIZE1	0.25	0.13	+	0.33	+
15d-AERIAL	0.12	0.11	+	0.14	+
14c-GENERALIST	0.19	0.11	-	0.17	+
7c-TERRESTRIAL	0.33	0.11	+	0.14	-
16b-ADUORLAR	0.43	0.10	-	0.13	-

3.2.3. Multiple-stressor analysis on experimental data

Periphyton-specific metrics

For the eight periphyton-specific metrics, the amount of variation explained (adjusted R^2) ranged between 0 and 0.63, and only the models of the first four metrics presented in Figure 18 were statistically significant ($p < 0.05$). Nutrient level, or the interaction between nutrients and sediment was not significant in any of the models, although for chl_richness_increaser nutrients had a significant positive effect and the interaction was marginally statistically significant (Figure 18). Sediment, on the other hand, had a positive effect on chl_pct_abund_increaser and a negative effect on chl_SQMCI as well as on chl_pct_abund_decreaser.

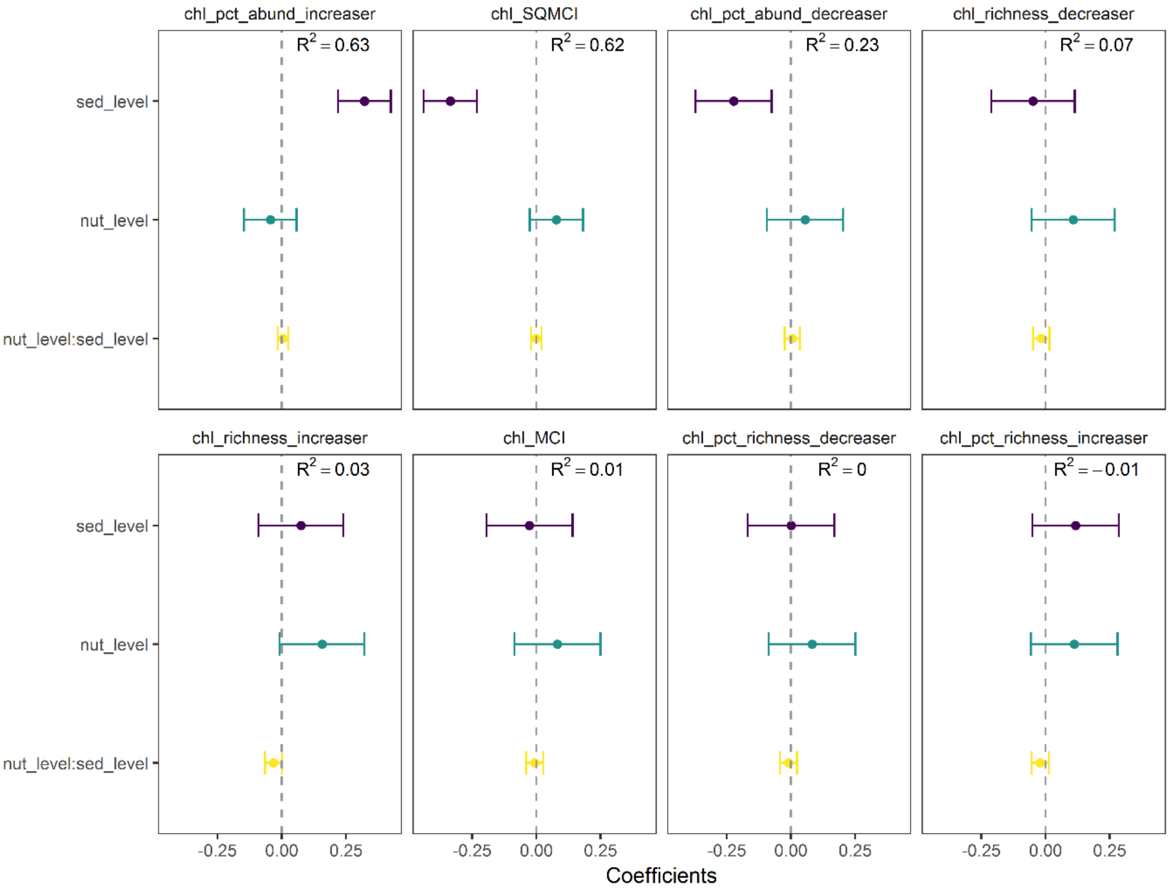


Figure 18. Regression coefficients (± 95% CI) of the predictors in the multiple linear regression models for eight periphyton-specific metrics built on experimental data. If the 95% CI crosses zero, the effect is not statistically significant. Metrics are ordered according to decreasing adjusted R^2 values. As metrics were scaled and centred before the analyses and we used nominal treatment values as predictors that represent a spectrum of nutrient concentrations and deposited sediment levels typically observed in New Zealand rivers, effect sizes (coefficients) are comparable within and among the models. See Table 2 for description of the metrics.

Sediment-specific metrics

For the eight sediment-specific metrics, the amount of variation explained (R^2) ranged between 0 and 0.54, and only the models of the first five metrics presented in Figure 19 were statistically significant ($p < 0.05$). Out of these, sed_pct_abund_increaser was positively affected by sediment and sed_pct_abund_decreaser was negatively affected (Figure 19). None of the metrics were affected by nutrients or the interaction between sediment and nutrients.

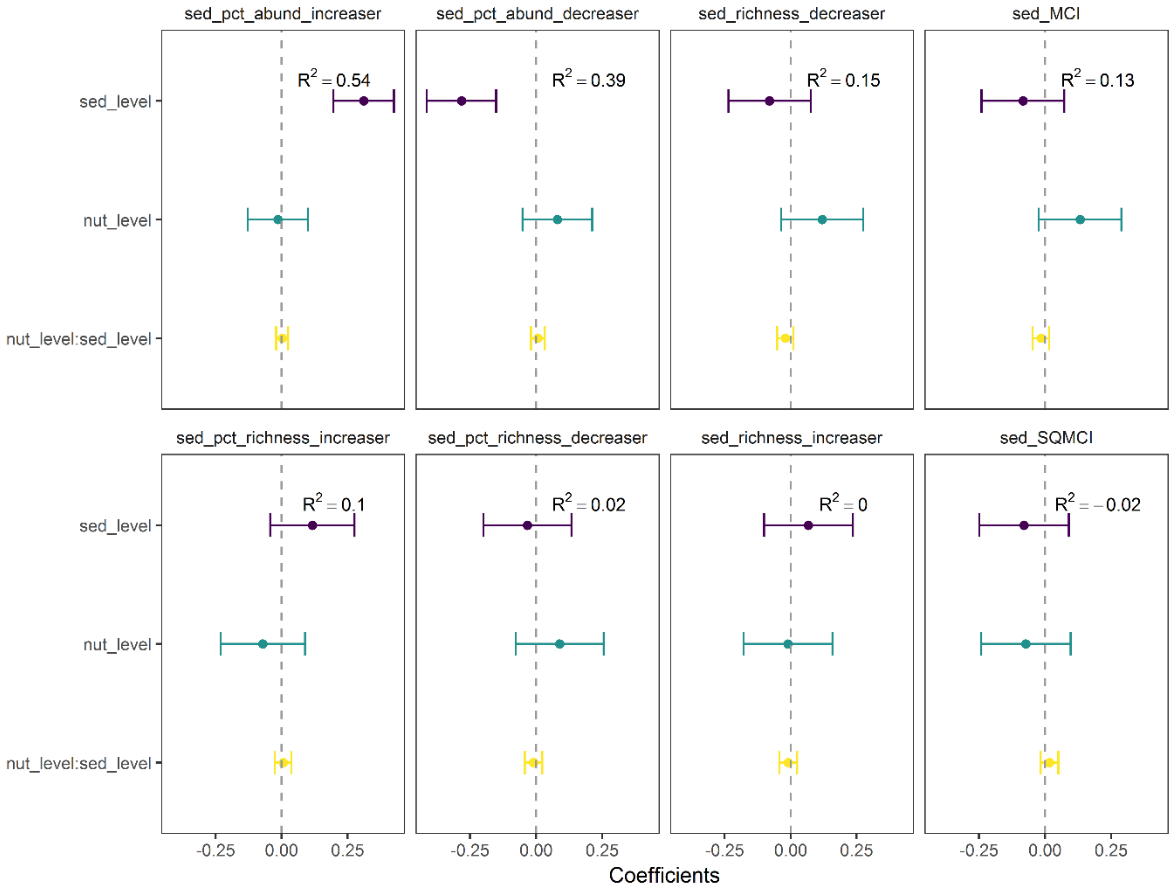


Figure 19. Regression coefficients (± 95% CI) of the predictors in the multiple linear regression models for eight sediment-specific metrics built on experimental data. Metrics are ordered according to decreasing adjusted R^2 values. See Figure 18 caption for more details.

After comparing the performance of periphyton-specific metrics and sediment-specific metrics, simple regression models were built to compare scores of periphyton-specific metrics and sediment-specific metrics directly. SQMCI-style metrics were only weakly correlated and MCI-style metrics were not correlated (Figure 20). On the other hand, the richness of periphyton-sensitive taxa was correlated with that of sediment-sensitive taxa. The richness of periphyton-tolerant and sediment-tolerant taxa were overall relatively small (Figure 20).

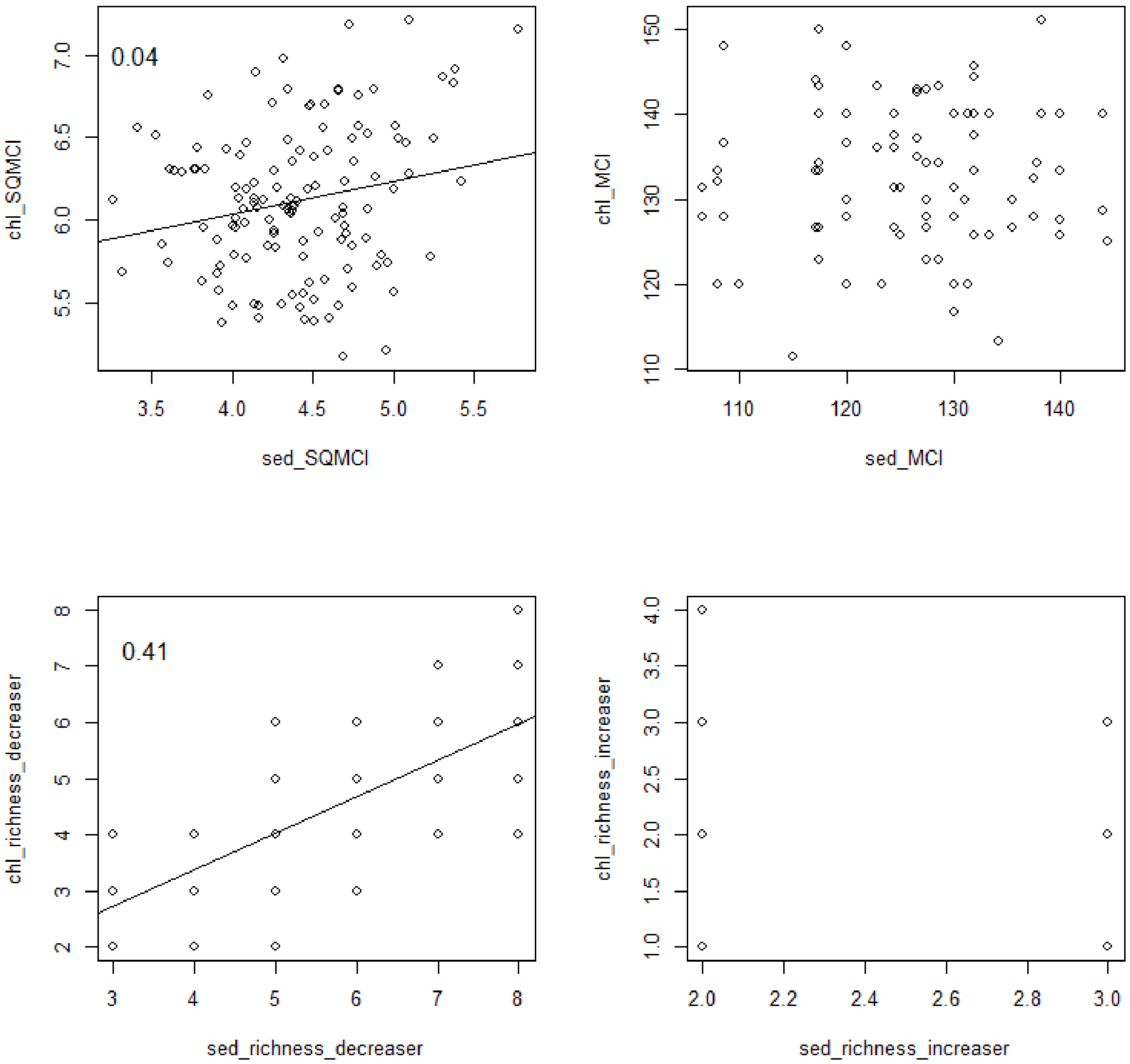


Figure 20. Scatterplots along with the simple linear regression models (if statistically significant) showing the relationship between periphyton-specific and comparable sediment-specific metrics based on the experimental dataset used for multiple-stressor analysis. The adjusted R^2 values are given at the top left of the panels.

Comparison with existing metrics

All five regression models were statistically significant ($p < 0.05$). The R^2 values of the models of the existing metrics ranged from 0.06 to 0.63 (Figure 21). The SQMCI and %EPT abundance were negatively affected by sediment, while EPT richness was positively affected by nutrients. None of the metrics were affected by sediment and nutrients, although for EPT richness sediment as well as the interaction were marginally statistically significant terms in addition to nutrients (Figure 21).

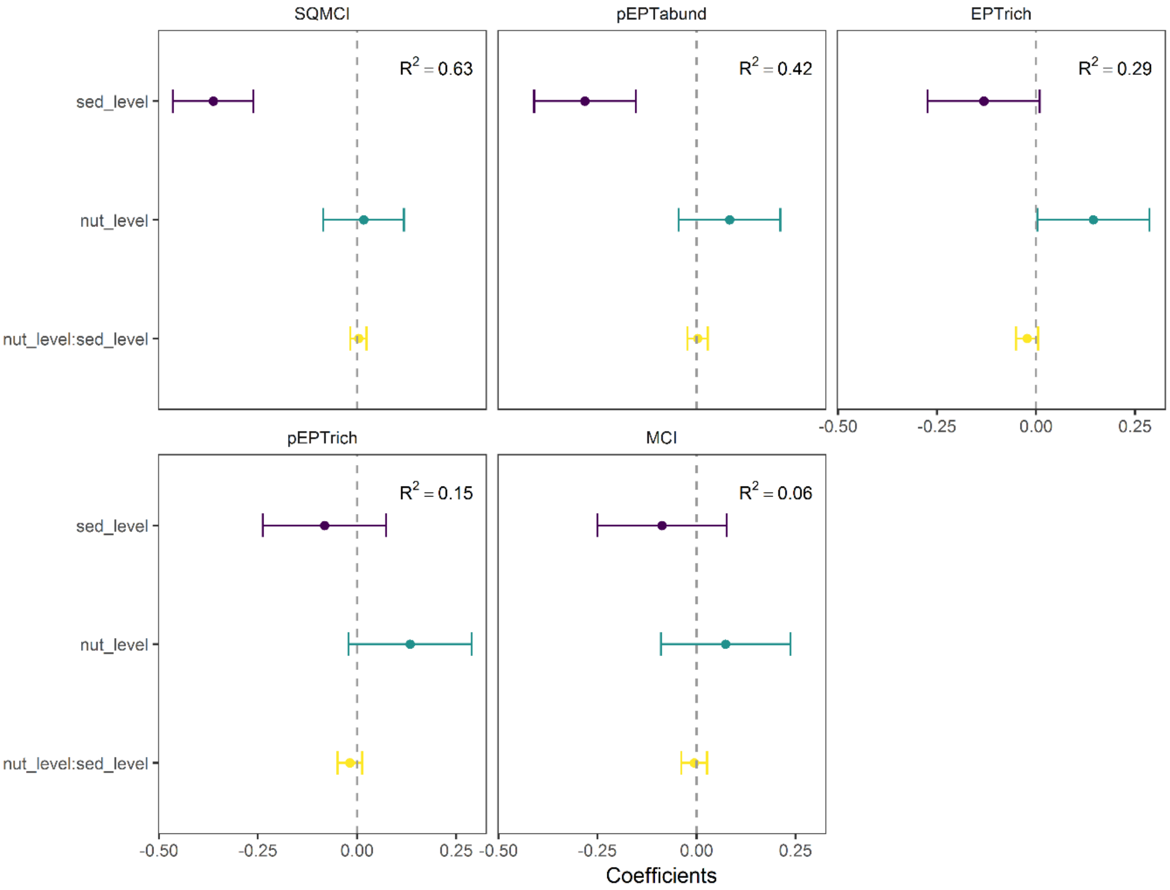


Figure 21. Regression coefficients (± 95% CI) of the predictors in the multiple linear regression models for existing metrics built on experimental data. Metrics are ordered according to decreasing adjusted R^2 values. See Figure 18 caption for more details.

Consequently, simple regression models were built between existing metric scores and scores of comparable periphyton-specific metrics for further comparison. SQMCI was highly correlated with chl_SQMCI (Figure 22) but not correlated with sed_QMCI (Figure 23). MCI on the other hand was correlated to both chl_MCI (Figure 22) and sed_MCI (Figure 23). Both EPT metrics were correlated with equivalent periphyton-specific and sediment-specific metrics but the correlation was higher for the latter (Figure 22 and Figure 23).

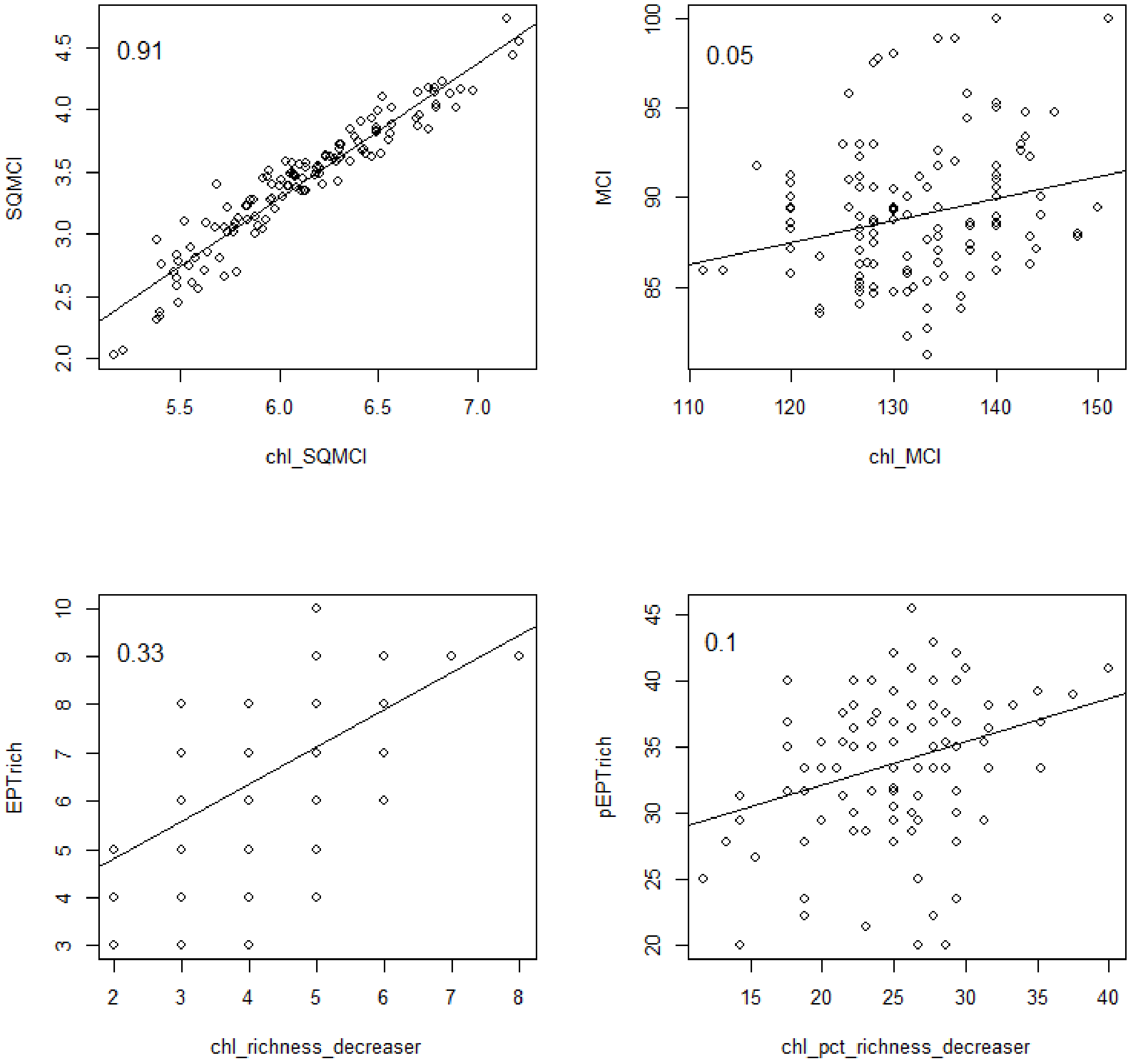


Figure 22. Scatterplots along with the simple linear regression models showing the relationship between existing metrics and comparable periphyton-specific metrics based on the experimental dataset used for multiple-stressor analysis. The adjusted R^2 values are given at the top left of the panels.

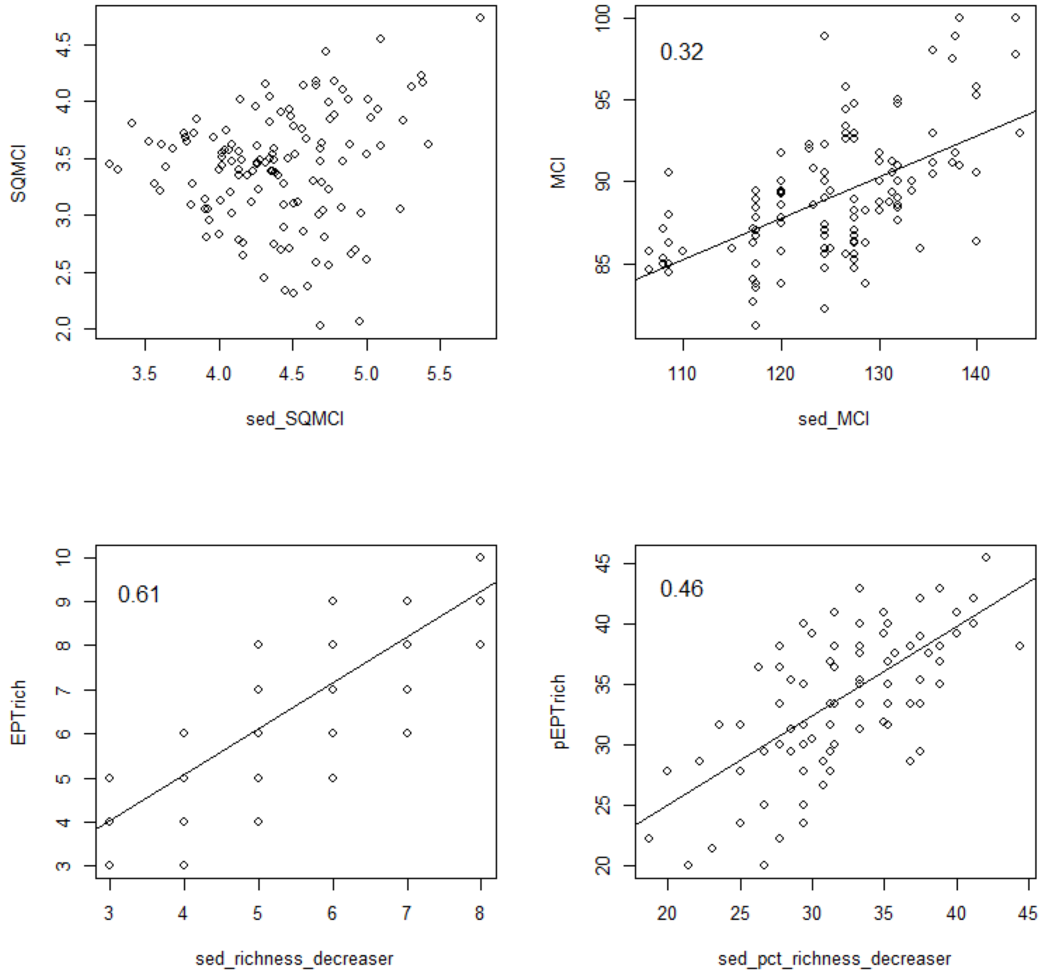


Figure 23. Scatterplots along with the simple linear regression models (if statistically significant) showing the relationship between existing metrics and comparable sediment-specific metrics based on the experimental dataset used for multiple-stressor analysis. The adjusted R^2 values are given at the top left of the panels.

Comparison with macroinvertebrate traits

Out of the total of 59 trait modalities, nutrient level was retained as a predictor in only a single model while sediment was retained in 23 models (Table 12). All these models were statistically significant ($p < 0.05$) and the R^2 values ranged between 0.04 and 0.53. None of the models retained both nutrients and sediment. Regression coefficients of the predictors along with R^2 values are presented in Appendix 9. We explored model output to identify trait metrics that may potentially also be useful as stressor-specific metrics in addition to our developed stressor-specific metrics. For nutrients, only 1d-SIZE4 may be stressor-specific, the model R^2 being very low, but our periphyton-specific metrics also did not respond to nutrients (Table 12). For sediment, on the other hand, many trait metrics may be stressor-specific with 14 traits having effect sizes of ≥0.20 among which the best were similar to the effect sizes of sed_pct_abund_increaser (0.31) and sed_pct_abund_decreaser (0.28) (Table 12).

Table 12. Summary of stressor-specific metrics and the trait metrics that responded to either nutrients or sediment according to multiple-stressor analysis of experimental data. The metrics within each group are ordered according to their standardised effect sizes. Response direction for nutrients and sediment, and model R^2 are also given.

		Nutrients		Sediment	
Metric	**R^2**	**Effect size**	**Direction**	**Effect size**	**Direction**
Periphyton-specific metrics					
chl_SQMCI	0.62			0.34	-
chl_pct_abund_increaser	0.63			0.32	+
chl_pct_abund_decreaser	0.23			0.22	-
Sediment-specific metrics					
sed_pct_abund_increaser	0.54			0.31	+
sed_pct_abund_decreaser	0.39			0.28	-
sed_MCI	0.13	0.13 (marg)	+		
Traits responsive to nutrients only					
1d-SIZE4	0.04	0.17	+		
8b-EGGCEMENT	0.14	0.14 (marg)	+		
5c-LDA3	0.11	0.14 (marg)	+		
Traits responsive to sediment only					
7a-SURFACE	0.22			0.33	-
13c-DEPOSIT	0.53			0.31	+
5a-LDA1	0.23			0.30	-
9b-DISSMEDIUM	0.49			0.28	-
7b-SUBMERGED	0.13			0.28	+
16b-ADUORLAR	0.28			0.27	-
1b-SIZE2	0.12			0.27	-
5e-LDA5	0.47			0.26	+
2d-DESC4	0.14			0.25	-
12b-FLATTENED	0.47			0.25	-
9a-DISSLOW	0.30			0.24	+
10c-BURROWER	0.47			0.23	+
1c-SIZE3	0.18			0.21	+
3a-SEMI	0.09			0.20	-
11c-HIGHFLEX	0.09			0.19	+
10a-SWIMMER	0.32			0.18	-
6a-SINGLE	0.25			0.18	+
3c-PLURIV	0.16			0.18	+
11b-LOWFLEX	0.16			0.18	-
13d-FILTERFEED	0.27			0.17	-
13b-SCRAPER	0.04			0.17	-
3b-UNIV	0.14			0.16	-
12a-STREAMLINED	0.26			0.15	+
12c-CYLINDRICAL	0.20			0.15 (marg)	+
2c-DESC3	0.08			0.14 (marg)	-
6b-HERMA	0.18			0.14 (marg)	-

4. DISCUSSION

4.1. How do the stressor-specific metrics perform?

Time-for-space analysis using field observational data collected from across New Zealand (covering a large variety of stream types) allowed testing of whether the stressor-specific metrics responded predominately to either algal proliferation or sedimentation; so that if used in concert, they may have diagnostic power to distinguish between the individual effects of the two stressors. Analysis of an experimental dataset from a mesocosm study provided an unbiased test of whether the metrics have diagnostic power when applied at a single stream site.

Our approach to validation of stressor-specific metrics was:
1. screen which metrics, out of a candidate set of metrics, have the strongest relationship with the respective stressor (single-stressor analysis on field observational data)
2. out of the reduced candidate set of metrics, test which metrics respond to a single stressor only, or at least more strongly to one than the other (multiple-stressor analysis of field observational and experimental data)
3. compare to what degree the best-performing stressor-specific metrics complement existing metrics (including results from all three analyses)
4. compare with trait metrics that might be alternative stressor-specific metrics (including results from all three analyses).

While the primary motivation and focus of this project is to test the usefulness of these metrics in resource management, our study also provides an opportunity to progress our understanding of the mechanistic effects of these two stressors on macroinvertebrate communities and the theory of the effects of multiple stressors. This opportunity arises because stressor-specific metrics can only be developed provided nutrients and sediment affect macroinvertebrates via different modes of action, and if there is a variety of species that respond to these stressors differently. As such, we also briefly discuss the likely underlying ecological reasons for the results observed.

4.1.1. Screening for suitability of the metrics

Single-stressor analysis using an independent field observational dataset showed some clear patterns with respect to the strength of the relationship with the stressors which informed decision on which metrics to exclude and which to retain.

Exclusion of _MCI_decreaser and _MCI_increaser

Overall, MCI/SQMCI-style metrics taking into account the indicator values of both decreasers and increasers (indicator values ranging from 1–10) were consistently among the best-performing metrics within and among periphyton-specific and sediment-specific metrics, and were considerably more strongly related to the stressors than the metrics that are based on decreasers or increasers only and do not take into account the degree of sensitivity or tolerance (i.e. no indicator values). The consistent improved performance of the MCI/SQMCI-style metrics suggests that the inclusion of increases and decreasers rather than differentiation of the degree of sensitivity or tolerance is an important feature. When more indicator taxa are used in metric calculation, an increase in precision occurs for a spatial dataset where other environmental variables also introduce large amounts of variation. As a result, we excluded _MCI_decreaser and _MCI_increaser.

Exclusion of _MCI_ plusnonind

The rational for assigning taxa that were non-responsive to the stressor gradient an intermediate indicator value was related to the expectation that a taxon that neither increases nor decreases in relative abundance is indifferent to the stressor condition. To be conservative, a value of 5 was assigned to these non-indicator taxa, the same as for increaser taxa that showed a weak tolerance to the stressor. While these taxa theoretically add no value to the prediction precision of the stressor effects, they may be able to stabilise unusually high or low metric scores due to sampling bias. Inclusion of the non-indicator taxa in metric calculation did, however, not improve metric performance and hence were excluded.

Exclusion of _MCI_log and _SQMCI_log

Our analyses showed that taxa sensitivities are not equally spread across the stressor gradients. Out of the 26 decreasers across the chlorophyll-*a* gradient, 13 had their thresholds in the lowest fifth of the threshold spectrum. Out of the 31 decreasers across the sediment cover gradient, 20 had their thresholds in the lowest fifth of the threshold spectrum. This suggests that the most sensitive taxa respond to relatively small increases in the levels of algal proliferation or sedimentation. This is possibly because they have not evolved under such conditions and hence have not developed the traits that allow them to cope with these environmental changes.

We expected to identify at least some increaser taxa as our gradient forest analysis was based on relative abundance data; hence, tolerant taxa necessarily increase even if they do not increase in absolute abundance. However, compared to decreaser taxa, only a few increaser taxa could be identified with a less obvious pattern in the distribution of their thresholds across the stressor gradient. Absence of traits that specifically evolved to cope with high stressor values may also explain why only a few increaser taxa could be identified.

With indicator value assignment according to 5 bins based on the raw-scale threshold spectrum, those 13 decreasers across the chlorophyll-*a* gradient and 31 decreasers across the sediment cover gradient have an indicator value of 10. With assignment according to 5 bins based on the log-scale threshold spectrum, many of these decreaser taxa had an indicator value of 9 or lower. Theoretically, both variants could be justified. The metric calculated with indicator values using the first variant should be able to better detect small changes in stressor conditions compared to reference, while the metric calculation with the other set of indicator values should be able to better discriminate between stressor conditions further along the stressor gradient. In our analyses, the metrics performed very similarly. We decided to exclude those that use the log-scale assignment because the raw-scale assignment is more intuitive and provides better discrimination in the lower range of the stressor gradient.

Exclusion of _MCI_weighted and _SQMCI_weighted

Use of a weighting factor in metric calculation had little effect on metric performance. Perhaps this was due to the fact that very few of the indicator taxa had a relatively strong relationship with the stressor gradient. It is also possibly that our measure of relationship strength was not a good indicator of the actual strength of the relationship. In favour of simplicity, we excluded all metrics that include a weighting factor in metric calculation.

Inclusion of _MCI and _SQMCI

Overall, there was only a small difference between the performance of the metrics calculated based on presence-absence data vs relative abundance data and this was not consistent across the two stressor gradients. Hence, we decided to keep MCI-style and SQMCI-style metrics.

Inclusion of decreaser and increaser metrics

While overall MCI/SQMCI-style metrics were better-performing compared to metrics that were not based on indicator values and also based on decreasers or increasers only, we did not want to exclude the latter type for the following reason. Number of sensitive taxa (_richness_decreaser) and number of tolerant taxa (_richness_increaser) or versions expressing a percentage (based on richness or abundance) compared to all taxa in the community can be more intuitive than metrics based on indicator values, particularly if one is not familiar with how indicator values had been assigned. As there was no consistency in which of the versions were the better-performing, we decided to keep them all (alike to having the three EPT metric versions) except for the metric expression the ratio of decreaser to increaser taxon richness (metric _richness_ratio_Decr_Incr) as it did not appear to be very useful.

4.1.2. Diagnostic power

Evidence from field observational data

Eight metrics selected using the single-stressor analysis were tested for their diagnostic power using multiple-stressor analyses. Analysis of the field observational data showed that among the periphyton-specific metrics, all but two (chl_SQMCI, chl_pct_abund_decreaser) responded to chlorophyll-*a* but not sediment cover. Similarly, among the sediment-specific metrics, all but one (sed_pct_abund_decreaser) responded to sediment cover but not chlorophyll-*a*, thereby providing evidence that these metrics have diagnostic power.

Also, direct comparison between the metric scores (using the dataset for metric development), showed that while there likely is an overlap of taxa that are sensitive to algal proliferation and sedimentation, MCI/SQMCI-style metric scores were less correlated. This suggests that different degrees of sensitivity to each of the stressors improve the diagnostic power of stressor-specific metrics.

Evidence from experimental data

None of the periphyton-specific metrics were influenced by nutrient levels. This was not surprising since the nutrient levels, despite their wide range, did not translate into different chlorophyll-*a* levels (Wagenhoff 2011). This result may suggest that the taxa we identified to be sensitive to algal proliferation do not respond to increased nutrient levels on their own, that is without nutrient levels translating into algal proliferation. Previous analyses of the experimental dataset had also shown that nutrient levels had only relatively weak effects on the density of common macroinvertebrate taxa or diversity indices, and these effects were predominately positive or subsidy-stress responses (Wagenhoff et al. 2012). The absence of strong negative effects on macroinvertebrate communities had been explained by nutrient levels not having translated into different levels of algal biomass but also possibly due to temporal and spatial limitations of the experiment. The short experimental period of 3 weeks may have not been long enough for nutrients to exert the full negative effects on macroinvertebrates that are typically observed in field surveys. Constant supply of macroinvertebrates and oxygenated water may have also reduced the chance for nutrients to exert strong effects on macroinvertebrate communities (Wagenhoff et al. 2012).

Two sediment-specific metrics (sed_pct_abund_increaser and sed_pct_abund_decreaser) were strongly affected by the sediment level. However, three periphyton-specific metrics (chl_pct_abund_increaser, chl_pct_abund_decreaser and chl_SQMCI) were also strongly affected by sediment level. This suggests, at first, that the stressor-specific metrics may not have diagnostic power at a site-specific level as they may simply both be affected in similar ways by the effects of sedimentation. Direct comparison between the metric scores, however, showed that while there likely is an overlap of taxa that are sensitive to algal proliferation and sedimentation in the Kauru River (from which organisms were sourced), MCI/SQMCI-style metric scores were different and should be able to discriminate between stressors. This pattern was shown by experimental and field observational datasets.

Overall, analyses of field observational and experimental data suggest that the stressor-specific metrics (or some of them) have diagnostic power and could discriminate between the effects of algal proliferation as a consequence of nutrient enrichment and the effects of sedimentation at both a site-specific level and on a broader landscape level (e.g. across a region).

4.1.3. How do the new metrics compare to existing metrics?

Primarily, we wanted to test whether stressor-specific metrics are diagnostic whereas existing metrics respond to both nutrient and sediment. Multiple-stressor analysis of field observational data confirmed our hypothesis and showed all existing metrics, except for EPT richness, were affected by both stressors whereas stressor-specific metrics predominantly responded to one stressor. This provides evidence for stressor-specific metrics offering complementary information to existing metrics. Due to increased nutrient levels not translating into increased levels of chlorophyll-*a* in the experiment, analysis of that data did not provide additional information on whether existing metrics are affected by both stressors.

Further, correlation between existing metrics scores and stressor-specific metric scores was explored to determine the degree of overlap. High correlation would suggest that there is a large degree of overlap of the indicator taxa and/or indicator values that contribute to the calculation of scores. By contrast, a weak correlation would suggest a low degree of overlap. Results from field observational data (single-stressor dataset) showed that periphyton-specific and equivalent existing metrics were relatively highly correlated. For example, the regression model of the SQMCI vs. chl_SQMCI had an R^2 of 0.82. Also, the correlation of EPT richness vs chl_richness_decreaser was relatively strong (R^2 of 0.73) suggesting overlap of indicator taxa, although there was a relatively wide spread of the number of EPT taxa for a given number of taxa sensitive to algal proliferation suggesting there were indicator taxa other than those belonging to the EPT orders. Similarly, when existing metric scores were correlated with equivalent sediment-specific metric scores, but there were also different taxa that informed the overall scores.

For experimental data, surprisingly, SQMCI and chl_SQMCI scores were highly correlated ($R^2 = 0.91$) while MCI and chl_MCI scores were only weakly correlated ($R^2 = 0.05$). The patterns were reversed for sediment-specific metrics, SQMCI and sed_SQMCI scores were not correlated while MCI and sed_MCI scores were weakly correlated ($R^2 = 0.32$). The latter result suggests that the sed_SQMCI and also, but to a less degree, the sed_MCI provide complementary information to the equivalent stream health indices. No conclusion was drawn from the surprising result for the chl_MCI and chl_SQMCI as nutrient enrichment did not translate into different levels of chlorophyll-*a*.

4.1.4. How do the metrics compare to trait metrics?

Evidence to decide which trait metrics may be useful as additional stressor-specific metrics was mainly drawn from multiple-stressor analysis on field observational data (presented in Table 12). Based on these results in combination with ecological knowledge and the results from multiple-stressor analysis on experimental data (for sediment) and single-stressor analysis on field observational data, we suggest which are likely the best stressor-specific trait metrics. Overall, more trait modalities were identified as sediment-specific than periphyton-specific and the former were also more strongly affected by the respective stressor. This may be related to algal proliferation having more varied modes of action affecting macroinvertebrates that may decrease the ability to detect relationships for individual traits, which are not independently distributed across the species (as a consequence of the trait syndrome, Menezes et al*.* 2010).

Periphyton-specific traits metrics

There were eight trait modalities, from seven traits, that were responsive to chlorophyll-*a* only, however the effect sizes (ranging from 0.12 to 0.14) were generally smaller than those for our periphyton-specific metrics (0.11 to 0.23). Among those the strongest evidence based on ecological knowledge is for 10d-ATTACHED (aquatic stage attached to substrate, negative response direction) as algal proliferation likely decreases opportunities for attachment. Body form modalities 12b-FLATTENED (dorso-ventrally or laterally flattened, -) and 12c-CYLINDRICAL (+) were also responsive in the single-stressor analysis but it is less obvious why the cylindrical form would be favoured over a flattened form with increasing algal proliferation. Experimental data was not useful to test any periphyton-specific metrics as increased nutrient levels did not translate into increased levels of chlorophyll-*a*.

Sediment-specific trait metrics

There were 16 trait modalities, from ten traits, that were responsive to sediment only, and effect sizes (0.12 to 0.29) were generally larger than those for our sediment-specific metrics (0.10 to 0.22). Among those, the strongest evidence based on ecological knowledge is for the following trait modalities:

- 13b-SCRAPER (-) as fine substrate likely decreases growth or quality of suitable periphyton types as food sources (also confirmed by experimental data)
- 10b-CRAWLERS (-) as unstable substrate is less suitable for invertebrates that crawl
- 2c-DESC3 (maximum number of descendants per reproductive cycle equal is larger than 1000 to 3000) as a larger number of descendants increases the chance of survival
- 3b-UNIV (univoltine, -) as a small number of reproductive cycles decreases the chance of survival (also confirmed by experimental data and by single-stressor analysis) and consequently 3a-SEMI (maximum number of reproductive cycles per year, semivoline, +, also confirmed by experimental data) and 3c-PLURIV (plurivoltine, +, also confirmed by single-stressor analysis)
- 7b-submerged (oviposition site submerged, +) as fine sediment may cover those eggs that were deposited at the water surface (also confirmed by experimental data and by single-stressor analysis).

Modalities 1b-SIZE2 (maximum potential size > 5-10mm, -) and 9a-DISSLOW (low dissemination potential, +) have also been confirmed by experimental data, and body form 12d-SPHERICAL (+) confirmed by single-stressor analysis but the ecological reasons are not clear.

4.2. Limitations of the stressor-specific metrics

We consider the largest limitation of the current stressor-specific metrics to be the relatively small number of 49 indicator taxa, out of which there were 32 indicator taxa for periphyton-specific metrics (26 decreasers, 6 increasers) and 37 indicator taxa for sediment-specific metrics (31 decreasers, 6 increasers). This is a considerably smaller number of indicator taxa than the 182 taxa that were given tolerance values to calculate the MCI, the well-performing and most widely-used macroinvertebrate metric in New Zealand.

The small number of indicator taxa is most likely a consequence of failing to detect the sensitivity (or possibly tolerance) of taxa, rather than the true absence of a response of these taxa to algal proliferation or sedimentation. Lack of a clear negative (or positive) response shape was the dominant reason for not including the taxon as an indicator taxon and this is likely due to the often low occurrences of many taxa across the dataset. We set ten occurrences across the dataset as a minimum requirement for taxa to enter the analysis. Given the large amount of variation in the spatial scale dataset, the chance of failing to detect the sensitivity is increased. The range of occurrences of the taxa among our set of indicator taxa and non-indicator taxa is very similar (10 to ~440, out of a total of 501 samples). However, half of the taxa in our non-indicator taxon list have less than 41 occurrences while half of the taxa in our indicator taxon list have less than 77 occurrences, suggesting that misclassification may have occurred.

We also stayed on the side of caution and that led to the rejection of several taxa as indicator taxa when there was mixed agreement among experts and the response shapes. Together these two factors decrease the likelihood of misclassifying taxa as indicator taxa that are in fact non-responsive, but increase the likelihood of misclassifying taxa as non-indicator taxa when they are in fact responsive to the stressor gradient. Finally, out of the 131 taxa that occurred at least once in our dataset, there were 44 taxa that did not enter the analysis because they occurred fewer than ten times. These taxa and also those that were completely absent from our dataset are likely generally rare across New Zealand and associated with specific stream types. Knowledge of their sensitivity and contribution to stressor-specific metric scores may not greatly improve diagnosis of stressor effects on a regional scale, but they may be important taxa for diagnosis of stressor effects at individual sites or for the stream types in which they occur. We conclude that there likely are more taxa that could be added to the indicator taxon list in the future that may improve the ability to detect and distinguish between the effects of sedimentation and algal proliferation.

While a larger indicator taxon list is likely to improve the diagnostic power of periphyton and sediment-specific metrics, periphyton-specific metrics will always, to a certain degree, respond to increasing sedimentation and sediment-specific metrics respond to increasing periphyton cover. This is due to considerable overlap of indicator taxa with similar response direction to both stressor gradients. Overlap could be a result of some of the mechanisms underlying the stressor effects being similar (e.g. both sediment and algae smother cobble habitat that species would have been adapted to). Furthermore, these metrics are also likely influenced by other stressors, in particular changes to flow, and also by natural environmental variables. The implications of these limitations are further discussed in Section 5 where we provide recommendations for application of the metrics.

5. RECOMMENDATIONS

Stressor-specific metrics and calculation of site scores

The stressor-specific metrics along with formulas recommended for use in resource management are presented in Table 13. None of the candidate metrics was consistently the best-performing metric across the different validation analyses, although stressor-specific MCI and SQMCI appeared to the most reliable in detecting and disentangling the effects of sedimentation and nutrient enrichment/eutrophication. Hence, we recommend calculating all eight types of metrics for each stressor. Because there are many indicator taxa that are sensitive to both stressors and consequently all metrics are, to a certain degree, responsive to nutrient enrichment and sedimentation, periphyton-specific and sediment-specific metrics are meant to be used together.

Metric site scores are calculated from a list of macroinvertebrate taxa collected at a stream site and identified to the level required for calculation of the existing metric MCI (Stark & Maxted 2007a, see Appendix 4 for updated list of taxa names). Decreaser and increaser taxa as well as indicator values for these taxa are presented in Table 14. Standard protocols for semi-quantitative sample collection and sample processing involving assignment of coded abundances (Stark et al. 2001) should be used, although the scores can also be calculated from quantitative samples and other enumeration methods. Number of sensitive/tolerant taxa, % Sensitive/tolerant taxa and stressor-specific MCI metrics, however, only require presence-absence of the taxa while coded abundances are required for calculation of % Sensitive/tolerant individuals and stressor-specific SQMCI metrics.

Table 13. Stressor-specific metrics for each of the two stressors, nutrient enrichment/algal proliferation and sedimentation. Note that the code of the metrics is preceded by either 'sed' or 'chl' to specify which stressor it is developed for.

Code	Name	Description
_MCI	Sediment or Periphyton MCI	_MCI = $\frac{\sum_{i=1}^{i=S} a_i}{S}$ x 20 where S = the total number of scoring taxa in the sample, and a_i = the indicator value for the ith taxon. The scaling factor of 20 has been added to distinguish between MCI site scores and site scores of the _SQMCI.
_SQMCI	Sediment or Periphyton SQMCI	_SQMCI = $\sum_{i=1}^{i=S} \frac{(n_i \times a_i)}{N}$ where S = the total number of scoring taxa in the sample, n_i = the coded abundance for the i^{th} scoring taxon, a_i = the indicator value for the i^{th} taxon and N = the total of the coded abundances of the scoring taxa for the entire sample. The coded abundance is the lower boundary of the respective abundance classes assigned to each taxon based on the counts: R (1-4), C (5-19), A (20-99), VA (100-499) and VVA (500+).
_richness_decreaser	Number of sensitive taxa	Number of decreaser taxa
_pct_richness_decreaser	% Sensitive taxa	Number of decreaser taxa divided by total number of taxa
_pct_abund_decreaser	% Sensitive individuals	Number of decreaser individuals divided by total number of individuals
_richness_increaser	Number of tolerant taxa	Number of increaser taxa
_pct_richness_increaser	% Tolerant taxa	Number of increaser taxa divided by total number of taxa
_pct_abund_increaser	% Tolerant individuals	Number of increaser individuals divided by total number of individuals

Table 14. Indicator taxa, either sensitive taxa (decreaser,−) or tolerant taxa (increaser, +), along with indicator values (IndVal) for calculation of periphyton-specific and sediment-specific metrics.

	Periphyton-specific metrics		**Sediment-specific metrics**	
Taxon	**Response**	**IndVal**	**Response**	**IndVal**
Ephemeroptera				
Acanthophlebia	−	10	−	10
Ameletopsis			−	10
Austroclima	−	10	−	10
Coloburiscus			−	10
Deleatidium	−	9	−	6
Ichthybotus	−	10		
Nesameletus	−	7	−	10
Zephlebia	−	7	−	10
Plecoptera				
Austroperla	−	10	−	10
Megaleptoperla			−	9
Stenoperla			−	10
Zelandobius	−	9		
Zelandoperla	−	10	−	10
Trichoptera				
Beraeoptera	−	10	−	10
Confluens			−	10
Costachorema			−	10
Helicopsyche	−	7	−	10
Hydrobiosis			−	6
Aoteapsyche	−	8	−	9
Orthopsyche	−	10		
Neurochorema			−	10
Olinga			−	9
Oxyethira	+	4		
Psilochorema			−	9
Pycnocentria	−	6	−	9
Pycnocentrodes	−	6	−	9
Triplectides	−	10		
Coleoptera				
Elmidae			−	6
Hydraenidae	−	10	−	10
Ptilodactylidae			−	9
Scirtidae	−	10	−	10
Diptera				
Aphrophila	+	5	−	10
Austrosimulium	−	8	+	5
Corynoneura			+	1
Empididae	+	5		

	Periphyton-specific metrics		**Sediment-specific metrics**	
Taxon	**Response**	**IndVal**	**Response**	**IndVal**
Lobodiamesa			−	10
Molophilus	−	10		
Orthocladiinae	+	5		
Paralimnophila	−	10		
Stictocladius	−	10	−	9
Megaloptera				
Archichauliodes			−	10
Mollusca				
Gyraulus	+	1	+	1
Latia	−	9	−	10
Potamopyrgus	−	6		
Sphaeriidae			+	1
Crustacea				
Paraleptamphopus	−	7		
Other				
Acarina	−	9		
Nematoda			+	1
Oligochaeta	+	5	+	5

Application of the metrics

During a recent workshop[4] hosted by Ministry for the Environment, it became clear that the primary need from a regional council perspective was for tools to help assess policy effectiveness (Clapcott & Wagenhoff 2018). The current MfE-funded macroinvertebrate research helps with assessing the effectiveness of regional plans to implement the NPS-FM, particularly in regard to the compulsory freshwater value of Ecosystem Health. New stressor-specific metrics will provide diagnostic tools to help identify the primary stressors responsible for degraded MCI scores in a multiple stressor environment. These diagnostic tools can be useful either for use at individual sites or broadly across the region. The former can inform what restoration measures or improvements at the site/catchment are likely to be most effective in improving ecosystem health. The latter, on the other hand, provides information across streams in the region that can inform and test policy effectiveness.

At the workshop, ideas to use stressor-specific metrics to hindcast ecological effects of nutrient enrichment and sedimentation using long-term invertebrate datasets were suggested as a means of ‘story-telling’ or relating effects to significant disturbance events or restoration efforts (Clapcott & Wagenhoff 2018). Furthermore, it was also

---

[4] The workshop was aimed at sharing recent and current macroinvertebrate project outputs, at getting feedback from regional councils on their primary needs for macroinvertebrate tools, and at identifying research gaps. Workshop participants included research providers, staff from several regional councils, Department of Conservation, and MfE.

suggested to identify impact and cessation thresholds of these metrics across stressor gradients (e.g. Wagenhoff et al. 2017c) to inform definition of regional objectives for deposited sediment and periphyton/nutrients (‘attributes’). Further support with respect to these specific ideas for application of the stressor-specific metrics was identified as an avenue for future research.

More specifically, the stressor-specific metrics are now ready to be applied by regional councils in case studies aimed at answering questions to inform resource management. For example, these metrics may be applied on a site-specific scale to:

1. track the likely effects of these two stressors at monitoring sites over time
2. detect early signs of ecological degradation
3. diagnose whether nutrient enrichment or sedimentation or both are the likely leading causes of degraded stream ecosystem health
4. estimate the ecological effects of a significant disturbance event, such as a large amount of sedimentation due to dam failure, liquefaction or potentially clear-cut forest harvest
5. measure restoration success.

These metrics may also be applied on a regional scale to:

1. track the likely effects of these two stressors across the region
2. diagnose whether nutrient enrichment or sedimentation or both are the likely leading causes of degraded stream ecosystem health in the region
3. test which land use practices have the largest effects on stream ecosystem health, and why
4. test policy effectiveness on stream ecosystem health, specifically with respect to measures that control or reduce the inputs of nutrients and fine sediments
5. define thresholds for nutrient concentrations, periphyton and sediment aimed at protection of ecosystem health, or thresholds for the metrics themselves to use within an ecosystem health assessment framework.

It is important to note these metrics, if used for diagnostic purposes, have to be used in conjunction as sediment-specific and periphyton-specific metrics are also influenced by increasing chlorophyll-*a* and sediment cover, respectively, due to considerable overlap of indicator taxa with similar response direction to both stressor gradients. Hence, for example, a declining trend of the Sediment MCI at a site is not necessarily indicative of increasing sedimentation but sedimentation is more likely the reason for impairment if the Periphyton MCI does not show such a strong declining trend. Also, the absolute metric scores should not be compared among sediment-specific and periphyton-specific metrics. Hence, in order to compare the magnitude of impairment attributable to each stressor, in most cases it will be necessary to determine what the reference metric score should be and quantify impact in relation to reference.

Furthermore, these metrics are also likely to be affected by other stressors and in particular to changes in flow. For example, water abstraction increases sediment deposition and algal proliferation. In a separate report, for the same overarching project, the New Zealand-specific Lotic-invertebrate Index for Flow Evaluation LIFENZ has been tested and confirmed to be a hydrologically-sensitive macroinvertebrate metric (Greenwood 2018). Use of the sediment-specific and periphyton-specific metrics in conjunction with the LIFENZ may help in disentangling the effects of sedimentation, nutrient enrichment and flow reduction. However, this has not been tested to date.

Finally, these metrics, as all macroinvertebrate metrics, are influenced by natural environmental gradients which is a result of the species being adapted to different environmental conditions. Hence, metric scores observed at reference sites will differ according to the specific natural environmental conditions (often called ‘stream type’) and equally, metric scores at impact sites cannot necessarily be compared among sites. The influence of natural environmental gradients will introduce variability in regional analyses where a time-for-space approach is being used but broad patterns may still be detectable. However, in order to address the questions listed above, and as already mentioned, in most cases it will be necessary to determine what the reference metric score should be and quantify impact in relation to reference.
Overall, we envision that application, validation and further improvement of the metrics are undertaken. We recommend that regional councils apply these metrics in case studies to answer question to inform resource management (examples given above) and develop frameworks of how to use these metrics as diagnostic tools or in ecosystem health assessment. Validation may involve testing whether metrics work for different stream types, in particular those that did not contribute to metric development. Finally, there is also scope for further improving the metrics at the same time (outlined below).

Further improvements of the metrics

Scope for improvement of the metrics is predominately via identifying further indicator taxa. We recommend the use of new spatial scale datasets that have become available post metric development. Ideally, the dataset (1) covers a wide range of stream types across the country, (2) covers wide stressor gradients with data points spread well across both focal gradients, (3) provides good estimates of prevailing deposited sediment levels and nutrient/periphyton biomass levels at a site (e.g. calculated from monthly samples vs. use of only a single estimate and due to use of standardised sampling protocols; this has mainly been a problem with sediment estimates in this project), and (4) consists of many data points to increase the number of taxa occurrences across the dataset of generally rare but likely sensitive species; increased occurrences improve model predictions. The difficulty of reliably identifying indicator taxa for periphyton-specific metrics is likely predominantly related to the multiple pathways via which algal proliferation can affect taxa. Possibly linking an upper percentile of monthly chlorophyll-*a* samples to each macroinvertebrate sample rather than an annual median may improve the likelihood of identifying indicator taxa.

We recommend considering alternative approaches in addition to our multiple-stressor model approach to identify indicator taxa and determine sensitivities, in particular for rare taxa. These may include single-stressor analyses on a spatial dataset (for which more data points would be available) and use of laboratory or mesocosm experiments.

During the workshop at MfE, the point was also raised whether there may be any value in looking at nutrient effects NOT via algal proliferation. In our analyses we had included DIN and DRP as predictors in addition to chlorophyll-*a*, hence potentially we could identify indicator taxa for their more direct effects. However, in a workshop held as part of the 2017 macroinvertebrate project, experts were not convinced that the response shapes were making much ecological sense. It may be however worth investigating effects of nutrients on macroinvertebrates via changes to the periphyton species community and food quality.

Finally, trait metrics appear to be useful to add to the list of candidate stressor-specific metrics. The recently updated macroinvertebrate trait database (Phillips & Smith 2018) is publicly available on the web, however calculation of the trait metrics is not as simple as calculation of the stressor-specific metrics. This is mainly due to trait information being presented mainly at the species level while macroinvertebrate taxa often are identified at the genus or even higher taxonomic levels. Instructions for how to assign trait affinity scores to higher taxonomic levels or provision of nationally-applicable affinity scores for higher levels in addition to instructions for how to calculate trait metrics would increase uptake of traits in resource management. In addition, it may be useful to identify new traits that are specifically tailored to detecting the effects of either sedimentation or nutrient enrichment, which would increase the strength of the stressor-response relationship and the diagnostic power of the traits.

6. ACKNOWLEDGEMENTS

We gratefully acknowledge data provided by regional councils via LAWA (www.lawa.org.nz) and the parallel MfE-funded sediment report (Depree et al. 2018). Also, various researchers approved access to their published and unpublished research datasets collected in partnership with project authors; our thanks to Christoph Matthaei, John Quinn, Kevin Collier, Richard Storey, Chris Hickey, Phillip Jellyman, Jay Piggott, Francis Magbanua, Elizabeth Graham, Francis Burdon, Robin Holmes, Katie Blackmore, Javad Ramezani, Katharina Lange, Romana Salis, Jan Macher, David Reid, Rebecca Eivers, Cale Riddle, Jason Eden, Aslan Wright-Stow and Colin Townsend.

Much progress of stressor-specific metric development had been achieved during the preceding MfE macroinvertebrate project and we thank scientists for having provided constructive feedback: Christoph Matthaei, John Quinn, Richard Storey, Kevin Collier, Roger Young, Martin Neale and Russell Death, and in particular Joanne Clapcott who also managed the preceding and current projects and made sure invertebrate tool development is made a priority. Hayden Rabel is thanked for statistical expertise in earlier metric development stages. We also thank Ministry for the Environment for their financial support of the preceding as well as the current project. We thank John Stark for providing an updated table of tolerance values for MCI calculation as well as for valuable discussions on metric development.

Finally, we thank several regional council staff as well as Carl Howarth and Isaac Bain from MfE who participated in a recent macroinvertebrate workshop (22 June 2018, hosted at MfE in Wellington) and put forward ideas on how these stressor-specific metrics could be used in future resource management and policy development. Joanne Clapcott organised and facilitated this workshop which was aimed at sharing recent and current project outputs and at identifying research gaps.

7. REFERENCES

APHA 2005. Standard methods for the examination of water and wastewater. Washington, DC, 21st ed., American Public Health Association.

Biggs BJF, Kilroy C 2000. Stream periphyton monitoring manual. Prepared for the New Zealand Ministry for the Environment. 226 p.

Booker D, Whitehead AL, Henderson RD, Yang J 2016. Nationally consistent methods for calculating water allocation indicators. Prepared for Ministry for the Environment. NIWA Client Report No. 2016023CH. National Institute of Water & Atmospheric Research Ltd, Christchurch, New Zealand. 45 p. plus appendices.

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

Booker DJ, Woods RA 2014. Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments. Journal of Hydrology 508(0): 227-239.

Bryce SA, Lomnicky GA, Kaufmann PR 2010. Protecting sediment-sensitive aquatic species in mountain streams through the application of biologically based streambed sediment criteria. Journal of the North American Benthological Society 29(2): 657-672. doi: 10.1899/09-061.1.

Burdon FJ, McIntosh AR, Harding JS 2013. Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. Ecological Applications 23(5): 1036-1047. doi:10.1890/12-1190.1.

Chevan A, Sutherland M 1991. Hierarchical partitioning. The American Statistician 45(2): 90-96.

Clapcott JE, Goodwin EO, Snelder TH, Collier KJ, Neale MW, Greenfield S 2016. Finding reference: a comparison of modelling approaches for predicting macroinvertebrate community index benchmarks. New Zealand Journal of Marine and Freshwater Research: 1-16.

Clapcott J, Wagenhoff A 2018. Summary of a workshop on macroinvertebrate metrics. Prepared for Ministry for the Environment. Cawthron Report No. 3204. 11 p. plus appendices.

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

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 2017b. Macroinvertebrate metrics for the National Policy Statement for Freshwater Management. Prepared for Ministry for the Environment. Cawthron Report No. 3073. Cawthron Institute, Nelson, New Zealand. 139 p. plus appendices.

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

Collier KJ, Smith BJ 2005. Effects of progressive catchment harvesting on stream invertebrates in two contrasting regions of New Zealand's North Island. Marine and Freshwater Research 56(1): 57-68. doi:10.1071/mf04105.

Cutler DR, Edwards TC, Jr., Beard KH, Cutler A, Hess KT 2007. Random forests for classification in ecology. Ecology 88(11): 2783-2792. doi:10.1890/07-0539.1.

Depree C, Clapcott J, Booker D, Franklin P, Hickey C, Wagenhoff A, Matheson F, Shelley J, Unwin M, Wadhwa S, Goodwin E, Mackman J, Rabel H 2018. Development of ecosystem health bottom-line thresholds for suspended and deposited sediment in New Zealand rivers and streams. Prepared for Ministry for the Environment. NIWA Client Report No. 2017076HN. NIWA, Hamilton, New Zealand. 190 p. plus appendices.

Eivers R 2006. The response of stream ecosystems to riparian buffer width and vegetative composition in exotic plantation forests. MSc thesis, University of Canterbury, Christchurch.

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

Extence CA, Balbi DM, Chadd RP 1999. River flow indexing using British benthic macroinvertebrates: A framework for setting hydroecological objectives. Regulated Rivers-Research and Management 15(6): 543-574.

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

Gray DP, Harding JS 2012. Acid Mine Drainage Index (AMDI): a benthic invertebrate biotic index for assessing coal mining impacts in New Zealand streams. New Zealand Journal of Marine and Freshwater Research 46(3): 335-352.

Greenwood MJ 2018. Testing a potential flow-specific invertebrate metric: LIFENZ. Prepared for Joanne Clapcott, Cawthron Institute, as part of MfE project: water quality and flows: states and trends. NIWA Client Report No. 2018263CH. National Institute of Water & Atmospheric Research Ltd, Christchurch, New Zealand. 42 p. plus appendices.

Greenwood MJ, Booker DJ, Smith BJ, Winterbourn MJ 2016. A hydrologically sensitive invertebrate community index for New Zealand rivers. Ecological Indicators 61: 1000-1010.

Haase R, Nolte U 2008. The invertebrate species index (ISI) for streams in southeast Queensland, Australia. Ecological Indicators 8(5): 599-613.

Harding JS, Jellyman PG 2015. Earthquakes, catastrophic sediment additions and the response of urban stream communities. New Zealand Journal of Marine and Freshwater Research 49(3): 346-355.

Holmes R 2008. Cumulative effects of multiple agricultural stressors on invertebrates in pasture streams. MSc thesis, University of Otago, Dunedin.

Hubler S, Huff DD, Edwards P, Pan Y 2016. The Biological Sediment Tolerance Index: Assessing fine sediments conditions in Oregon streams using macroinvertebrates. Ecological Indicators 67: 132-145.

Lange K, Townsend CR, Matthaei CD 2014. Can biological traits of stream invertebrates help disentangle the effects of multiple stressors in an agricultural catchment? Freshwater Biology 59(12): 2431-2446.

Leathwick JR, Snelder T, Chadderton WL, Elith J, Julian K, Ferrier S 2011. Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications. Freshwater Biology 56(1): 21-38.

Mac Nally R 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between and reconciliation of 'predictive' and 'explanatory' models. Biodiversity and Conservation 9(5): 655-671.

Macher JN, Salis RK, Blakemore KS, Tollrian R, Matthaei CD, Leese F 2016. Multiple-stressor effects on stream invertebrates: DNA barcoding reveals contrasting responses of cryptic mayfly species. Ecological Indicators 61: 159-169.

Magbanua FS, Townsend CR, Blackwell GL, Phillips N, Matthaei CD 2010. Responses of stream macroinvertebrates and ecosystem function to conventional, integrated and organic farming. Journal of Applied Ecology 47(5): 1014-1025.

Matthaei CD, Weller F, Kelly DW, Townsend CR 2006. Impacts of fine sediment addition to tussock, pasture, dairy and deer farming streams in New Zealand. Freshwater Biology 51(11): 2154-2172.

Menezes S, Baird DJ, Soares AMVM 2010. Beyond taxonomy: a review of macroinvertebrate trait-based community descriptors as tools for freshwater biomonitoring. Journal of Applied Ecology 47(4): 711 - 719.

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

Murphy JF, Jones JI, Pretty JL, Duerdoth CP, Hawczak A, Arnold A, Blackburn JH, Naden PS, Old G, Sear DA, Hornby D, Clarke RT, Collins AL 2015. Development of a biotic index using stream macroinvertebrates to assess stress from deposited fine sediment. Freshwater Biology 60(10): 2019-2036.

Phillips N 2004. Stream biomonitoring using species traits. Water & Atmosphere 12(4): 8-9.

Phillips N, Reid D 2012a. Biological trait analysis: application to the Waikato Regional Council Monitoring Programme Document No: 2113749.

Phillips N, Reid D 2012b. Biological traits: Application to the Auckland Council River Ecology Monitoring Programme. Technical Report 2012/001.

Phillips N, Smith B 2018. New Zealand Freshwater Macroinvertebrate Trait Database. Report prepared for NIWA. National Institute of Water & Atmospheric Research Ltd, Hamilton. 25 p.

Quinn JM, Hickey CW 1990. Magnitude of effects of substrate particle size, recent flooding, and catchment development on benthic invertebrates in 88 New Zealand rivers. New Zealand Journal of Marine and Freshwater Research 24(3): 411-427.

R Core Team 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reid DJ, Quinn JM, Wright-Stow AE 2010. Responses of stream macroinvertebrate communities to progressive forest harvesting: Influences of harvest intensity, stream size and riparian buffers. Forest Ecology and Management 260(10): 1804-1815.

Relyea CD, Minshall GW, Danehy RJ 2012. Development and validation of an aquatic Fine Sediment Biotic Index. Environmental Management 49(1): 242-252.

Smith AJ, Bode RW, Kleppel GS 2007. A nutrient biotic index (NBI) for use with benthic macroinvertebrate communities. Ecological Indicators 7(2): 371-386.

Snelder TH, Biggs BJF 2002. Multiscale River Environment Classification for water resources management. Journal of the American Water Resources Association 38(5): 1225-1239.

Stark J, Boothroyd I, Harding J, Maxted J, Scarsbrook M 2001. Protocols for sampling macroinvertebrates in wadeable streams. New Zealand Macroinvertebrate

Working Group Report No. 1. Prepared for the Ministry for the Environment, Sustainable Management Fund Project No. 5103. 57 p.

Stark JD 1985. A macroinvertebrate community index of water quality for stony streams Water & Soil Miscellaneous Publication 87: 1-52.

Stark JD 1998. SQMCI: A biotic index for freshwater macroinvertebrate coded-abundance data. New Zealand Journal of Marine and Freshwater Research 32(1): 55-66.

Stark JD, Maxted JR 2007a. A user guide for the Macroinvertebrate Community Index. Prepared for the Ministry for the Environment. Cawthron Report No.1166. 58 p.

Stark JD, Maxted JR 2007b. A biotic index for New Zealand's soft-bottomed streams. New Zealand Journal of Marine and Freshwater Research 41(1): 43-61.

Storey R, Parkyn S, Smith B, Croker G, Franklin P 2009. Effects of development on zero-order streams in the Waikato Region. Environment Waikato Technical Report 2009/22.

Townsend CR, Uhlmann SS, Matthaei CD 2008. Individual and combined responses of stream ecosystems to multiple stressors. Journal of Applied Ecology 45(6): 1810-1819.

Wagenhoff A 2011. Multiple-stressor effects along gradients of deposited fine sediment and dissolved nutrients in streams. Unpublished PhD thesis, University of Otago, Dunedin.

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

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

Wagenhoff A, Clapcott JE, Lau KEM, Lewis GD, Young RG 2017a. Identifying congruence in stream assemblage thresholds in response to nutrient and sediment gradients for limit setting. Ecological Applications 27(2): 469–484.

Wagenhoff A, Liess A, Pastor A, Clapcott JE, Goodwin EO, Young RG 2017b. Thresholds in ecosystem structural and functional responses to agricultural stressors can inform limit setting in streams. Freshwater Science 36(1): 178–194.

Wagenhoff A, Liess A, Pastor A, Clapcott JE, Goodwin EO, Young RG 2017c. Thresholds in ecosystem structural and functional responses to agricultural stressors can inform limit setting in streams. Freshwater Science 36(1): 178-194

Winterbourn MJ, Gregson KLD, Dolphin CH 2006. Guide to the aquatic insects of New Zealand. 4 ed, Entomological Society of New Zealand. 108 p.

Zuur AF, Ieno EN, Elphick CS 2010. A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1(1): 3-14.

Zweig LD, Rabeni CF 2001. Biomonitoring for deposited sediment using benthic invertebrates: A test on 4 Missouri streams. Journal of the North American Benthological Society 20(4): 643-657.

8. APPENDICES

Appendix 1. Tolerance values for MCI and SQMCI metric calculation, developed for hard-bottom streams (HB) and soft-bottom streams (SB). In this report, only those for hard-bottom streams were used. The table is an updated version of the table provided in Stark & Maxted (2007a). Minor updates have been provided by John Stark (personal communication, January 2017).

Taxon	HB	SB
COELENTERATA		
Hydra	3	1.6
PLATYHELMINTHES	3	0.9
RHABDOCOELA	3	0.9
BRYOZOA	-	4.0
NEMATODA	3	3.1
NEMATOMORPHA	3	4.3
NEMERTEA	3	1.8
OLIGOCHAETA	1	3.8
POLYCHAETA	-	6.7
HIRUDINEA	3	1.2
TARDIGRADA	-	4.5
CRUSTACEA		
Amphipoda	5	5.5
Cladocera	5	0.7
Copepoda	5	2.4
Halicarcinus	-	5.1
Helice	-	6.6
Isopoda	5	4.5
Mysidae	-	6.4
Ostracoda	3	1.9
Paracalliope	5	-
Paraleptamphopus	5	-
Paranephrops	5	8.4
Paranthura	-	4.9
Paratya	5	3.6
Tanaidacea	4	6.8
INSECTA		
Ephemeroptera		
Acanthophlebia	7	9.6
Ameletopsis	10	10.0
Arachnocolus	8	8.1
Atalophlebioides	9	4.4
Austroclima	9	6.5
Austronella	7	4.7
Coloburiscus	9	8.1
Deleatidium	8	5.6
Ichthybotus	8	9.2
Isothraulus	8	7.1
Mauiulus	5	4.1
Neozephlebia	7	7.6
Nesameletus	9	8.6
Oniscigaster	10	5.1
Rallidens	9	3.9
Odonata		
Aeshna	5	1.4
Anisoptera	5	6.0
Antipodochlora	6	6.3
Austrolestes	6	0.7
Hemianax	-	1.1
Hemicordulia	5	0.4
Ischnura	-	3.1
Procordulia	6	3.8
Uropetala	5	0.4
Xanthocnemis	5	1.2
Hemiptera		
Anisops	5	2.2
Diaprepocoris	5	4.7
Microvelia	5	4.6
Saldidae	5	3.9
Sigara	5	2.4
Coleoptera		
Antiporus	5	3.5
Berosus	5	-
Copelatus	5	3.7
Dytiscidae	5	0.4
Elmidae	6	7.2
Enochrus	5	2.6
Hydraenidae	8	6.7
Hydrophilidae	5	8.0
Liodessus	5	4.9
Onychohydrus	5	-
Podaena	8	-
Ptilodactylidae	8	7.1
Rhantus	5	1.0
Scirtidae	8	6.4
Staphylinidae	5	6.2
Neuroptera		
Kempynus	5	-
Diptera		
Aphrophila	5	5.6
Austrosimulium	3	3.9
Calopsectra	4	-
Ceratopogonidae	3	6.2
Chironomidae	2	3.8
Chironomus	1	3.4
Corynoneura	2	1.7
Cryptochironomus	3	-
Neolimnia	3	5.1
Nothodixa	4	9.3
Orthocladiinae	2	3.2
Paradixa	4	8.5
Paralimnophila	6	7.4
Parochlus	8	-
Paucispinigera	6	7.7
Pelecorhynchidae	9	-
Peritheates	7	-
Podonominae	8	6.4
Polypedilum	3	8.0
Psychodidae	1	6.1
Scatella	7	-
Sciomyzidae	3	3.0
Stictocladius	8	-
Stratiomyidae	5	4.2
Syrphidae	1	1.6
Tabanidae	3	6.8
Tanypodinae	5	6.5
Tanytarsini	3	4.5
Tanytarsus	3	-
Thaumaleidae	9	8.8
Tipulidae	5	3.4
Zelandotipula	6	3.6
Trichoptera		
Alloecentrella	9	-
Beraeoptera	8	7.0
Confluens	5	7.2
Conuxia	8	-
Costachorema	7	7.2
Cryptobiosella	9	-
Diplectrona	9	-
Ecnomidae	8	-
Ecnomina	8	9.6
Edpercivalia	9	6.3
Helicopsyche	10	8.6
Hudsonema	6	6.5
Hydrobiosella	9	7.6
Hydrobiosis	5	6.7
Hydrochorema	9	-
Hydropsyche - Aoteapsyche	4	6.0
Hydropsyche - Orthopsyche	9	7.5
Kokiria	9	-
Neurochorema	6	6.0

Tolerance values table, continued

Taxon	**HB**	**SB**
Siphlaenigma	9	-
Tepakia	8	7.6
Zephlebia	7	8.8
Plecoptera		
Acroperla	5	5.1
Austroperla	9	8.4
Cristaperla	8	-
Halticoperla	8	-
Megaleptoperla	9	7.3
Nesoperla	5	5.7
Spaniocerca	8	8.8
Spaniocercoides	8	-
Stenoperla	10	9.1
Taraperla	7	8.3
Zelandobius	5	7.4
Zelandoperla	10	8.9
Megaloptera		
Archichauliodes	7	7.3

Taxon	**HB**	**SB**
Lepidoptera		
Hygraula	4	1.3
Collembola	6	5.3
ACARINA	5	5.2
ARACHNIDA		
Dolomedes	5	6.2
MOLLUSCA		
Ampullariidae	3	1.6
Glyptophysa = Physastra	5	0.3
Gundlachia = Ferrissia	3	2.4
Gyraulus	3	1.7
Hyridella = Echyridella	3	6.7
Latia	3	6.1
Lymnaeidae	3	1.2
Melanopsis = Zemelanopsis	3	1.9
Physa = Physella	3	0.1
Potamopyrgus	4	2.1
Sphaeriidae	3	2.9

Taxon	**HB**	**SB**
Culex	3	-
Culicidae	3	1.2
Diptera		
Diptera indet.	3	2.9
Dixidae	4	7.1
Dolichopodidae	3	8.6
Empididae	3	5.4
Ephydridae	4	1.4
Eriopterini	9	7.5
Harrisius	6	4.7
Hexatomini	5	6.7
Limnophora	3	4.5
Limonia	6	6.3
Lobodiamesa	5	7.7
Maoridiamesa	3	4.9
Mischoderus	4	5.9
Molophilus	5	6.3
Muscidae	3	1.6
Nannochorista	7	-
Neocurupira	7	-

Taxon	**HB**	**SB**
Oecetis	6	6.8
Oeconesidae	9	6.4
Olinga	9	7.9
Trichoptera		
Oxyethira	2	1.2
Paroxyethira	2	3.7
Philorheithrus	8	5.3
Plectrocnemia	8	6.6
Polyplectropus	8	8.1
Psilochorema	8	7.8
Pycnocentrella	9	-
Pycnocentria	7	6.8
Pycnocentrodes	5	3.8
Rakiura	10	-
Synchorema	9	-
Tiphobiosis	6	9.3
Triplectides	5	5.7
Triplectidina	5	-
Zelandoptila	8	7.0
Zelolessica	10	6.5

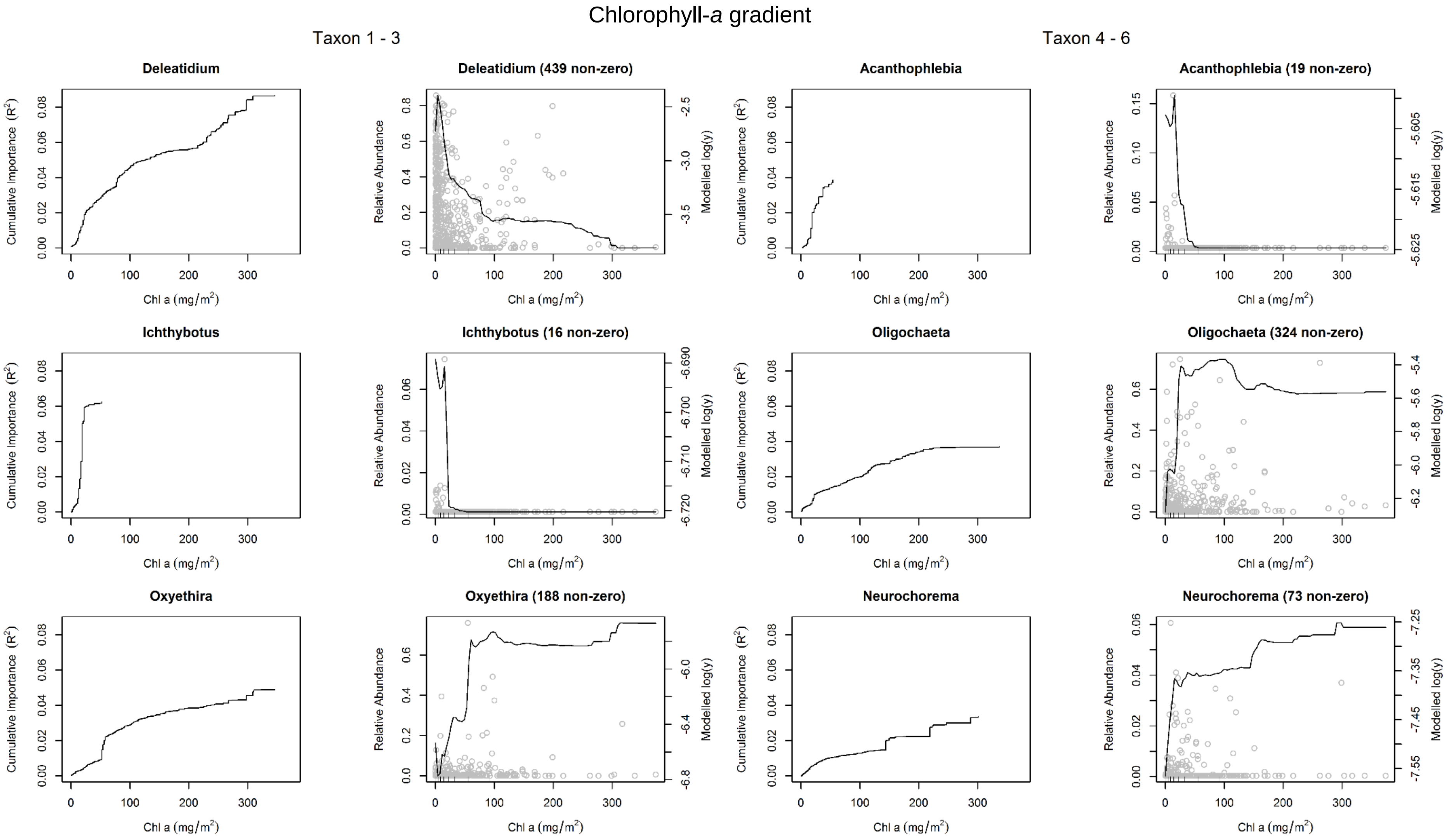
Appendix 2. Indicator taxa list, ordered alphabetically, showing output of analyses used to develop tolerance scores for sediment and periphyton biomass. Output includes response group (decreaser, increaser) identified for stressors chlorophyll-*a* (CHLA) and % sediment cover (SEDCOVER), total turnover (R^2_{fp}), the threshold calculated from the species turnover functions (at 25% of total turnover), number of occurrences in the dataset (N = 501) and the random forest model R^2.

Taxon	**Stressor**	**Response group**	R^2_{fp}	**Threshold**	**No. of occurrences**	R^2
Acanthophlebia	CHLA	decreaser	0.039	15	19	0.76
Acarina	CHLA	decreaser	0.006	28	77	0.08
Aoteapsyche	CHLA	decreaser	0.018	39	307	0.22
Aphrophila	CHLA	increaser	0.028	36	240	0.39
Austroclima	CHLA	decreaser	0.029	7	165	0.53
Austroperla	CHLA	decreaser	0.015	4	68	0.58
Austrosimulium	CHLA	decreaser	0.003	43	204	0.23
Beraeoptera	CHLA	decreaser	0.027	2	48	0.55
Deleatidium	CHLA	decreaser	0.087	28	439	0.57
Empididae	CHLA	increaser	0.029	28	93	0.32
Gyraulus	CHLA	increaser	0.006	140	30	0.14
Helicopsyche	CHLA	decreaser	0.002	47	43	0.32
Hydraenidae	CHLA	decreaser	0.020	6	98	0.53
Ichthybotus	CHLA	decreaser	0.062	13	16	0.48
Latia	CHLA	decreaser	0.003	22	30	0.10
Molophilus	CHLA	decreaser	0.005	2	23	0.05
Nesameletus	CHLA	decreaser	0.010	49	115	0.42
Oligochaeta	CHLA	increaser	0.037	21	324	0.43
Orthocladiinae	CHLA	increaser	0.028	43	313	0.37
Orthopsyche	CHLA	decreaser	0.018	9	26	0.48
Oxyethira	CHLA	increaser	0.049	50	188	0.22
Paraleptamphopus	CHLA	decreaser	0.008	56	25	0.47
Paralimnophila	CHLA	decreaser	0.005	13	19	0.18
Potamopyrgus	CHLA	decreaser	0.033	64	402	0.38
Pycnocentria	CHLA	decreaser	0.009	75	191	0.18
Pycnocentrodes	CHLA	decreaser	0.006	64	307	0.23
Scirtidae	CHLA	decreaser	0.016	2	16	0.20
Stictocladius	CHLA	decreaser	0.011	2	13	0.52
Triplectides	CHLA	decreaser	0.005	13	33	0.10
Zelandobius	CHLA	decreaser	0.001	24	77	0.03
Zelandoperla	CHLA	decreaser	0.018	9	95	0.33
Zephlebia	CHLA	decreaser	0.008	60	57	0.70
Acanthophlebia	SEDCOVER	decreaser	0.000	13	19	0.76
Ameletopsis	SEDCOVER	decreaser	0.007	11	12	0.30
Aoteapsyche	SEDCOVER	decreaser	0.037	15	307	0.22
Aphrophila	SEDCOVER	decreaser	0.008	13	240	0.39

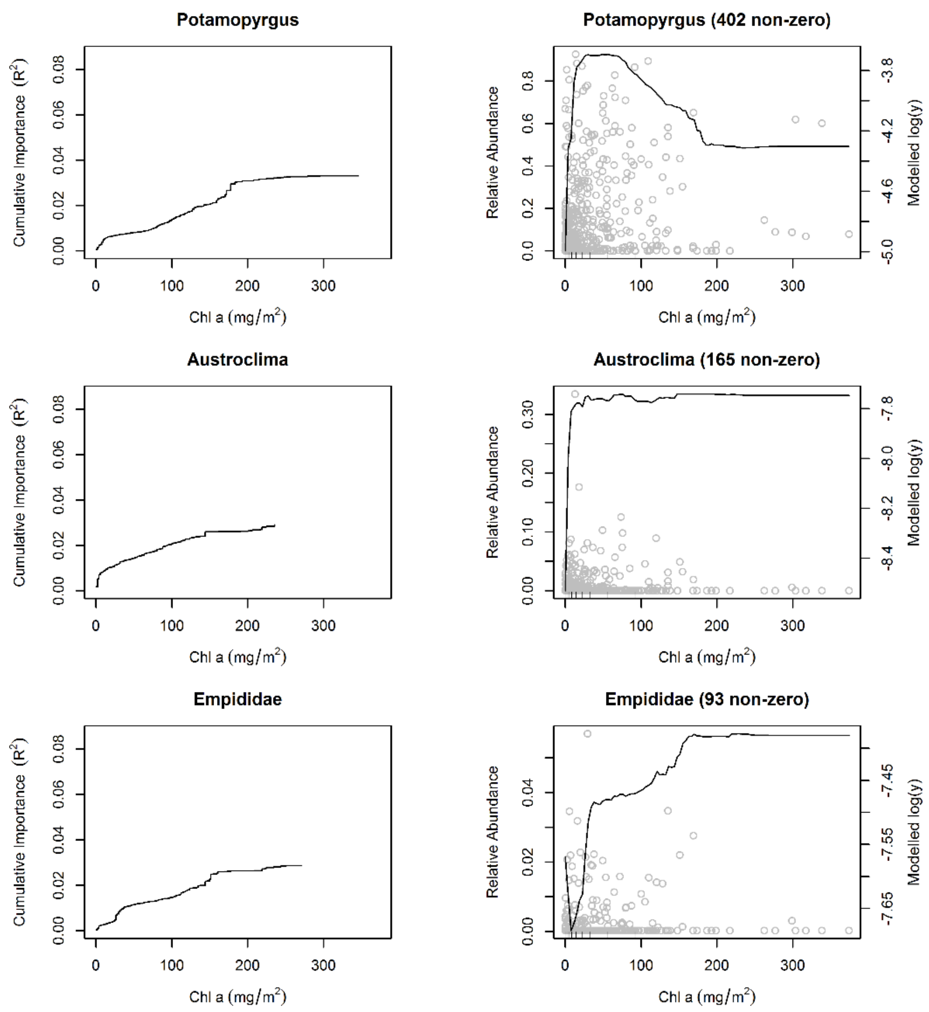
Indicator taxa list, continued

Taxon	**Stressor**	**Response group**	R^2_{fp}	**Threshold**	**No. of occurrences**	R^2
Archichauliodes	SEDCOVER	decreaser	0.012	13	260	0.51
Austroclima	SEDCOVER	decreaser	0.013	14	165	0.53
Austroperla	SEDCOVER	decreaser	0.004	13	68	0.58
Austrosimulium	SEDCOVER	increaser	0.022	18	204	0.23
Beraeoptera	SEDCOVER	decreaser	0.001	4	48	0.55
Coloburiscus	SEDCOVER	decreaser	0.004	13	183	0.37
Confluens	SEDCOVER	decreaser	0.002	9	10	0.61
Corynoneura	SEDCOVER	increaser	0.153	83	21	0.54
Costachorema	SEDCOVER	decreaser	0.004	8	89	0.19
Deleatidium	SEDCOVER	decreaser	0.065	55	439	0.57
Elmidae	SEDCOVER	decreaser	0.026	55	437	0.47
Gyraulus	SEDCOVER	increaser	0.041	83	30	0.14
Helicopsyche	SEDCOVER	decreaser	0.014	13	43	0.32
Hydraenidae	SEDCOVER	decreaser	0.010	9	98	0.53
Hydrobiosis	SEDCOVER	decreaser	0.016	51	372	0.26
Latia	SEDCOVER	decreaser	0.002	14	30	0.10
Lobodiamesa	SEDCOVER	decreaser	0.001	8	16	0.18
Megaleptoperla	SEDCOVER	decreaser	0.005	19	51	0.35
Nematoda	SEDCOVER	increaser	0.187	82	45	0.69
Nesameletus	SEDCOVER	decreaser	0.004	9	115	0.42
Neurochorema	SEDCOVER	decreaser	0.001	8	73	0.42
Oligochaeta	SEDCOVER	increaser	0.003	15	324	0.43
Olinga	SEDCOVER	decreaser	0.004	18	126	0.23
Psilochorema	SEDCOVER	decreaser	0.008	15	247	0.23
Ptilodactylidae	SEDCOVER	decreaser	0.010	15	12	0.55
Pycnocentria	SEDCOVER	decreaser	0.003	19	191	0.18
Pycnocentrodes	SEDCOVER	decreaser	0.018	18	307	0.23
Scirtidae	SEDCOVER	decreaser	0.004	13	16	0.20
Sphaeriidae	SEDCOVER	increaser	0.050	82	78	0.42
Stenoperla	SEDCOVER	decreaser	0.002	9	64	0.32
Stictocladius	SEDCOVER	decreaser	0.020	15	13	0.52
Zelandoperla	SEDCOVER	decreaser	0.005	7	95	0.33
Zephlebia	SEDCOVER	decreaser	0.013	9	57	0.70

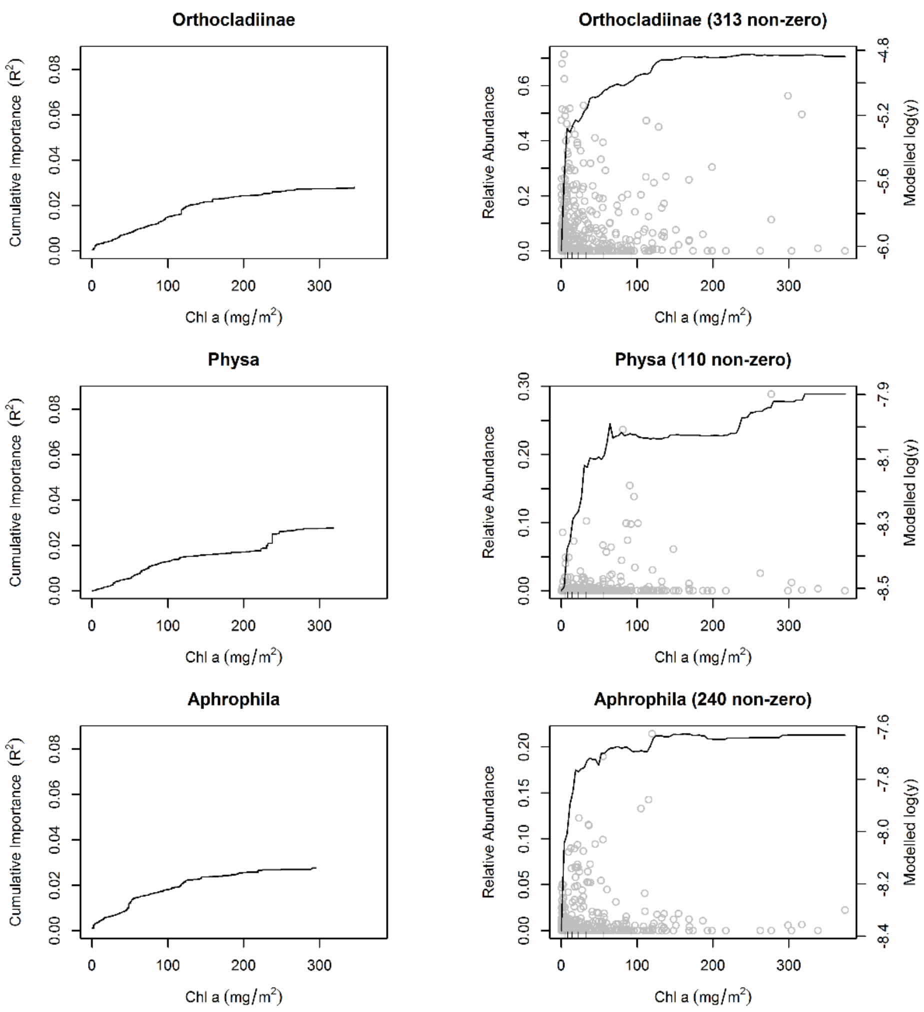
Appendix 3. (Following pages) Species turnover functions (left side) and random forest partial dependence plots (fitted functions) with overlay of data points (right side) across the chlorophyll-*a* and % sediment cover gradients for all taxa that had a random forest model $R^2 > 0$. For each stressor gradient, the taxa are ordered by their maximum cumulative importance.


Chlorophyll-a gradient
Taxon 1 - 3
Taxon 4 - 6
Deleatidium
Deleatidium (439 non-zero)
Acanthophlebia
Acanthophlebia (19 non-zero)
Ichthybotus
Ichthybotus (16 non-zero)
Oligochaeta
Oligochaeta (324 non-zero)
Oxyethira
Oxyethira (188 non-zero)
Neurochorema
Neurochorema (73 non-zero)
Cumulative Importance (R^2)
0.00
0.02
0.04
0.06
0.08
Relative Abundance
Modelled log(y)
-2.5
-3.0
-3.5
-5.605
-5.615
-5.625
-6.690
-6.700
-6.710
-6.720
-5.4
-5.6
-5.8
-6.0
-6.2
-6.0
-6.4
-6.8
-7.25
-7.35
-7.45
-7.55
0.0
0.2
0.4
0.6
0.8
0.05
0.10
0.15
0.02
0.04
0.06
0
100
200
300
Chl a (mg/m^2)

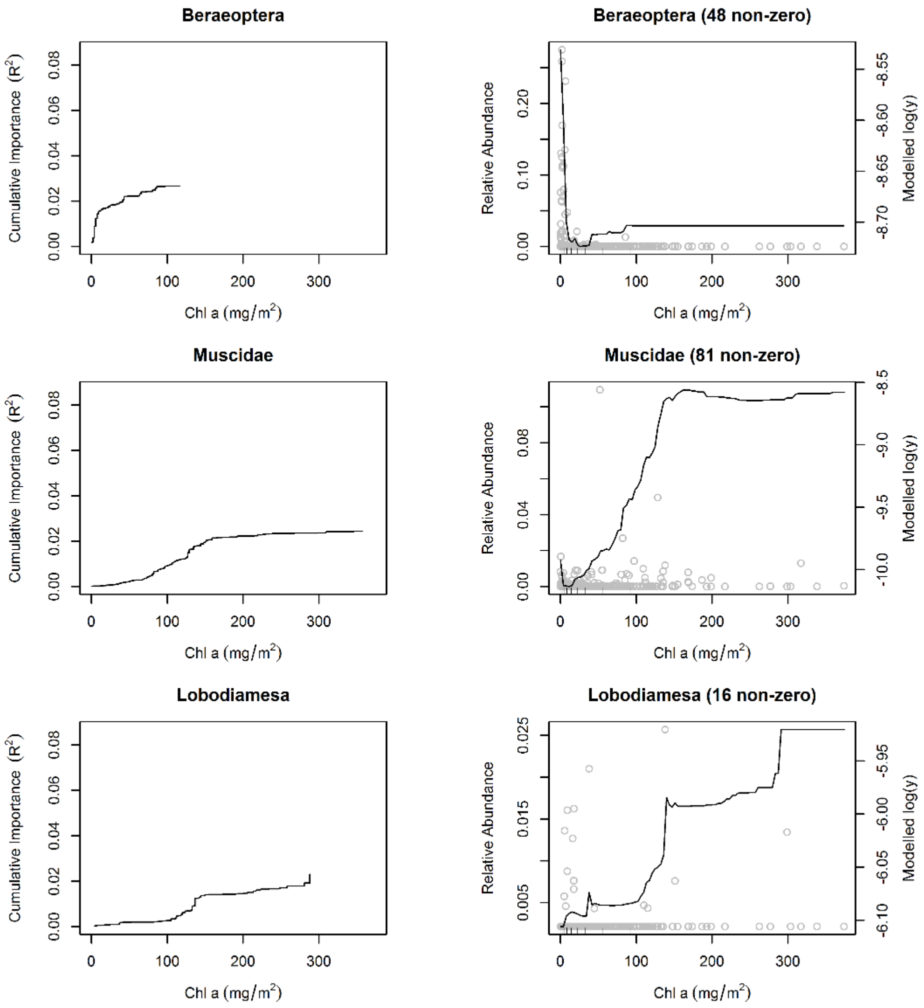
Taxon 7 - 9



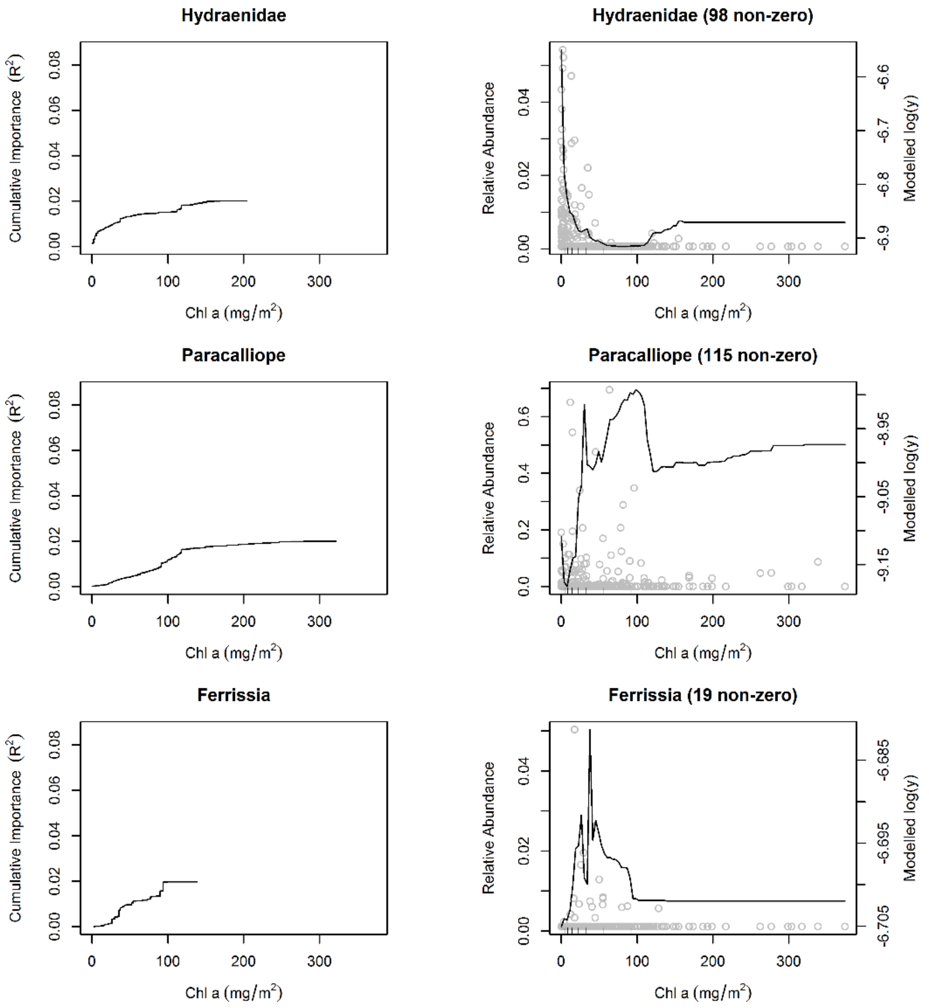
Taxon 10 - 12



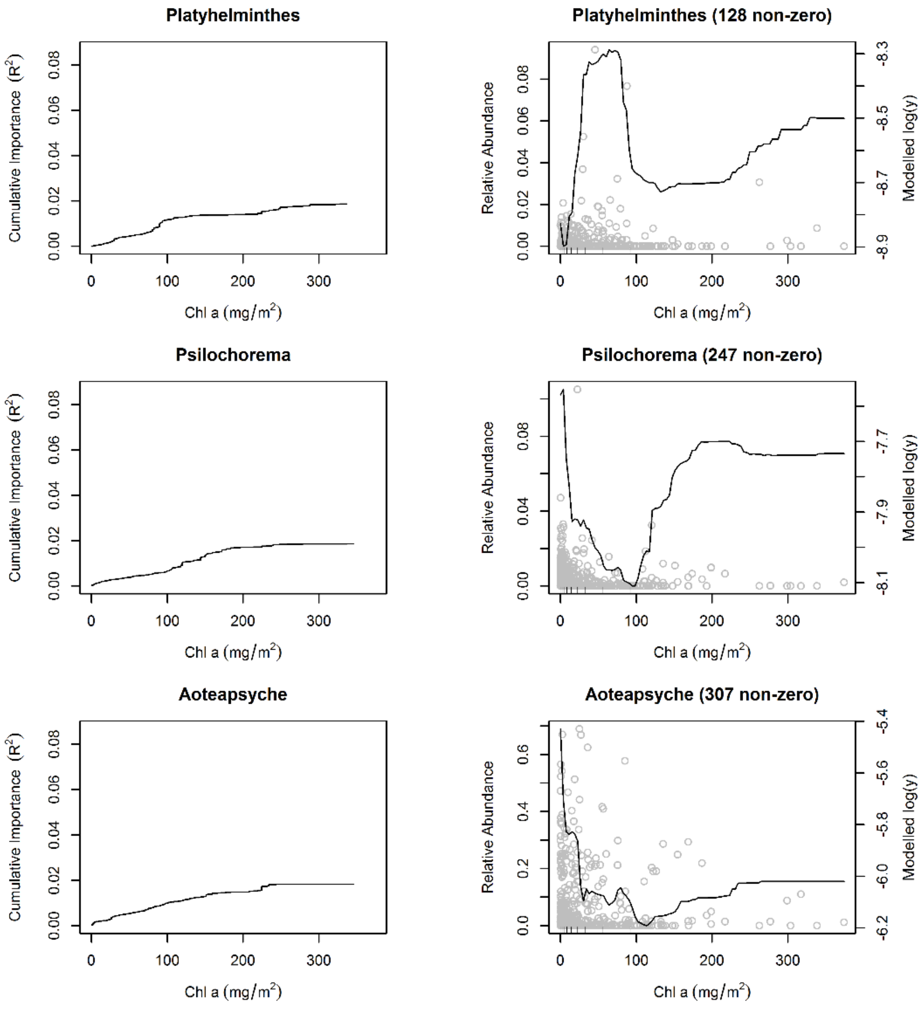
Taxon 13 - 15



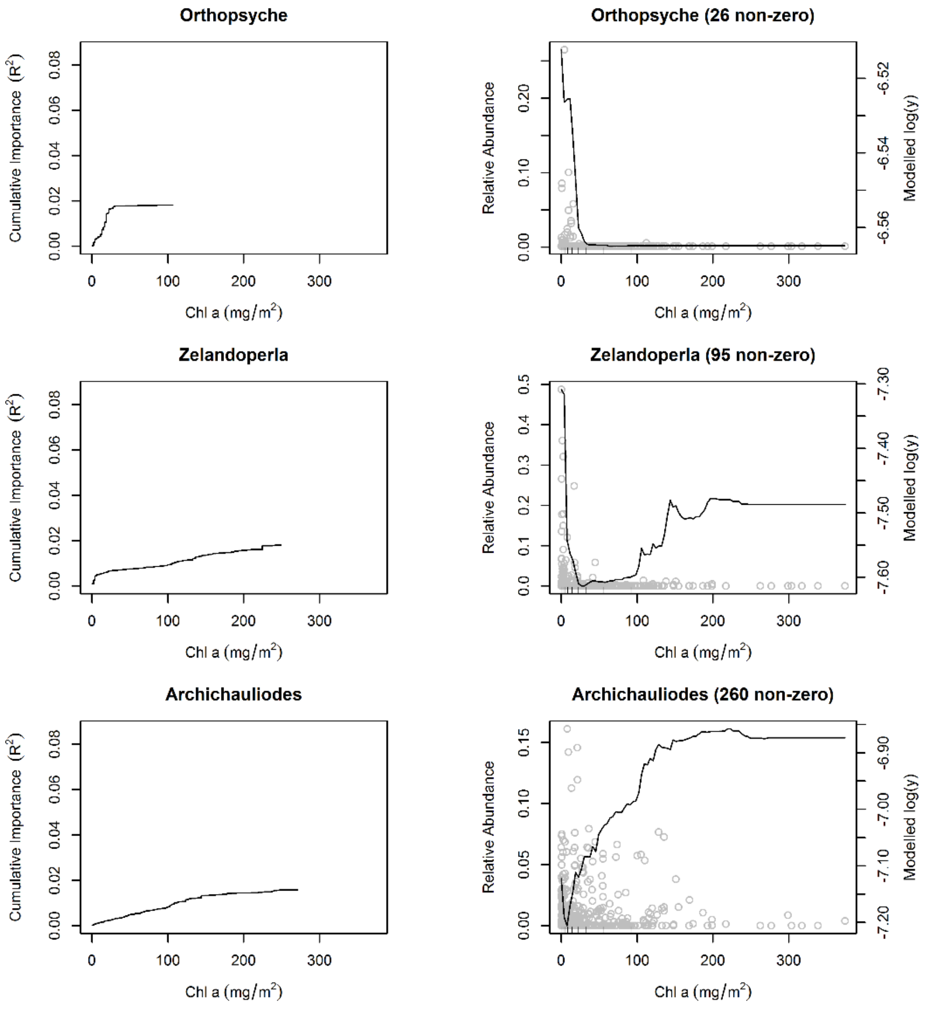
Taxon 16 - 18



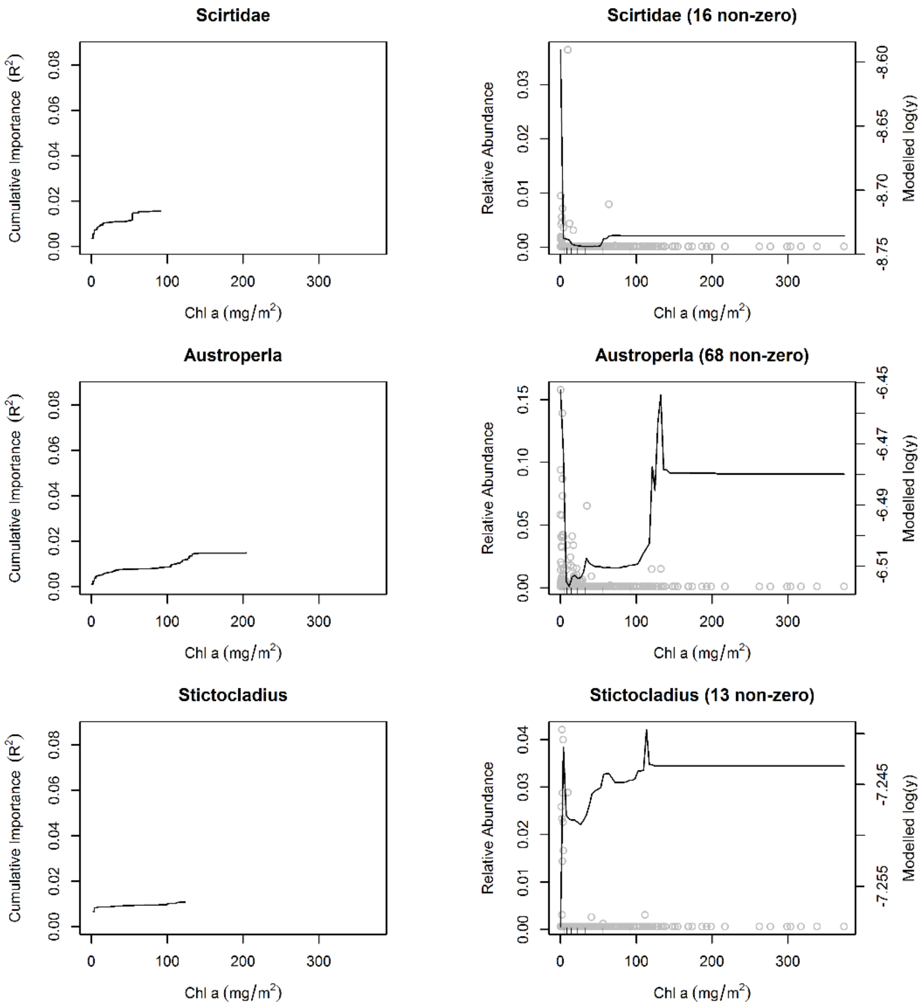
Taxon 19 - 21



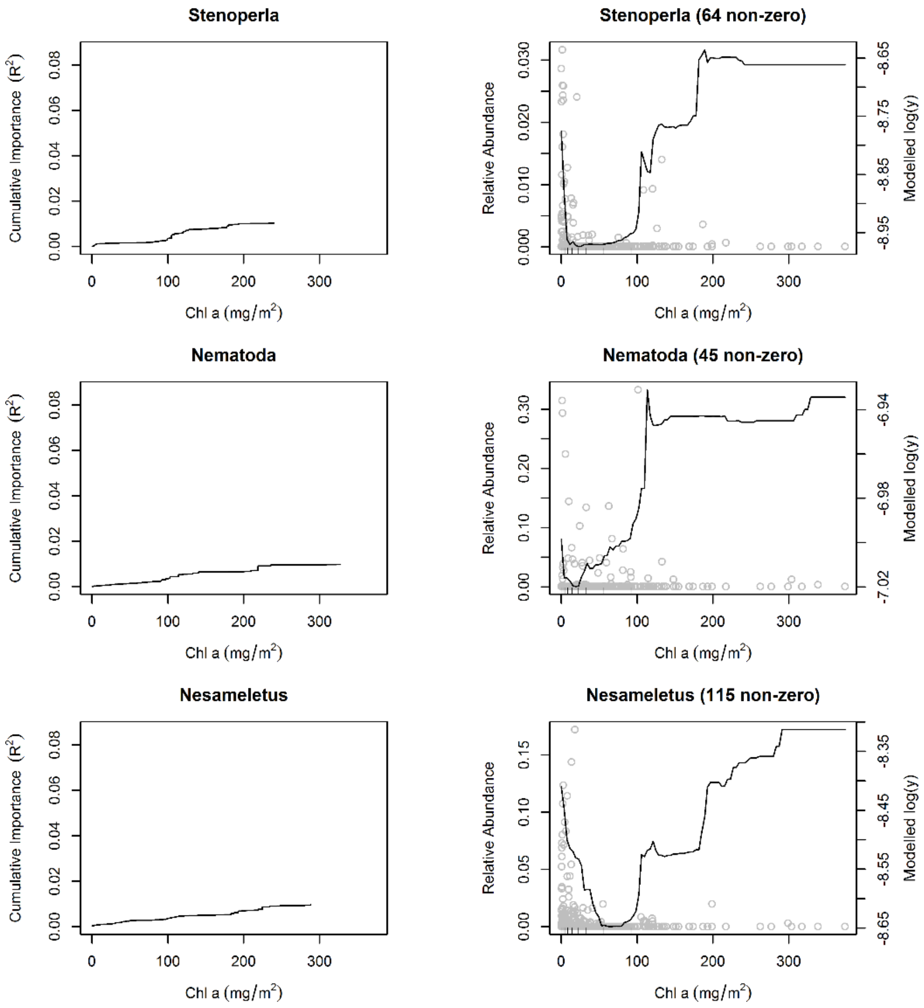
Taxon 22 - 24



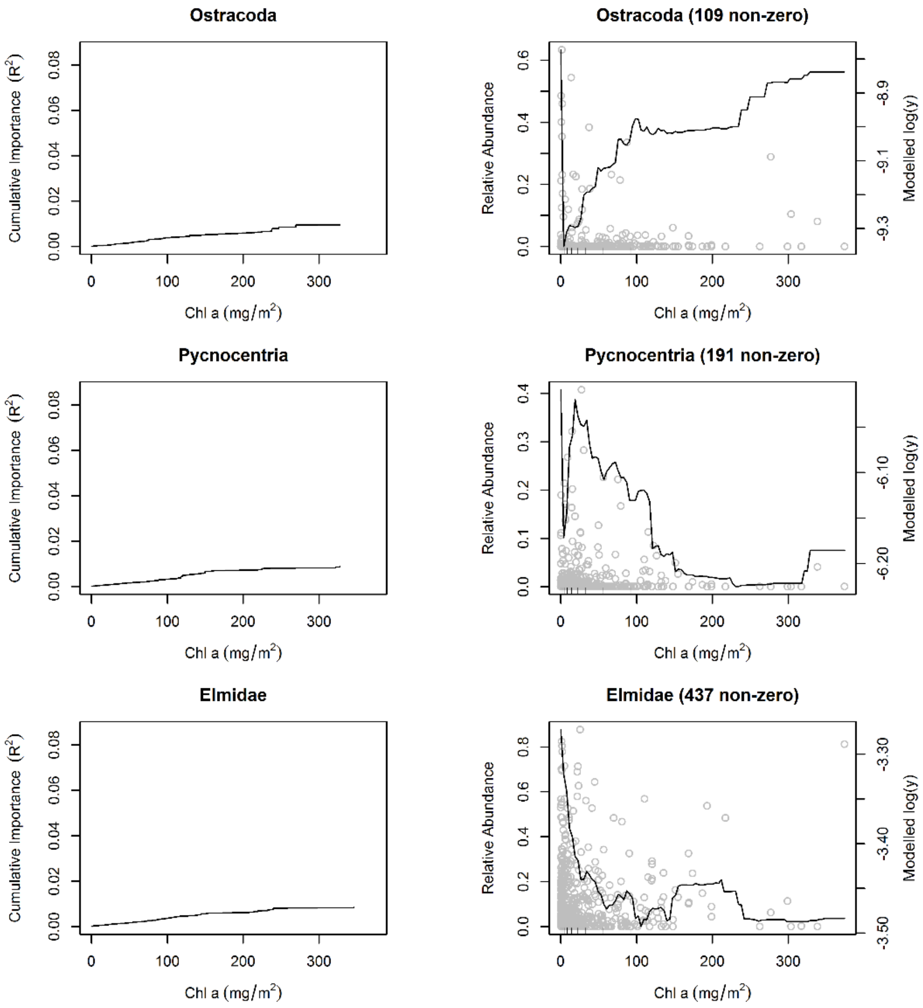
Taxon 25 - 27



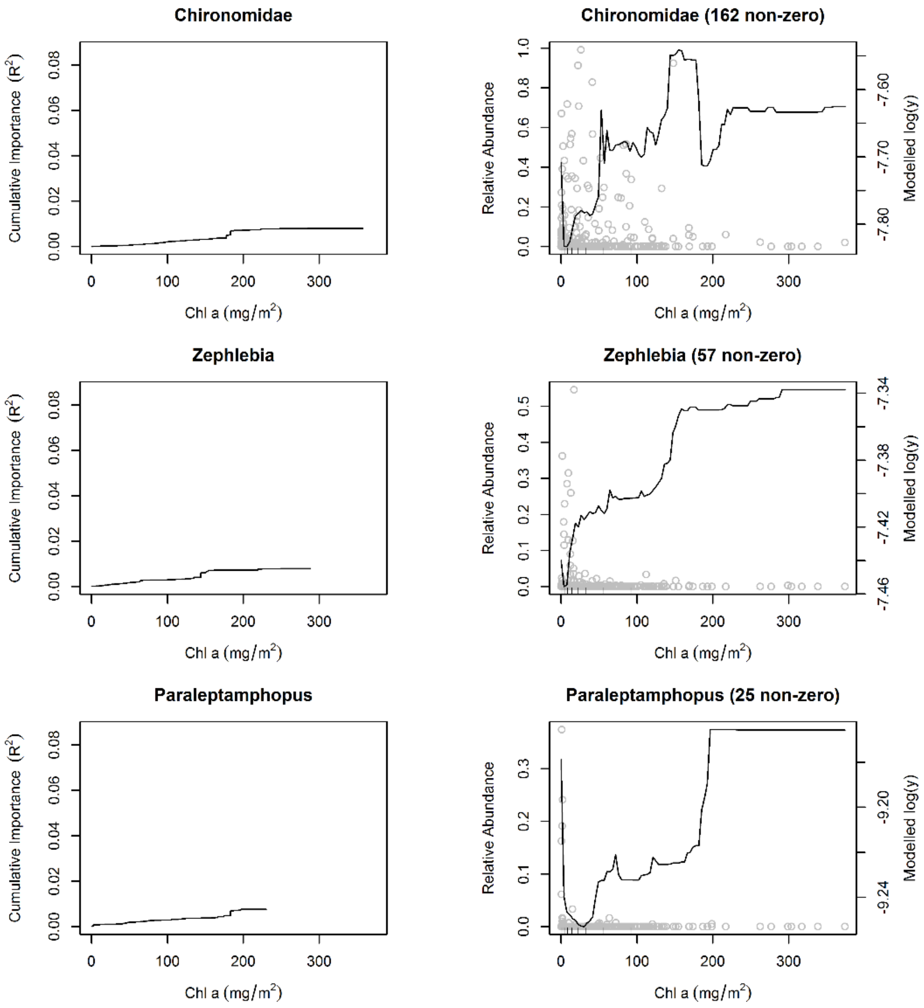
Taxon 28 - 30



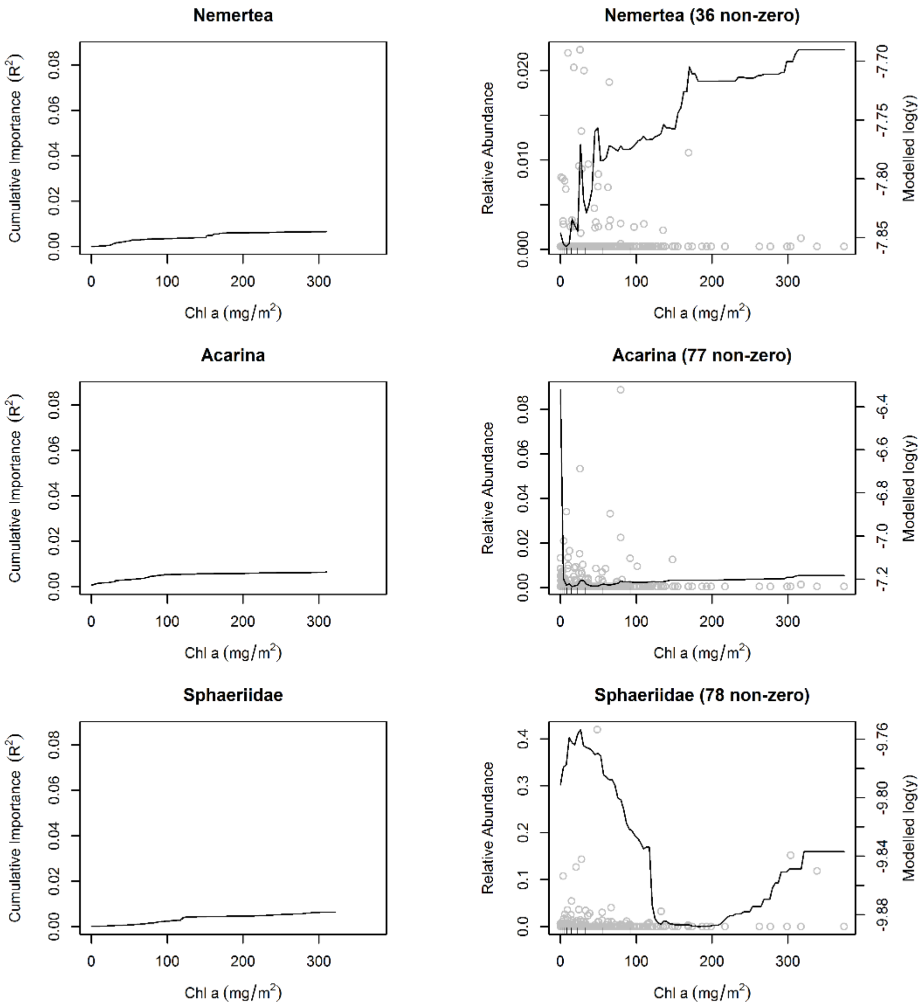
Taxon 31 - 33



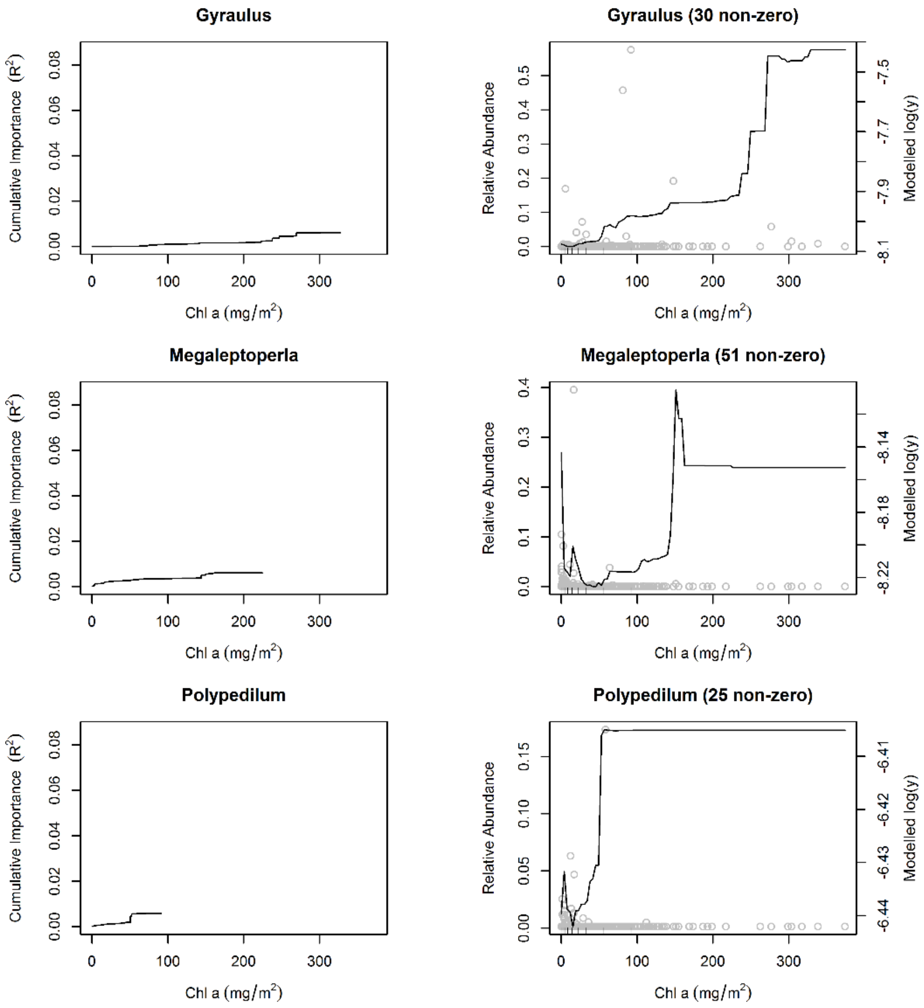
Taxon 34 - 36



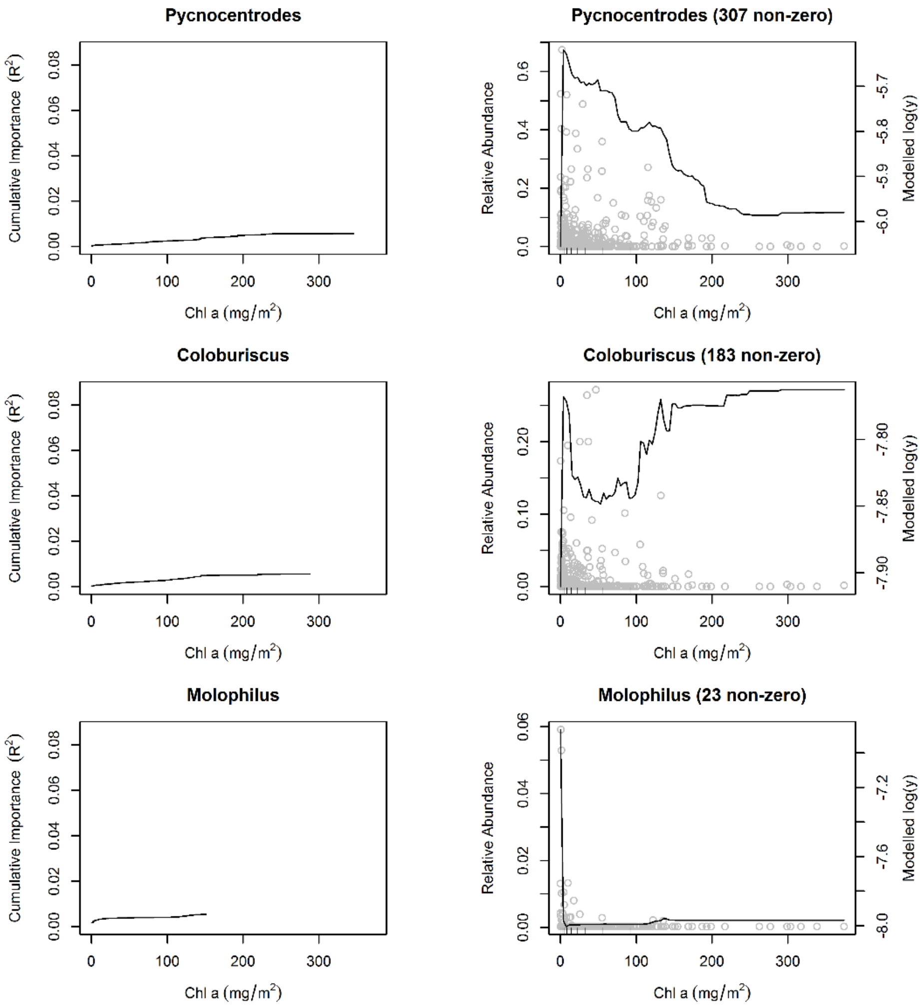
Taxon 37 - 39



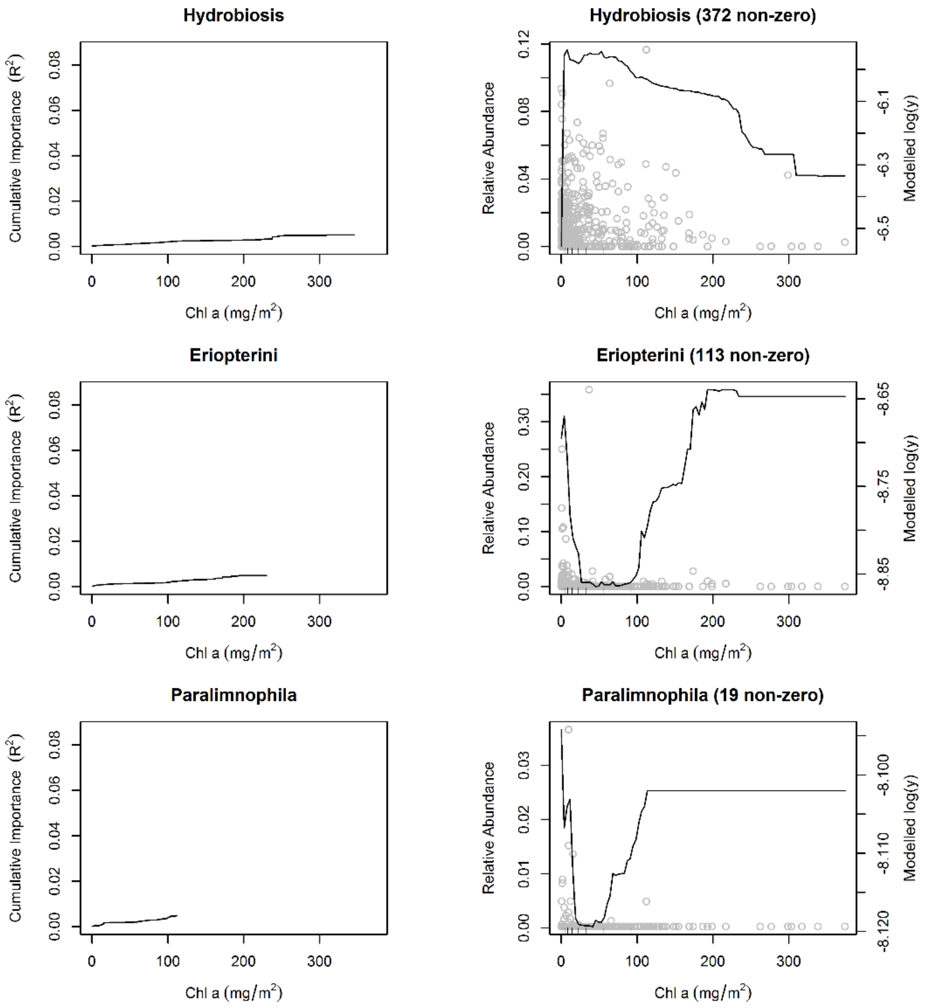
Taxon 40 - 42



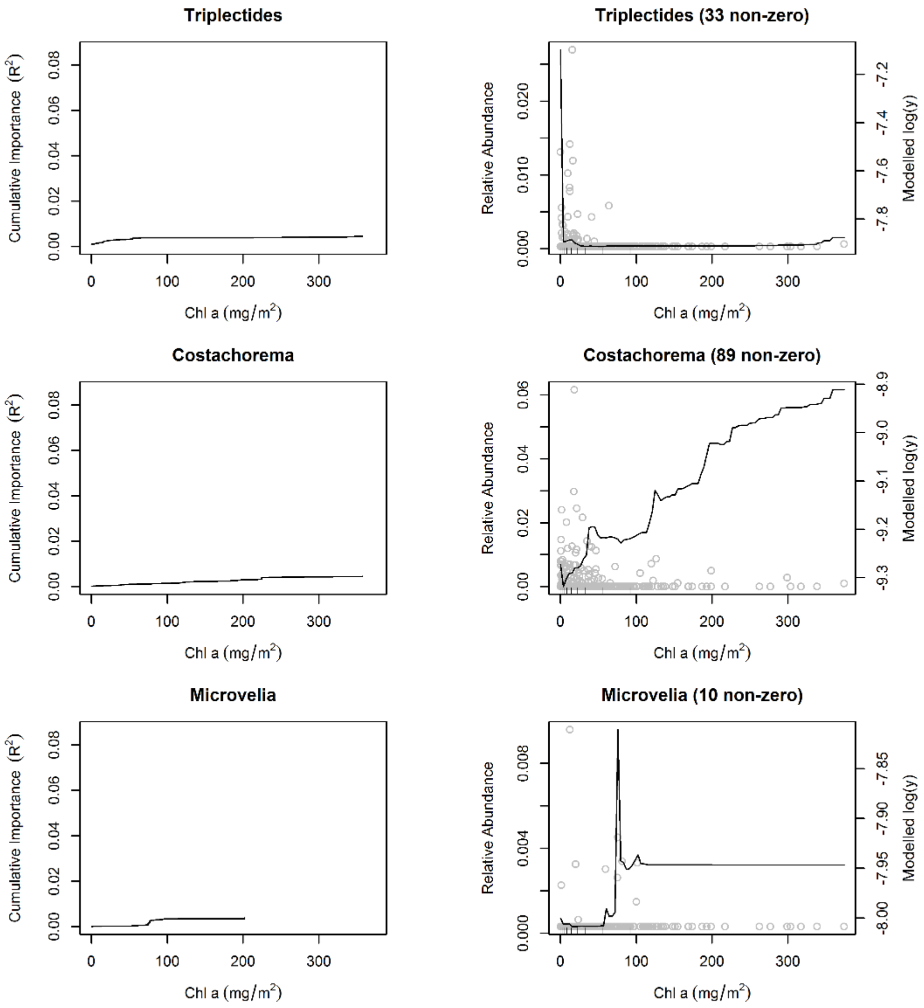
Taxon 43 - 45



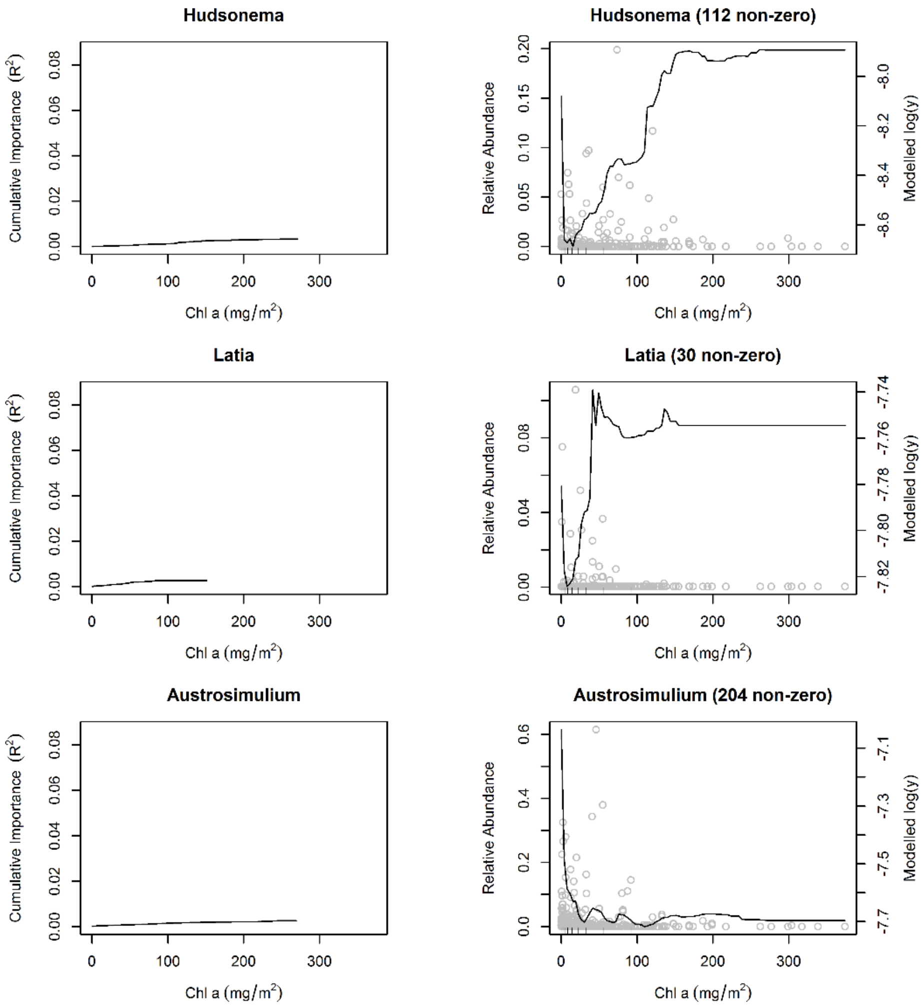
Taxon 46 - 48



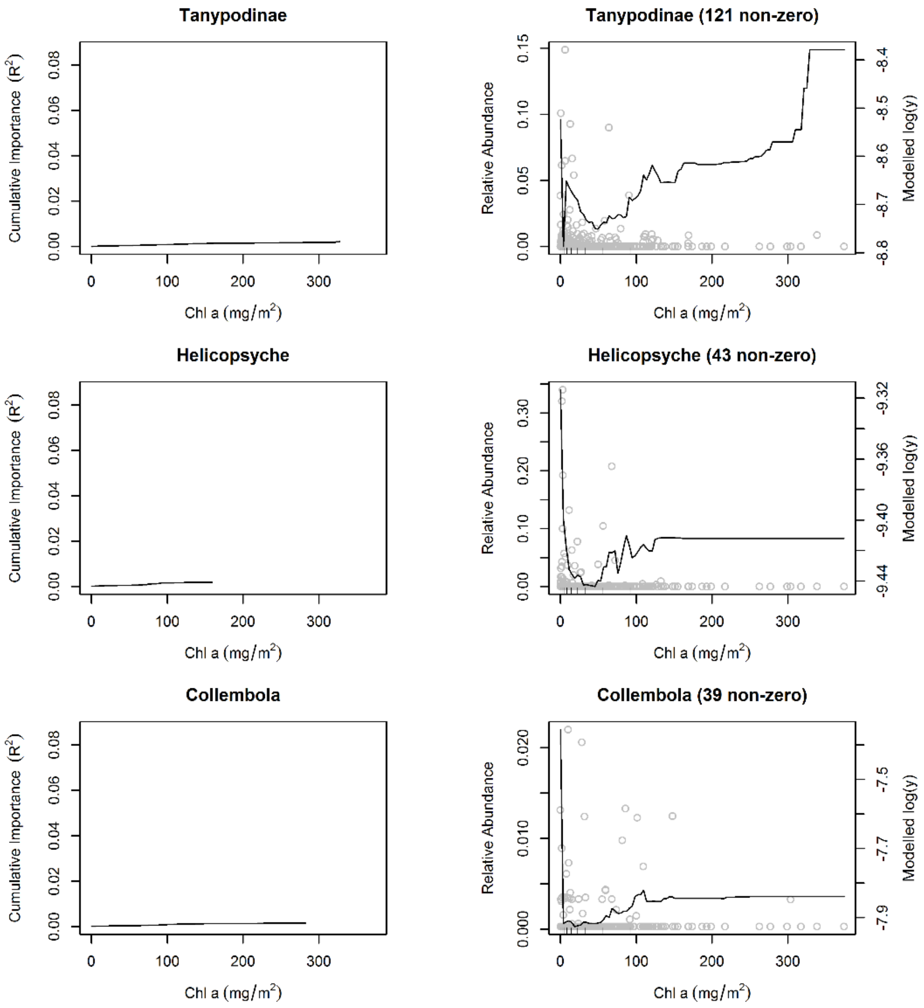
Taxon 49 - 51



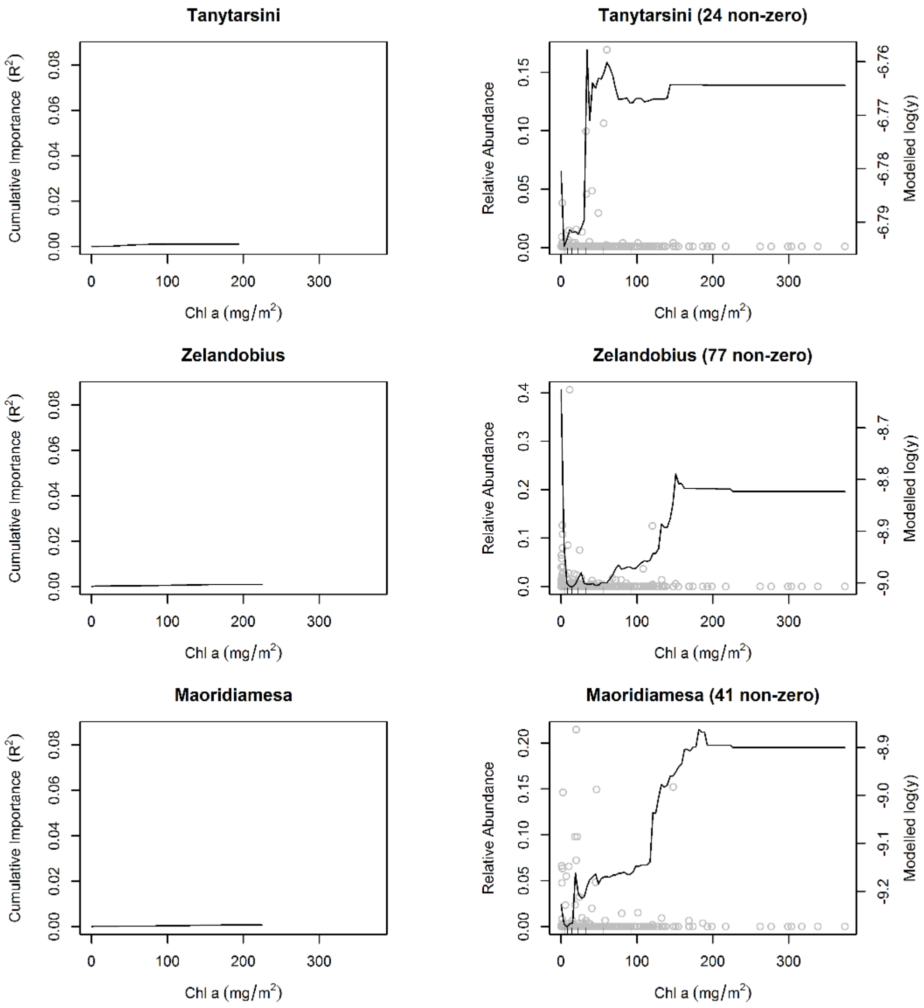
Taxon 52 - 54



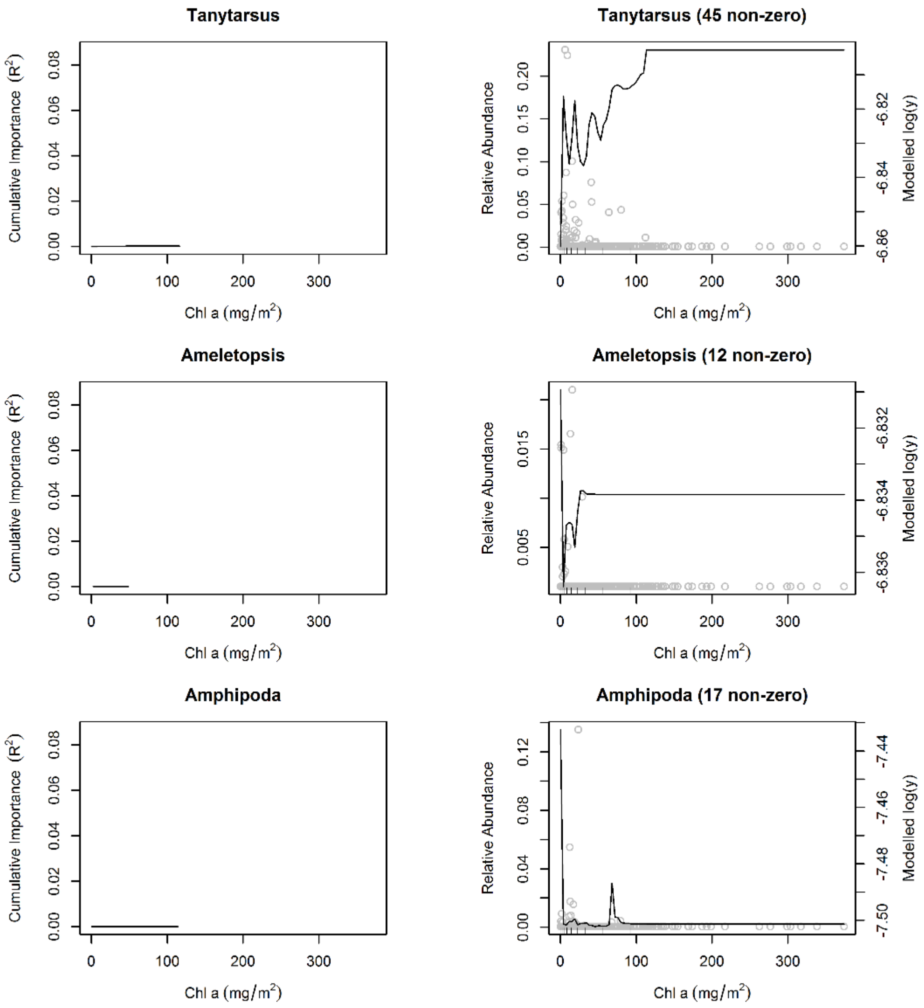
Taxon 55 - 57



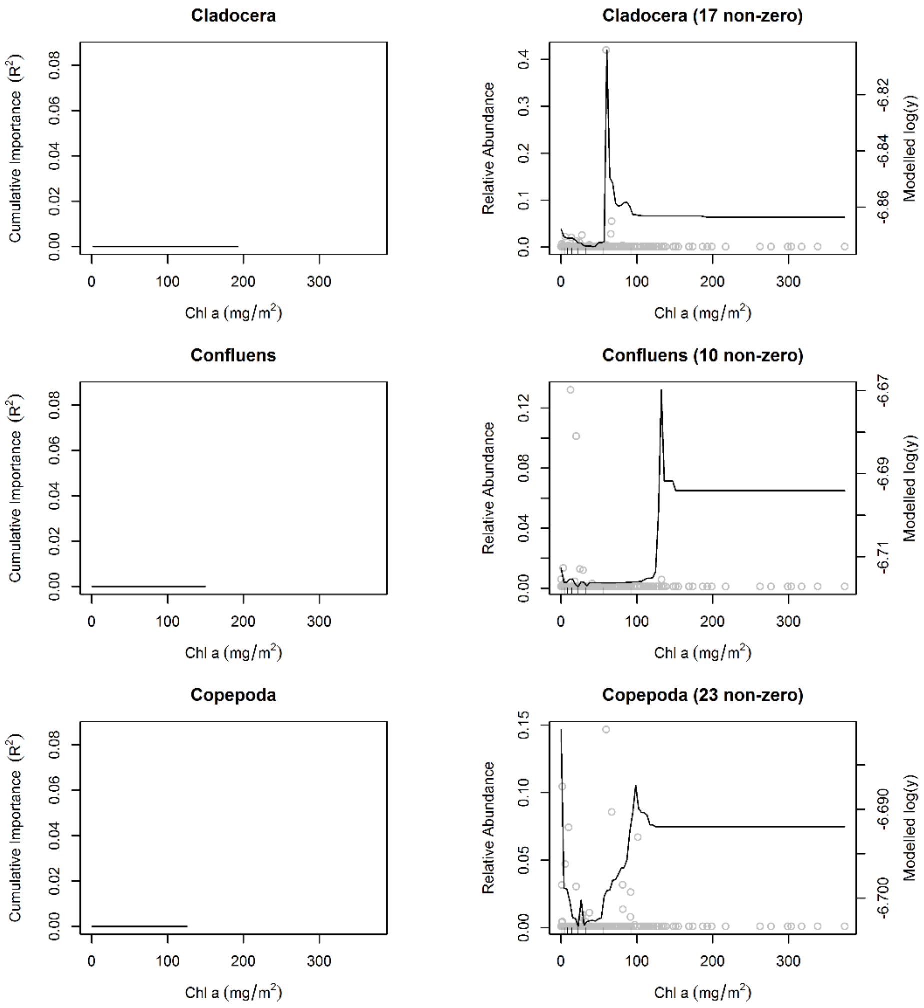
Taxon 58 - 60



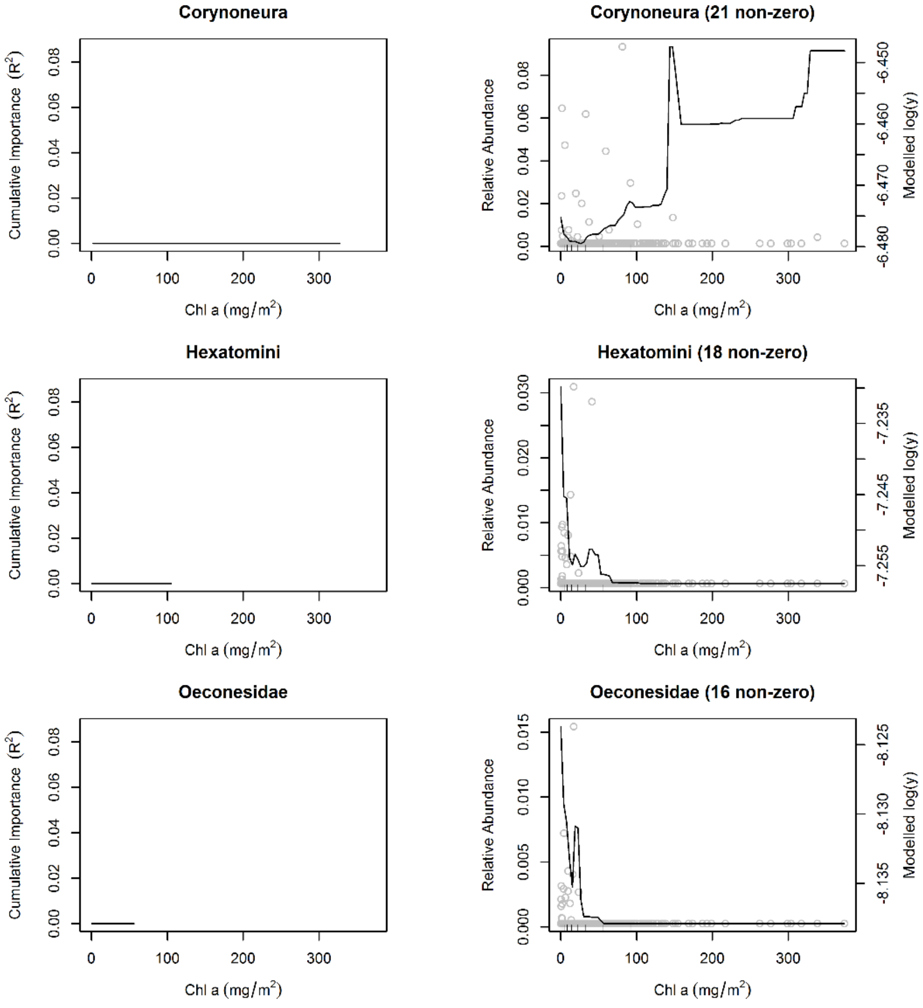
Taxon 61 - 63



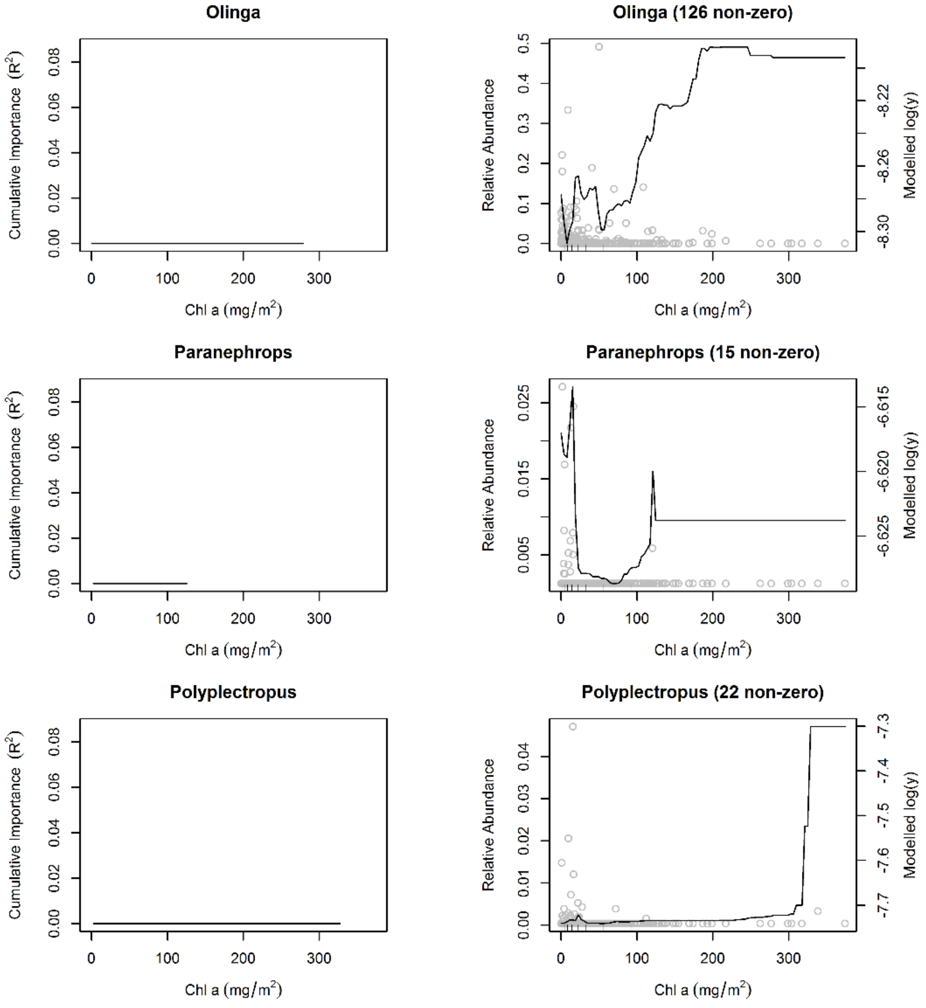
Taxon 64 - 66



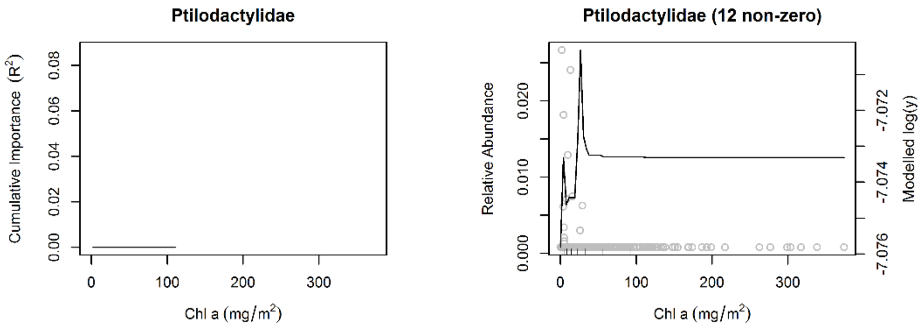
Taxon 67 - 69

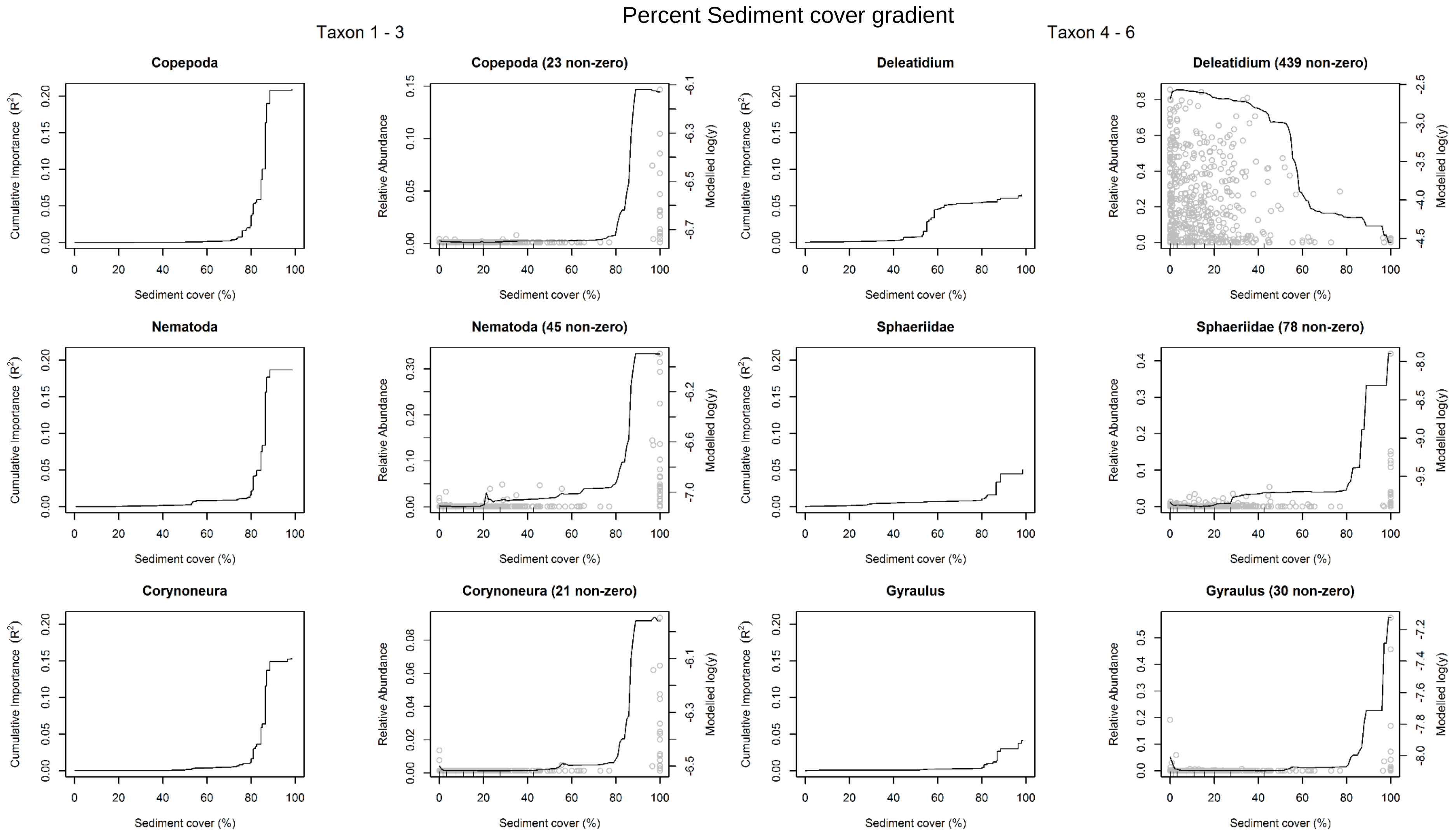


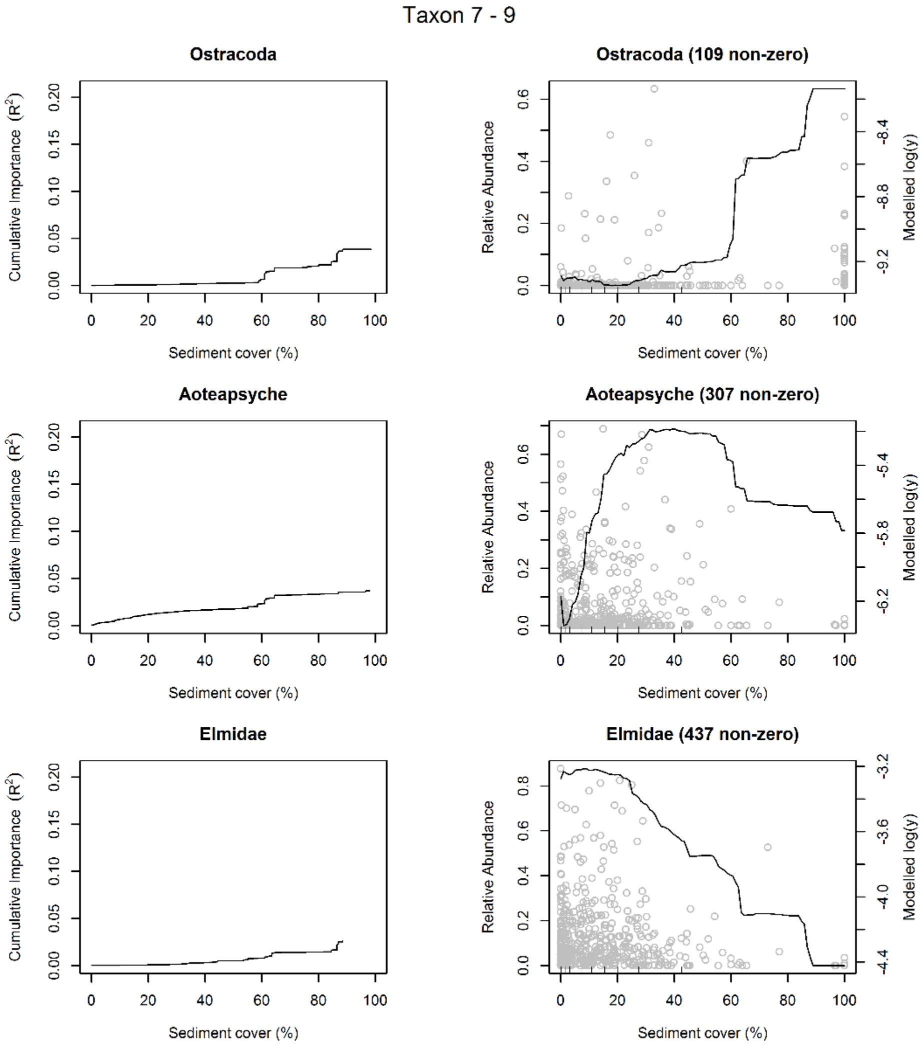
Taxon 70 - 72

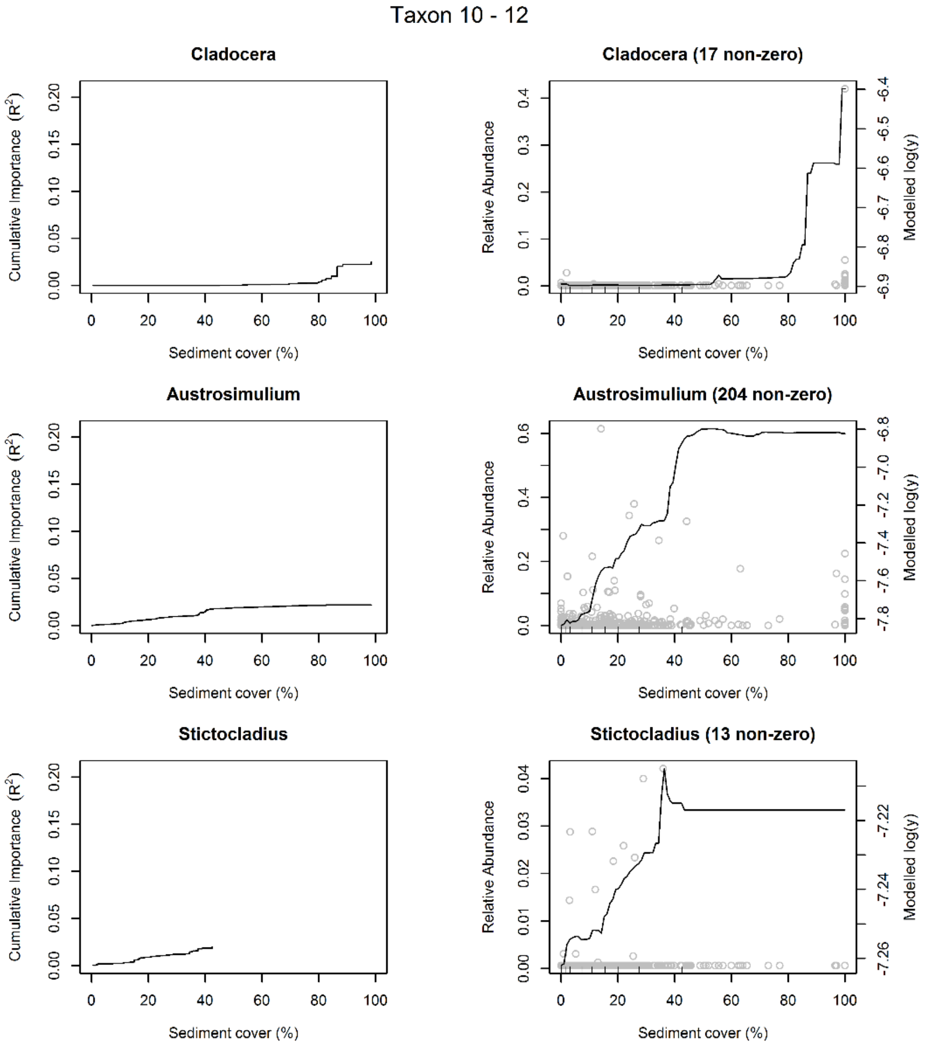


Taxon 73 - 73

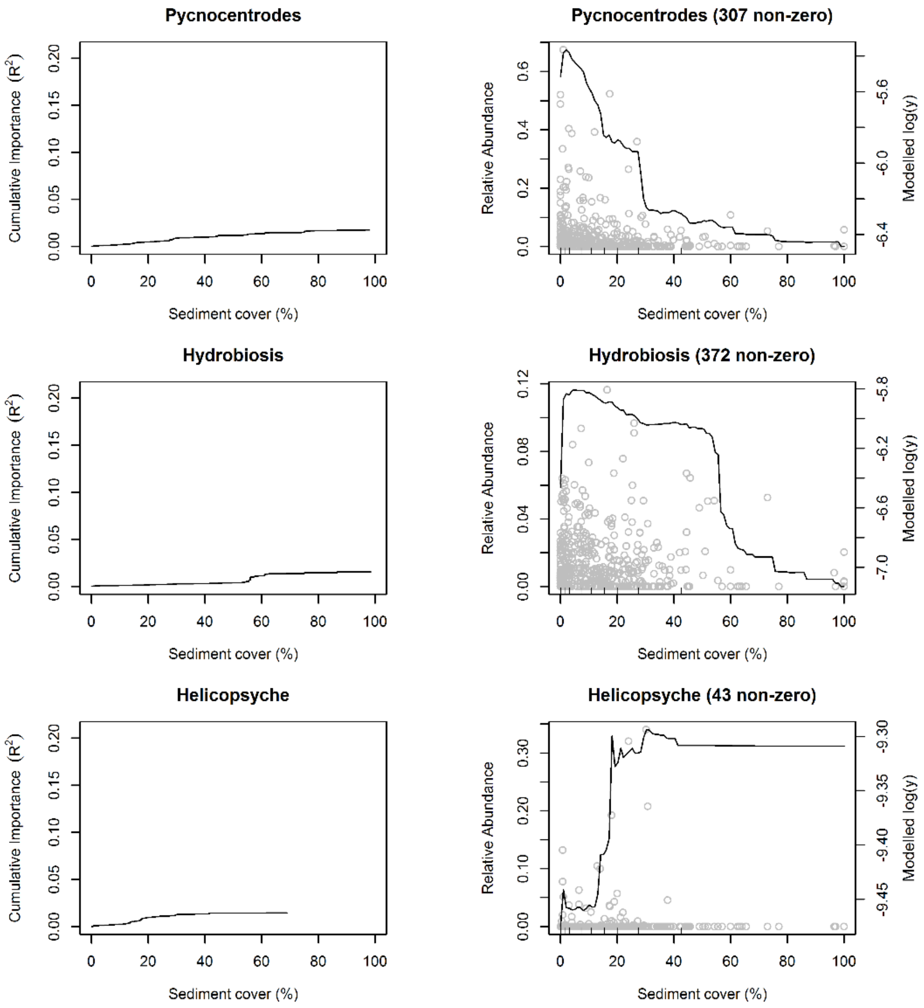

Ptilodactylidae
Cumulative Importance (R^2)
0.00
0.02
0.04
0.06
0.08
0
100
200
300
Chl a (mg/m^2)
Ptilodactylidae (12 non-zero)
Relative Abundance
0.000
0.010
0.020
Modelled log(y)
-7.076
-7.074
-7.072
0
100
200
300
Chl a (mg/m^2)


Percent Sediment cover gradient
Taxon 1 - 3
Taxon 4 - 6
Copepoda
Copepoda (23 non-zero)
Deleatidium
Deleatidium (439 non-zero)
Nematoda
Nematoda (45 non-zero)
Sphaeriidae
Sphaeriidae (78 non-zero)
Corynoneura
Corynoneura (21 non-zero)
Gyraulus
Gyraulus (30 non-zero)
Cumulative Importance (R²)
Relative Abundance
Modelled log(y)
Sediment cover (%)
0
20
40
60
80
100
0.00
0.05
0.10
0.15
0.20
-6.1
-6.3
-6.5
-6.7
0.8
0.6
0.4
0.2
0.0
-2.5
-3.0
-3.5
-4.0
-4.5
0.30
0.20
0.10
-6.2
-6.6
-7.0
0.4
0.3
0.1
-8.0
-8.5
-9.0
-9.5
0.08
0.06
0.04
0.02
0.5
-7.2
-7.4
-7.6
-7.8

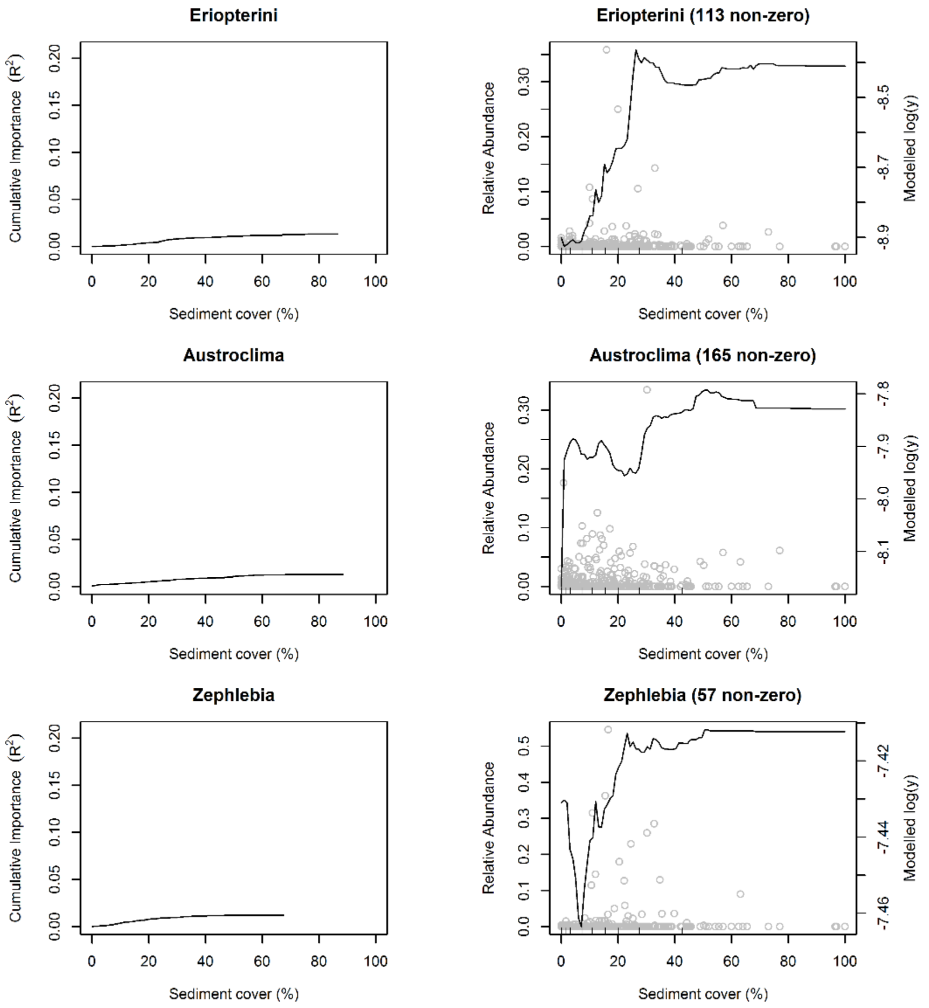

Taxon 7 - 9
Ostracoda
Ostracoda (109 non-zero)
Aoteapsyche
Aoteapsyche (307 non-zero)
Elmidae
Elmidae (437 non-zero)
Cumulative Importance (R^2)
Relative Abundance
Modelled log(y)
Sediment cover (%)
0
20
40
60
80
100
0.00
0.05
0.10
0.15
0.20


Taxon 10 - 12
Cladocera
Cladocera (17 non-zero)
Austrosimulium
Austrosimulium (204 non-zero)
Stictocladius
Stictocladius (13 non-zero)
Cumulative Importance (R^2)
Relative Abundance
Modelled log(y)
Sediment cover (%)
0
20
40
60
80
100
0.00
0.05
0.10
0.15
0.20

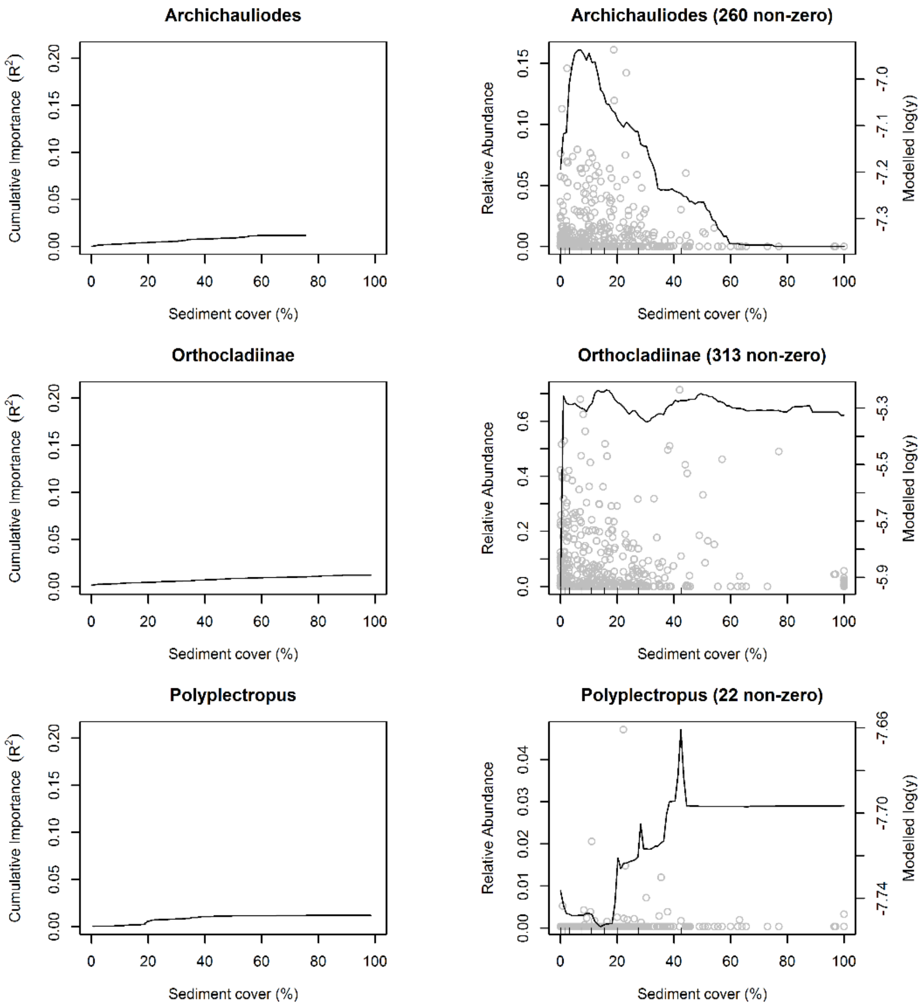
Taxon 13 - 15



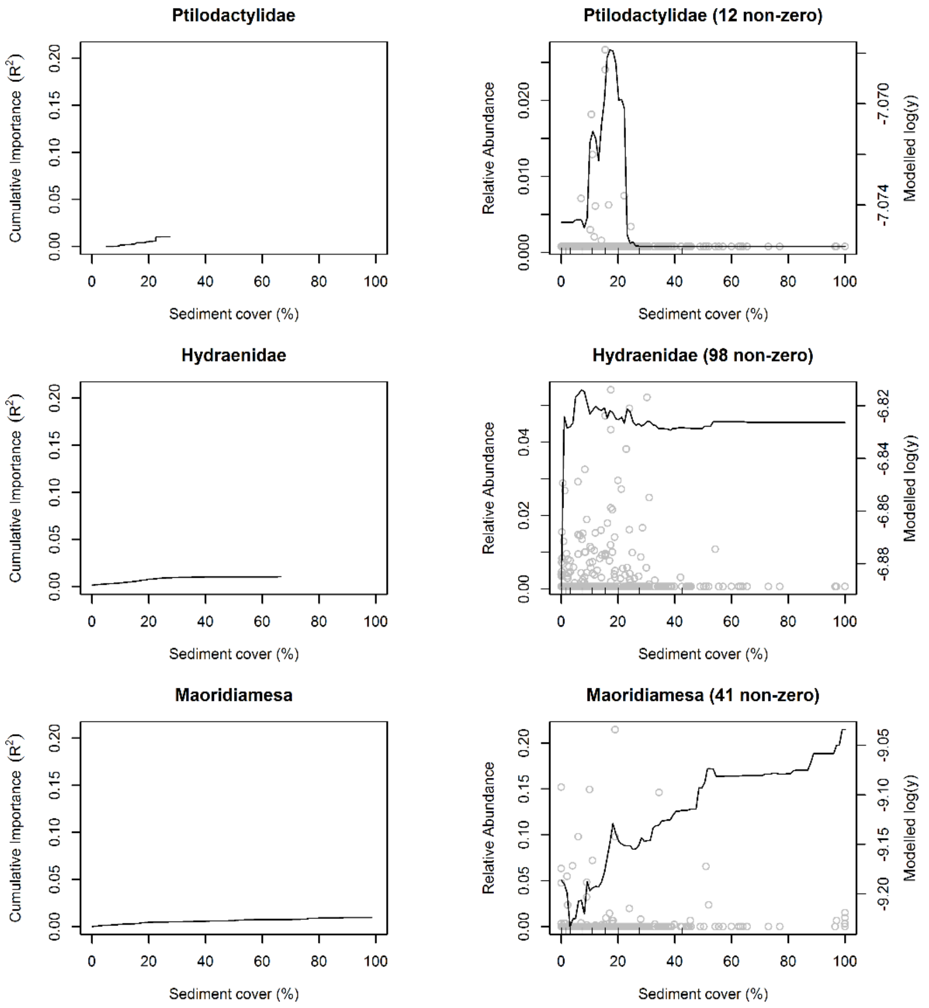
Taxon 16 - 18



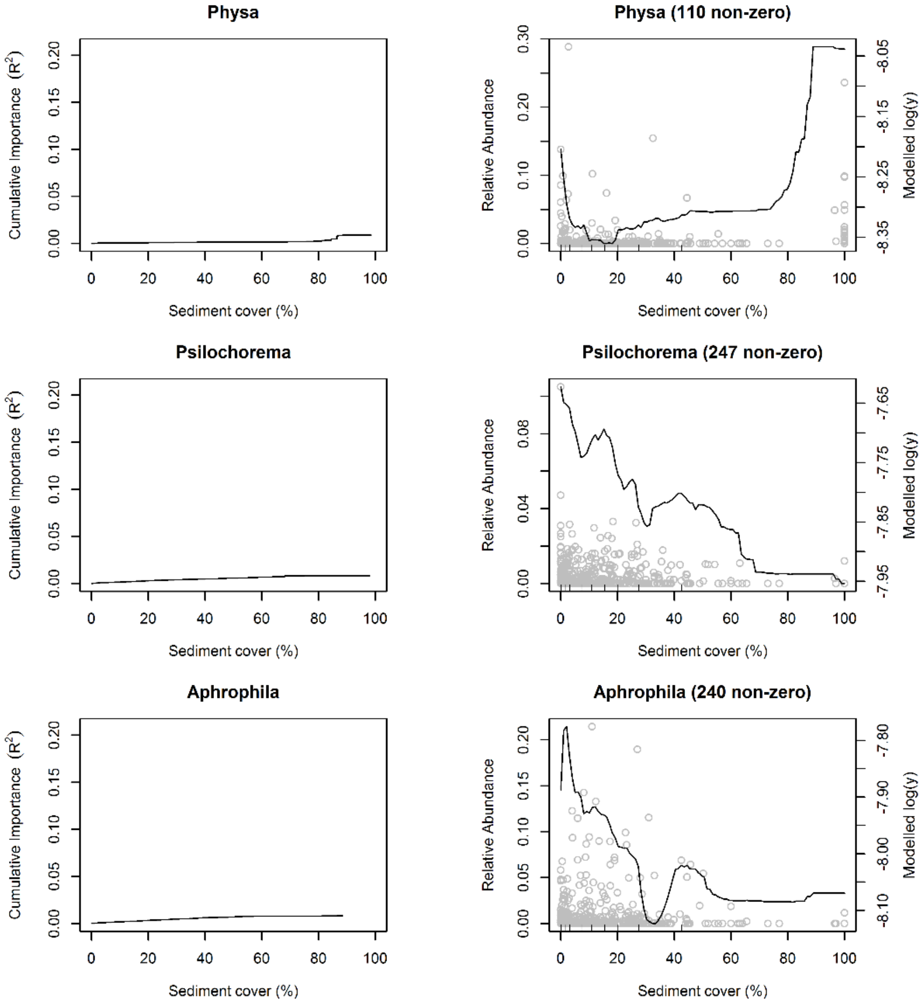
Taxon 19 - 21



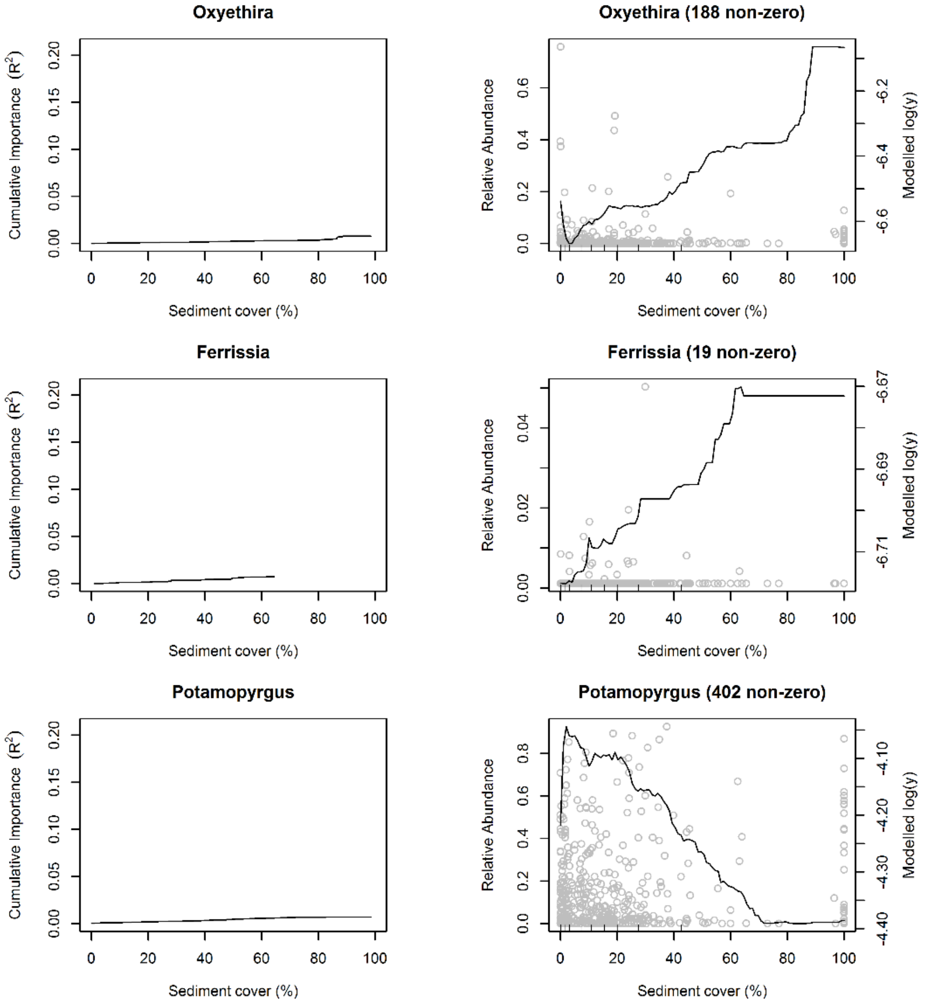
Taxon 22 - 24



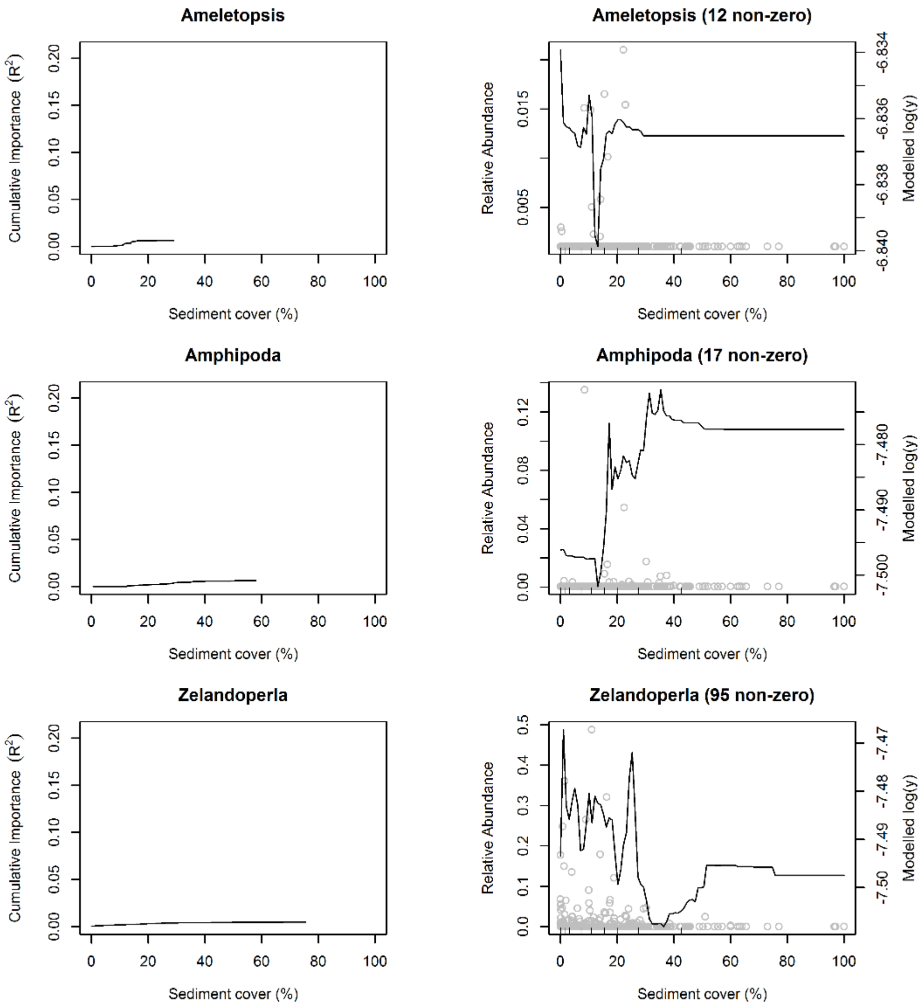
Taxon 25 - 27



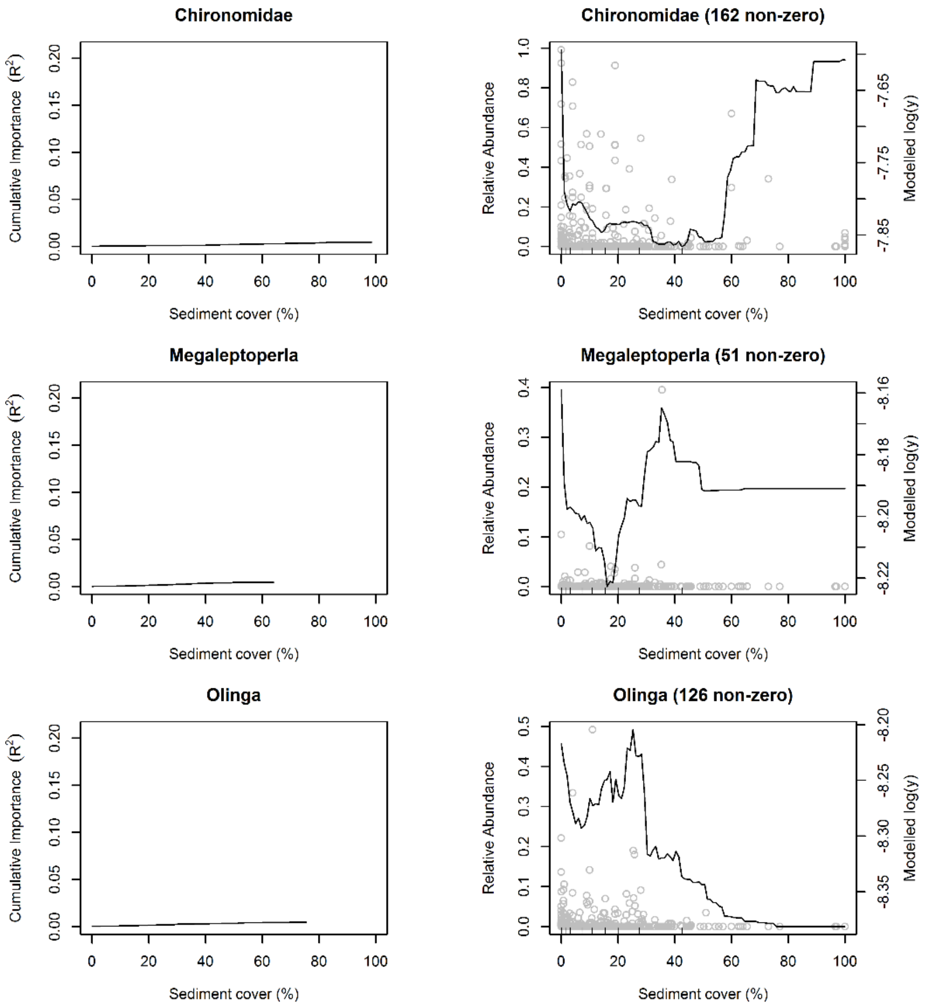
Taxon 28 - 30



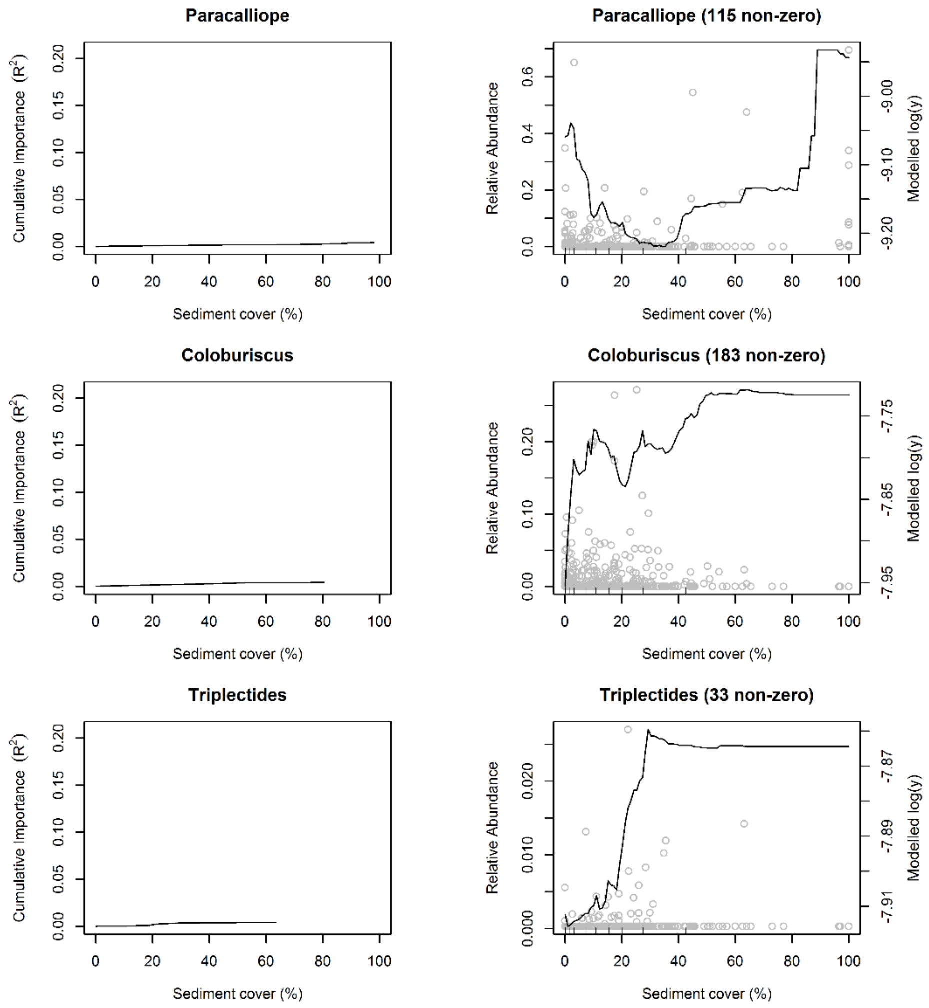
Taxon 31 - 33



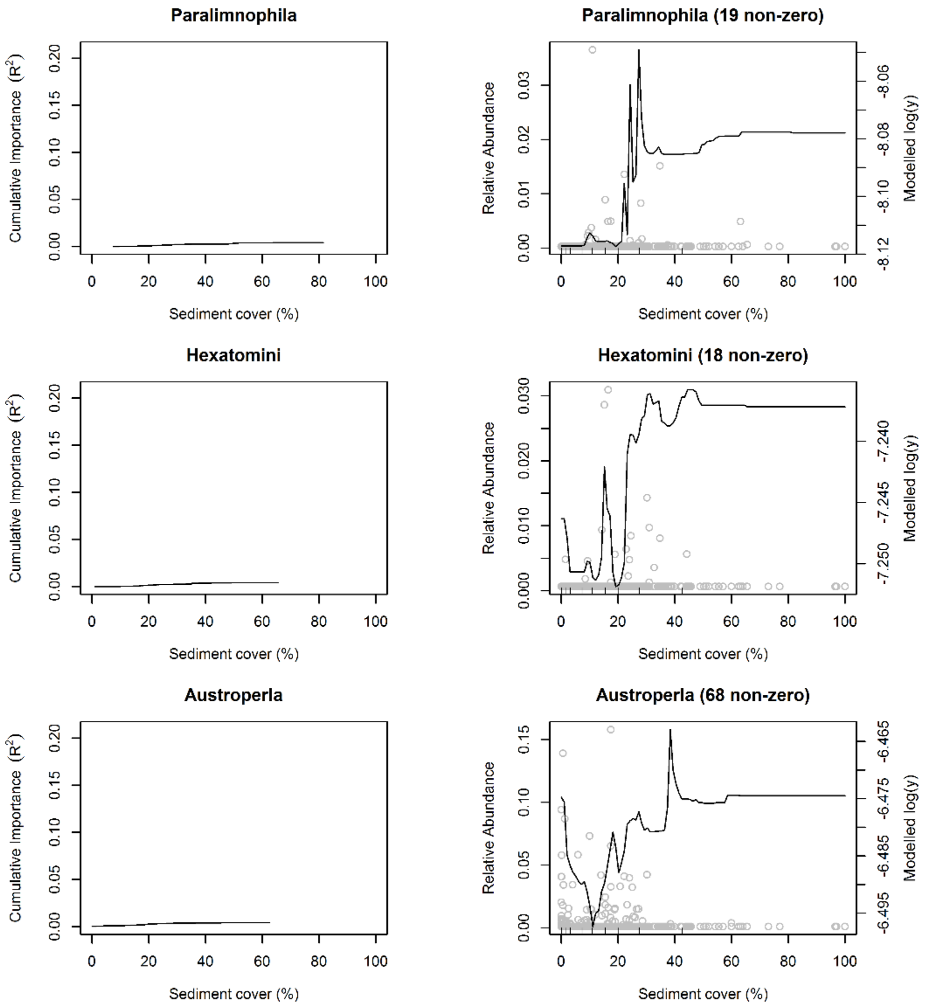
Taxon 34 - 36



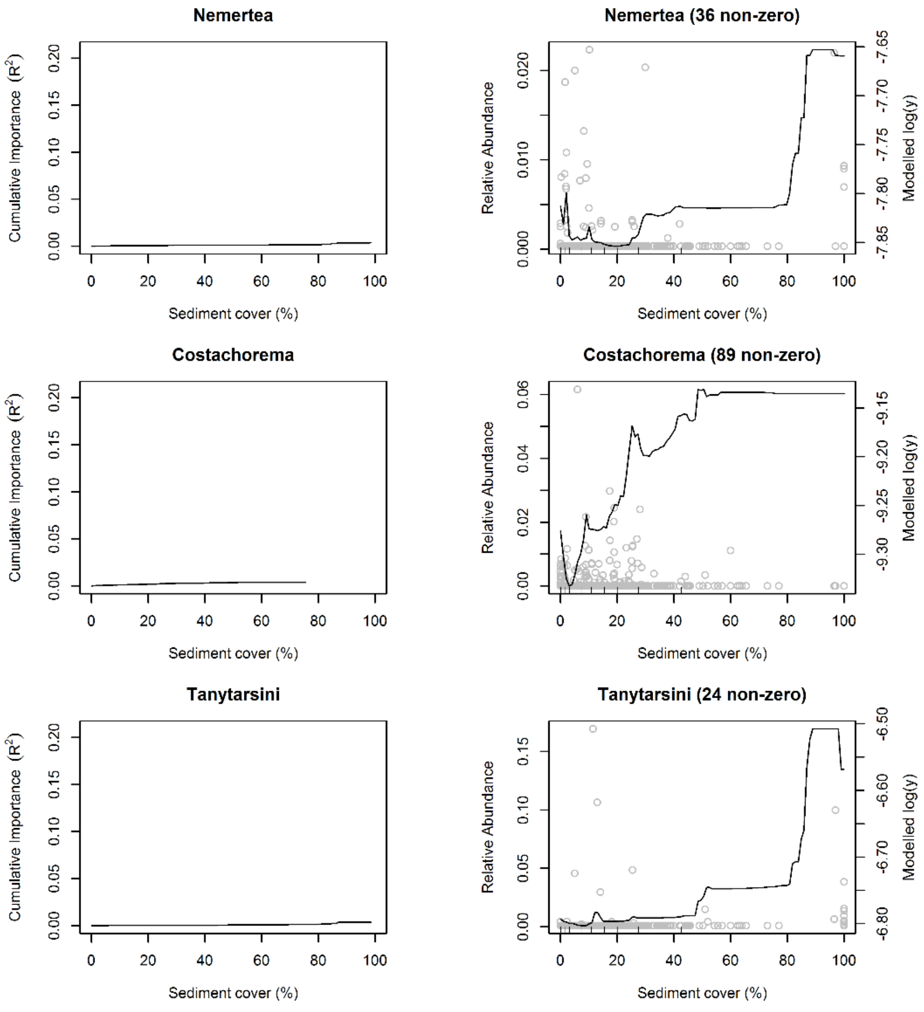
Taxon 37 - 39



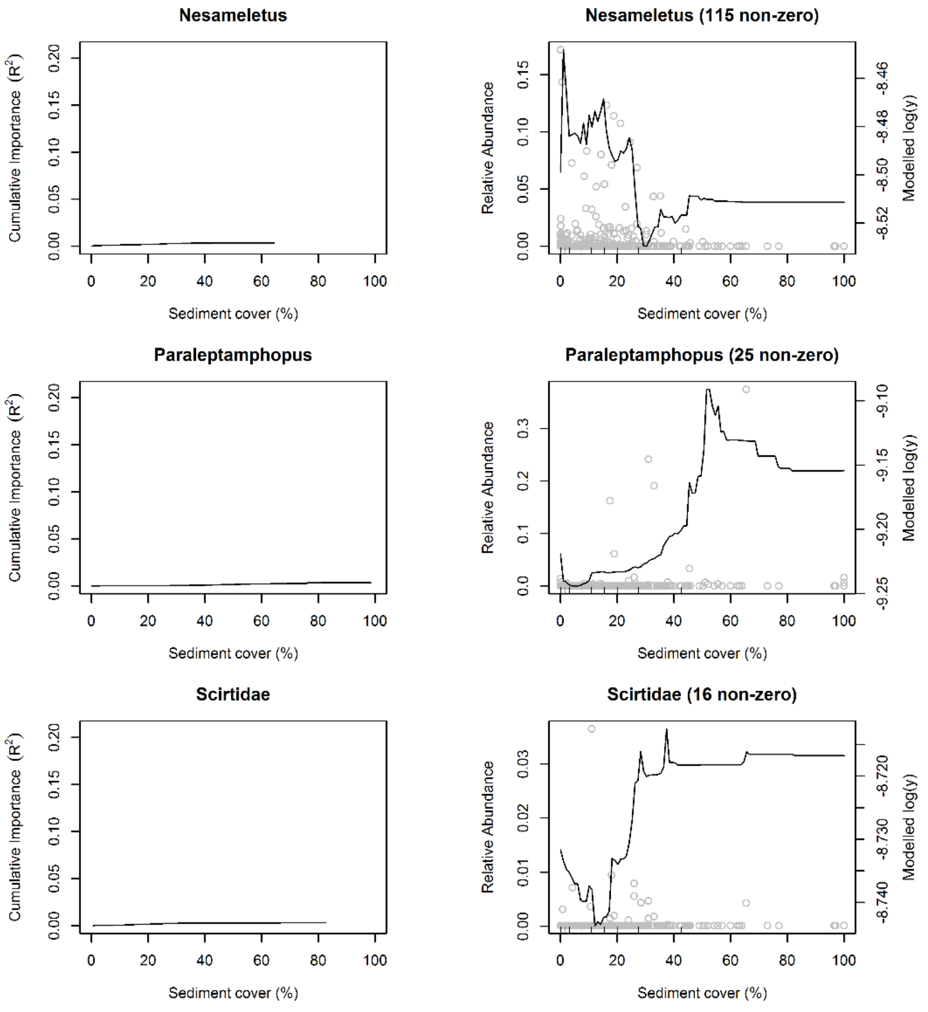
Taxon 40 - 42

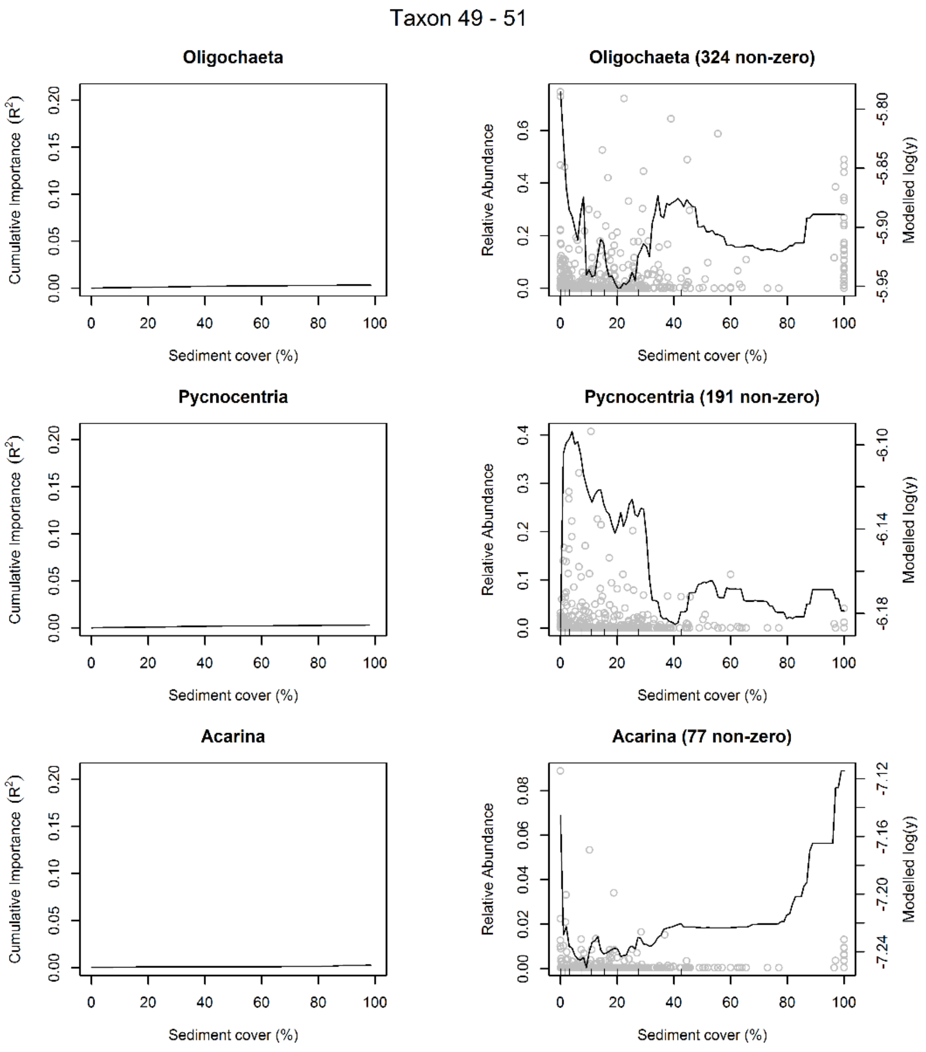


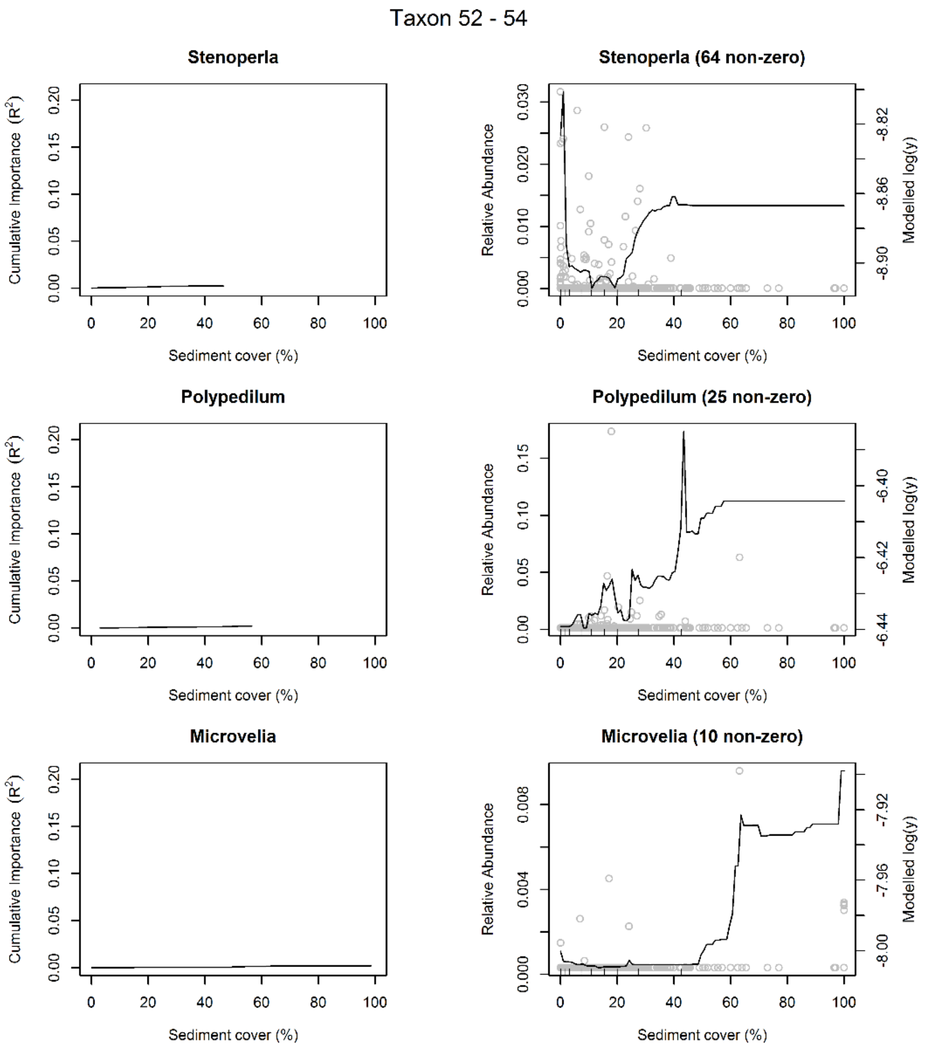
Taxon 43 - 45



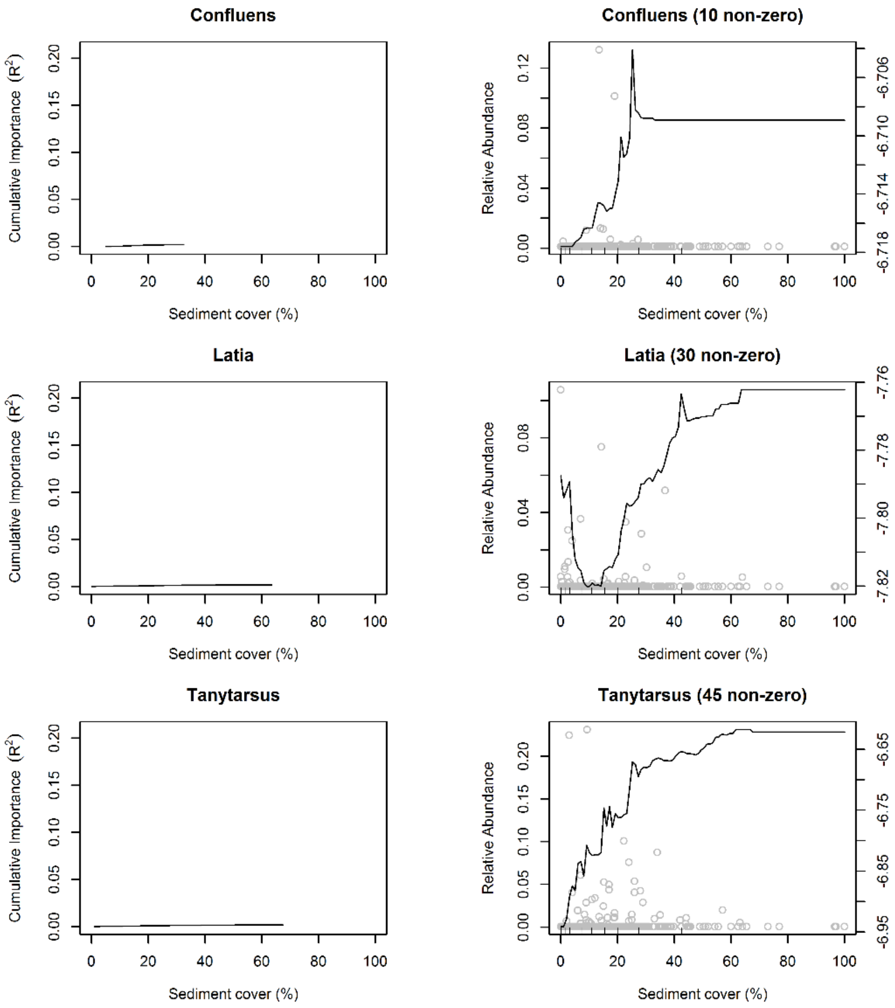
Taxon 46 - 48



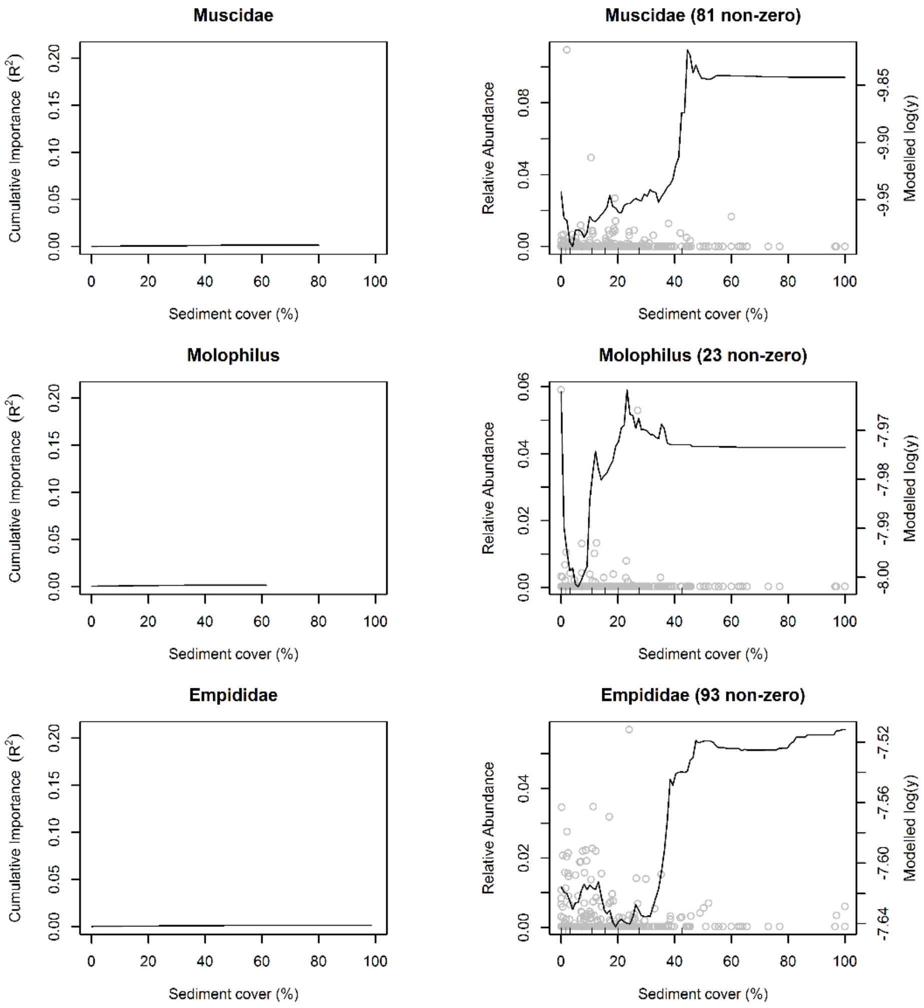

Taxon 49 - 51
Oligochaeta
Cumulative Importance (R^2)
0.00 0.05 0.10 0.15 0.20
0 20 40 60 80 100
Sediment cover (%)
Oligochaeta (324 non-zero)
Relative Abundance
0.0 0.2 0.4 0.6
Modelled log(y)
-5.95 -5.90 -5.85 -5.80
Sediment cover (%)
Pycnocentria
Cumulative Importance (R^2)
Sediment cover (%)
Pycnocentria (191 non-zero)
Relative Abundance
0.0 0.1 0.2 0.3 0.4
Modelled log(y)
-6.18 -6.14 -6.10
Sediment cover (%)
Acarina
Cumulative Importance (R^2)
Sediment cover (%)
Acarina (77 non-zero)
Relative Abundance
0.00 0.02 0.04 0.06 0.08
Modelled log(y)
-7.24 -7.20 -7.16 -7.12
Sediment cover (%)


Taxon 52 - 54
Stenoperla
Cumulative Importance (R^2)
0.00 0.05 0.10 0.15 0.20
0 20 40 60 80 100
Sediment cover (%)
Stenoperla (64 non-zero)
Relative Abundance
0.000 0.010 0.020 0.030
Modelled log(y)
-8.90 -8.86 -8.82
Sediment cover (%)
Polypedilum
Cumulative Importance (R^2)
Sediment cover (%)
Polypedilum (25 non-zero)
Relative Abundance
0.00 0.05 0.10 0.15
Modelled log(y)
-6.44 -6.42 -6.40
Sediment cover (%)
Microvelia
Cumulative Importance (R^2)
Sediment cover (%)
Microvelia (10 non-zero)
Relative Abundance
0.000 0.004 0.008
Modelled log(y)
-8.00 -7.96 -7.92
Sediment cover (%)

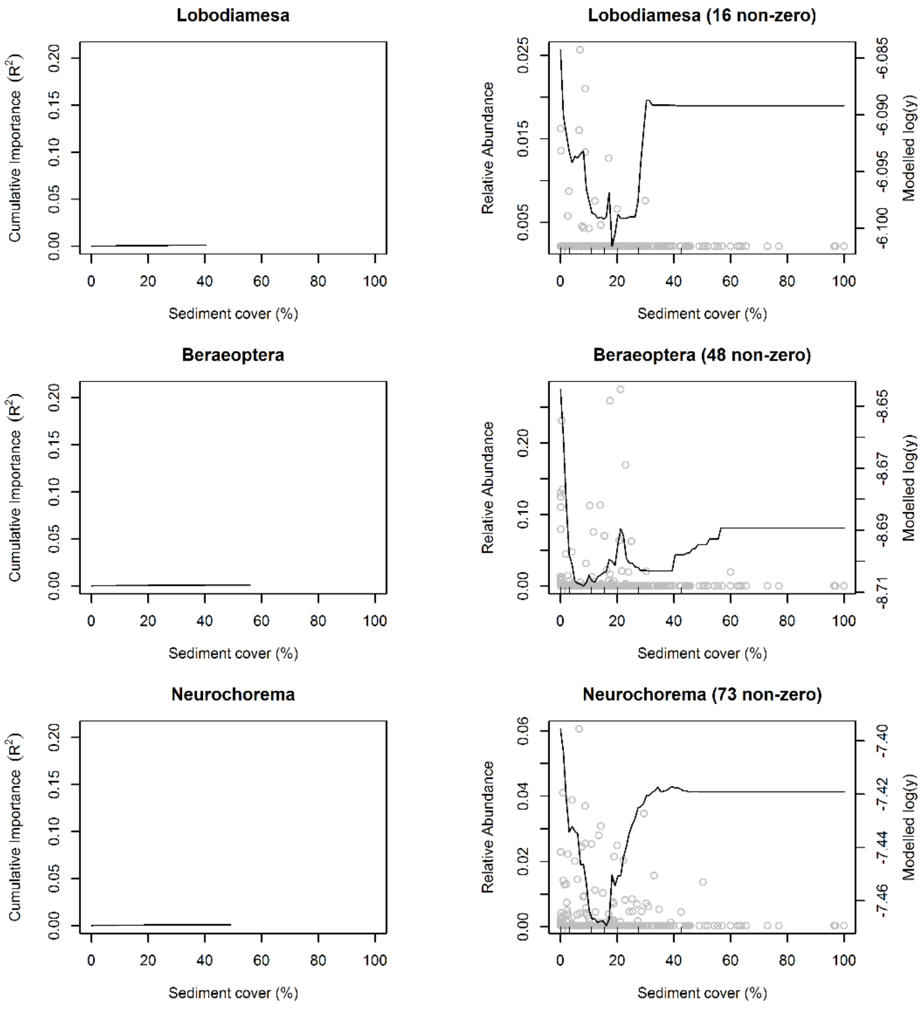
Taxon 55 - 57



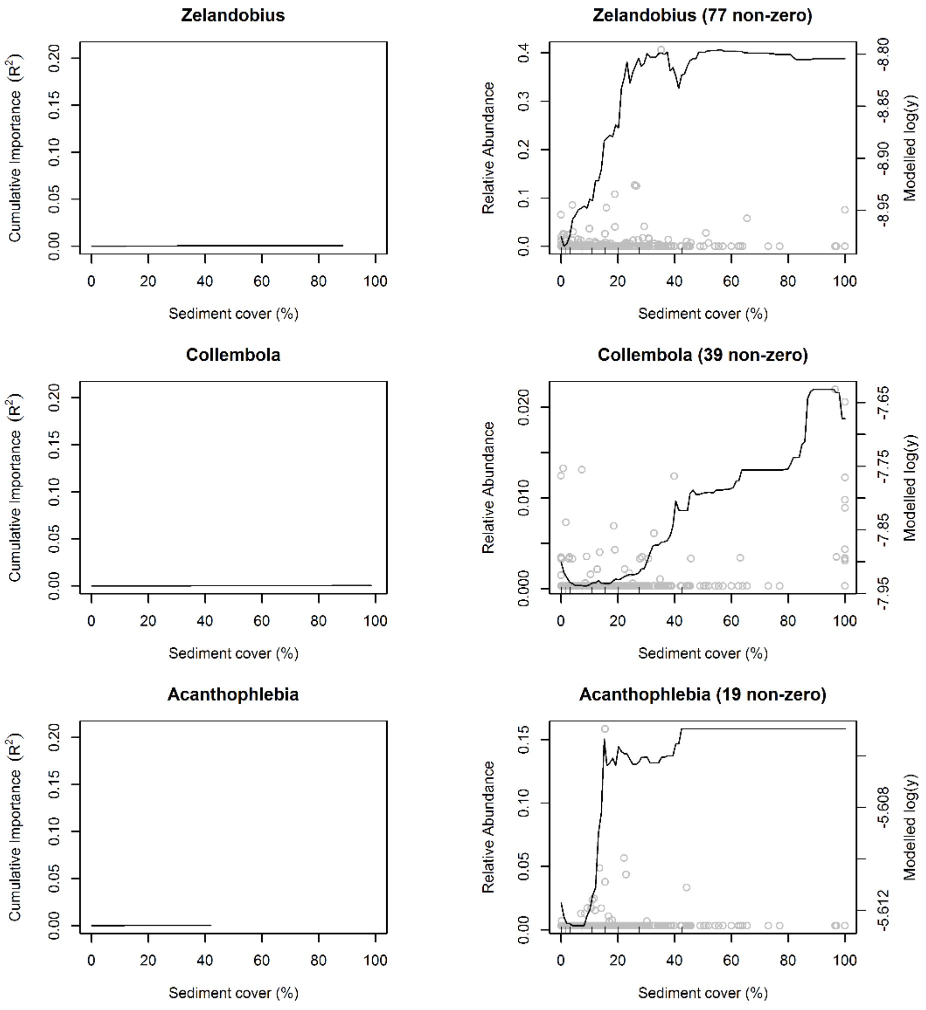
Taxon 58 - 60



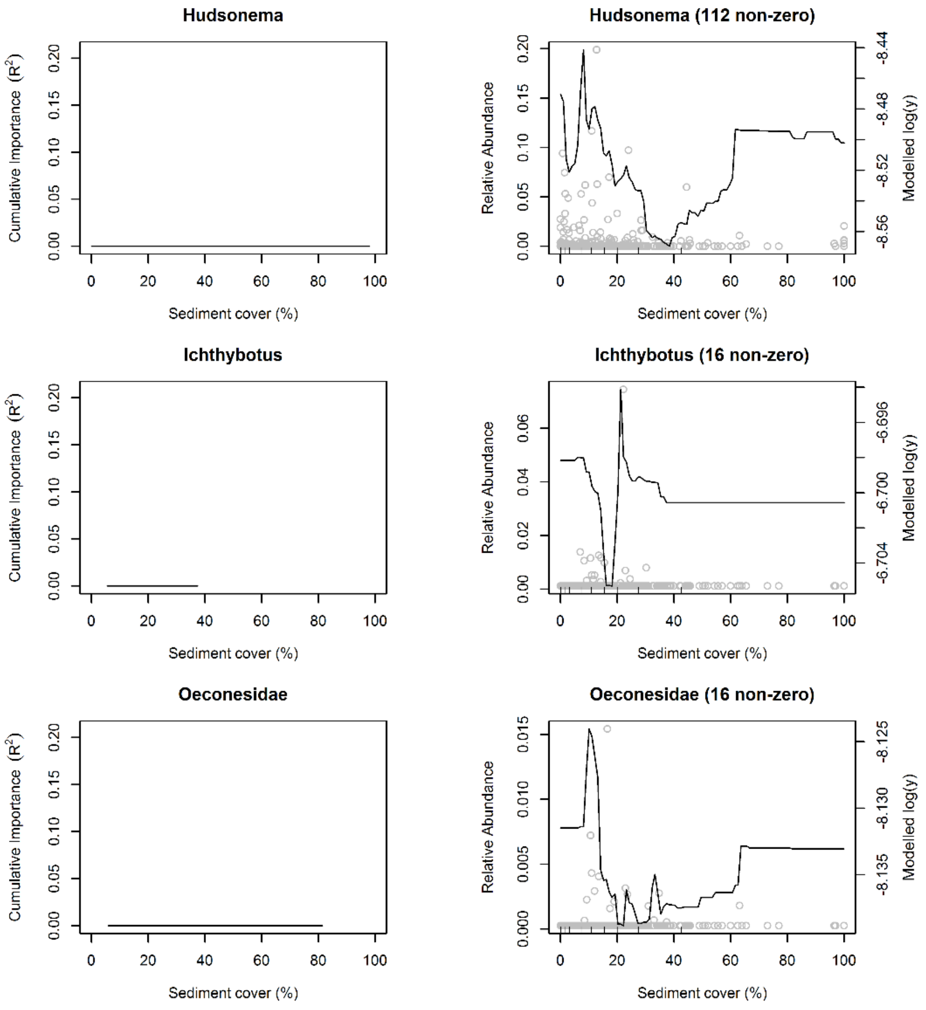
Taxon 61 - 63



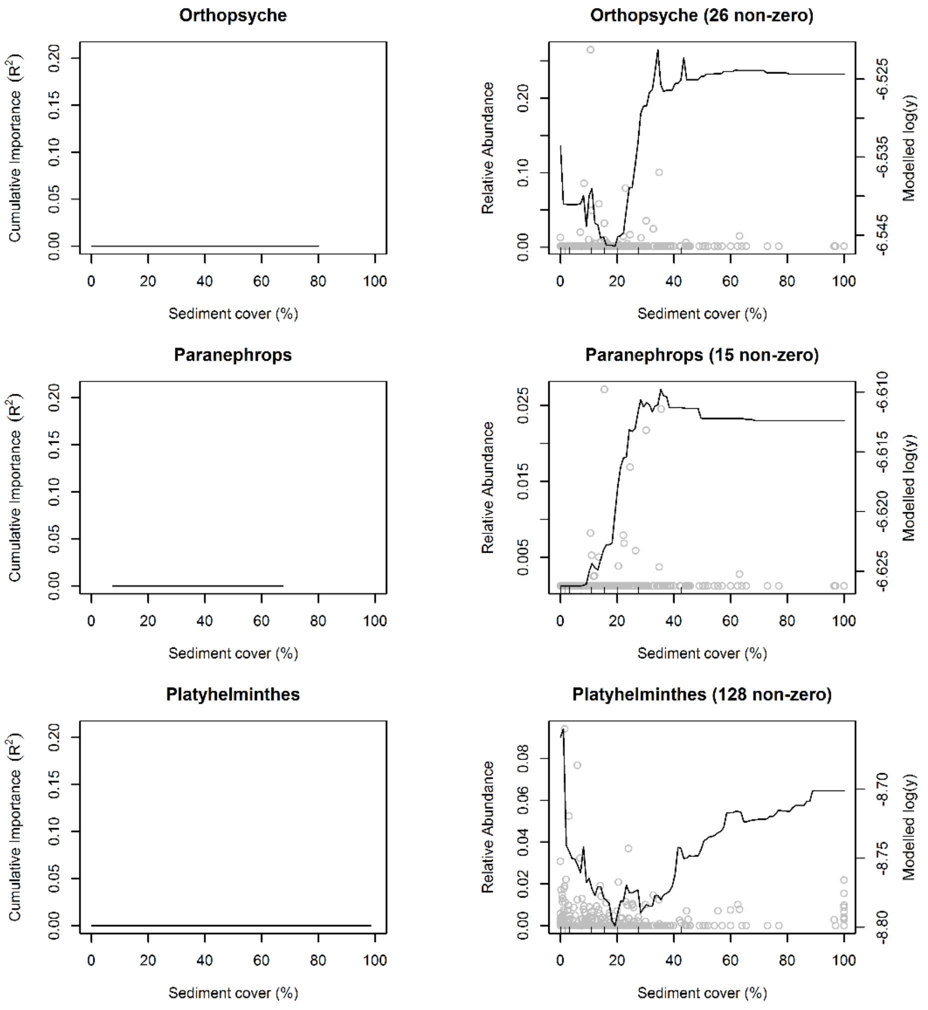
Taxon 64 - 66



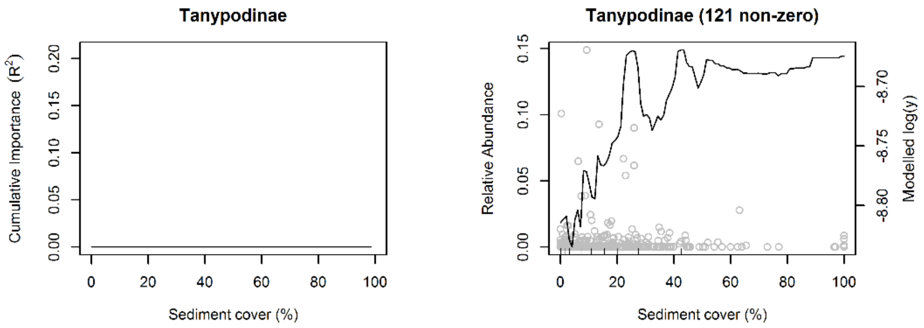
Taxon 67 - 69



Taxon 70 - 72



Taxon 73 - 73


Tanypodinae
Cumulative Importance (R^2)
0.00
0.05
0.10
0.15
0.20
0
20
40
60
80
100
Sediment cover (%)
Tanypodinae (121 non-zero)
Relative Abundance
0.00
0.05
0.10
0.15
Modelled log(y)
-8.80
-8.75
-8.70
0
20
40
60
80
100
Sediment cover (%)

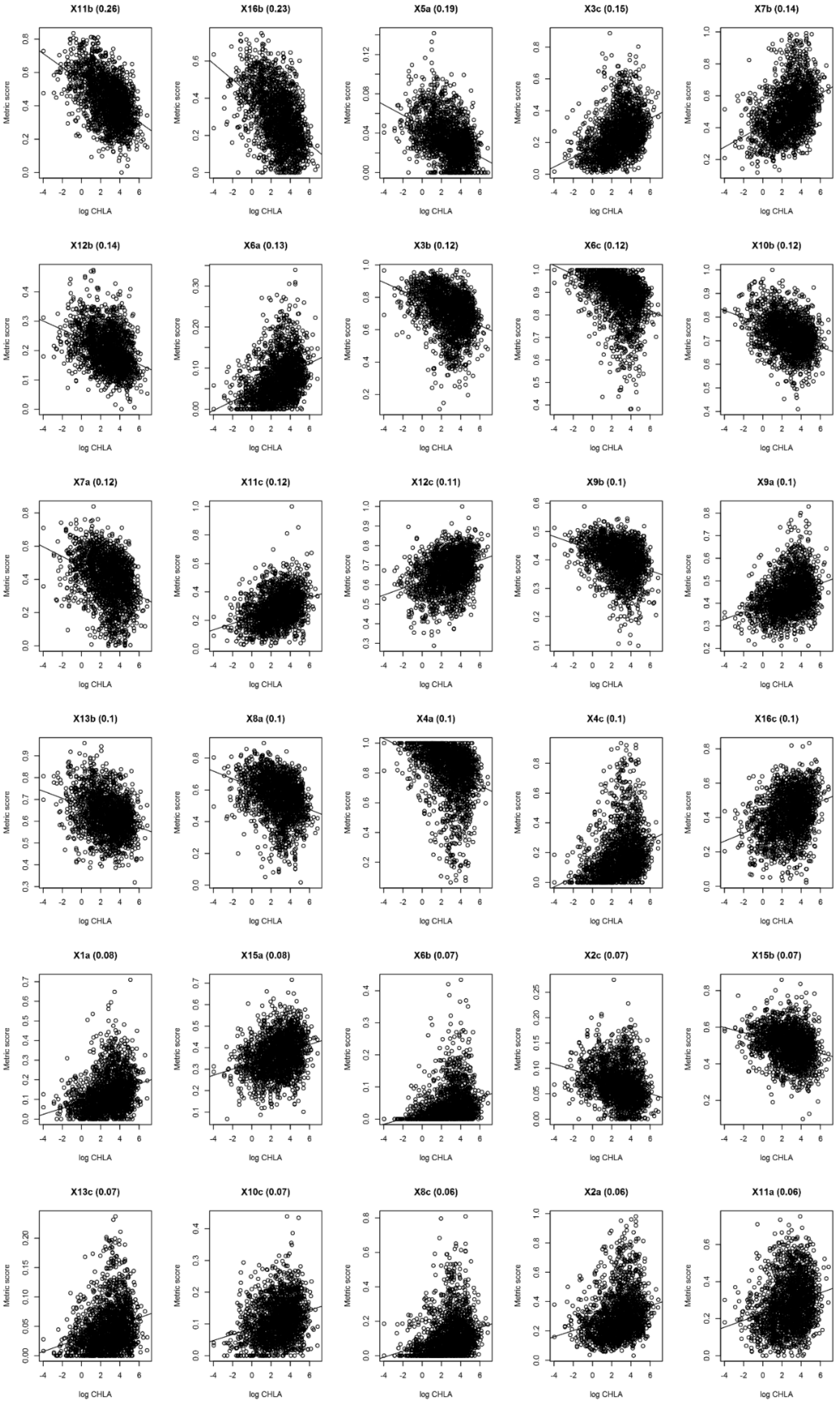
Appendix 4. Model results and names of the taxa, ordered alphabetically, that did not show a clear response shape to chlorophyll-*a* (CHLA) or sediment cover (SEDCOVER) and/or were not expected to be responsive to the stressor gradient by the experts. See Appendix 2 for details on model results presented.

Taxon	**Stressor**	**Response group**	R^2_{fp}	**Threshold**	**No. of occurrences**	R^2
Ameletopsis	CHLA	none	0.000	49	12	0.30
Amphipoda	CHLA	none	0.000	114	17	0.07
Archichauliodes	CHLA	none	0.016	43	260	0.51
Chironomidae	CHLA	none	0.008	97	162	0.53
Cladocera	CHLA	none	0.000	193	17	0.14
Collembola	CHLA	none	0.001	65	39	0.02
Coloburiscus	CHLA	none	0.005	36	183	0.37
Confluens	CHLA	none	0.000	150	10	0.61
Copepoda	CHLA	none	0.000	125	23	0.52
Corynoneura	CHLA	none	0.000	327	21	0.54
Costachorema	CHLA	none	0.004	79	89	0.19
Elmidae	CHLA	none	0.008	67	437	0.47
Eriopterini	CHLA	none	0.005	41	113	0.19
Ferrissia	CHLA	none	0.020	34	19	0.37
Hexatomini	CHLA	none	0.000	105	18	0.06
Hudsonema	CHLA	none	0.003	65	112	0.03
Hydrobiosis	CHLA	none	0.005	69	372	0.26
Lobodiamesa	CHLA	none	0.023	118	16	0.18
Maoridiamesa	CHLA	none	0.001	52	41	0.16
Megaleptoperla	CHLA	none	0.006	13	51	0.35
Microvelia	CHLA	none	0.004	73	10	0.08
Muscidae	CHLA	none	0.024	80	81	0.15
Nematoda	CHLA	none	0.010	90	45	0.69
Nemertea	CHLA	none	0.007	32	36	0.08
Neurochorema	CHLA	none	0.034	34	73	0.42
Oeconesidae	CHLA	none	0.000	56	16	0.14
Olinga	CHLA	none	0.000	279	126	0.23
Ostracoda	CHLA	none	0.010	71	109	0.44
Paracalliope	CHLA	none	0.020	58	115	0.38
Paranephrops	CHLA	none	0.000	125	15	0.14
Physa	CHLA	none	0.028	54	110	0.36
Platyhelminthes	CHLA	none	0.019	52	128	0.11
Polypedilum	CHLA	none	0.006	36	25	0.13
Polyplectropus	CHLA	none	0.000	327	22	0.14
Psilochorema	CHLA	none	0.019	73	247	0.23
Ptilodactylidae	CHLA	none	0.000	110	12	0.55
Sphaeriidae	CHLA	none	0.006	80	78	0.42
Stenoperla	CHLA	none	0.010	95	64	0.32

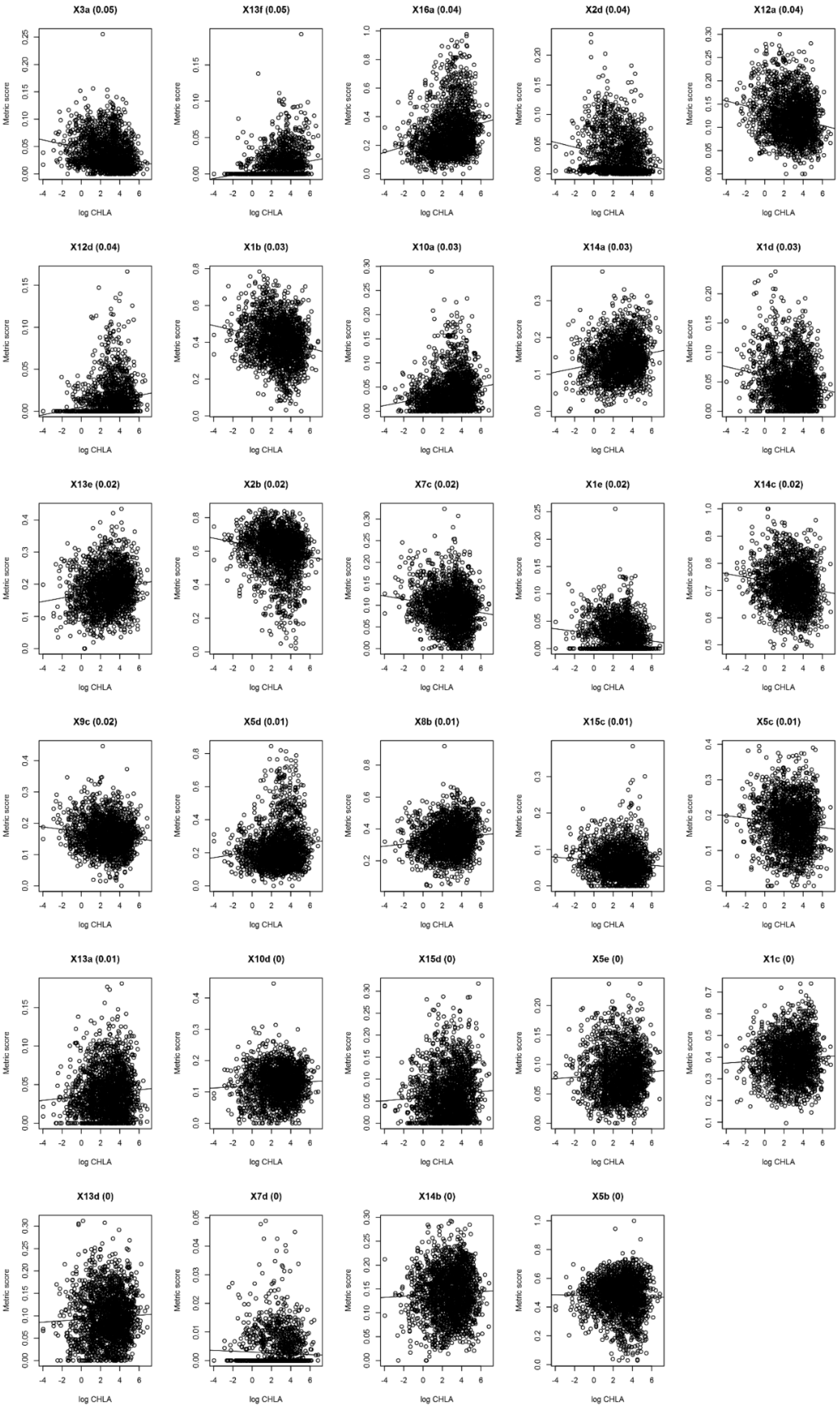
Appendix 4, continued

Tanypodinae	CHLA	none	0.002	65	121	0.04
Tanytarsini	CHLA	none	0.001	32	24	0.08
Tanytarsus	CHLA	none	0.000	9	45	0.02
Acarina	SEDCOVER	none	0.003	23	77	0.08
Amphipoda	SEDCOVER	none	0.006	18	17	0.07
Chironomidae	SEDCOVER	none	0.005	40	162	0.53
Cladocera	SEDCOVER	none	0.025	82	17	0.14
Collembola	SEDCOVER	none	0.001	46	39	0.02
Copepoda	SEDCOVER	none	0.209	81	23	0.52
Empididae	SEDCOVER	none	0.001	12	93	0.32
Eriopterini	SEDCOVER	none	0.013	17	113	0.19
Ferrissia	SEDCOVER	none	0.007	23	19	0.37
Hexatomini	SEDCOVER	none	0.004	18	18	0.06
Hudsonema	SEDCOVER	none	0.000	98	112	0.03
Ichthybotus	SEDCOVER	none	0.000	38	16	0.48
Maoridiamesa	SEDCOVER	none	0.010	10	41	0.16
Microvelia	SEDCOVER	none	0.002	49	10	0.08
Molophilus	SEDCOVER	none	0.001	7	23	0.05
Muscidae	SEDCOVER	none	0.001	11	81	0.15
Nemertea	SEDCOVER	none	0.004	28	36	0.08
Oeconesidae	SEDCOVER	none	0.000	82	16	0.14
Orthocladiinae	SEDCOVER	none	0.012	16	313	0.37
Orthopsyche	SEDCOVER	none	0.000	80	26	0.48
Ostracoda	SEDCOVER	none	0.038	61	109	0.44
Oxyethira	SEDCOVER	none	0.008	44	188	0.22
Paracalliope	SEDCOVER	none	0.004	23	115	0.38
Paraleptamphopus	SEDCOVER	none	0.004	42	25	0.47
Paralimnophila	SEDCOVER	none	0.004	22	19	0.18
Paranephrops	SEDCOVER	none	0.000	68	15	0.14
Physa	SEDCOVER	none	0.009	78	110	0.36
Platyhelminthes	SEDCOVER	none	0.000	99	128	0.11
Polypedilum	SEDCOVER	none	0.002	19	25	0.13
Polyplectropus	SEDCOVER	none	0.012	19	22	0.14
Potamopyrgus	SEDCOVER	none	0.007	22	402	0.38
Tanypodinae	SEDCOVER	none	0.000	99	121	0.04
Tanytarsini	SEDCOVER	none	0.004	63	24	0.08
Tanytarsus	SEDCOVER	none	0.002	16	45	0.02
Triplectides	SEDCOVER	none	0.004	19	33	0.10
Zelandobius	SEDCOVER	none	0.001	17	77	0.03

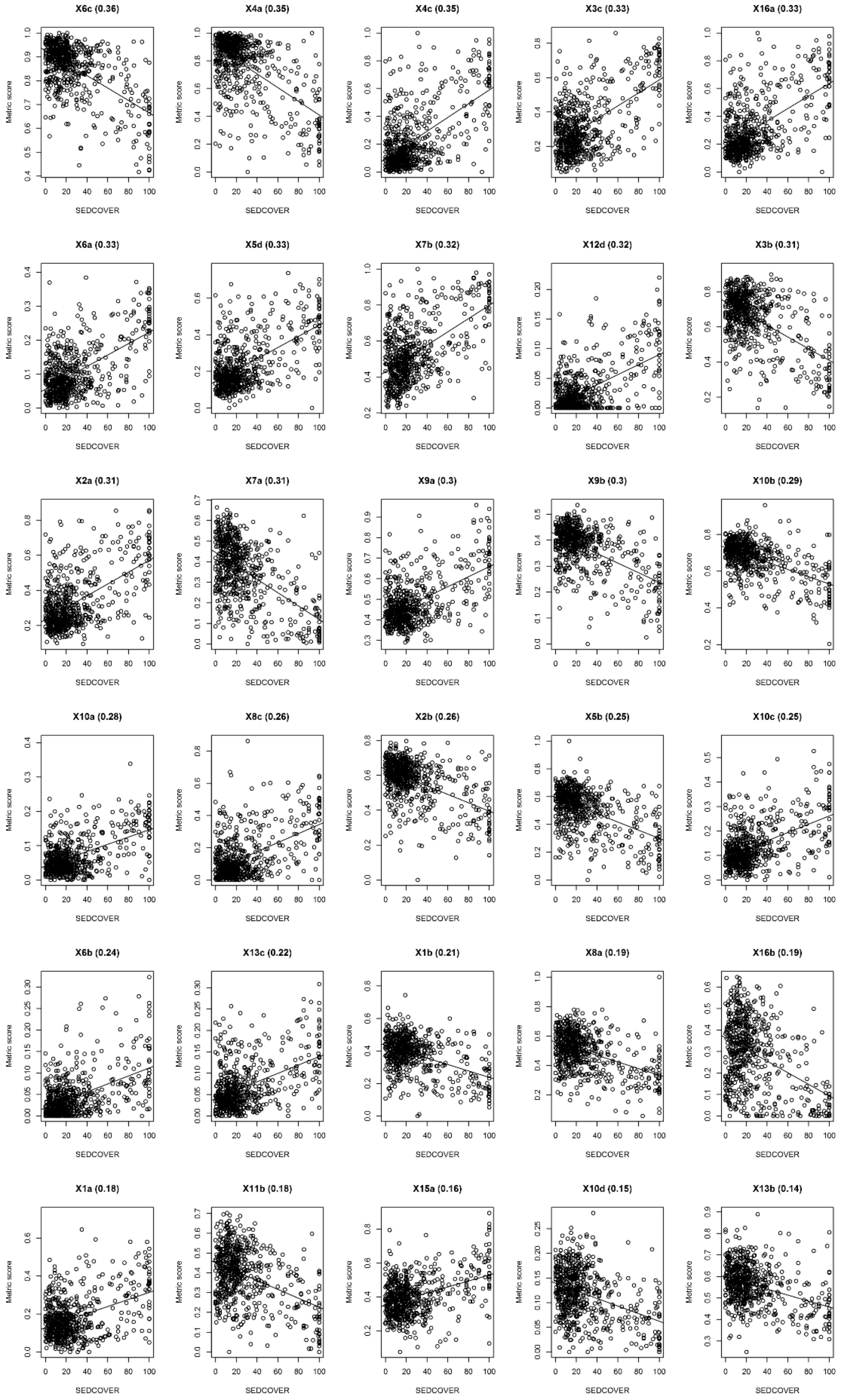
Appendix 5. Scatterplots along with the simple linear regression models for all modalities of the 16 traits across the chlorophyll-*a* (CHLA in mg/m^2) gradient built on field observational data. The metrics are ordered according to their adjusted R^2 values (R^2 given in parentheses).



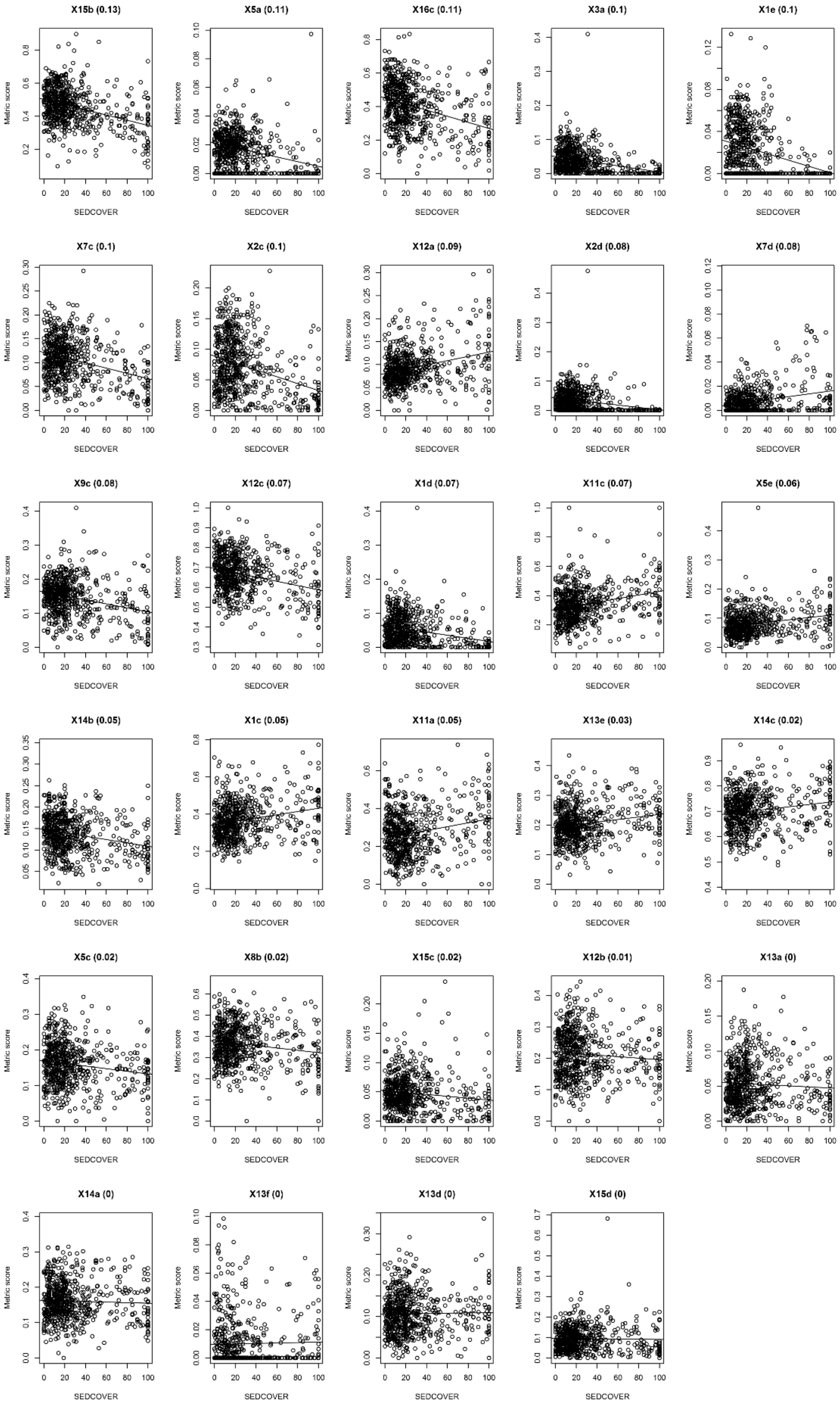
Appendix 5, continued



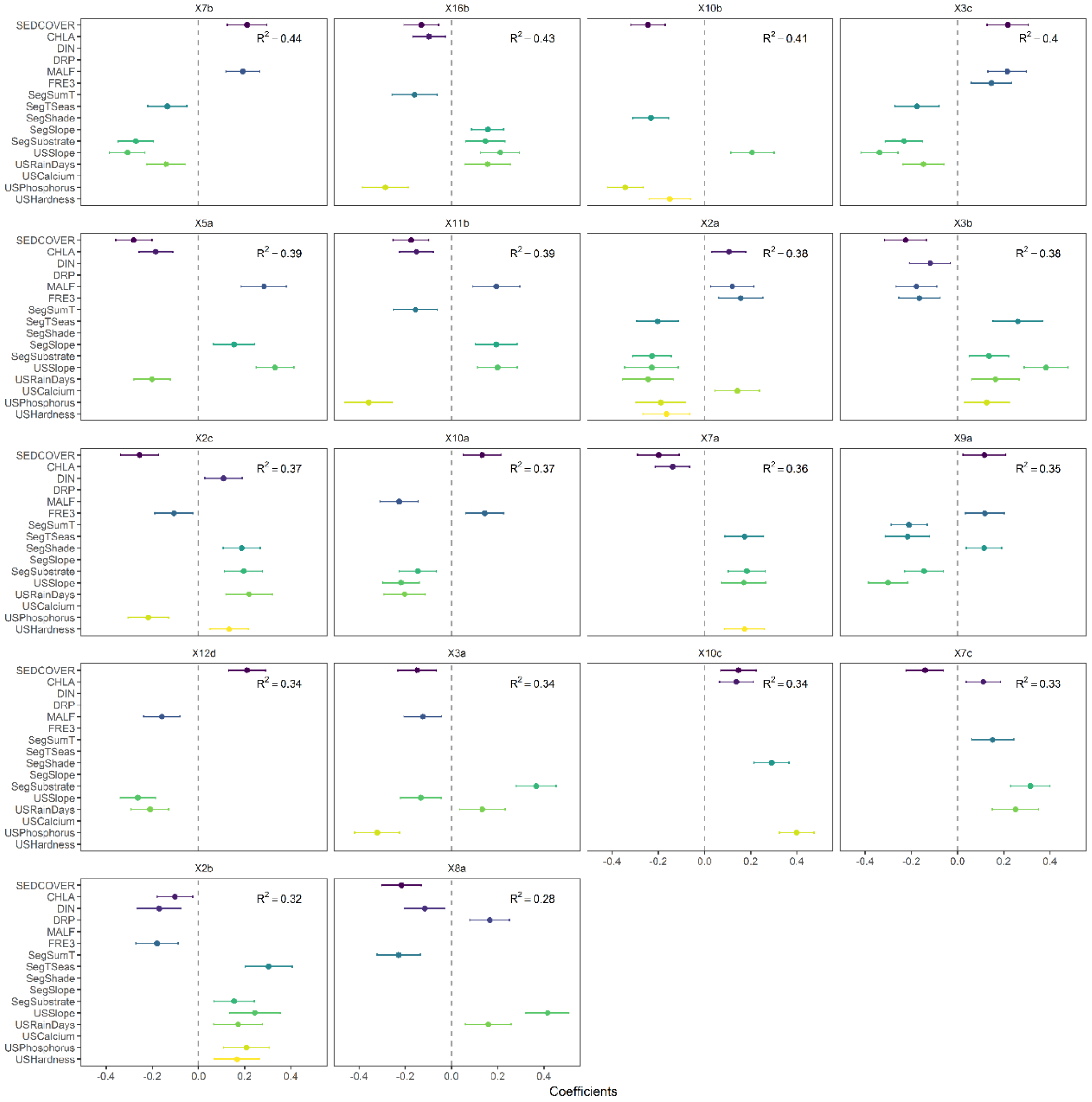
Appendix 6. Scatterplots along with the simple linear regression models for all modalities of the 16 traits across the % sediment cover (SEDCOVER) gradient built on field observational data. The metrics are ordered according to their adjusted R^2 values (R^2 given in parentheses).



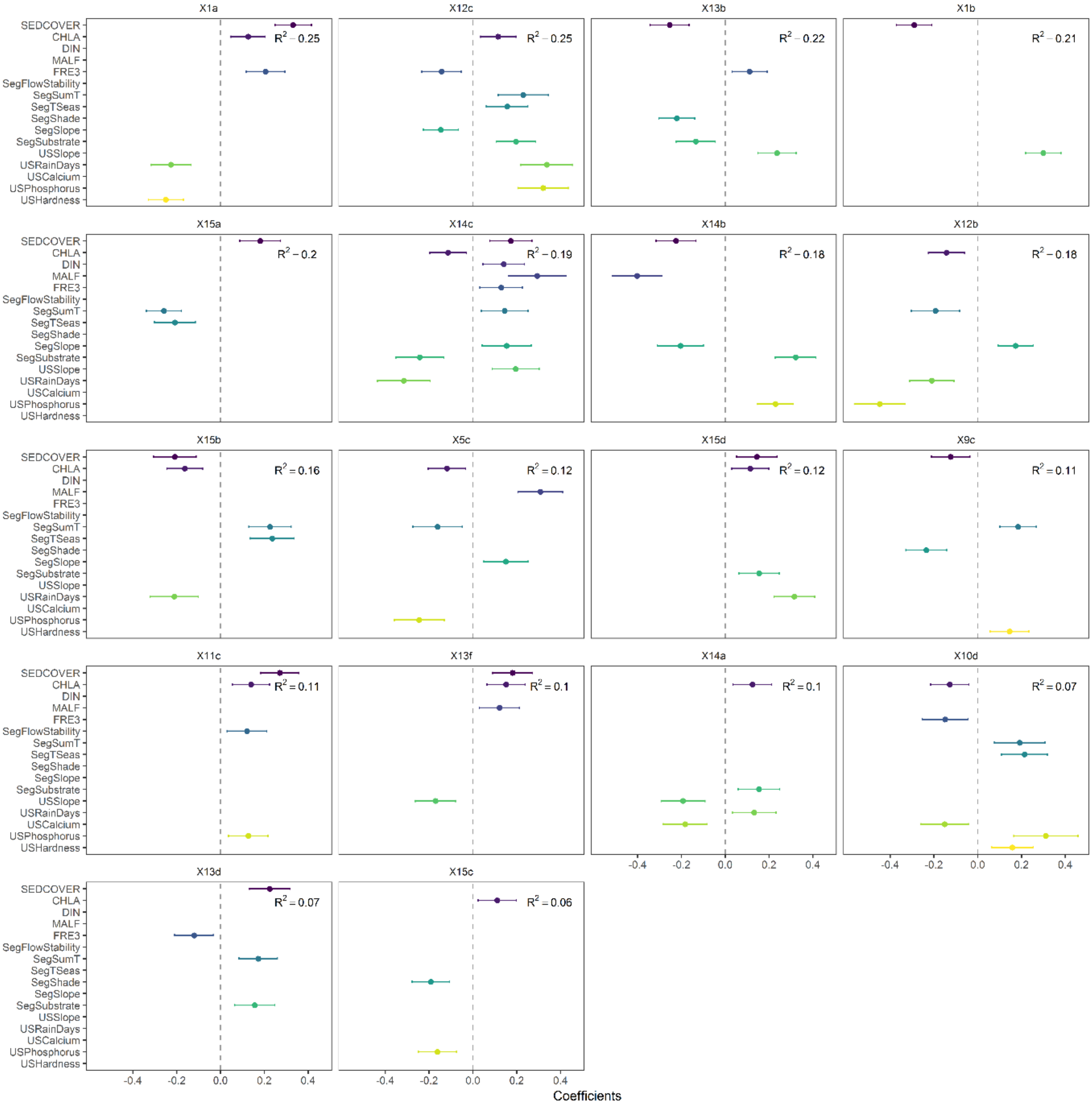
Appendix 6, continued



Appendix 7. Regression coefficients (± 95% CI) of the predictors in the multiple linear regression models for all trait metrics for which at least CHLA or SEDCOVER was a predictor in the model built on field observational data. Metrics are ordered according to decreasing R^2 values. See Figure 11 caption for more details.



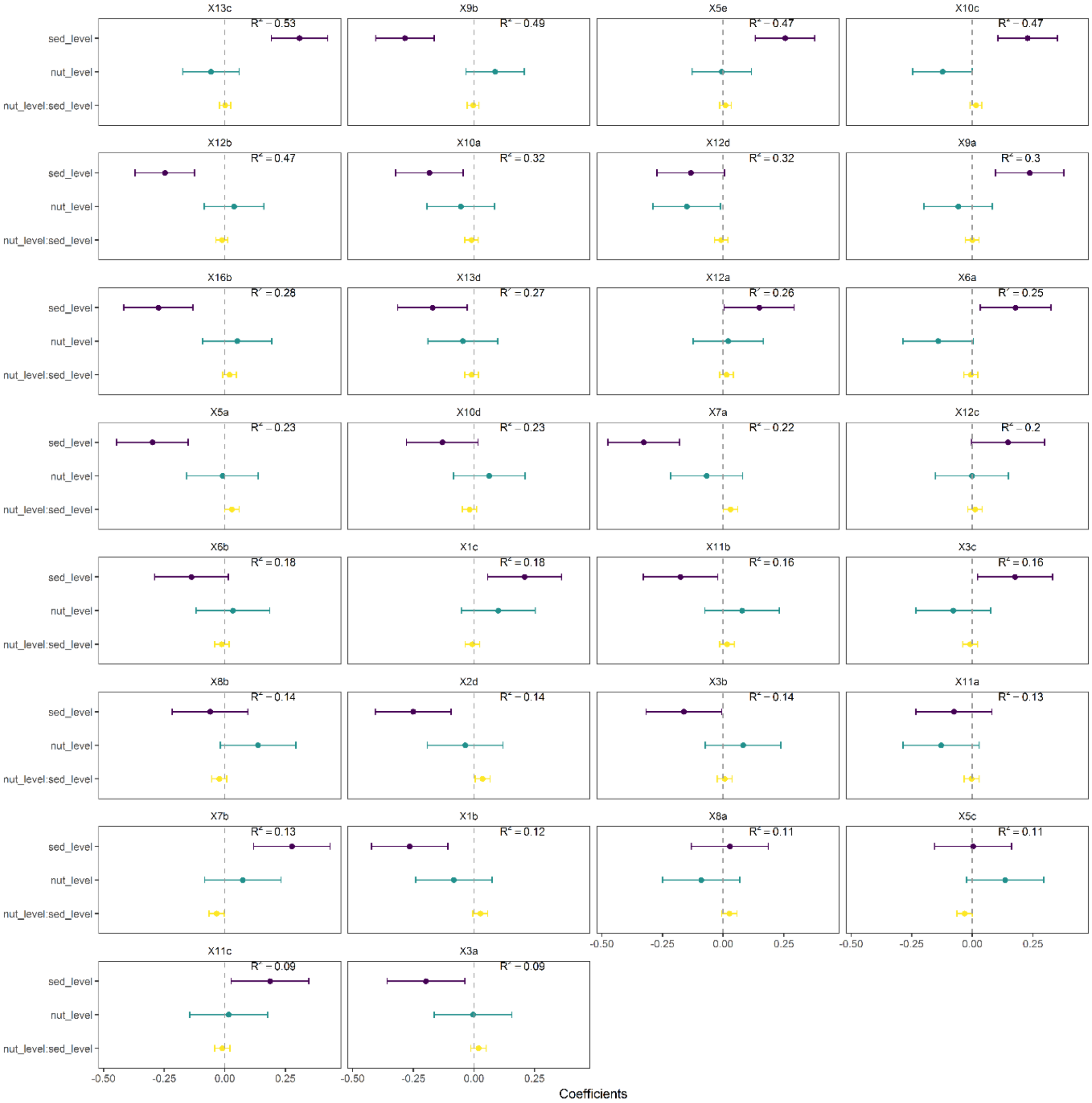
Appendix 7, continued



Appendix 8. Relative importance of predictor variables determined by hierarchical variance partitioning from linear models for all trait metrics for which at least CHLA or SEDCOVER was a predictor in the model built on field observational data. Metrics are ordered according to decreasing R^2 values. See Figure 11 caption for more details.

Appendix 8, continued

Appendix 9. Regression coefficients (± 95% CI) of the predictors in the multiple linear regression models for all 59 macroinvertebrate trait modalities built on experimental data. Metrics are ordered according to decreasing R^2 values. See Figure 18 caption for more details.



Appendix 9, continued