Relationships between multiple land-use pressures and individual and combined indicators of stream ecological integrity

Joanne Clapcott, Roger Young, Eric Goodwin, John Leathwick and David Kelly

newzealand.govt.nz

Department of
Conservation
Te Papa Atawhai

CONTENTS

Relationships between multiple land-use pressures and individual and combined indicators of stream ecological integrity

Joanne Clapcott[1], Roger Young[1], Eric Goodwin[1], John Leathwick[2] and David Kelly[3]

[1] Cawthron Institute, Private Bag 2, Nelson 7010, New Zealand.
Email: Joanne.Clapcott@cawthron.org.nz

[2] National Institute for Water and Atmospheric Research, Private Bag 1115, Hamilton 3216, New Zealand.

[3] Department of Conservation, Research and Development, PO Box 357, Westport 7866, New Zealand.

Abstract

The identification and prioritisation of rivers for their conservation management is a key issue facing the Department of Conservation, and effective tools for doing so are needed. This study investigated the relationships between indicators of river health and measures of human pressure, and evaluated the usefulness of land-use pressure measurements for predicting the ecological integrity (EI) of New Zealand streams. Fourteen metrics of stream health based on fish, invertebrate, water quality and ecosystem process data were used and tested against three land-use pressure gradients (native vegetation cover, impervious cover and predicted nitrogen loading). We used boosted regression trees (BRT) to quantify the shape of the response of each metric to the pressure gradients. The strongest metrics were oxidised nitrogen concentrations, $\delta^{15}N$ of primary consumers and the macroinvertebrate community index, while the weakest overall indicators were gross primary productivity, fish richness and invertebrate richness. Monotonic decreases in response to native vegetation clearance and impervious cover were generally found, while distinct non-monotonic responses were observed for some indicators in relation to predicted nitrogen loading. Environmental variability, summarised by Freshwater Environments of New Zealand groups, significantly improved the explanatory power of individual BRT models. However, the direct use of environmental variables provided even better model performance. Output from the BRT analyses was used to predict observed and expected values for each metric for all stream reaches in New Zealand. Ten metrics calculated using the observed and expected data were then combined to produce a multi-metric index (MMI) of EI. Ecological response curves for the MMI suggested distinct thresholds for change in EI in response to land-use pressure—a potentially important implication for sustainable water resource planning and management.

Keywords: river health, land use, indicators, ecological response curves, multi-metric index, environmental variability

1. Introduction

1.1 Background

One of the key themes in the New Zealand Government's Sustainable Water Programme of Action[1] was to identify Waters of National Importance (WONI) with respect to their values for natural heritage conservation, recreation, culture and heritage, irrigation, energy generation, industrial use and tourism. In 2006, the Department of Conservation (DOC) and Landcare Research took on the task of identifying aquatic systems of national importance for freshwater natural heritage. They developed a series of pressure measures (e.g. land-use intensity, percentage catchment clearance) derived from several national databases. The purpose was to assess the ecological integrity (EI) of river (Chadderton et al. 2004) and wetland (Ausseil et al. 2008) ecosystems. Unlike traditional site-specific biotic indices or water quality indicators, these measures are both spatially explicit, comprehensive and scaleable from stream reach to the nation. However, there are few empirical data from which to characterise relationships between measures of pressure and EI to support or interpret the measures. A working draft that produces pressure estimates as part of the WONI model (Leathwick & Julian 2007) uses a series of conceptual models (Fig. 1)[2] to identify ways in which the EI of freshwaters might respond across a gradient of land-use pressure. Therefore, there is a need to quantify these responses using empirical rather than conceptual models. This research was initiated to address this need.
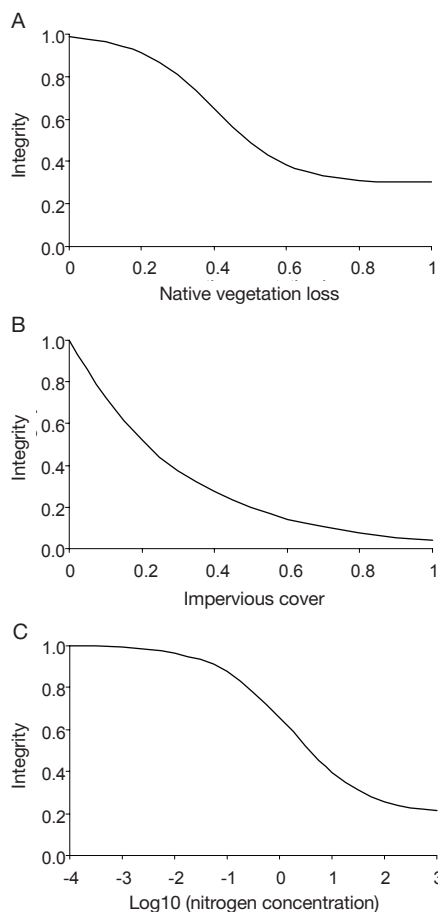
Figure 1. Examples of conceptual ecological response curves that relate variation in pressure factors with ecological condition (A.—Native vegetation loss, B.—Impervious cover, C.—Nitrogen load) used in the production of pressure estimates for New Zealand river catchments.

1.2 New Zealand rivers

In New Zealand, the ecological status of every river is influenced to some degree by human pressures. From channelised urban streams to streams in national parks, the ecology of each stream will reflect a level of human influence. Each stream will also be influenced by natural environmental variability, with streams ranging from cool-dry, volcanic, hill-country streams to warm-wet, low-elevation, sedimentary streams. Hence, the contemporary ecological status of New Zealand's rivers can be described by a combination of the natural background environment and the quantification of human pressures.

1.3 Defining ecological integrity (EI)

In 2006, a technical advisory group was established to develop a working definition of EI for rivers and wetlands in New Zealand. Drawing strongly from conceptual development in the literature (Karr 1981, 1996; Angermeir 1994; Barbour et al. 2000) and years of experience in a New Zealand setting, the group defined ecological integrity as:

The degree to which the physical, chemical and biological components (including composition, structure, and process) of an ecosystem and their relationships are present, functioning and maintained close to a minimally impacted reference condition.

(Schallenberg et al. 2011)

---

[1] www.mfe.govt.nz/issues/water/prog-action/index.html (viewed 10 March 2011).

[2] Based on figure in Leathwick & Julian (2007); see also discussion on the influence of land use on stream ecosystems.

From this definition, it was evident that EI can be viewed as an emergent property of complex systems encompassing components such as pristineness, nativeness, diversity and resilience. Following this logic, assessment of EI should incorporate knowledge of each of its components, and a list of indicators was proposed—each of which potentially represents a component (Table 1).

The Schallenberg et al. (2011) definition of ecological integrity further suggests that ecosystems can vary in their degree of EI depending on their deviation from a reference condition. They proposed that the most dominant factor leading to deviation from a reference condition would be human-induced disturbance. Human land-use pressures or stressors present a press-type disturbance (Resh et al. 1988; Lake 2000). Press disturbances lead to an increasing divergence from the reference condition in both ecosystem structure and function and where they are ongoing events, change may arise sharply but be maintained over a long-time frame.

1.4 Determining reference condition

To assess the departure from a natural state requires knowledge of the original natural, or reference, condition. According a site reference status is based on it meeting minimal land-use criteria or in-stream chemical, physical or biological criteria (Maxted et al. 2000); the use of best professional judgement (Pont et al. 2006; Whittier et al. 2007); the interpretation of its historical condition using, for example, photographs or sediment cores (Hughes et al. 1998); or the assessment of its ambient condition (e.g. top 5% of all sites becomes least disturbed sites;

Table 1.   Indicators of the ecological integrity (EI) of rivers proposed by Schallenberg et al. (2011).

GENERAL PROPERTY OF EI	INDICATOR	EXAMPLES OF MAIN STRESSORS THAT MAY BE DETECTED
Nativeness	Native fish (% Native, no. of introduced species, O/E*)	Invasion and introduction
	Presence of invasive macrophytes/algae	
Pristineness	Macroinvertebrate community composition (MCI†, %EPT‡)	Multiple disturbances
Structural	Fish IBI§	Invasion and introduction
Functional	Ecosystem metabolism Wood decomposition rates BOD $\delta^{15}N$ of primary consumers	Eutrophication, habitat degradation, flow abstraction Eutrophication, change in land use Organic enrichment Specific N and P enrichment
Physicochemical	Water clarity, turbidity Nutrient concentrations Water temperature, dissolved oxygen	Eutrophication, sedimentation Eutrophication, sedimentation Riparian and catchment clearance, abstraction
Diversity	Macroinvertebrate taxonomic richness, diversity, O/E richness	Multiple disturbances
	Abiotic structure (habitat template)	Change in physical template, abstraction, irrigation
Resilience	Presence/absence of key indicator taxa Ecosystem metabolism	Multiple disturbances Change in physical template

*   O/E = Observed : expected.
†   MCI = Macroinvertebrate community index.
‡   EPT = Ephemeroptera, Plecoptera and Trichoptera, excluding Hydroptilidae.
§   IBI = Index of biotic integrity.

Stoddard et al. 2008). In New Zealand, reference sites have been categorised as pristine sites, near pristine sites, best management sites and local benchmark sites (Boothroyd et al. 2002). Most commonly, in the development of MMIs, the mean or range in values for a group of reference sites is used to determine the reference condition for a test site.

Recently, empirical models have been used to predict reference condition from the relationship between human pressures and ecological response. For example, Angradi et al. (2009) used the *y*-intercept in a quantile regression model to predict reference values for fish indices for large rivers in the absence of human pressures.

For many stream types in New Zealand, there is a distinct lack of information on reference condition due to the widespread development of landscapes for human use (e.g. lowland streams). Therefore, assessing EI requires either accepting a nearest-neighbour reference condition or predicting a reference condition based on an established relationship between land uses, environmental variability and ecological response for that stream type. The observed condition (O) could then be compared to expected condition (E) to assess EI. Ultimately, future ecological condition under differing land-use scenarios could also be predicted in a similar way (Van Sickle et al. 2004).

1.5 The influence of land use on stream ecosystems

Land use provides a measure of human pressure that is commonly used to estimate in-stream condition based on the adage that 'in every respect, the valley rules the stream' (Hynes 1975). There is a substantial body of literature that addresses the effects of human land use on the physical (Roth et al. 1996; Norris et al. 2007), chemical (Herlihy et al. 1998; Harris 2001) and biological (Harding et al. 1999; Van Sickle et al. 2004) properties of rivers and streams. Correlative relationships are often used to illustrate how an increase in catchment development leads to a decline in stream integrity (Allan et al. 1997; Burcher et al. 2007; Death & Collier 2009; Young & Collier 2009). Unfortunately, most studies are limited in their geographical extent or in their assessment of stream integrity, focusing on a single property or group of organisms (e.g. fish) to infer the integrity of the whole system. Furthermore, looking for strictly linear correlations can impose a restrictive view on what can be complex relationships between human pressures and stream integrity, particularly if the study is based on a small set of samples that inadequately capture the broad range of background environmental conditions.

Recently, broad-scale assessments of nations and entire continents have been instigated to address the concern of the limited spatial coverage of data. Results confirm degradation in stream quality as a result of human pressures; for example, in southern Australia, over 85% of stream reaches were significantly modified (Norris et al. 2007) and in the USA, over 42% of selected sites were classified as having poor biological condition (67% were less than 'good'; Paulsen et al. 2008). To date, most of these broad-scale studies have adopted a categorical approach whereby streams are grouped into land-use categories (e.g. non-impacted, impacted). Using this approach to assess EI would result in a simple ordinal classification such as 'high' EI or 'low' EI and does not provide the avenue to define a deviation from a natural state as outlined in our definition. Hence, a gap in the current literature is an assessment of EI as a continuous response to a full gradient of land-use pressure and across a broad range of environmental conditions.

Meta-analyses involving multiple indicators of stream health are increasingly common as agencies take advantage of an accumulation of data collected for state-of-the-environment reporting (e.g. Marchant et al. 2006; Carlisle et al. 2008). As databases increase in scope, there is the opportunity to quantify the response of EI to human pressures for use in broad-scale resource and conservation planning. The challenge then becomes how to adequately investigate complex interactions among predictors and response variables. Attempts to do so often use some type of component analysis that groups land-use information. Recent advances in statistical

modelling techniques, such as classification and regression tree models, provide a means to better understand complex relationships without the need to reduce or summarise information. Interactive relationships can be characterised and used in nationwide predictions of response variables (for example, the distribution of freshwater fish in New Zealand; Leathwick et al. 2008a).

1.6 Natural variability and Freshwater Ecosystems of New Zealand (FENZ)

An obstacle to applying predictive models of stream integrity based on broad-scale land use is the potential for environmental variability at varying spatial scales to dominate patterns. Hence, it is necessary to take into account how conditions vary naturally before it is possible to measure departure from a natural state. Stream typing or bioregional classification is widely used to ensure that sites are compared to relevant reference conditions. Whether it be ecoregions (Harding & Winterbourn 1997; Herlihy et al. 1998; Paulsen et al. 2008), bioregions (Pont et al. 2006; Leathwick et al. 2008b), drainage basins (Norris et al. 2007) or stream classes such as upland-lowland-coastal (King et al. 2005; Fellows et al. 2006), the grouping of sites into similar types has provided a good framework to evaluate stream conditions. By contrast, Hawkins et al. (2000) have suggested that broad-scale classification is best used to ensure that all 'regions' are representatively sampled rather than as a categorical assessment of stream condition, particularly when adequately measured environmental variables can be used to predict reference values independent of stream type.

In New Zealand, the entire spatial coverage of rivers and streams has been classified by segment in two complementary geographical information layers—the River Environment Classification (REC; Snelder et al. 2004) and the Freshwater Ecosystems of New Zealand (FENZ, previously FWENZ; Leathwick et al. 2008b). Such classifications can be used to provide a categorical framework for the broad-scale assessment of stream condition, while the information on which the classifications are based can alternatively be used to provide a more continuously varying description of environmental conditions. Hence, the classification of stream reaches in New Zealand could be used in two ways:

- To determine the representative coverage of indicators of EI (and, therefore, the spatial limits of model predictions based on those data)
- To provide environmental measures to investigate the relationship between indicators of EI and natural environmental variability

In this study, we address natural variability by quantifying the relationship between indicators of stream integrity and land use at a national scale using data that incorporate an almost full range of FENZ stream types. We also test the FENZ classification of streams by investigating whether natural variability, as a function of stream 'type', significantly influences the response of indicators to land-use pressures, compared with use of the underlying continuous environmental variables that were used to construct that classification.

1.7 Types of indicators

There has been a strong focus on assessing biotic endpoints to infer the ecological condition of rivers based on the proviso that biological integrity will directly reflect physical and chemical integrity (Barbour et al. 2000). This is probably due in part to the widespread use of macroinvertebrate and fish data in river health assessments and because of a general understanding of what an increase or decrease in the relative abundance of species and community composition means. Indices of biological integrity often include measures of diversity (richness and composition), tolerance and functional traits of organisms in addition to abundance (Karr 1981). The quantitative Macroinvertebrate Community Index (QMCI), for example, can be calculated from the abundance of organisms with varying tolerances of organic pollution (Stark 1998).

Arguments have recently been made for the assessment of the functional integrity of ecosystems (Bunn & Davies 2000; Gessner & Chauvet 2002), but the widespread uptake of indicators of ecosystem processes has been limited so far. Recent research has helped to provide insight into the response of functional indicators to disturbance and natural gradients of variability (Anderson & Cabana 2005; Roberts et al. 2007; Clapcott & Barmuta 2010) and has provided methodology for including functional indicators in holistic assessments of stream integrity (Gessner & Chauvet 2002; Young et al. 2008). As such, indicators of ecosystem processes now form part of regular river health assessments in Australia (Fellows et al. 2006).

Measuring a suite of indicators provides a more holistic assessment than measuring a single biotic endpoint (see section 1.8). Equally, the measurement of numerous indicators allows us to assess all components of EI including nativeness, pristineness, diversity and resilience as defined by Schallenberg et al. (2011) (Table 1).

1.8 The MMI approach

Karr (1981) was the first to suggest the combination of fish metrics to create an index of biological integrity (IBI). The IBI has formed the cornerstone of multi-metric index (MMI) development and has been widely adopted in river assessment, with single taxa most commonly being the focus. Multi-metric indexes have been developed using fish (in New Zealand—Joy & Death 2004; in the USA—Hughes et al. 1998, Stoddard et al. 2008; and in Europe—Pont et al. 2006), macroinvertebrates (in New Zealand—Collier 2008; in the USA—Maxted et al. 2000, Stribling et al. 2008; and in Europe—Vlek et al. 2004), diatoms (Cao et al. 2007), and periphyton (Hill et al. 2000). The Rapid Bioassessment Protocols (Plafkin et al. 1989) adopted by the USEPA and the River Habitat Survey (Raven et al. 1998) used in the UK could also be considered to follow the MMI approach because they assess multiple components of stream habitat. However, their use does not necessarily require a development phase as is common for other MMIs. As reviewed by Stoddard et al. (2008), the development of an MMI involves several steps:

1. Collect data from sites designated *a priori* as having various environmental conditions, e.g. reference versus impacted (see section 1.4 for what is meant by 'reference' in the EI context)
2. Classify candidate metrics into classes including richness, evenness and/or diversity, composition, function and/or habit and tolerance
3. Evaluate the range of values and reproducibility of data in terms of a signal-to-noise ratio
4. Adjust for varying responses across natural gradients, e.g. stream size
5. Test for redundancy between candidate metrics
6. Determine scores by comparison to reference conditions and sum metrics

The exploration of an MMI in this study differs from this approach because it does not involve the collection of new data and is, instead, based on the meta-analysis of previously compiled indicators of stream condition. Furthermore, it is not based on the identification of reference sites *a priori*; instead, we look at the response of indicators across full disturbance gradients rather than differences between categorical groupings of sites. Finally, and perhaps most importantly, the inclusion of candidate metrics is determined by a recognition that conceptual inclusiveness (i.e. the need to measure EI components of pristineness, diversity, nativeness and resilience as outlined in Schallenberg et al. (2011)) needs to be balanced by management-focussed indicators (i.e. measures that have been and can be widely adopted and communicated).

Combined, candidate metrics would provide a multi-assemblage, multi-component index of EI.

1.9 Why combine different groups of organisms?

There have been an increasing number of studies measuring a combination of stream components as methodologies have been developed and biological indicators more widely applied (e.g. Smith & Storey 2001; Griffith et al. 2005; Carlisle et al. 2008; Johnson & Hering 2009). Specifically, recent comparative studies of multiple indicators have shown how different groups of organisms provide complementary information on ecological condition. For example, fish and macroinvertebrate indices show differing sensitivities to land use (Flinders et al. 2008), and such differences have also been documented when diatoms were included in similar studies (Hering et al. 2006; Johnson & Hering 2009). The investigations by Carlisle et al. (2008) showed how a single group evaluation indicated impaired conditions on average much less often than when several groups were used to assess stream condition.

If different organisms respond similarly to the same pressure then this strengthens the inference of change. On the other hand, if organisms respond differently to the same pressure or respond to different pressures, then the responses could be used to assign varying grades of pressure and/or elucidate the interactive effects of different pressures (Davies & Jackson 2006; Johnson & Hering 2009).

While we might be tempted to ask 'which group is best?' when selecting indicators of EI (see Downes et al. (2002) or Resh (2008) for a review of advantages and disadvantages of organism-specific assessments), clearly there are benefits in using multiple indices of stream condition, especially when there are likely to be multiple gradients of environmental and land-use pressure variables influencing EI.

1.10 Study goals and rationale

The goals of this study were to characterise the responses of indicators of the ecological integrity of rivers to land-use pressure gradients and then to use this information in the development of a multi-metric index. To address the primary goal, we:

- Compiled a national database of indicators of ecological integrity
- Conducted field sampling to fill gaps in the database
- Investigated the response of indicators (i.e. developed ecological response curves)
- Investigated the universality of the responses by examining the response across different stream types
- Identified how additional environmental variability contributes to characterising the relationship between land use and EI

To address the second goal, we:

- Modelled the reference condition for each candidate metric
- Calculated an O/E measure for each candidate metric
- Determined the weighting functions for candidate metrics for combination in the MMI
- Characterised the ecological response curve of an MMI to disturbance gradients

2. Methods

2.1 Characterising ecological response curves

There is a wealth of potential indicators of EI. A recent American study to develop an index of biological integrity based on fish alone identified over 230 potential indices (Whittier et al. 2007). The modelling approach we adopted was relatively computationally complex and so, by necessity, only a limited number of indicators could be investigated. The selection of indicators was guided by data availability, support in the relevant literature and a pilot analysis that identified significant relationships between indicators and individual land-use pressures (Scarsbrook 2008). In the future, with development and improved knowledge of variability, alternative indicators could also be examined using the methods applied here.

The response variables we investigated were grouped into four groups of indicators—water quality, ecosystem processes, invertebrate and fish metrics. Water quality, fish and macroinvertebrate data were compiled from existing datasets, whereas measures of ecosystem processes were compiled from a combination of existing data and field sampling in 2008. Within each indicator group, the quantity and spatial distribution of sample sites varied substantially.

2.1.1 Water quality metrics

Water quality data came from a Ministry for the Environment-funded compilation of regional and national State of the Environment data covering the period of 1996–2002 (Larned et al. 2003). Median values for dissolved reactive phosphorus (*DRP*), nitrates and nitrites (*NOx*), and black disk clarity (*Clarity*) were calculated for each site (Table 2, Fig. 2).

Table 2. Description of the indicators of river ecological integrity including the mean and range of values used in this study. 'Group' is the number of 100-level FENZ groups in the database. 'Coverage' is the percentage of stream reaches at a national scale represented by the sampled Groups.

ECOLOGICAL INDICATOR		*N*	MEAN (RANGE)	GROUP (*N*)	COVERAGE (%)
Water quality					
Clarity	Black disk clarity (m)	525	2.08 (0.10, 10.43)	16	78
DRP	Dissolved reactive phosphorus (ppm)	578	0.020 (0.001, 0.646)	12	78
NOx	Nitrates-nitrites (ppm)	360	0.605 (0.01, 8.45)	16	74
Invertebrates					
Cycle	Percentage of macroinvertebrate taxa reproducing only once (%)	2626	82.5 (0, 100)	40	92
EPT	Percentage of taxa being Ephemeroptera, Plecoptera, Trichoptera (%)	2626	50.5 (0, 100)	40	92
ITR	Invertebrate taxonomic richness	2626	17.0 (1, 53)	40	92
MCI	Macroinvertebrate Community Index	2666	115.9 (20, 180)	38	91
Fish					
F-IBI	Fish index of biological integrity	6858	34.26 (0, 60)	35	95
Native	Percentage of native fish taxa (%)	6858	75.3 (0, 100)	35	95
FTR	Fish taxonomic richness	6858	2.62 (1, 11)	35	95
Ecosystem processes					
Cotton	Cellulose decomposition potential (*k*/dd)	108	0.004 (0.001, 0.026)	9	55
Del15N	Delta15N of primary consumers (‰)	73	6.75 (1.03, 20.99)	5	31
ER	Ecosystem respiration (gO$_2$ m^{-2} d^{-1})	156	9.936 (0.34, 55.78)	13	65
GPP	Gross primary productivity (gO$_2$ m^{-2} d^{-1})	156	4.603 (0.078, 35.99)	13	65

Figure 2.   Maps of New Zealand showing location of sample data sites for each of the four groups of indicators—ecosystem processes (functional indicators), water quality, invertebrates and fish.

2.1.2 Invertebrate metrics

Macroinvertebrate data were compiled from a variety of sources, including regional councils, NIWA, Cawthron Institute, and Otago, Canterbury and Massey Universities. The dataset contained 136 taxa identified to the level of taxonomic resolution of the Macroinvertebrate Community Index (*MCI*) (Stark et al. 2001). Any taxa not in the MCI list were not included in analyses. Invertebrate data were used to calculate several metrics commonly used in stream health assessment including the number of taxa present or invertebrate taxonomic richness (*ITR*), percentage of total taxa belonging to Ephemeroptera, Plecoptera and Trichoptera (*EPT*) excluding Hydroptilidae (Maxted et al. 2003), and the percentage of taxa reproducing only once in a life cycle (*Cycle*) (Doledec et al. 2006). MCI scores were calculated by assigning pollution tolerance scores to 86 specific taxa (average values were used for several pairs or small groups of taxa that were amalgamated into single entities to preserve consistency across the different data sets making up the broader data collection) (Table 2) .

2.1.3 Fish metrics

Fish data were obtained from the NZ Freshwater Fish Database (Richardson 1989) and included samples collected using electric fishing methods post-1980 (Table 2). Fish taxonomic richness (*FTR*) was calculated as the number of fish taxa present (sites with no taxa were excluded from analyses) and percentage of native fish taxa (*Native*) was also calculated. Finally, an index of biotic integrity (*F-IBI*) developed for New Zealand freshwater fish was calculated for each site using the method of Joy & Death (2004).

2.1.4 Ecosystem process metrics

Ecosystem process data were compiled from published and unpublished datasets (R.G. Young, Cawthron Institute) taken over a range of seasons from 1994 to 2008. Data were collected using consistent methodology as described in Clapcott et al. (2009). Ecosystem process metrics included gross primary productivity (*GPP*), ecosystem respiration (*ER*), cellulose decomposition potential (*Cotton*) and the $\delta^{15}N$ of primary consumers (*Del15N*) (Table 2).

2.1.5 Environmental data sources

As part of the development of the River Environment Classification (REC) database (Snelder et al. 2004), every stream segment in New Zealand (i.e. the portion between two confluences) was given a unique numerical code. This 'NZREACH' number was used to collate a spatially-referenced dataset of environmental variables from the FENZ database of pressures and predictors for each site with indicator data (Leathwick et al. 2008b). Variables are described in Table 3. Additionally, sites were assigned a 100-level FENZ classification for some analyses (see boosted regression tree analysis below, in section 2.1.6). This classification provides a biologically optimised, hierarchical grouping structured predominantly by similarities in climate, glacial influence, slope, and in-stream habitat. At the 100-group level, all 576 180 stream segments in New Zealand are divisible into 100 classes. Sites were unevenly distributed among these classes for each group of metrics (see Fig. 2).

2.1.6 Boosted regression tree analysis

Sixteen indicators of stream EI were initially chosen to investigate relationships with land-use pressures. Strong correlations between some water quality measures gave good reason to reduce the indicator list to the 14 listed in Table 2 for the statistical analyses. Excluded indices were total nitrogen (*TN*; correlated to *NOx*, $R^2 = 0.982$, $P < 0.001$, $n = 320$) and total phosphorus (*TP*; correlated to *DRP*, $R^2 = 0.894$, $P < 0.001$, $n = 446$). Despite significant correlations between some other indicators (*MCI* and *EPT*, $R^2 = 0.640$, $P < 0.001$, $n = 2339$; *MCI* and *Cycle*, $R^2 = 0.662$, $P < 0.001$, $n = 2339$; *F-IBI* and *FTR*, $R^2 = 0.667$, $P < 0.001$, $n = 6858$; *F-IBI* and *Native*, $R^2 = 0.572$, $P < 0.001$, $n = 6858$; *GPP* and *ER*, $R^2 = 0.705$, $P < 0.001$, $n = 156$), they were retained in the analysis as they were seen as representing differing components of EI. Remaining indicators provided measures of water quality, invertebrate community composition, fish community composition and ecosystem processes.

Statistical analyses, using boosted regression trees (BRTs), were used to quantify non-linear relationships between predictor and response variables, with automatic detection of interactions between predictors (Friedman 2001). By combining additive regression modeling with boosting techniques, the BRT method provides an estimate from numerous—often thousands—of models, with results including a measure of the comparative strength of association between the response variable and predictor variables (percentage deviance explained) and a cross-validation coefficient (CV) indicating the degree to which the model fits the holdout data. BRT analysis also indicates the form of the non-parametric relationship (e.g. linear, curvilinear, sigmoidal). The model development for BRT analysis is discussed in detail in the literature (Friedman 2001; Elith et al. 2008; Hastie et al. 2009).

Table 3. Description of the land-use pressure gradients (first three descriptors) and additional environmental variables, including the mean and range of values, used in this study.

ENVIRONMENTAL DESCRIPTOR		MEAN (RANGE)
VegL	Loss of native vegetation in the catchment (%)	0.66 (0, 1)
LogN	Stream nitrogen loading (ppm), log-transformed	–0.01 (–1.40, 1.59)
IC	Impervious cover in the catchment (%)	0.05 (0, 1)
SegHabitat	Weighted average of proportional cover of local habitat using categories of: 1 = still; 2 = backwater; 3 = pool; 4 = run; 5 = riffle; 6 = rapid; 7 = cascade	4.0 (1.7, 4.6)
SegSubstrate	Weighted average of proportional cover of bed sediment using categories of: 1 = mud; 2 = sand; 3 = fine gravel; 4 = coarse gravel; 5 = cobble; 6 = boulder; 7 = bedrock	3.7 (1.4, 5.7)
SegLowFlow	Mean annual 7-day low flow (m^3/s), fourth-root transformed	1.06 (1, 2.23)
SegFlowStability	Annual low flow/annual mean flow (ratio)	0.17 (0.01, 0.52)
SegSumT	Summer air temperature (°C)	16.01 (12.5, 18.8)
SegTSeas	Winter air temperature (°C), normalised with respect to SegSumT	0.62 (–1.3, 3.0)
SegShade	Riparian shade (proportional)	0.39 (0, 0.80)
SegSlope	Segment slope (°), square-root transformed	1.44 (1, 4.42)
USRainDays	Days/year with rainfall in the catchment greater than 25 mm	9.8 (1.6, 25.1)
USAvgT	Average air temperature (°C) in the catchment, normalised with respect to SegSumT	0.19 (-4.04, 1.64)
USSlope	Average slope in the catchment (°)	10.3 (0.05, 26.77)
USCalcium	Average calcium concentration of rocks in the catchment, 1 = very low to 4 = very high	1.74 (0.66, 3.0)
USHardness	Average hardness of rocks in the catchment, 1 = very low to 5 = very high	3.01 (1.0, 4.0)
USPhosphorus	Average phosphorus concentration of rocks in the catchment, 1 = very low to 5 = very high	2.87 (1.0, 4.92)
USPeat	Area of peat in upstream catchment (proportional)	0.001 (0, 0.13)
USLake	Area of lake in upstream catchment (proportional)	0.001 (0, 0.17)
USGlacier	Area of glacier in upstream catchment (proportional)	0.00001 (0, 0.01)
DSDist2Coast	Distance to coast (km)	40.89 (0.1, 258.8)
DSDist2Lake	Distance to a downstream lake (km); where no lake is present, set to 500, which is greater than the maximum reach length	454.23 (0.53, 500)
DSAvgSlope	Average downstream slope (°), square-root transformed	0.86 (0, 5.44)
DSMaxSlope	Maximum downstream slope (°)	5.55 (0, 25.65)

The response of each EI indicator to environmental variables was investigated using three BRT models. Firstly, the response to three land-use pressures was examined (Model A), i.e. loss of native vegetation (*VegL*), percentage impervious cover (*IC*) and nitrogen load log-transformed (*LogN*). Fitted values from Model A were then used as a fixed offset in a second model (Model B) to examine the degree to which remaining variation could be explained by 21 additional environmental descriptors (e.g. *SegShade, USSlope, DSDist2Coast*; see Table 3). Finally, a third model (Model C) was used to test the ability of the FENZ 100-group categorical classification (*Group*) to capture this background environmental signal by substituting the continuous environmental predictors with a vector indicating the FENZ 100-group membership.

For Model A and Model C, tree complexity was set at three to allow for higher-order interactions between *Group* and/or the three land-use pressures; data were divided into five, seven or ten folds depending on the size of the dataset (ten folds means the holdout data was 10% of all data) and the learning rate was tuned to ensure approximately 1000 trees were added to the final model. For Model B, tree complexity was set at seven (owing to the higher number of explanatory

variables that could interact), folds were five or ten and the learning rate was tuned to yield approximately 1000 trees (see Appendix 1 for full list of BRT model parameters). BRT analyses were carried out in R, version 2.7.2, using the 'gbm' library of Ridgeway (2006) supplemented by scripts from Elith et al. (2008).

The response variables represent various data distributions including count data that typically had Poisson distributions (e.g. *FTR*, *ITR*); proportional data bounded by 0 and 1 (e.g. *EPT*, *Cycle*, *Native*); and indices that, whilst often bounded between 0 and a maximum, had data with normal distributions (e.g. *MCI*, *F-IBI*). This allowed for two model approaches for non-normally distributed data, using a Poisson-error model or using a Gaussian-error model following transformation of the response variable. For each applicable variable, both approaches were examined and results were reported for the best model based on goodness-of-fit and examination of residual errors. Left-and right-bounded data (e.g. *Native*) were arcsine transformed and models fitted using a Gaussian error.

2.2 Developing a multi-metric index (MMI)

Each group of ecological indicators was assessed separately with the aim of including at least one metric from each group, i.e. at least one each for fish, invertebrates, water quality and ecosystem processes. Firstly, the indicator that best responded to the three land-use pressures was included, which meant that *NOx*, *MCI*, *F-IBI* and *Del15N* were incorporated. Next, where a second indicator from each group showed a significant response but was also seen to represent an alternative component of EI, it was also included, which meant that *Clarity*, *Cycle*, *Native* and *Cotton* were incorporated. Finally, based on the results of regional analyses for *ER* and *GPP* (Clapcott et al. 2010) and the potential for improved performance with increased datasets, *GPP* and *ER* were also included, as they were seen to represent components of EI that were not assessed by other indicators. In all, 10 out of 14 potential metrics were investigated as part of a multi-metric index.

2.2.1 Predicting observed and expected values

Whilst characterising ecological response curves (see section 2.1.6), a model was fitted (Model A) that examined the indicator response to three disturbance gradients (*VegL*, *IC*, *LogN*). Fitted values from Model A were then used as a fixed offset in a second model (Model B) to examine the degree to which remaining variation could be explained by additional environmental descriptors (e.g. *SegShade*, *USSlope*, *DSDist2Coast*).

The outputs from these two models were used to predict expected (E) and observed (O) metric values for each stream reach in New Zealand in the following way. Specifically, the output from Model A was used to estimate a reference value in the absence of land-use pressure. This was assumed to occur at sites where *VegL* was less than 5% and *IC* was equal to zero (a restriction was not placed on *LogN* because values were seen to incorporate natural spatial variability). The number of samples used to estimate reference values ranged from $n = 6$ for *ER* to $n = 698$ for *F-IBI*. The output from Model B was used to provide an estimate of the average departure from the reference condition due to environmental variability, i.e. an estimate of the spatial variation of reference values on a national scale. The average metric value in the absence of land-use pressure from Model A and the mathematical functions describing natural variability from Model B were then used to predict E reference values for 568 600 stream reaches which equate to approximately 425 000 km of mapped river network in New Zealand. When model-predicted reference values fell outside the range of possible values (e.g. greater than 60 for *F-IBI*), values were set to equal the minimum or maximum in the sample range. Finally, fitted functions from Model B were used to predict the O contemporary value for all stream reaches in New Zealand using all metric values from Model A and the departure from reference functions from Model B.

In an attempt to make the metrics comparable and combinable, O values were divided by E values to produce O/E (sometimes 'OE') scores. For indicators that increased in response to increasing land-use pressure, the inverse of the O/E score was calculated (i.e. 1/(O/E). Furthermore, calculations that involved combining metrics were done using log-transformed variables to correct for differences in distributions between O/E and 1/(O/E) metrics, and back-transformed for future analysis (e.g. BRT analysis of MMI). Predicted O values, E values and O/E scores for each metric were analysed, basic statistics were calculated, outliers and frequency distributions were scrutinised, and plots were visually checked. The geographical variation in O/E scores was also visually checked to determine whether meaningful spatial patterns occurred at a national scale. The ecological response curves of O/E scores were plotted against land-use disturbance gradients to establish whether different metrics responded differently to the pressures and, hence, when combined, could provide a complementary assessment across full disturbance gradients. This step was taken because the findings of the first part of this study indicated that there was the potential for different indicators to respond differently to the same levels of land-use stress.

2.2.2 Combining metrics

Next, two alternative ways of deriving MMI scores were developed:

1. O/E scores for candidate metrics were averaged (*MMIavg*). Prior to the aggregation that preceded calculating an average, the average score per category was determined to ensure equal weighting among categories with two metrics (i.e. water quality, fish, invertebrates) with the ecosystem process category that had four metrics. See Equation 1 below.

2. O/E scores for candidate metrics were weighted (*MMIwt*) prior to the aggregation that preceded averaging. Weighting was based on two factors—the strength (percentage deviance explained) and precision (cross-validation coefficient) of initial BRT models; and the national data coverage of measured values (see Equation 2 below). For example, the *F-IBIOE* score was multiplied by the average of the deviance explained (0.113, from Table 4) and the cross-validation coefficient (0.338, from Table 4) in the BRT model, and then multiplied by national representation of stream types (0.95, from Table 2), such that the *F-IBIOE* contribution = *F-IBIOE**(0.113+0.338)/2)*0.95 = 0.214. The relative contribution per category to the *MMIwt* score was 0.37 for invertebrates, 0.34 for water quality, 0.18 for fish and 0.11 for ecosystem processes. Finally, *MMIwt* values were scaled to an equivalent within the original range of MMI scores, i.e. 0 to 1, by multiplying values by the sum of weighting factors (i.e. 2.56). It is anticipated that as the indicator database grows, these weightings could be revisited. Similarly, it is worth noting that new indices could be used interchangeably in the creation of an MMI score or together with current indices to strengthen the inference of change.

   Equation 1: *MMIavg* = ((*F-IBIOE* + *NativeOE*)/2 + (*MCIOE* + *CycleOE*)/2 + (*NOxOE* + *ClarOE*)/2 + (*EROE* + *GPPOE* + *Del15NOE* + *CottonOE*)/4))/4

   Equation 2: *MMIwt* = ((*F-IBIOE**0.214 + *NativeOE**0.199)/2 + (*MCIOE**0.435 + *CycleOE**0.406)/2 + (*NOxOE**0.464 + *ClarOE**0.327)/2 + (*EROE**0.125 + *GPPOE**0.060 + *Del15N*OE*0.188 + *CottonOE**0.141)/4))/4

The data distributions of resulting MMIs were examined as well as their response to the three land-use pressure gradients and spatial patterns at a national scale.

Table 4. Proportion of deviance explained (%) and cross-validation coefficients (CV; %) from boosted regression tree models of indicators of ecological integrity. Model A examines the response to three land-use pressures; Model B uses the output from Model A as a fixed offset and examines the response to environmental variables; and Model C examines the response to three land-use pressures and 100-level FENZ group.

ECOLOGICAL INDICATOR	MODEL A		MODEL B		MODEL C	
	%	CV (SE)*	%	CV (SE)*	%	CV (SE)*
Water quality						
Clarity	29.94	53.9 (4.2)	54.81	73.8 (2.4)	34.21	57.8 (4.6)
DRP	20.28	33.6 (3.1)	58.60	74.4 (4.0)	27.48	37.6 (6.2)
NOx	56.96	68.4 (7.4)	75.55	86.4 (3.1)	60.35	72.2 (9.0)
Invertebrates						
Cycle	31.64	56.7 (2.3)	42.40	65.6 (1.6)	35.33	60.2 (2.0)
EPT	22.54	47.9 (1.4)	40.33	63.4 (1.3)	25.72	50.6 (1.8)
ITR	5.60	24.2 (1.4)	36.37	60.5 (1.7)	6.38	25.9 (2.3)
MCI	35.60	60.1 (1.3)	48.87	70.1 (0.7)	40.10	63.1 (1.4)
Fish						
F-IBI	11.32	33.8 (0.9)	53.22	73.1 (0.7)	35.00	59.1 (0.9)
Native	10.03	31.8 (0.9)	47.53	69.1 (0.5)	28.00	52.9 (0.8)
FTR	7.44	25.9 (1.2)	42.34	63.8 (0.7)	21.87	44.3 (1.3)
Processes						
Cotton	12.91	38.4 (9.6)	54.62	79.5 (4.1)	10.03	28.6 (14.8)
Del15N	48.19	72.9 (5.7)	86.13	93.4 (2.2)	38.19	68.1 (2.9)
ER	10.27	28.2 (6.0)	33.79	57.5 (5.5)	11.45	33.3 (77.0)
GPP	0.25	18.1 (4.5)	25.92	47.6 (9.3)	2.95	20.3 (6.5)

* Note: mean and standard error estimated within model building.

3. Results

3.1 Ecological response curves

3.1.1 Water quality metrics

Three land-use pressures explained 57% of the deviance in *NOx* data, 30% of the deviance in *Clarity* and 20.3% of the deviance in *DRP* data (Model A). Cross-validation estimates of model performance indicated that the model for *NOx* performed best (Table 4). The dominant influencing land use varied for each of the three metrics, as did the form of the response to each land-use predictor (Fig. 3). For example, *LogN* strongly influenced *Clarity*, which showed a general monotonic decrease in response to increasing *LogN* (Fig. 3).

The addition of environmental predictors (Model B) increased the explanatory power of models for all water quality metrics; in particular, more than doubling the proportion of deviance explained for *DRP* to 58.6% (Table 4). Environmental variables that contributed most to improving the model for *DRP* were the ratio of low-to-mean flow (*SegFlowStability*) (21.7% of the additional deviance explained by the inclusion of environmental predictors was attributed to this variable), average phosphorus concentration of rocks in the catchment (*USPhosphorus*) (12%) and proportional sediment size (*SegSubstrate*) (10.1%). For *Clarity*, the variables were average air temperature in the catchment (*USAvgT*) (9.8%), mean annual 7-day low flow (*SegLowFlow*) (9.8%) and winter air temperature (*SegTSeas*) (9.7%). For *NOx*, the variables were *SegTSeas* (25.6%), *USAvgT* (15.4%) and average slope in the catchment (*USSlope*) (15.4%) (see also Appendix 2).

Clarity

DRP

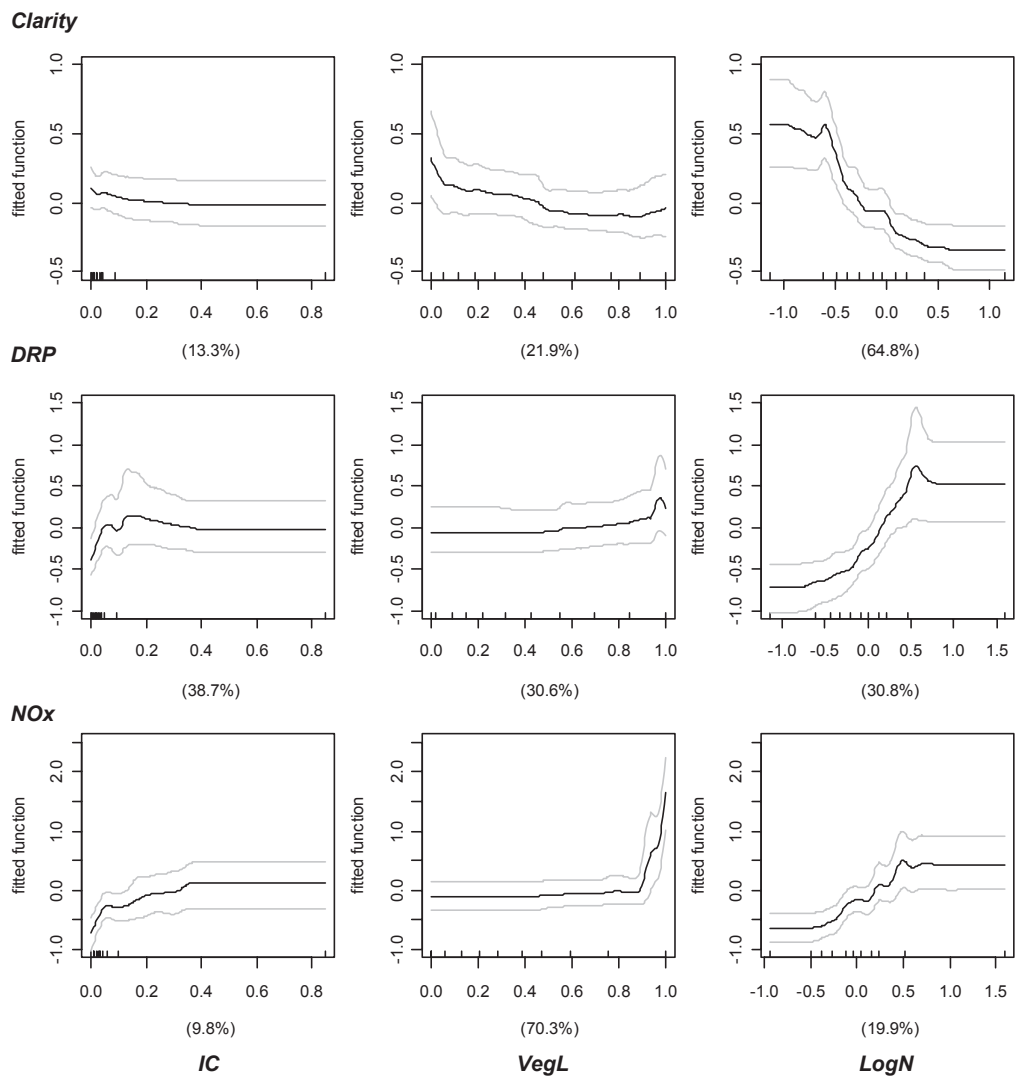*NOx*

IC   VegL   LogN

Figure 3. BRT-fitted functions for the response of water quality metrics to the three land-use pressures for *Clarity*, *DRP* and *NOx*. Percentages show the relative contribution of each pressure to explaining deviance in the metric data. Rug plots inside the *x*-axes show the distribution of sites, in deciles. Functions are plotted with a loess-smoothing span of 0.3 and 95% confidence intervals.

Substituting the FENZ group membership (*Group*) for the continuous environmental variables (Model C) yielded greater predictive power than that observed for Model A for all responses, but less than that observed for Model B (Table 4). *Group* explained 27.3% of all deviance in *Clarity* data, 26.8% for *DRP* and 17.5% for *NOx*. The relative contribution of land-use predictors in explaining the deviance for each of the three water quality metrics was similar to that observed for previous models. For example, for the Model C for *NOx*, deviance in the data was explained by *VegL* (53%), *LogN* (18.6%) and *IC* (10.9%), compared with the Model A for *NOx*, where deviance in the data was explained by *VegL* (70.3%), *LogN* (19.9%) and *IC* (9.8%).

3.1.2   Invertebrate metrics

Three land-use pressures explained between 5.6% (*ITR*) and 36% (*MCI*) of the deviance in invertebrate metric data (Table 4). *VegL* was the dominant explanatory variable for *MCI*, *Cycle* and *EPT*, and each of these metrics displayed a general monotonic increase in response to increasing *VegL* (Fig. 4). For *ITR*, *LogN* was the most influential land-use pressure, with a non-monotonic response. However, there was low model confidence in the output for *ITR* (Table 4, Appendix 1).
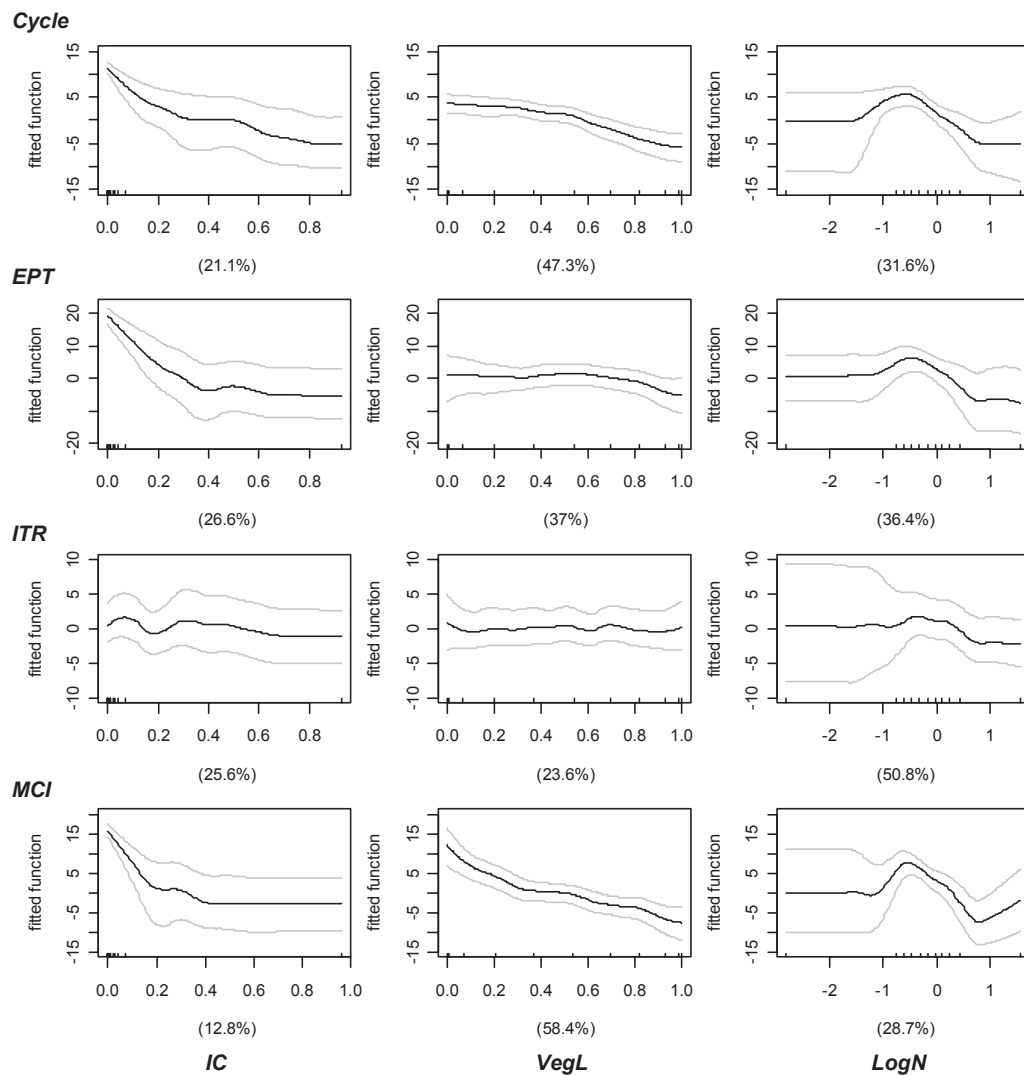
Figure 4. BRT-fitted functions for the response of invertebrate metrics to the three land-use pressures for *Cycle*, *EPT*, *ITR* and *MCI*. Percentages show the relative contribution of each pressure to explaining deviance in the metric data. Rug plots inside the *x*-axes show the distribution of sites, in deciles. Functions are plotted with a loess-smoothing span of 0.3 and 95% confidence intervals.

Addition of environmental variables (Model B) increased the proportion of deviance explained for all invertebrate metrics (Table 4). The number of high rainfall days per year (*USRainDays*), the ratio of low-to-mean flow (*SegFlowStability*) and summer air temperature (*SegSumT*) were important explanatory variables in invertebrate metric models (see Appendix 2).

Group contributed a relatively high proportion of the total deviance explained in Model C for all metrics—47.9% for *ITR*, 43.4% for *EPT*, 29.5% for *MCI* and 28.5% for *Cycle*. However, only minor increases in the total proportion of deviance explained were achieved by the inclusion of *Group* compared with the inclusion of environmental variables (Table 4). For example, the total proportion of deviance explained for *MCI* was 36% for Model A, 48.9% for Model B and 40% for Model C. This suggests that while the FENZ grouping is informative about the nature of a stream segment, there is additional environmental detail available that is not represented by *Group*.

3.1.3 Fish metrics

Three land-use pressures explained a small proportion of the deviance observed in fish metric data—11.3% for *F-IBI*, 10% for *Native*, and 7.4% for *FTR* (Model A). *LogN* was the main contributing pressure for all three fish metrics, which showed general, albeit noisy, increases in values between –1.0 and 1.0 log-nitrogen (Fig. 5). The inclusion of 21 environmental variables noticeably increased the explanatory power of the BRT models to 53.2% for *F-IBI*, to 47.5% for *Native* and to

F-IBI

Native

FTR

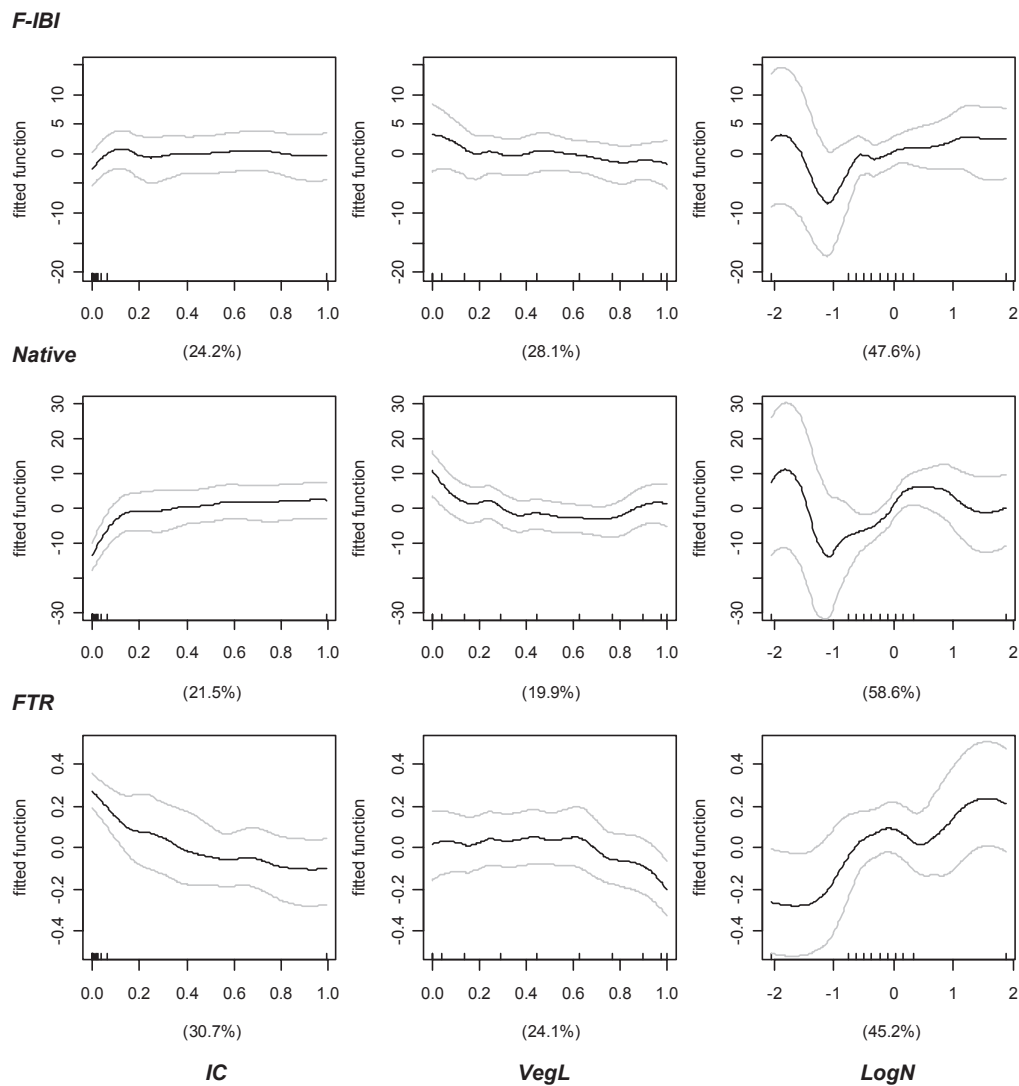|		
IC	**VegL**	**LogN**

Figure 5.   BRT-fitted functions for the response of fish metrics to the three land-use pressures for *F-IBI*, *Native* and *FTR*. Percentages show the relative contribution of each pressure to explaining deviance in the metric data. Rug plots inside the *x*-axes show the distribution of sites, in deciles. Functions are plotted with a loess-smoothing span of 0.3 and 95% confidence intervals.

42.33% for *FTR* (Model B). The main contributing environmental variable for *F-IBI* and *Native* was downstream distance to the coast (*DSDist2Coast*) (22.4% and 18.3%, respectively), and for *FTR*, it was the maximum slope downstream (*DSMaxSlope*) (19.1%) (see also Appendix 2).

The inclusion of *Group* more than doubled the proportion of deviance explained in Model C compared with Model A for each fish metric (Table 4). *Group* contributed more than the combination of all three land-use pressures in explaining the deviance in the data for each fish metric; i.e. 76.5% for *F-IBI*, 71% for *Native* and 58.7% for *FTR*.

3.1.4 Ecosystem process metrics

Three land-use pressures explained between 0.2% (*GPP*) and 48.2% (*Del15N*) of the deviance in process metric data (Model A; Table 4). *LogN* was the most important pressure for *Del15N* and *Cotton* data. While *Del15N* generally increased between values of –0.5 and 0.5 *LogN*, *Cotton* had a non-monotonic response to increasing nitrogen pressure (Fig. 6). The most influential pressure varied for other process metrics, but model validation (Table 4) as well as large confidence intervals (Fig. 6) suggested these were weak models for explaining deviance in *GPP* and *ER* data (Table 4).
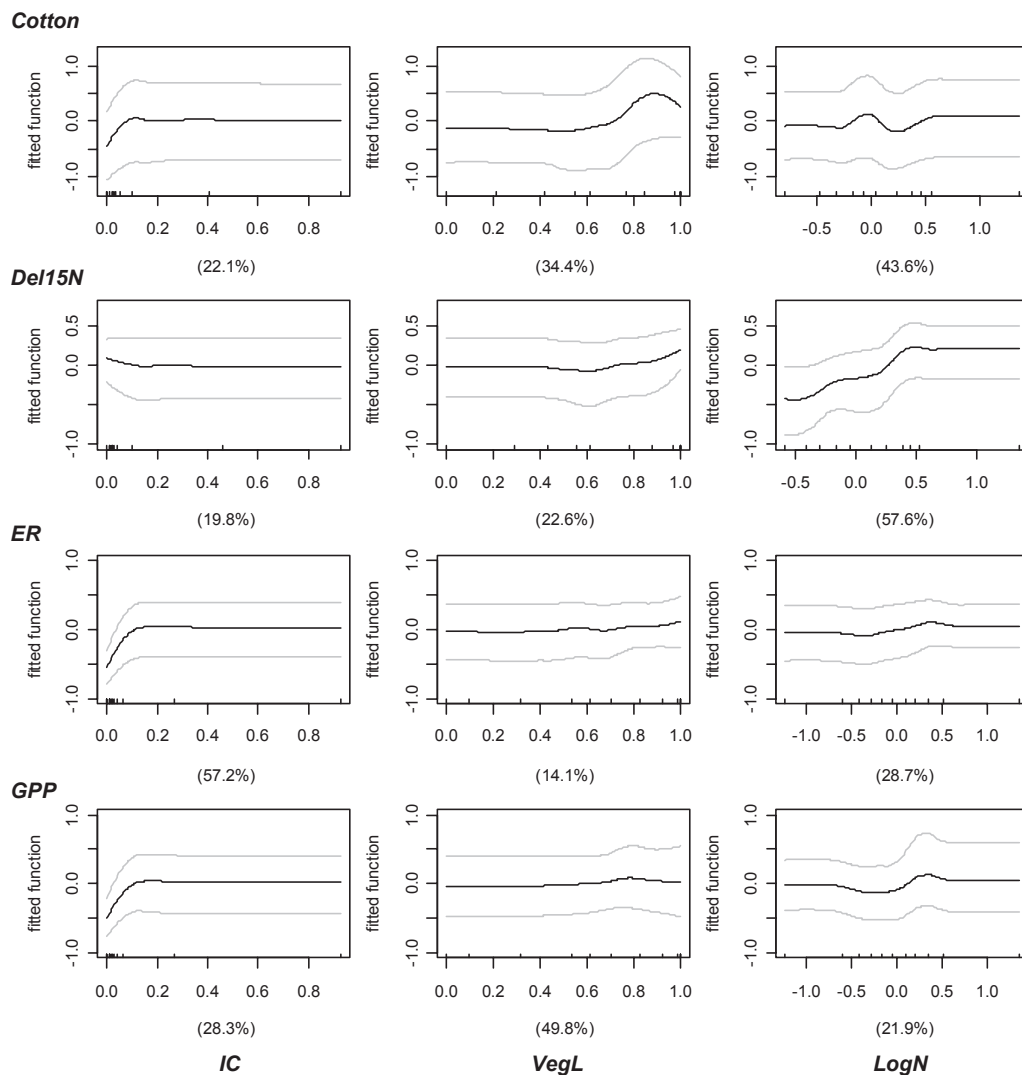
Figure 6.   BRT-fitted functions for the response of ecosystem process metrics to the three land-use pressures for *Cotton*, *Del15N*, *ER* and *GPP*. Percentages show the relative contribution of each pressure to explaining deviance in the metric data. Rug plots inside the *x*-axes show the distribution of sites, in deciles. Functions are plotted with a loess-smoothing span of 0.3 and 95% confidence intervals.

The inclusion of 16 environmental variables (Model B) improved the explanatory power of the models for all process metrics, noticeably more than tripling the deviance explained for most metrics and increasing the proportion of deviance explained by *Del15N* from 48.2% to 86.1% (Table 4). A range of segment and upstream variables contributed to each model. For *GPP* and *ER*, they were mean annual 7-day low flow (*SegLowFlow*) (47.9% and 13.1%, respectively) and the average hardness of rocks in the catchment (*USHardness*) (14.1% and 22.4%); for *Del15N*, they were segment slope (*SegSlope*) (32.3%) and winter air temperature (*SegTSeas*) (31%); and for *Cotton*, they were summer air temperature (*SegSumT*) (18.3%) and rock hardness (*USHardness*) (15%) (see also Appendix 2).

The inclusion of *Group* resulted in a decrease in the proportion of deviance explained in Model C compared with Model A for *Del15N* and *Cotton*, and *Group* explained a small proportion of the total deviance in these models (6.2% and 8.1%, respectively). In contrast, small increases in the total proportion of deviance explained were attributable to *Group* for *GPP* and *ER* (Table 4), with *Group* explaining 19% and 14.7% of the total deviance, respectively.

3.2 Predicted observed and expected values

3.2.1 Water quality metrics

For *NOx*, modelled-expected data had low variation around the mean, and the distribution of modelled-O data was similar to that of the sample data (Fig. 7,). On average, modelled-O values for *NOx* were 72% greater than modelled-E values, suggesting a significant divergence from the predicted reference condition at a national scale (Table 5). In contrast, modelled-E values for *Clarity* showed a large range in values. There also appeared to be two distinct populations in the modelled-O data, with one hump associated with a considerable range in reference data and another hump associated with lower values (Fig. 8). This p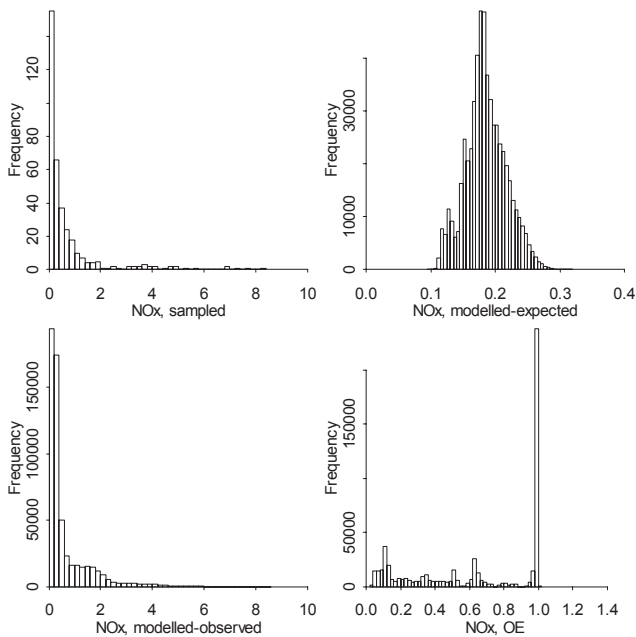attern was common across several candidate metric distributions (e.g. *MCI*, *Cycle*, *Del15N* and *Cotton*). The mean value for modelled-O *Clarity* was 62% less than that for modelled-E *Clarity*, again suggesting a significant decline in EI at a national scale (Table 5).

Both *ClarityOE* and *NOxOE* had large ranges in O/E scores. The proportion of stream reaches at a national scale with scores greater than 0.9 was 45% for *NOxOE* and c. 28% for *ClarityOE* (Table 6). However, there were also a large proportion of sites with O/E scores less than 0.3—27.6% for *NOx* and 21.5% for *Clarity* (see also Fig. 8).

The O/E scores for water quality metrics showed meaningful patterns at a national scale. Values of *NOxOE* were highest on conservation land and lowest in areas of intense land-use development such as Waikato, Canterbury and Southland (Fig. 9). Notably, large rivers reflected the status of their headwaters. In contrast, there was a distinct disparity between high and low values of *ClarityOE*, with most land not administered by DOC having a marked reduction in integrity (Fig. 10). Compared with other metrics, *ClarityOE* painted the bleakest picture for the EI of streams at a national scale.

Figure 7. Distribution of sample data, modelled-expected values, modelled-observed values (ppm) and observed-to-expected (O/E) scores for the *NOx* water quality metric.
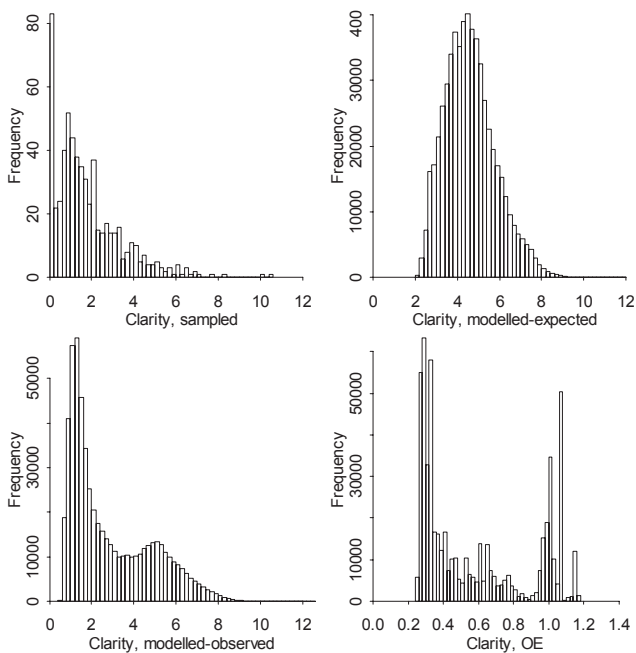

Figure 8. Distribution of sample data, modelled-expected values, modelled-observed values (ppm) and observed-to-expected (O/E) scores for the *Clarity* water quality metric.

3.2.2 Invertebrate metrics

Comparison of the distributions of modelled values with the sample data showed a bimodal distribution in observed values for both *MCI* and *Cycle* metrics (Figs 11 & 12). There was a tendency for models to under-predict extremes in invertebrate metric values, but predicted data showed the same general distribution as sample data (Figs 11 & 12). The mean difference between modelled-O and -E values at a national scale was 18.7 for *MCI* (16% decrease from E) and 9.5 for *Cycle* (11% decrease from E) (Table 5).

The ranges in *MCIOE* and *CycleOE* values were 0.48 to 1.00 and 0.24 to 1.16, respectively. The proportion of stream reaches at a national scale with O/E scoressgreater than 0.9 were c. 42% for *MCIOE* and c. 57% for *CycleOE* (Table 6). The O/E scores showed

Table 5.   Mean and range in modelled-expected and modelled-observed values for ten candidate metrics of ecological integrity compared with sample data. 'Reference' is the metric value at low pressure used in the calculation of modelled-expected.

METRIC	REFERENCE	EXPECTED	OBSERVED	SAMPLE DATA
NOx (mg/L)	0.072	0.185 (0.09, 0.32)	0.665 (0.11, 8.56)	0.605 (0.01, 8.45)
Clarity (m)	4.8	4.62 (2.06, 11.8)	2.85 (0.55, 12.5)	2.08 (0.10, 10.43)
MCI	138.4	136.6 (97.5, 161.9)	117.9 (55.5, 162.1)	115.9 (20, 180)
Cycle (%)	93.5	92.53 (62.5, 100)	83.05 (16.8, 100)	82.5 (0, 100)
F-IBI	37.8	35.5 (0, 60)	29.8 (0, 60)	34.26 (0, 60)
Native (%)	83.8	92.0 (0, 100)	79.0 (0, 100)	75.3 (0, 100)
ER (gO_2 m^{-2} d^{-1})	6.50	6.257 (5.07, 8.28)	9.103 (4.59, 24.83)	9.936 (0.34, 55.78)
GPP (gO_2 m^{-2} d^{-1})	3.04	2.27 (1.55, 5.13)	2.66 (1.2, 39.49)	4.603 (0.078, 35.99)
Del15N (‰)	5.22	5.18 (5.09, 5.31)	5.62 (2.25, 19.0)	6.75 (1.03, 20.99)
Cotton (*k*/dd) x100	0.16	0.17 (0.12, 0.20)	0.27 (0.11, 1.53)	0.38 (0.02, 2.58)

Table 6.   Mean and range in modelled observed-to-expected (O/E) scores for ten candidate metrics and two summary multi-metric indexes (MMI) of ecological integrity, including the proportion of the national stream network with a range of O/E scores.

METRIC	O/E SCORE	PROPORTION OF STREAM NETWORK WITH O/E SCORES (%)			
		< 0.3	0.3–0.6	0.6–0.9	> 0.9
NOxOE	0.64 (0.03, 1.02)	27.6	15.4	13.0	45.0
ClarityOE	0.58 (0.24, 1.16)	21.5	37.7	13.1	27.7
MCIOE	0.86 (0.48, 1.00)	<0.1	0.4	57.7	41.9
CycleOE	0.90 (0.24, 1.00)	<0.1	0.5	42.8	56.7
F-IBIOE	0.81 (<0.1, >10)	5.8	9.6	39.8	44.8
NativeOE	16.40 (<0.1, >10)	1.9	4.3	13.6	80.2
EROE	0.92 (0.32, 1.11)	<0.1	12.7	25.1	62.2
GPPOE	0.97 (0.11, 1.37)	1.1	9.4	20.4	69.1
Del15NOE	1.04 (0.28, 2.27)	<0.1	13.9	21.9	64.2
CottonOE	0.74 (0.13, 1.10)	3.6	34.0	13.2	49.2
MMIavg	0.76 (0.04, 3.79)	1.9	26.3	38.9	32.9
MMIwt	0.78 (0.18, 1.99)	0.1	23.6	39.9	36.4

meaningful geographic patterns for both invertebrate metrics, with high values for conservation land and for upland streams in general. The lowest O/E scores were found in areas of extensive agricultural development and, compared with *MCIOE*, *CycleOE* also appeared more sensitive to urban development, e.g. Auckland, Hamilton and Christchurch (Figs 13 & 14).

3.2.3   Fish metrics

For *F-IBI*, modelled-O data had a similar distribution to sample data, with a large proportion of 0 values (Fig. 15). The proportion of 0 values in modelled-E data was much reduced, providing a sensible prediction of reference condition at a national scale (Fig. 15). The mean difference between modelled-O and modelled-E *F-IBI* was 5.7 units (a 19% decrease) (Table 5). In contrast, model predictions for O and E *Native* values were influenced by large values close to or at 100% (Fig. 16). On average, there was a 16% decrease in modelled-O compared with modelled-E *Native* values (Table 5).

The range in O/E scores for both fish metrics was less than 0.1 to greater than 10. The mean value for *NativeOE* was 16.4 because of the presence of 7577 sites with very high O/E scores (where modelled-O was greater than 10 and modelled-E was less than 1); however, the median value for *NativeOE* was 1.01. The proportion of stream reaches at a national scale with O/E scores greater than 0.9 was c. 45% for *F-IBIOE* and c. 80% for *NativeOE* (Table 6). These unusual distributions in scores are probably, in part, due to the non-monotonic response to pressures used in modelling these metrics.

Figure 9.   Geographic variation in observed-to-expected (O/E) scores for *NOx*. Scores are mapped using equal intervals between maximum and minimum values.
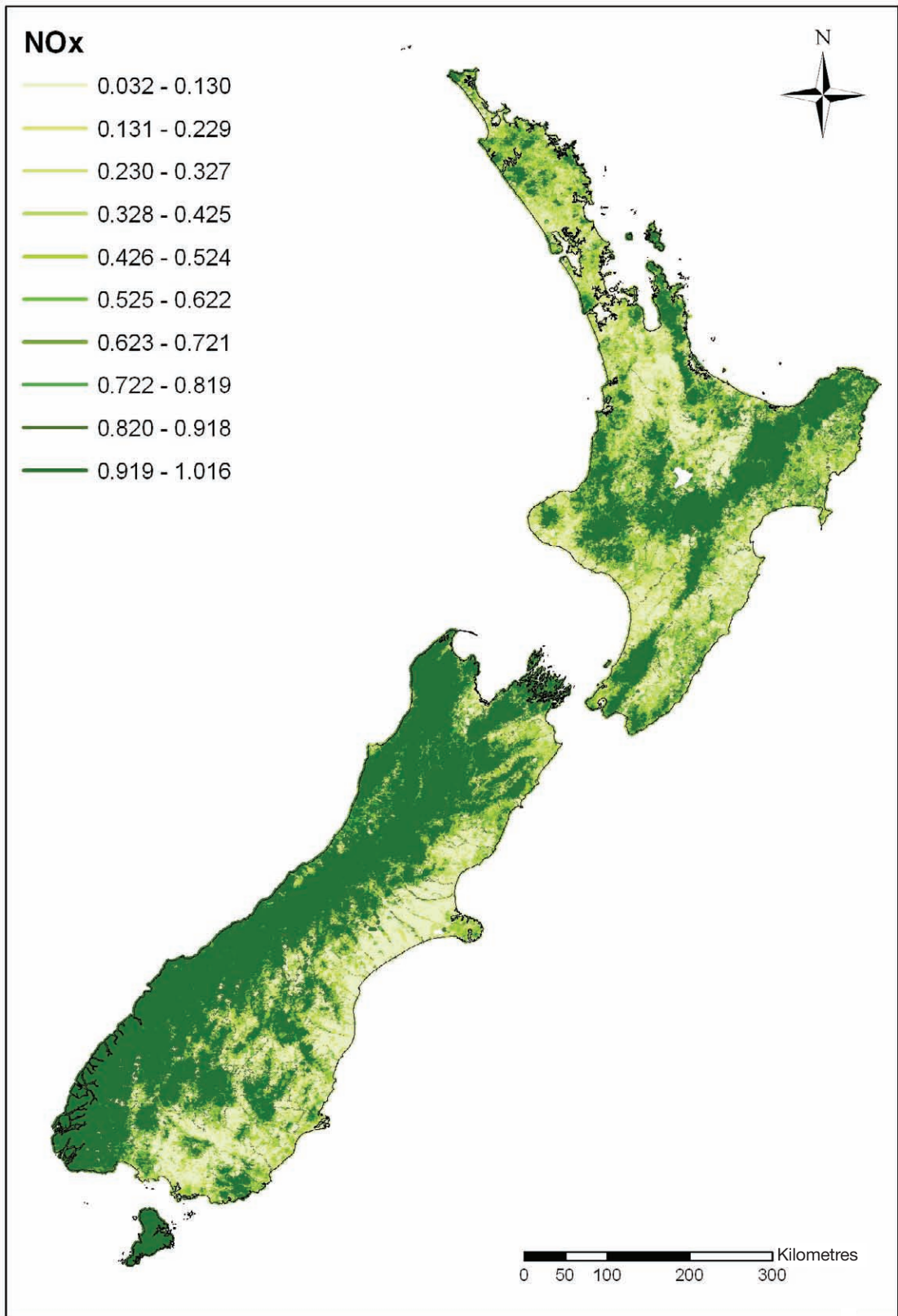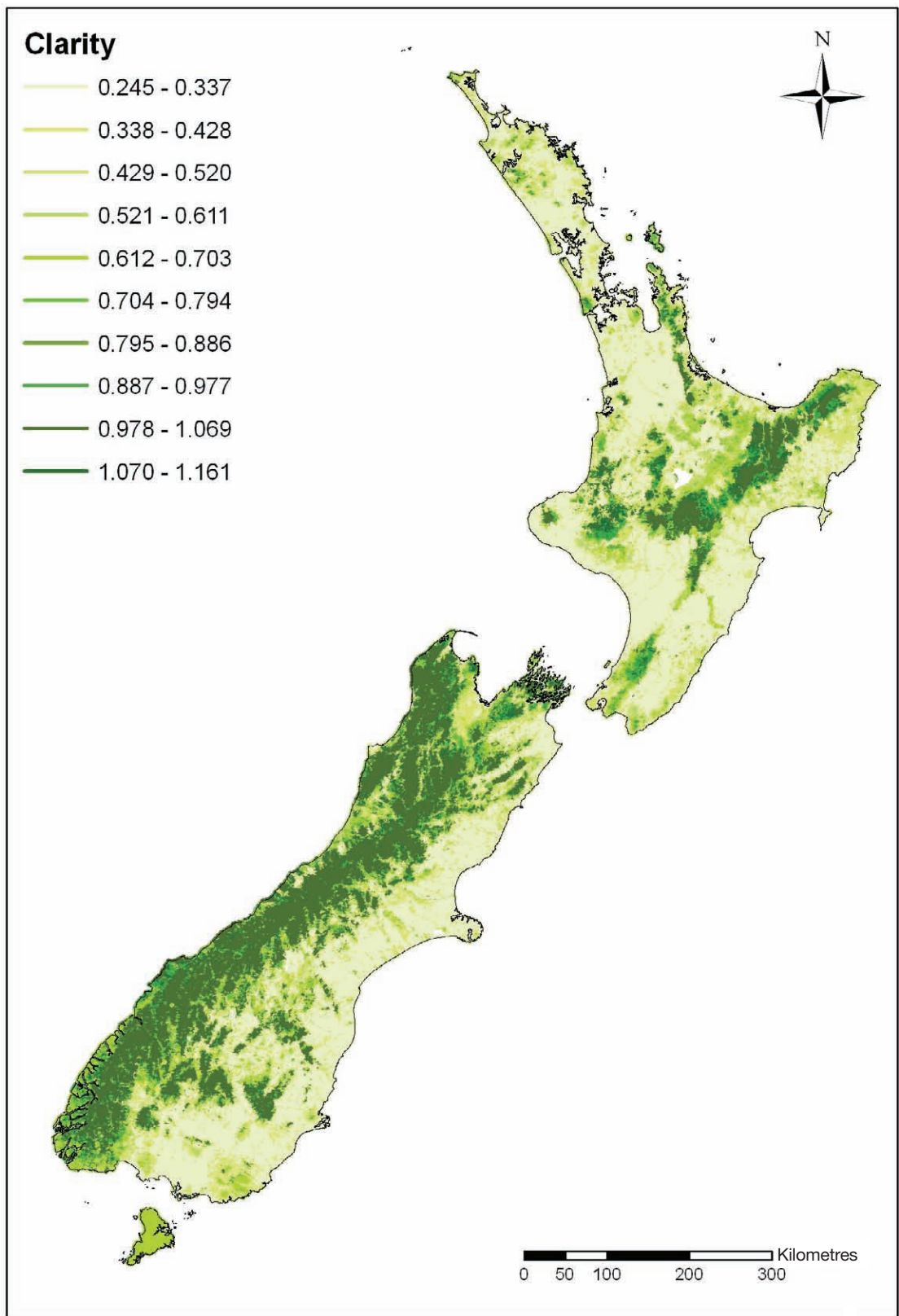
Figure 10.   Geographic variation in observed-to-expected (O/E) scores for *Clarity*. Scores are mapped using equal intervals between maximum and minimum values

Clapcott et al.—Relationship between land-use pressures and indicators of stream ecological integrity
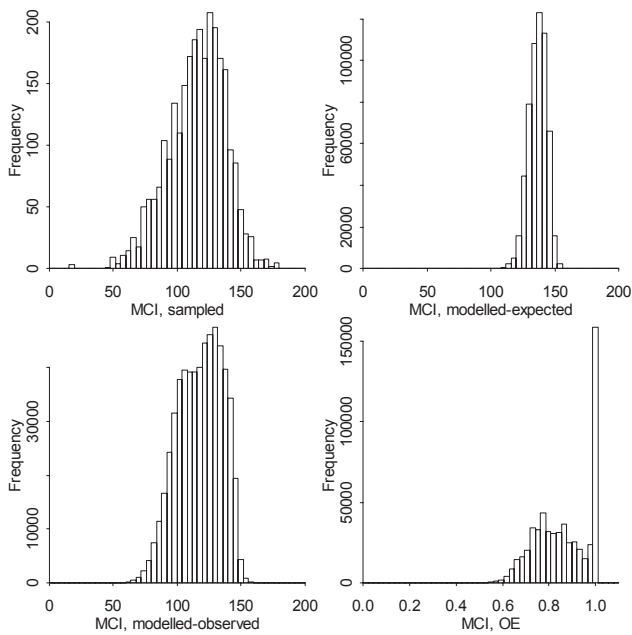
Figure 11. Distribution of sample data, modelled-expected values, modelled-observed values and observed-to-expected (O/E) scores for the *MCI* invertebrate metric.
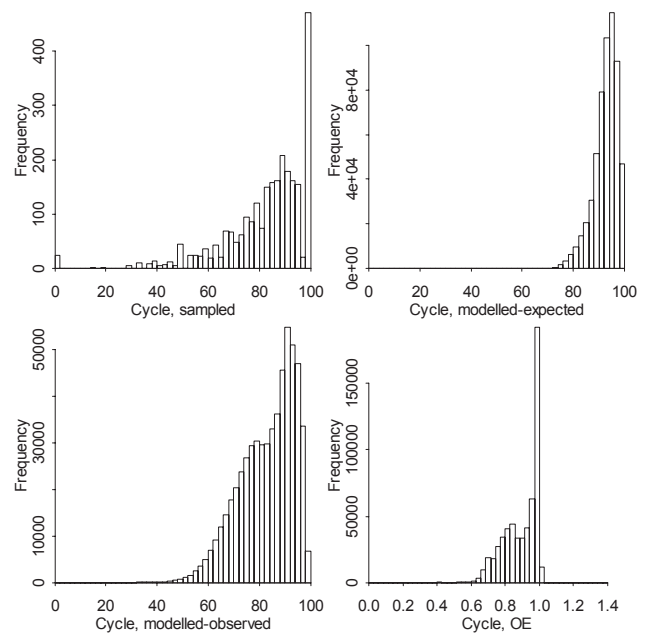

Figure 12. Distribution of sample data, modelled-expected values, modelled-observed values (%) and observed-to-expected (O/E) scores for the *Cycle* invertebrate metric.

There were few intuitive patterns in the geographic distribution of O/E scores for fish metrics. Surprisingly low values of *F-IBIOE* were evident in areas of Otago and the Southern Alps/ Kā Tiritiri o te Moana (Fig. 17). Similarly, the variation in *NativeOE* suggested lowland areas had higher integrity than upland areas (Fig. 18).

3.2.4 Ecosystem process metrics

Model predictions resulted in similar data distribution patterns for all four measures of ecosystem processes (Figs 19, 20, 21 & 22). Modelled-E data had low variation around the mean and the distribution of modelled-O data was similar to that of the sample data (Figs 19, 21 & 22), except for *Del15N* (Fig. 20). The greatest difference between mean modelled-O and modelled-E values was a 37% increase in *Cotton* values. All process metrics showed an increase in modelled-O values compared with modelled-E values, with *ER* increasing by 31%, *GPP* increasing by 15% and *Del15N* by 8% (Table 5). There was a wide range in all process metric O/E scores from close to 0 to greater than 2 (Table 6). Three process metrics (*EROE*, *GPPOE*, *Del15NOE*) showed more than 60% of sites with O/E scores greater than 0.9 (Table 6).

Overall, the O/E scores for ecosystem process metrics showed meaningful geographic patterns at a national scale. *CottonOE* (Fig. 23) and *Del15NOE* (Fig. 24) had similar distributions to *MCIOE* with high values for conservation land and lower values for intensely developed agricultural areas. However, *CottonOE* showed more variation among closely associated catchments and, like *NOxOE*, suggested poor integrity throughout the Bay of Plenty region. Similarly, *GPPOE* (Fig. 25) and *EROE* (Fig. 26) appeared highly variable at smaller spatial scales. The lowest values of *EROE* were in high-density urban areas.

3.3 Multi-metric index response to land-use pressures

The range in values for *MMIavg* (0.04 to 3.79) was much greater than for *MMIwt* (0.18 to 1.99), although both metrics had similar mean values (Table 6). Both metrics also quantified a high number of sites as having O/E scores greater than 0.9 (c. 30%, Table 6). *MMIwt* was the most responsive summary metric to land-use pressure, with a BRT model explaining 90.5% (CV = 0.95) of the deviance in *MMIwt* data. In contrast, 85.6% (CV = 0.92) of deviation in *MMIavg* data was explained by the three land-use pressures.

Figure 13. Geographic variation in observed-to-expected (O/E) scores for *MCI*. Scores are mapped using equal intervals between maximum and minimum values

Clapcott et al.—Relationship between land-use pressures and indicators of stream ecological integrity

Cycle

	0.245 - 0.321
	0.322 - 0.396
	0.397 - 0.472
	0.473 - 0.547
	0.548 - 0.623
	0.624 - 0.699
	0.700 - 0.774
	0.775 - 0.850
	0.851 - 0.925
	0.926 - 1.001

Figure 14. Geographic variation in observed-to-expected (O/E) scores for *Cycle*. Scores are mapped using equal intervals between maximum and minimum values.

Figure 15. Distribution of sample data, modelled-expected values, modelled-observed values and observed-to-expected (O/E) scores for the *IBI* fish metric; 219 sites with *F-IBIOE* >1.5 were excluded from the plot.

Figure 16. Distribution of sample data, modelled-expected values, modelled-observed values (%) and observed-to-expected (O/E) scores for the *Native* fish metric; 2007 sites with *NativeOE* >2 were excluded from the plot.

The geographical distribution of the summary metric for EI, i.e. *MMIwt,* showed meaningful patterns across New Zealand (Fig. 27). There were high values for conservation land and low values for areas of urban and agricultural development; in particular, Waikato, Manawatu, Canterbury and Southland showed large areas of low EI. However, despite their relatively low contribution to *MMIwt,* it appears that metrics with less meaningful geographic distributions (i.e. *F-IBIOE, NativeOE* and *GPPOE*) contributed to increased variation in EI at a localised scale, especially in areas of Otago and the Southern Alps/Kā tiritiri o te Moana.

IBI

— <0.001 - 0.117
— 0.118 - 0.235
— 0.236 - 0.352
— 0.353 - 0.470
— 0.471 - 0.587
— 0.588 - 0.704
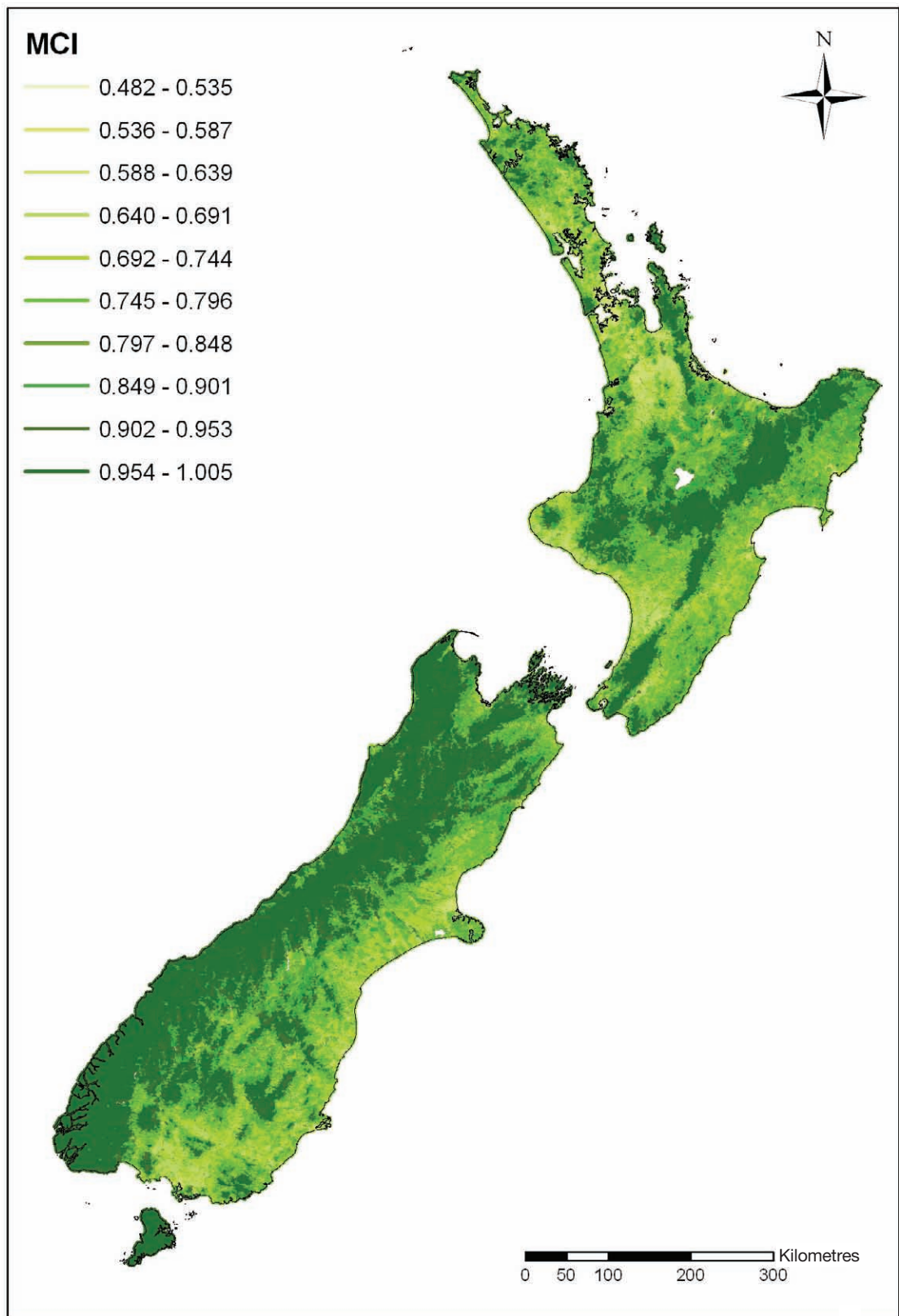— 0.705 - 0.822
— 0.823 - 0.939
— 0.940 - 1.056
— 1.057 - >10

Figure 17. Geographic variation in observed-to-expected (O/E) scores for *F-IBI*. Scores are mapped using equal intervals between maximum and minimum values.
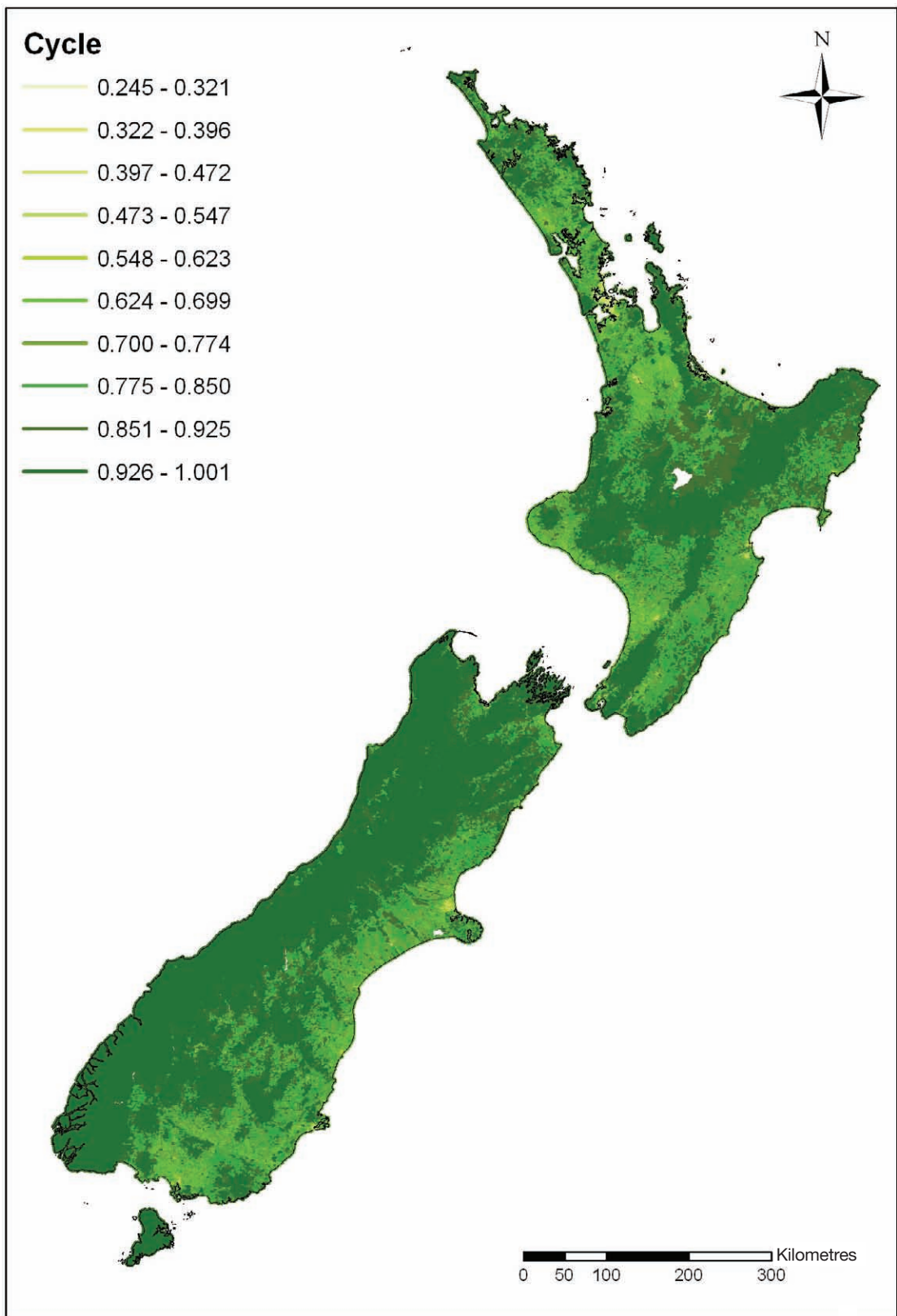
Figure 18.   Geographic variation in observed-to-expected (O/E) scores for *Native*. Scores are mapped using equal intervals between maximum and minimum values.
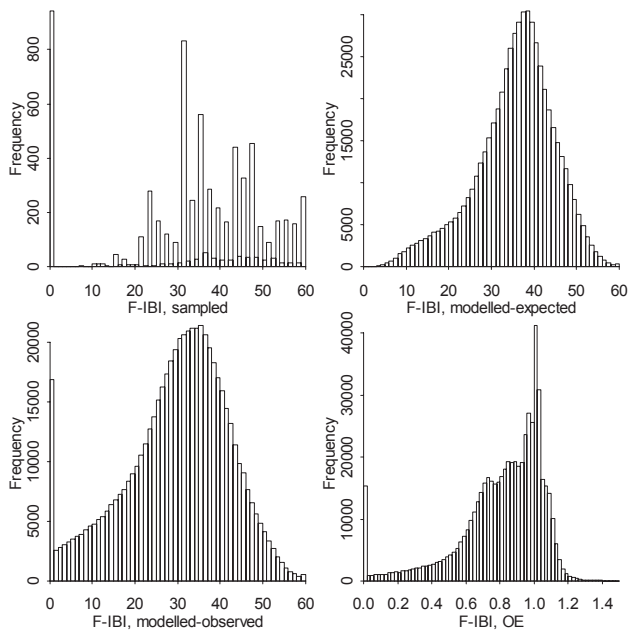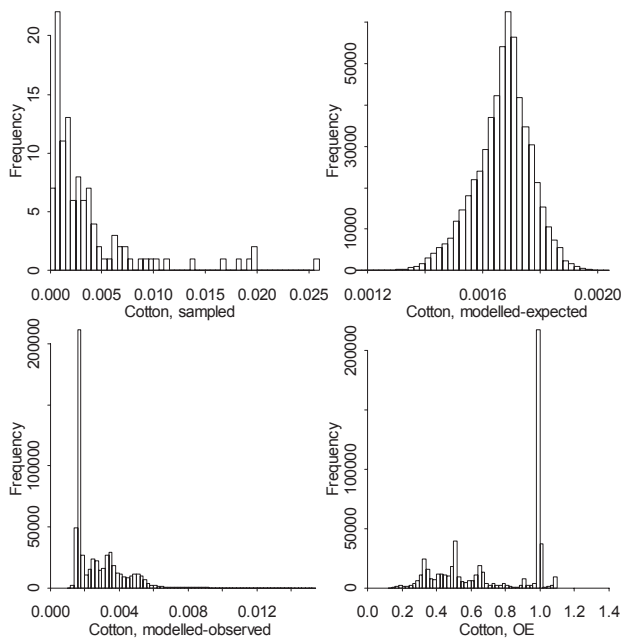
Figure 19. Distribution of sample data, modelled-expected values, modelled-observed values (*k*/dd) and observed-to-expected (O/E) scores for the *Cotton* process metric.
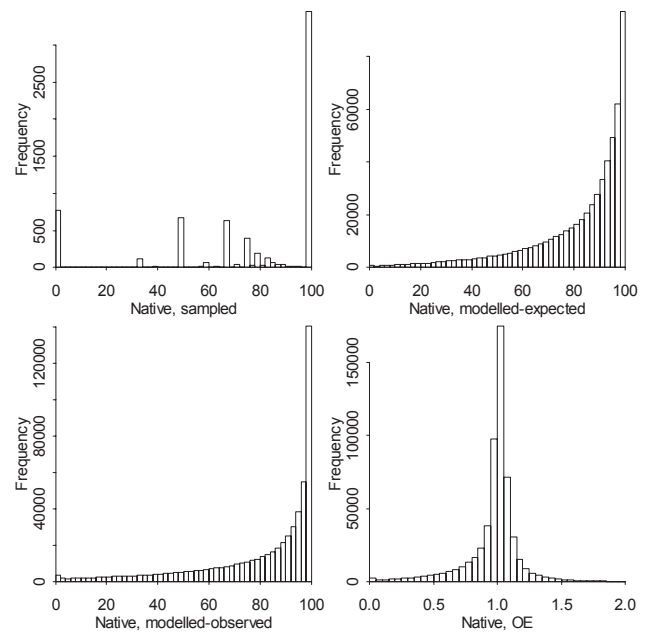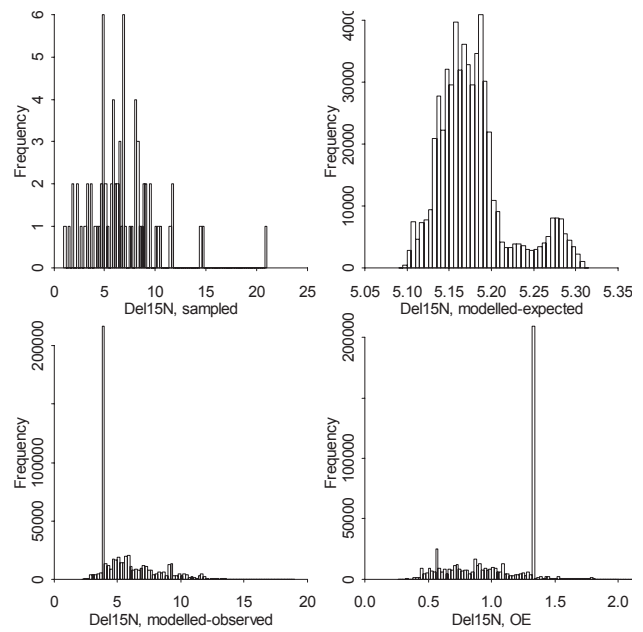

Figure 20. Distribution of sample data, modelled-expected values, modelled-observed values (%) and observed-to-expected (O/E) scores for the *Del15N* process metric.
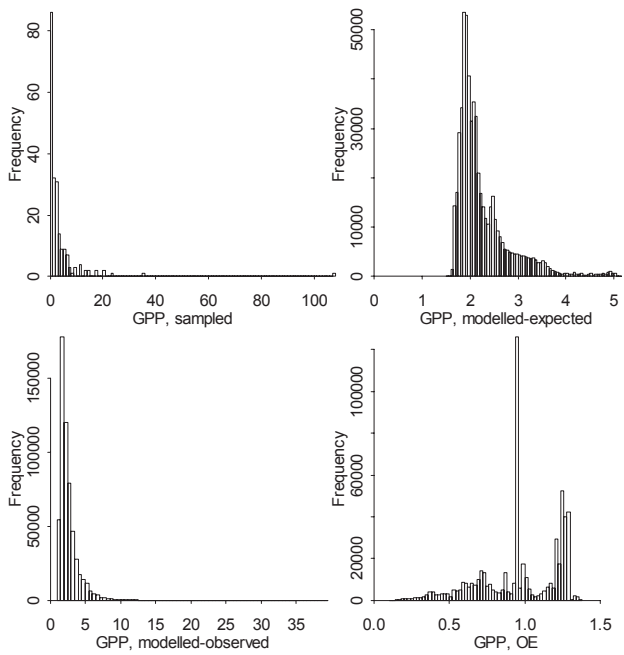
Figure 21. Distribution of sample data, modelled-expected values, modelled-observed values (gO$_2$/m^2/d) and observed-to-expected (O/E) scores for the *GPP* process metric.
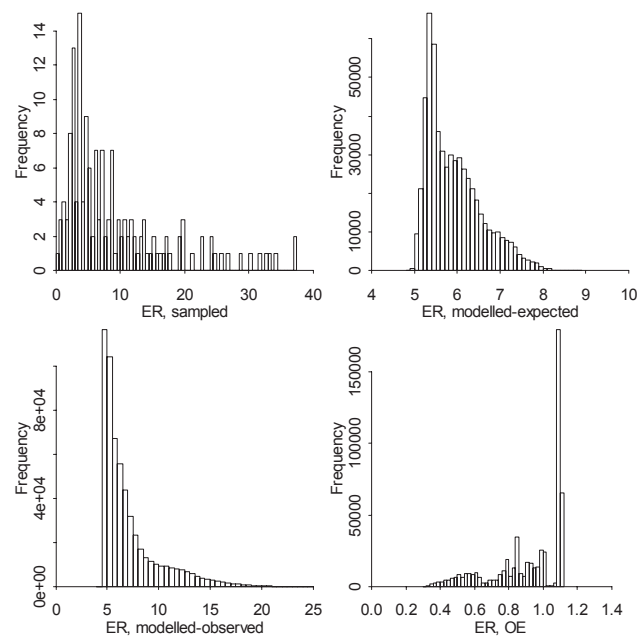
Figure 22. Distribution of sample data, modelled-expected values, modelled-observed values (gO$_2$/m^2/d) and observed-to-expected (O/E) scores for the *ER* process metric; two sites where *ER* >40 are excluded from the plot.
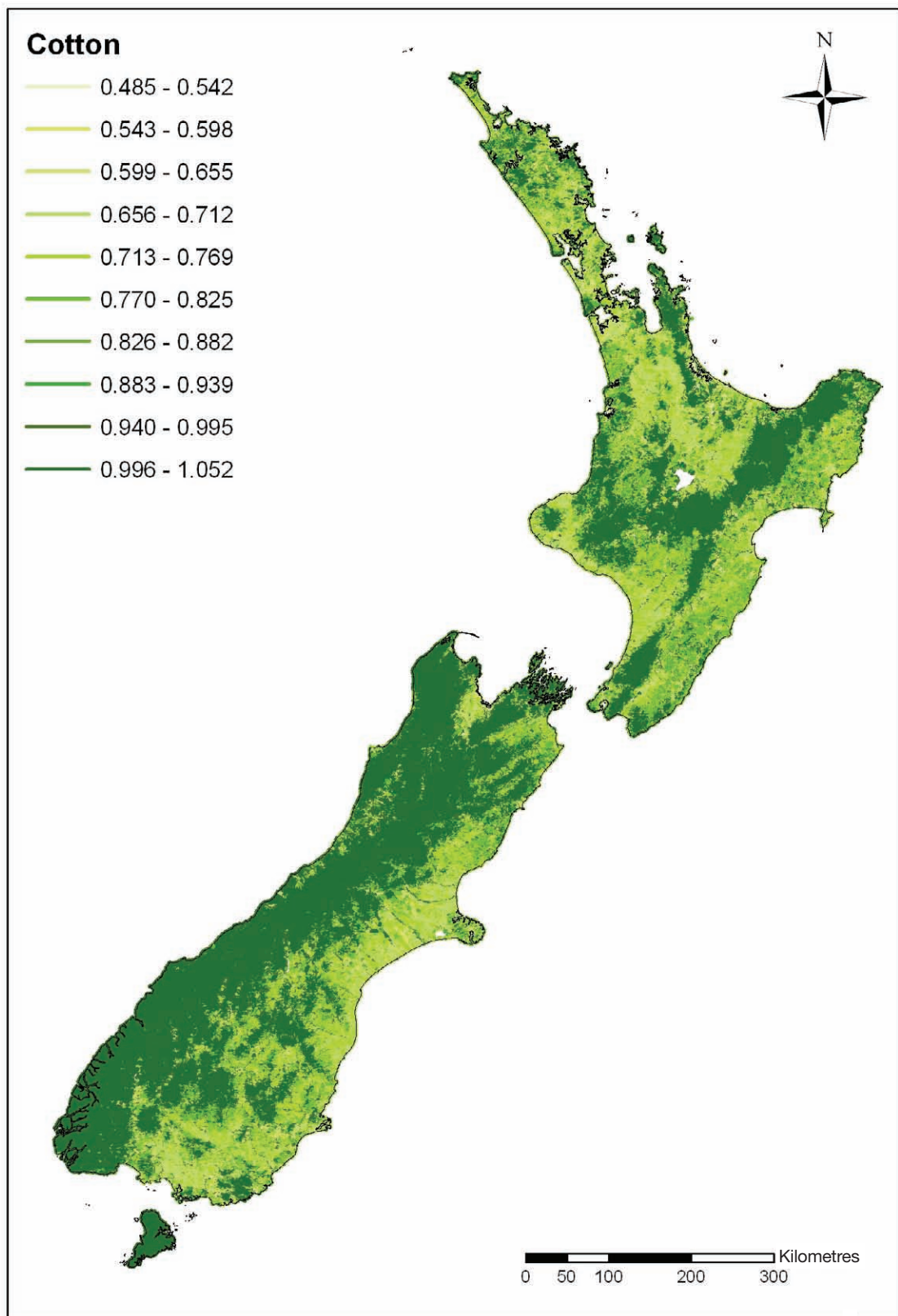
Figure 23. Geographic variation in observed-to-expected (O/E) scores for *Cotton*. Scores are mapped using equal intervals between maximum and minimum values.
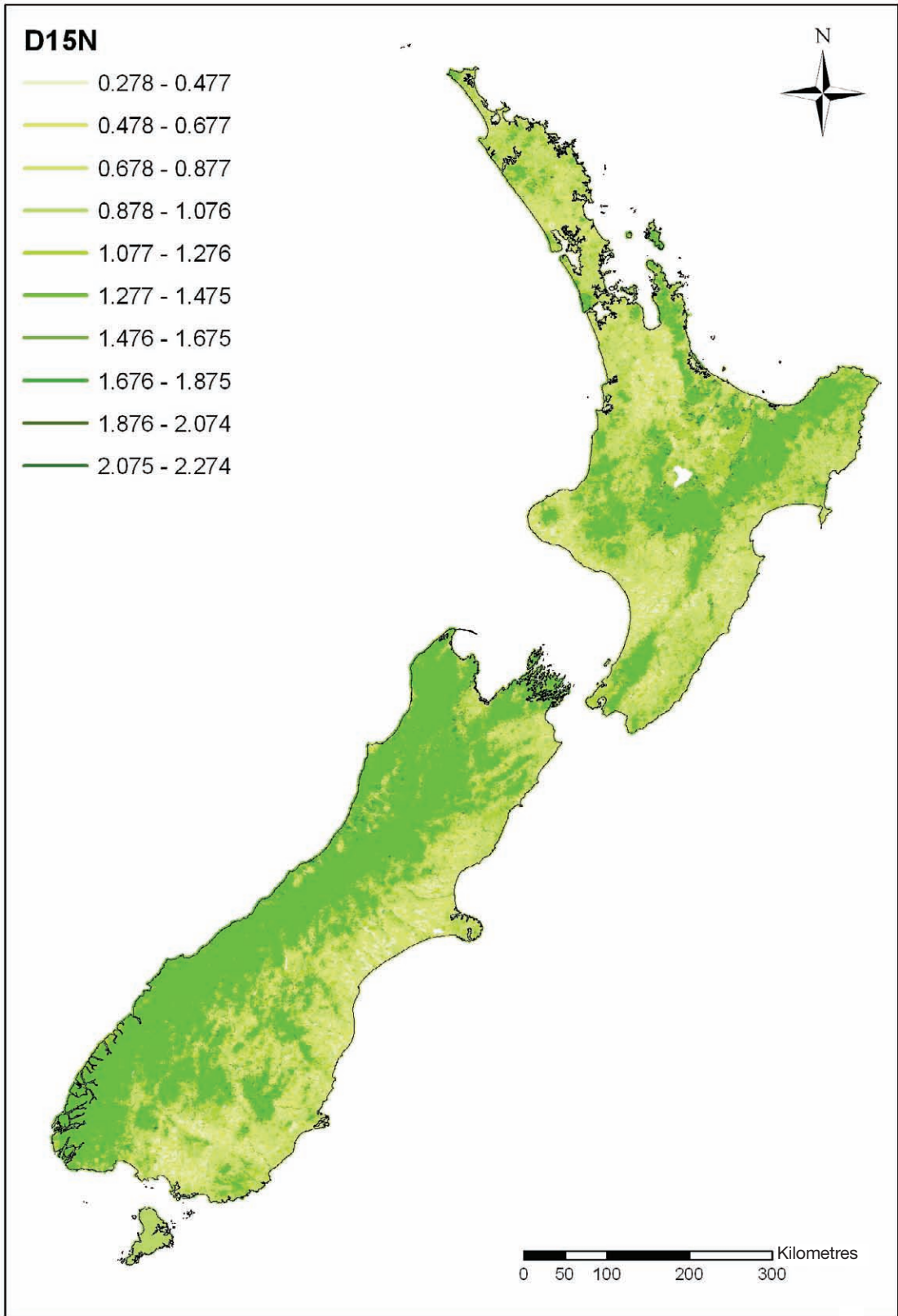
Figure 24. Geographic variation in observed-to-expected (O/E) scores for *Del15N*. Scores are mapped using equal intervals between maximum and minimum values.
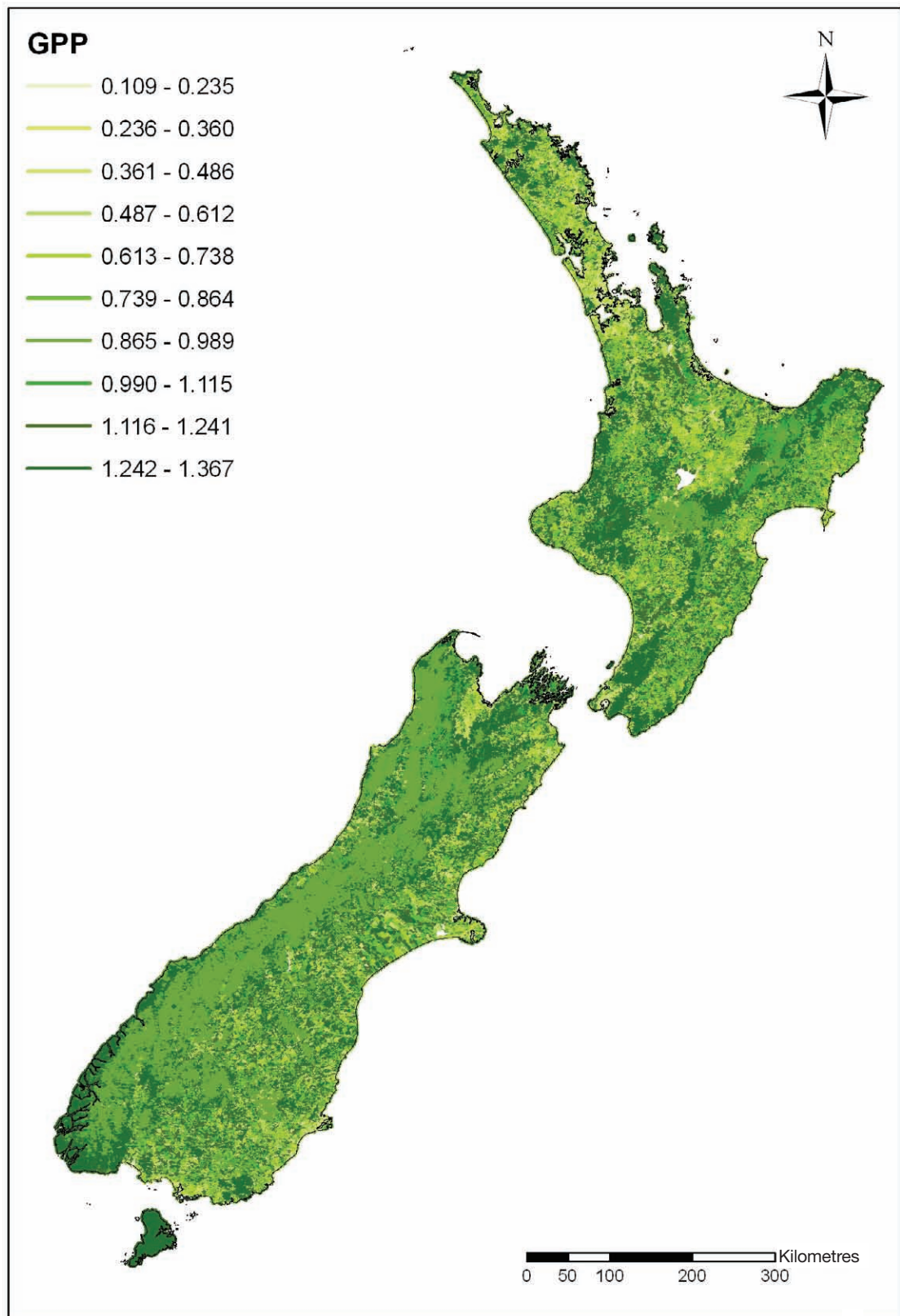
Figure 25. Geographic variation in observed-to-expected (O/E) scores for *GPP*. Scores are mapped using equal intervals between maximum and minimum values.
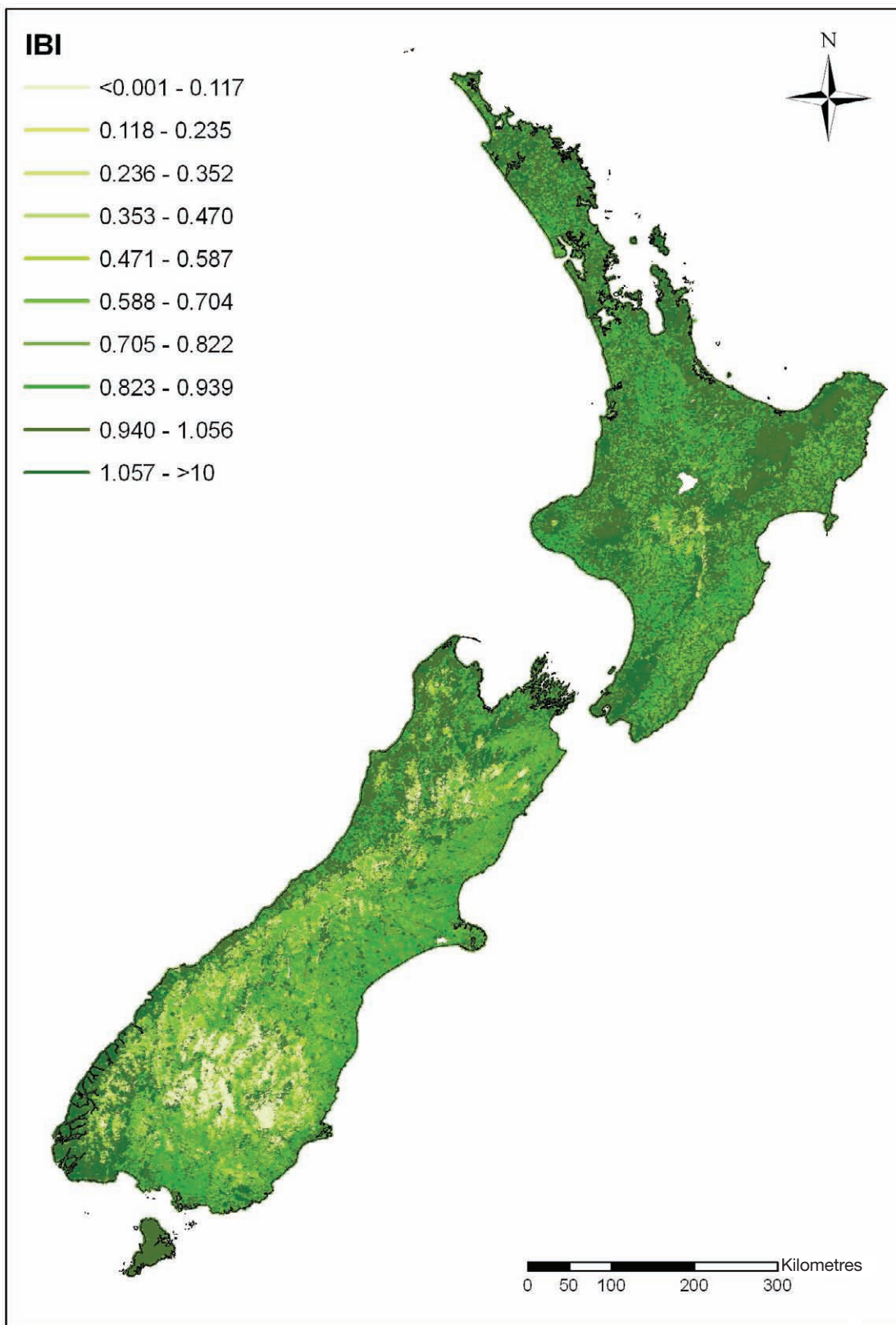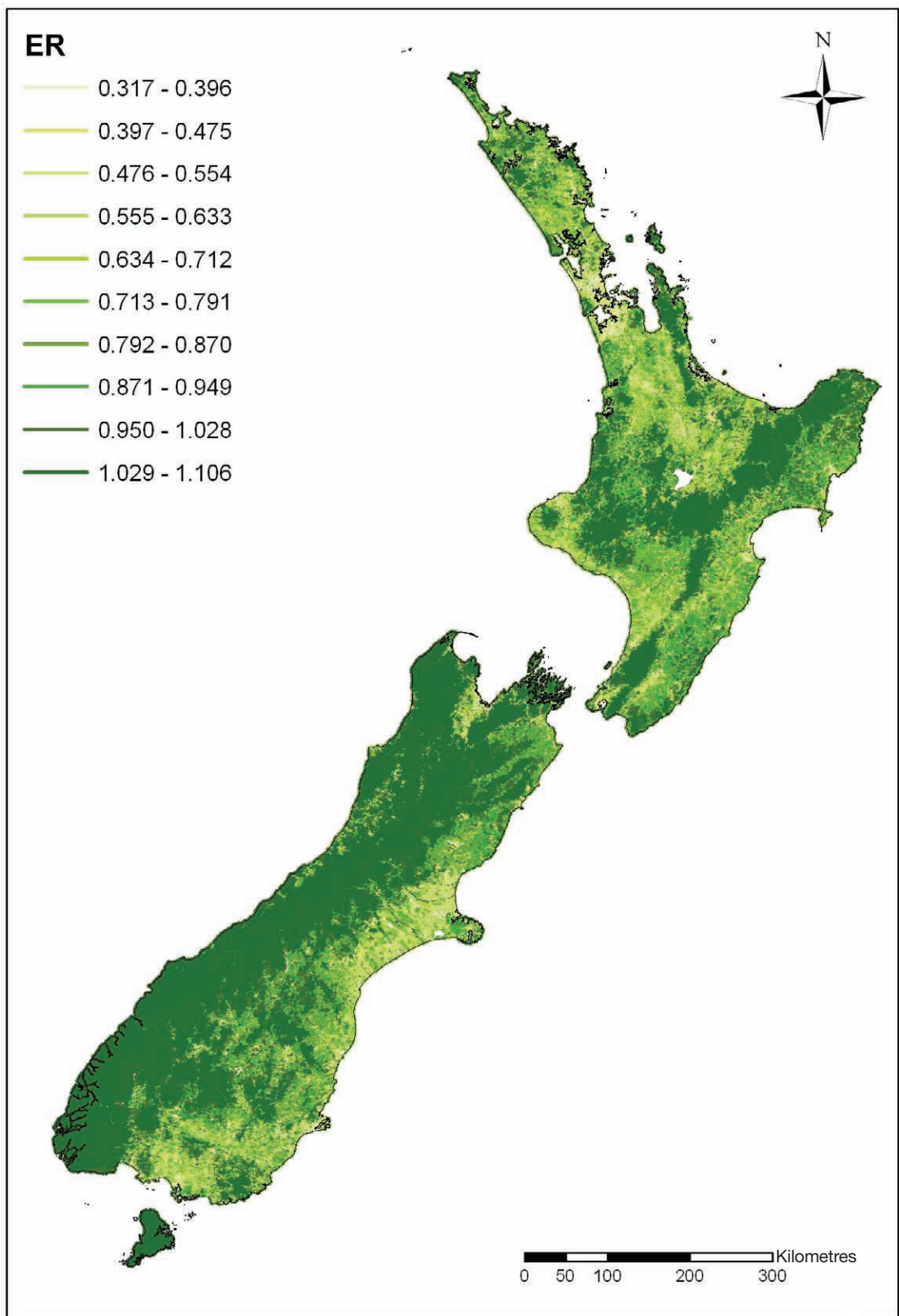
Figure 26. Geographic variation in observed-to-expected (O/E) scores for *ER*. Scores are mapped using equal intervals between maximum and minimum values.
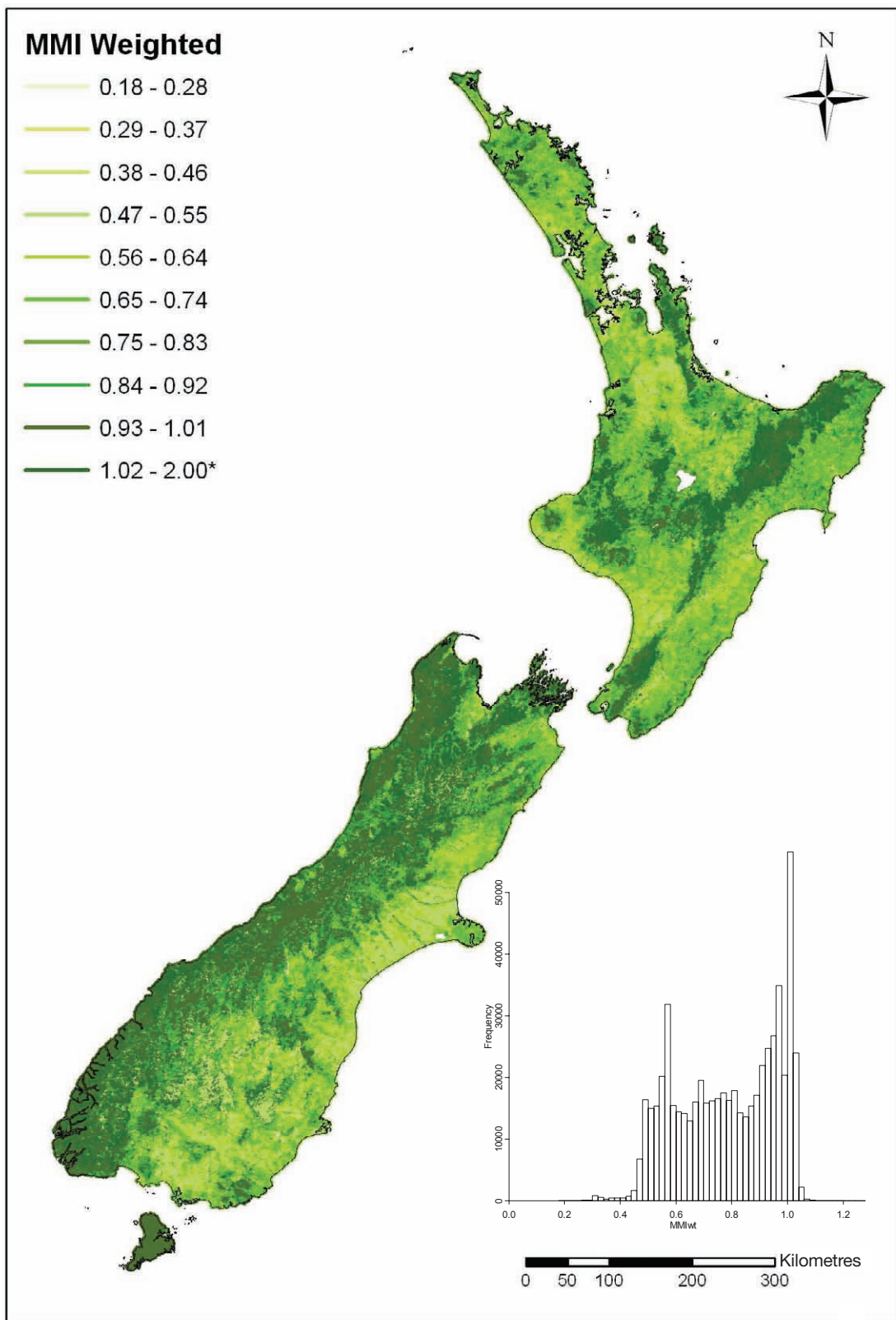
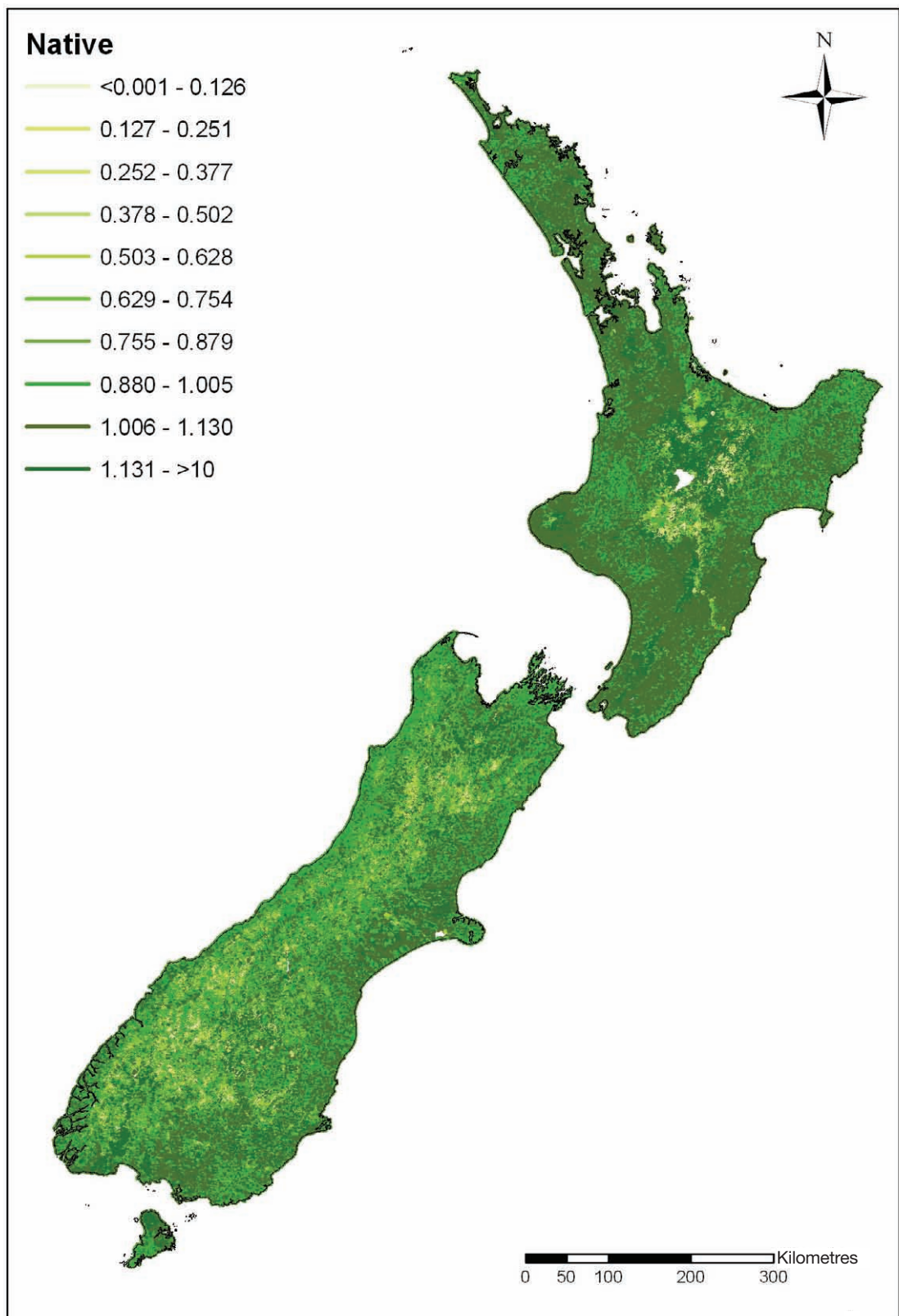Figure 27.   Geographic variation in values for the multi-metric index of ecological integrity (*MMIwt*). Scores are mapped using equal intervals between maximum and minimum values, except for the final group (*), which includes a greater range in values. A histogram is included to illustrate the data distribution.

Clapcott et al.—Relationship between land-use pressures and indicators of stream ecological integrity

4. Discussion

4.1 Quantifying the relationships between human pressures and indicators of river integrity

Our results demonstrate the nature of the response of 14 different indicators of stream EI to three land-use pressures. The nature of each response is characterised by the form of the relationship in terms of direction and shape, and it is also characterised by the strength of association. In contrast to previous studies, which have demonstrated a categorical difference between the integrity of 'impacted' and 'non-impacted' streams, this study demonstrates the response along a gradient of land-use stressors. This information is valuable for understanding the response of stream ecosystems across the broad range of pressures in a contemporary New Zealand landscape.

Water quality

Out of all of the indicators of EI investigated, the nitrate+nitrite concentration (*NOx*) in water showed the strongest response to human pressure and varied significantly from the mean value in response to all three pressures investigated. Steep monotonic increases in *NOx* accompanied increases in predicted nitrogen loading, native vegetation loss and impervious cover. The response of *NOx* to nitrogen loading provides an external validation of the model used to predict in-stream nitrogen loads (i.e. CLUES). However, vegetation loss rather than nitrogen loading best explained deviance in the *NOx* data. This probably illustrates the strong correlation between *LogN* and *VegL* at a national scale, with greatest divergence between the two pressures occurring at greater than 90% vegetation loss where the influence of land-use intensity becomes pronounced in *NOx* data. Hence, 90% native vegetation loss suggests a distinct threshold for the change in the response of *NOx* to land use.

But what does the nitrate+nitrite concentration of water tell us about the EI of streams? Previous studies have shown strong correlations between nutrient concentrations and biological response in streams (Sheeder & Evans 2004; Gücker et al. 2006; Soranno et al. 2008; Townsend et al. 2008). Furthermore, lower trophic levels have been shown to have a stronger response to nutrient pressure gradients than higher trophic levels. For example, Johnson & Hering (2009) observed that nutrient enrichment was most correlated to diatom metrics, then macrophytes, then invertebrates, then fish. This may illustrate a lagged response to disturbance through the food chain, or—more likely—that higher trophic levels are more influenced by habitat (top-down) than nutrients (bottom-up) (Niyogi et al. 2007; Johnson & Hering 2009). Either way, nutrient concentrations as spot measurements or as nutrient accumulation in consumers (i.e. *Del15N*) appear to provide a good indication of nutrient effects on stream ecosystems.

The three land-use pressures investigated explained more than 20% of deviance in the *DRP* data. *DRP* showed a significant monotonic increase in response to *LogN* and general increases across a truncated range in *IC* and *VegL*. Like other water quality metrics, *DRP* appeared most sensitive to small amounts of *IC* in a catchment and also responded when *VegL* exceeded 50%. The response of *DRP* was similar to *NOx*, reflecting the correlation between these variables.

Almost 30% of the deviance in *Clarity* data was explained by the three land-use pressures investigated. *Clarity* showed a steep monotonic decrease in response to *LogN* and a gradual monotonic decline in response to *IC* and *VegL*. A strong link has been demonstrated between land use and *Clarity*, with the most likely causal pathway being increased delivery of in-stream sediment resulting from disturbance of the landscape (Quinn & Stroud 2002). The ecological effects of reduced clarity are many. At a low level, reductions in water clarity may influence fish feeding and distribution, while at higher levels, effects include changes to plant production rates and to the sensitive organs of fish and invertebrates (e.g. Hicks & Barrier 1996; Hay et al. 2006).

Furthermore, *Clarity* can provide an indication of the likelihood of deposition of sediment on the riverbed with associated effects on food and habitat quality. *Clarity* may provide an early indication of land-use effects on EI, but may also indicate immediate effects in terms of reduced biotic capacity, i.e. feeding and production. An apparent threshold for a change in the response of *Clarity* was 3.16 g/m³ predicted nitrogen load, although it is difficult to discern whether this is a natural threshold or the result of the data distribution; i.e. few sites had less than 3.16 g/m³. There were also potential thresholds in the response of *Clarity* to vegetation loss at 10% and 50% native vegetation loss.

For all water quality metrics, the form of the response curves suggests a significant difference between values of metrics at 'low' and 'high' levels of all three land-use pressures.

Invertebrate metrics

The three land-use pressures explained approximately 35% of the deviance in *MCI* and *Cycle* data. There was a significant negative relationship for both metrics in response to native vegetation removal and impervious cover, as seen in previous studies (e.g. Quinn et al. 1997; Doledec et al. 2006). While both metrics had strong monotonic relationships with the pressures, there were also apparent thresholds for a change in the ecological response curves at 50% *VegL* for *Cycle* and 30% and 80% *VegL* for *MCI*. These broad-scale responses appear applicable at regional levels; for example, a recent study suggested a 20% loss of vegetation cover as a lower threshold for 'clean' water quality based on an assessment of *MCI* in Waikato streams (Death & Collier 2009).

There was a distinct non-monotonic response in all four invertebrate metrics to *LogN*, even within the range of stressor values for which there was confidence. This pattern may illustrate a subsidy-stress response, where an initial increase in *LogN* may provide favourable conditions for invertebrates (e.g. enhancing food and habitat availability) until a threshold is reached and further increases in *LogN* lead to detrimental effects (e.g. through competition, eutrophication, exceeded tolerance levels). This idea is not new and has been shown to be applicable in stream systems (Townsend et al. 1997b; Quinn 2000). The results of this study support such concepts but do not test the causal mechanisms for the observed patterns. However, the findings do confirm that non-monotonic patterns (e.g. unimodal patterns) are real in the environment and, therefore, previously suggested increases or decreases in invertebrate metrics in response to land-use intensity (e.g. Harding & Winterbourn 1995; Quinn et al. 1997; Townsend et al. 1997a) are likely to be due to viewing the response across a limited range in the pressure gradient. The level of stress at the inflection point of the unimodal curves for all invertebrate metrics was approximately 3.16 g/m³ predicted nitrogen load.

Fish metrics

The relationships between the fish metrics and the three land-use pressures were consistently weak, with less than 12% of the deviance in the data explained by the pressure gradients. LogN made the strongest contribution to any relationship and fish metric values tended to increase in a monotonic fashion over the range of stressor values where there was confidence. The marked improvement in the explanatory power of the BRT models after including environmental variables indicated that factors such as distance to the coast and maximum downstream slope are critical in influencing fish community composition, as has been widely recognised in the past (Hayes et al. 1989; Minns 1990; Joy et al. 2000). The F-IBI was specifically developed to take account of the fact that the number of species will decrease with increasing distance from the sea and elevation (Joy & Death 2004). However, downstream distance to the coast was the main contributing variable for the F-IBI, suggesting that this variable still confounds interpretation of the F-IBI. Therefore, the positive response of the F-IBI over the range of LogN values where there was confidence may be due to the tendency in the dataset for elevated nitrogen loading to occur at low-elevation sites close to the coast, rather than being a direct impact of nitrogen loading on the fish community.

Ecosystem process metrics

There was less confidence in the model outputs for process metrics compared with other indicators. This probably reflects the small sample sizes of process metric datasets compared with those for other metrics. Also, the difference in the strength and form of relationships observed in the full national analysis of process metrics, compared with a regional analysis (Clapcott et al. 2010), suggests there is an unaccounted source of deviance in the national dataset. Strong temporal variation in the metrics may be one source of this error that could be corrected in future analyses (see also section 4.3). Poor contribution of *Group* to the model could also be related to the narrow range of FENZ stream types in the dataset.

In general, process metrics showed a positive response to an increase in all three land-use pressures. In particular, it appeared that the threshold for response to vegetation clearance was approximately 40% for all metrics. Exceptions to positive responses were a decrease in *Del15N* in response to increasing *IC*, and a non-monotonic response by *Cotton* to increasing *VegL* and *LogN*. The unimodal patterns observed for *Cotton* are likely to reflect the multiple pathways in which changes in land use can affect the processing of in-stream organic matter. Other studies have noted antagonistic responses as a result of land-use change, such as the opposing influences of decreased shredder density v. increased nutrient availability (Huryn et al. 2002; Niyogi et al. 2003), or increased stream acidification v. a change in litter composition (Gessner & Chauvet 2002). As such, the *Cotton* metric has the potential to provide an assessment of a range of ecosystem interactions, if there is an understanding of confounding response pathways (e.g. Nyogi et al. 2007).

The steep monotonic increase in *Del15N* in response to *LogN* was also observed in recent studies investigating the potential of the $\delta^{15}N$ of primary consumers to provide a time-integrated measure of nitrogen loading in-stream (Lake et al. 2002; Anderson & Cabana 2006). Similarly, it has been shown that the $\delta^{15}N$ of primary consumers also provides a good estimation of the degree of denitrification occurring in a watershed (Diebel & Zanden 2009). Hence, *Del15N* appears to provide a good indicator of the integrity of nutrient processes and the current lack of confidence in the national model is likely to improve by expanding the small dataset.

4.1.1 Comparison of indicator response

Based on the percentage deviance explained and cross-validation coefficients in the BRT results from Model A (response to three land-use pressures), the *NOx* concentration of water and the $\delta^{15}N$ signature of primary consumers showed the greatest response to the three land-use pressures investigated. It could, therefore, be argued that these indicators provide the best measures of human pressure. However, it is important to assess each indicator individually, taking into account the confidence of the model as illustrated in the functional plots; sample size; which of the pressures were most correlated with the indicators; and the range of the pressure gradient over which the indicators were responding. Furthermore, in terms of national applicability, it is important to consider the spatial coverage of the data used to build the model, as indicated by the coverage of contributing FENZ groups.

NOx values vary significantly from mean *NOx* in response to all three pressures. Combined, the pressures explain more than 56% of the variability in the data and cross-validation coefficient of the BRT model is almost 70%. *NOx* values represent 16 FENZ groups. Together, these FENZ groups account for 74% of stream segments in New Zealand. In contrast, while the cross-validation coefficient is high for the *Del15N* model, contributing data are representative of only 31% of New Zealand streams, limiting the national applicability of this model. Following this rationale, the invertebrate metrics *MCI* and *Cycle* are also good indicators of stream integrity at a national scale because the dataset includes a wide range of stream types and the metrics have a significant response to the pressure gradients. In contrast, the poorest correlations were observed for fish indicators and process metrics because variation in the fish metrics is not explained well by the three land-use pressures (i.e. weak models) and process metrics are limited in their spatial representation and some also have a weak response to the pressures.

The shape of the response to the three land-use pressures further grouped indicators based on their suitability as indicators of EI. Unless complimentary measures are available to elucidate the patterns or there is a good mechanistic understanding of the pattern, non-monotonic responses in indicators are generally seen as problematic for assessing ecosystem health (Downes et al. 2002). Non-monotonic responses to the pressure that explained most of the deviance in the data were observed for taxonomic richness metrics, fish metrics and *Cotton*. However, care must be taken not to dismiss indicators based on non-monotonic responses when such a response is outside of the range in values for which there is confidence. For example, the response of fish metrics to *LogN* is (monotonic) positive within the bounds of –1 and 1 *LogN*, which is where most sites are located.

In addition to the direction of the response, the shape of response provides valuable information. For most metrics, some degree of non-linearity is present that illustrates the likelihood of land-use thresholds. Thresholds are points where even small changes in environmental conditions will lead to large changes in response variables (Suding & Hobbs 2009). These are important because they show potential limits that could be used as standards to protect the EI of river systems, or as land-use criteria for the identification of impairment categories. Identification of thresholds requires an understanding of the patterns and the identification of broad-scale drivers, both of which are outcomes of this study. However, the analytical approach used in this study does not provide a formal test of threshold values. The results suggest potential thresholds exist at less than 10% impervious cover, 0.1 and 3.2 g/m^3 nitrogen and between 20% and 90% vegetation loss, depending on the indicator.

Previous studies have suggested that specific thresholds of land-use change may exist from which we can predict a change in the response of ecological indicators. For example, Wang et al. (2001) illustrated how a change in impervious cover from 8% to 12% in a catchment leads to major changes in stream condition based on fish community metrics. Similarly, Donohue et al. (2006) suggested that the ecological status of rivers based on a biological index was impaired when catchments were greater than 69% agricultural. In contrast, Gergel et al. (2002) suggested 20% urbanisation and 50% agricultural development were thresholds for change in habitat, biological and water quality indicators. In comparison, Anderson & Cabana (2006) recently suggested that as little as 5% catchment development was a threshold that indicated nitrogen saturation and the delivery of excessive nitrogen to waterways, affecting nitrogen processing. These studies highlight the potential for critical thresholds in land-use change, and that these thresholds may differ depending on geographical location and environmental characteristics. Our BRT analyses provide a preliminary identification of thresholds by examination of model output.

4.1.2 How useful are measures of pressure at a catchment scale?

We believe measures of anthropogenic pressure have the potential to be very useful. Consistently, catchment-scale measures of land use are reported as having strong predictive relationships with stream health (Roth et al. 1996; Allan et al. 1997; Herlihy et al. 1998; Young & Huryn 1999; Gergel et al. 2002; Weijters et al. 2009). Our results support these studies, as we examined responses across a full gradient in catchment vegetation removal, predicted nitrogen loading and impervious cover. The fact that there were significant relationships evident between catchment-scale measures of land-use pressure and some indicators of EI supports the usefulness of broad-scale assessments. Among the three pressures, loss of native vegetation was the primary explanatory variable in the distribution of *NOx*, *Cycle*, *EPT*, *MCI* and *GPP* data; nitrogen load was the primary explanatory variable for *Clarity*, *ITR*, *F-IBI*, *Native*, *FTR*, *Cotton* and *Del15N* data; and impervious cover was the primary explanatory variable for *DRP* and *ER* data. There are correlations between pressure variables (Scarsbrook 2008) and yet, when used together in BRT analysis, different pressures complement each other, providing improved understanding of responses. For example, *NOx* responds best to *VegL*, with a marked response above 90% *VegL*; however, at lower levels of this stress, the relationship between *NOx* and *LogN* is likely to be more useful in elucidating patterns in *NOx* values.

Catchment-scale measures of land use provide a good tool for broad-scale planning, but do not provide an understanding of the causal pathways for responses. Studies that have attempted to establish causal links using surveys suggest in-stream drivers are affected by land use through a series of pathways (Allan et al. 1997; Poff 1997; Burcher et al. 2007). For example, Lenat & Crawford (1994) illustrated how invertebrate and fish composition responded to differences in land use as the result of changes in sediments, nutrients and contaminants. Hence, a mechanistic understanding of the correlative patterns observed between indicators of EI and land-use stress requires knowledge of smaller-scale land-use patterns.

Previous studies (Snyder et al. 2003; Strayer et al. 2003; Weigel et al. 2003) have suggested that more localised measures of land use can provide better measures of stress than catchment-level measures, at least for some in-stream responses. For example, Weigel et al. (2003) showed how invertebrate composition was most correlated with catchment-scale land use, whereas relative abundance and presence/absence of specific taxa appeared more influenced by reach-scale variables. Similarly, Strayer et al. (2003) observed a stronger relationship between macroinvertebrate richness and riparian-scale land use compared with catchment-scale land use. In contrast, Snyder et al. (2003) showed that catchment-scale land use had a greater influence on fish communities than reach- or local-scale land use. These contrasting results suggest benefit might be gained by quantifying land-use pressures at spatial scales other than the watershed, especially if the intention is to describe response pathways and, potentially, identify filters for minimising the impact of catchment-scale land use at a local scale (*sensu* Gergel et al. 2002). Improved spatial resolution in catchment land use may help improve the predictive relationship between pressures and stream integrity.

Consideration could also be given to the spatial arrangement and juxtaposition of land uses within a catchment (King et al. 2005) as well as the temporal change in land use, i.e. the ghost of land use past (Harding et al. 1998). Based on the relationship between nitrate and macroinvertebrate assemblages and four different measures of land use, King et al. (2005) illustrated how spatial autocorrelations and arrangement can affect the predictability of in-stream condition. Results suggested that the percentage of cropland in the catchment had more influence on invertebrate communities in large watersheds than in smaller watersheds. Similarly, invertebrates were affected more by certain arrangements and the proximity of developed land in the catchment. Weighted land-use indices have been widely investigated and consideration of patch configuration suggested (e.g. O'Neill et al. 1997; Van Sickle & Johnson 2008). The FENZ database provides sufficient spatial resolution to investigate land-use configuration at the stream segment scale rather than catchment scale. This information could be investigated in future analyses.

4.1.3 Accounting for environmental variability

Pont et al. (2006) suggested that the most useful measures of stream health are insensitive to natural variability. But the BRT approach showed that when we have measures (or at least estimates) of natural variability, these can be used to elucidate patterns in the response to pressure across the landscape, provided that there are no confounding correlations (e.g. nitrogen and *F-IBI*). For all indicators, the inclusion of environmental predictors doubled, on average, the sensitivity and precision of models in comparison with the initial response of indicators to land-use pressures. Environmental variables that increased the deviance explained in subsequent models were indicator specific. As expected, downstream variables were significant contributors to explaining the deviance in fish indicator data—in particular, distance to the coast. For macroinvertebrate metrics, catchment rainfall, segment flow stability, segment slope and temperature were significant environmental predictors, while for indicators of water quality and ecosystem processes, a combination of temperature, slope and flow variables significantly improved the ability of models to explain deviance in the data. Unaccounted deviance in the data may, perhaps, be explained by additional environmental variables.

4.1.4 How useful are broad-scale classifications of stream categories?

The inclusion of stream-type categories significantly improved the explanatory power of individual BRT models. This suggested that FENZ classifications provide a meaningful summary of how natural environmental variability may further influence stream health. Interpretation of the results suggests that where *Group* has a relatively large contribution to the proportion of deviance explained, the response to pressure will vary by stream type. Hence, stream types could be used to determine the average response for particular stream categories. The grouping of streams is often used as a way to ensure degree of departure from reference is tested against suitable reference conditions (e.g. Herlihy et al. 1998; Norris et al. 2007). It appears that the *Group* classification used in this study (i.e. FENZ) is useful for the identification of ecoregions and potential assessments of EI within those classifications. It provides a tool for small-scale studies, local authority investigations or research. For example, FENZ classification could be used to identify a reference stream to assess the restoration success of a stream of the same stream type. However, at a national scale, there may be no need to stratify reference condition by *Group* because the environmental data used to classify groups are available, and this detailed data can be used directly in modelling to calibrate the relationship between EI, land use and natural variability.

4.2 Assessment of predictive models

Reference values

Values from low-pressure sites were used to predict reference condition throughout the country. These values were comparable with suggested guideline values for assessing stream health (ANZECC 2000). The reference value for *NOx* was 10-fold less than that recommended for good water quality in the ANZECC 2000 water quality guidelines, although the range in subsequent modelled-E values exceeded the recommended value for upland streams (i.e. 0.167 mg/L). Water quality guidelines (ANZECC 2000) further suggested that *Clarity* should not be less than 1.6 m for swimming or should not be reduced by 20%, to maintain aesthetic values. The reference value in this study for *Clarity* was 4.8 m and modelled-expected values ranged from 2.06 m to 11.8 m; all values were much greater than 1.6 m. Therefore, reference values for water quality metrics provided a much higher baseline than currently recommended Australasian guidelines.

The reference value for *MCI* was c. 138, much greater than the cut-off of 120 recommended by Stark (1998) as indicating pristine condition, although, across the country, modelled-E values did range from 97 to 162. There are no recommended guidelines for the percentage of invertebrate species with a univoltine life cycle. However, the reference value for *Cycle* was 93%, much higher than the 45% of species observed by Doledec et al. (2006) in minimally impacted tussock streams. Again, these comparisons suggest that the offset used in modelling E values for invertebrate metrics provided a high reference baseline.

Intuitively, a reference value for the proportion of native fish in a stream would be 100%. However, the reference value for *Native* in this study was 83% and, at a national scale, a full range (i.e. 0–100%) was expected. Similarly, in the development of a fish IBI, the highest value of 60 represents a pristine ecosystem (Joy & Death 2004), and, in this study, the reference *F-IBI* from streams with low stress was 38. These values suggest that the models used to predict fish metrics at a national scale do not currently capture a representative picture of fish communities in New Zealand.

For most ecosystem process metrics, guidance on the appropriateness of reference values can be taken from the literature. The reference value for *GPP* (3.04 $gO_2 m^{-2} d^{-1}$) was less than the 3.5 $gO_2 m^{-2} d^{-1}$ recommended by Young et al. (2008) as indicative of a healthy stream, whereas the reference value for *ER* (6.5 $gO_2 m^{-2} d^{-1}$) was greater than the 5.8 $gO_2 m^{-2} d^{-1}$ recommended by those authors. However, recent research suggests that a much greater range in *ER* values may be observed at relatively non-impacted sites, especially those subject to temporal variability

(Roberts et al. 2007; Young & Clapcott 2010). There is, therefore, potential for developing seasonally corrected reference values for the future assessment of stream integrity using *GPP* and *ER*. The assessment of organic matter decomposition using *Cotton* already corrects for seasonal variation due to temperature. The reference value for *Cotton* in this study (0.0016 k/dd) was very low in the authors' experience. Comparing these rates with those from other studies is problematic because of differences in application; however, an international standard approach is currently being developed (authors, in collaboration with Scott Tiegs, Oakland University, California). Similarly, there is currently no recommended value for the $\delta^{15}N$ of primary consumers in healthy ecosystems. However, the reference value in this study (5.2‰) is very similar to that observed for primary consumers in relatively un-impacted streams on the West Coast of New Zealand (Harding et al. 2004), in Canada (Anderson & Cabana 2006) and the USA (Diebel & Zanden 2009). These values, combined with the low range in modelled-expected values, suggest that the *Del15N* reference value provides a high and consistent baseline for assessing deviation due to land-use pressures.

Data distributions

Models that predict low variability about the mean expected value result in relatively standard reference values that are applicable throughout the country (e.g. *NOx*, *Del15N*, *Cotton*, *ER* and *GPP*). In contrast, for metrics where models predict a wide range in E values, reference will vary among different regions and stream types (e.g. *Clarity*, *MCI*, *Cycle*). In both circumstances, FENZ groupings could be used to calculate the mean and range in expected reference for each bioregion. However, what both these types of distributions do not take into account is the contribution of temporal variability. As suggested, some metrics may be subject to natural temporal variability that would result in predicted reference conditions varying throughout a year. The current analysis has been conducted on compiled datasets that include samples collected across seasons. Therefore, for any metric that shows a strong response to land use, it is likely that land use is more influential than any effect of temporal variability. However, weaker models might be improved in the future by accounting for temporal variation in the datasets (e.g. *GPP* and *ER*).

It makes sense that models did not predict outlying values for some measures because the modelled values were driven by the response to three land-use stressors. In reality, there are numerous other stressors that influence ecological responses that were not taken into account in this study, such as point-source impacts or flow regulation. These additional pressures are likely to result in the outlying values observed in sample data that were not seen in modelled-O data. It is likely that improved model predictions would be achieved by the inclusion of additional pressure variables.

The geographical patterns observed in the modelled output were sensible for measures of water quality, invertebrates and *Del15N*. This probably reflects the strong influence of three land-use pressures (*VegL*, *LogN*, *IC*) in shaping the patterns in metric responses, as previously illustrated by the percentage deviance explained in Model A. These patterns further support the strong evidence in the literature of a link between catchment-scale land-use pressure, and water quality (e.g. Townsend et al. 1997a; Larned et al. 2003) and macroinvertebrate community composition (e.g. Allan 2004; Death & Collier 2009).

Metrics with weaker models appeared to predict distributions with much more variation at smaller spatial scales, such as those seen for *CottonOE*, *EROE* and *GPPOE*. With initial consideration, this suggests that environmental variability may have more influence on these metrics than land-use pressures. However, an earlier study showed that these metrics were less variable across stream types when compared with fish and invertebrate metrics (Clapcott et al. 2010). Therefore, it is more likely that these patterns reflect weak models rather than unaccounted natural variability, probably because of the limited national coverage of the datasets.

The high variability within modelled fish metrics and their spatial distributions illustrate weak predictive models, but also an interesting pattern in terms of distance to coast. The lack of a

significant response between fish metrics and land cover is likely to be confounded by the fact that native vegetation cover tends to be greater in inland waters, whereas native fish distributions have strong longitudinal gradients driven by diadromy (McDowall 1993). Hence, the response of fish metrics to land-use pressure is confounded by distance to coast. Percentage native fish is a component of the *F-IBI*, but this metric was developed with a linear correction for distance to coast (Joy & Death 2004). However, it appears that this covariate is yet to be fully accounted for in a national model. This is further supported by the fact that boosted regression tree models showed *DSDist2Coast* significantly increased the predictability of *F-IBI* and *Native* metrics in addition to land-use pressures (Table 4, Appendix 2). The fish model results highlight the potential for correlation between land use and environmental variables to confound predictions and indicate that the O/E approach may not be suitable for some metrics.

Cumulative assessment

The results of this second part of the study further illustrate how different MMI components can provide contrasting or complementary information in the assessment of EI. A qualitative evaluation based on water quality metrics alone would suggest that, on average, 25% of stream reaches in New Zealand have 'poor' ecological integrity (i.e. O/E scores less than 0.3). An evaluation based on macroinvertebrate metrics would suggest that less than 1% of New Zealand's stream segments have poor EI, while little more than 1% of segments would be considered to have poor EI based on an evaluation of ecosystem process metrics. Similarly, Carlisle et al. (2008) found limited concordance between the number of 'impaired' sites predicted by independent evaluations of fish, invertebrate and diatom metrics. These contrasting assessments suggest assigning ordinal classifications, based on a prescribed value for example (e.g. >0.9 = high EI), may not be the best way to judge EI. In this case, analysis of thresholds in the response curves of metrics may provide a better indication of where changes in 'grades' of EI occur. Alternatively, the contrasting evaluations may show that water quality metrics provide an early warning indicator of degradation in EI, whereas the impact of land-use in other metrics may take some years to manifest (Harding et al. 1998).

The possibility that assessments of differing in-stream components are likely to provide contrasting qualitative assessments of ecosystem integrity is not a new concept. Studies testing multiple indicators have shown how one indicator responds better to a certain stress than another (e.g. Griffith et al. 2005), provides a better signal in response over time (e.g. Stribling et al. 2008), provides a faster response over time (e.g. Adams et al. 2002), or provides less spatial variability in response (e.g. Hughes et al. 2000). In summary, few metrics are sensitive to all stressors, so the argument has been made that a combination of indicators is required to assess multiple stressors and all components of EI (Hughes et al. 2000; Resh 2008). Hence, the next challenge is to increase our understanding of redundancy and complementarities among metrics, i.e. what does a response in one metric tell us in comparison with the response of another about the overall EI of a system?

4.2.1 A combined EI metric

The results of this study show that the MMI calculated by weighting the contribution of component metrics based on their response to land-use pressure and on national coverage provides a meaningful assessment of EI. However, the MMI is strongly driven by those metrics that respond best to the three land-use pressures investigated and is limited in its assessment of all components of EI.

The form of the response to the three land-use pressures appears most similar to the metrics with greatest influence, i.e. those of water quality and invertebrates. However, the influence of other metrics is also apparent, especially those of fish. It appears that weaker models (such as those for fish) introduce noise into the MMI predictions, both in the form of the response and in geographical patterns. This noise could be addressed by further down-weighting the contribution of fish metrics to the MMI until more robust models are developed (e.g. through reanalysis with new pressures or development of new fish metrics). Alternatively, metrics with weak models

Clapcott et al.—Relationship between land-use pressures and indicators of stream ecological integrity

could be removed from the MMI, but this would result in a metric that is less representative of EI as previously defined. Importantly, at this stage, the analytical approach developed in this study provides a framework for improving predictions of EI as more data become available.

In this study, we attempted to standardise component 'scores' by calculating O/E metrics, where closer to or greater than 1 represents good ecological condition. Since O/E scores for the MMI represent EI, those close to 1 represent high EI. Based on a qualitative evaluation of the MMI, the results suggest 36% of stream segments in New Zealand have high EI (>0.9) and less than 1% of stream segments have low EI (<0.3). However, we did not develop a method of assigning qualitative scores in this study and suggest that this may be best achieved by identifying thresholds of change in ecological response curves of component metrics, i.e. different degrees of impairment based on the differing sensitivity of component metrics. For example, the form of the response of several component metrics to *LogN* suggests a distinct threshold for change at nitrogen loads of 0.32 g/m^3. Interestingly, 36% of streams in New Zealand have predicted nitrogen loads less than this value (for 9% of sites, it is less than 0.1 g/m^3). Commonly, the development of an MMI involves scoring component metrics prior to aggregation. This makes the MMI a valuable tool for state of the environment reporting. Scores are assigned based on divergence from reference or natural breaks in the data distribution, e.g. scores 6, 3, 1 (Fore et al. 1996) or 1 to 10 (Hughes et al. 1998). Scores usually result in categorical groupings that make trends in EI readily communicable—e.g. Good, Fair, Poor. This approach could also be further examined in future studies.

4.2.2 Quantifying conceptual response curves

The shape of the ecological response curves for *MMIwt* had similarities and differences in comparison with existing conceptual response curves for pressure estimates (i.e. Leathwick & Julian 2007) (Fig. 28). Based on the model *MMIwt* data, the ecological response to nitrogen loads was a distinct non-monotonic shape with a dip and then a return to EI values around 1 at –1 and then –0.5 *LogN*, respectively (0.1 to 0.32 g/m^3, respectively), then a steep decline to approximately 0.5 EI at 0.5 *LogN* (3.16 g/m^3) (Fig. 28). However, because of limitations in data availability, we can have confidence in the shape of the response only between –0.5 and 0.5 *LogN*, where the pattern is similar to the conceptual response curve. The conceptual response curve suggests a steep decline in EI from 0.9 to 0.4 beginning at –1 *LogN* and continuing to 2 *LogN* (0.1 to 100 g/m^3, respectively) before a gradual decline to 0.3 EI at 3 *LogN* (1000 g/m^3).

The ecological response to *VegL* shows a gradual linear decrease from 1 to 0.5 EI, with a steeper rate of decrease in EI up to 10% vegetation removal and after 90% vegetation removal (Fig. 28). In comparison, the conceptual response curve suggested a relatively slow decline in EI followed by a steep response between 20% and 50% vegetation loss and then a slow decline again until EI is reduced to 0.3 at 100% vegetation removal.
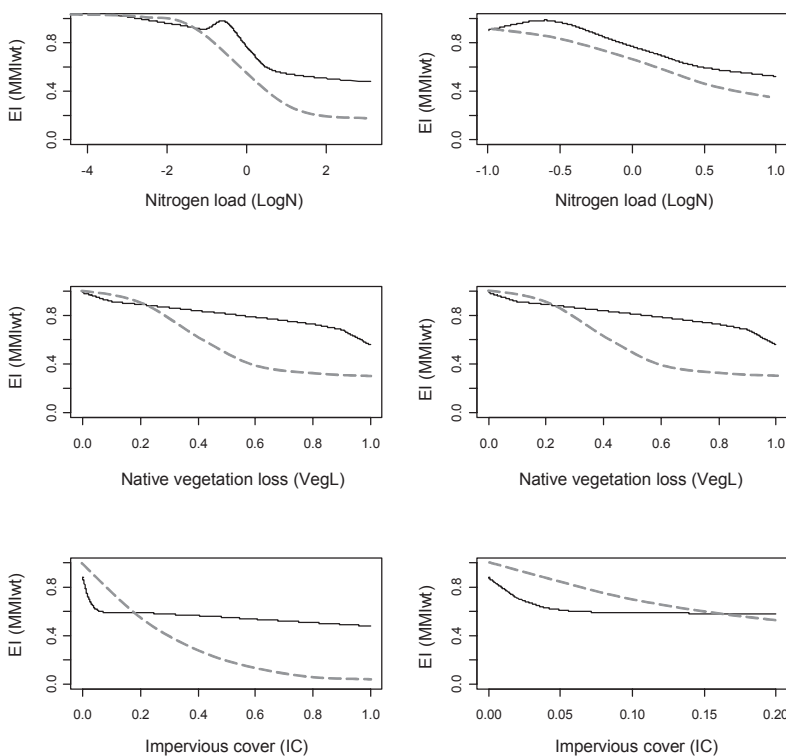

Figure 28. Ecological response curves across a full gradient in land-use pressure estimates (left column) and across a range for which there is confidence in the model predictions (right column). Functions are plotted with a loess-smoothing span of 0.3. Dotted lines are conceptual ecological response curves from Leathwick & Julian (2007).

The ecological response to increasing impervious cover was a rapid decline in EI from 1 to 0.6 as *IC* increased to 5%, followed by a continued but slow decline to approximately 0.45 EI at 1 00% impervious cover (Fig. 28). Most confidence can be placed in the shape of the response up to 20% impervious cover based on the distribution of initial model-building data and the BRT confidence intervals determined during the empirical characterisation of response curves. In contrast, the conceptual response curve for *IC* predicts a rapid decrease in EI, where EI declines to 0.2 after 50% impervious cover and reaches 0 at 100% impervious cover.

Notably, none of the ecological response curves characterised in this study showed EI falling below 0.4 in relation to increases in the three land-use pressures investigated. This may highlight a limitation in the MMI development or, more likely, identifies that other pressures not taken into account in this study (such as invasive species, flow modification or industrial discharges) contribute to decreasing the EI of rivers.

These comparisons show that the conceptual response curves used by Leathwick & Julian (2007) provided a good estimation of the response of EI to land use pressures. The empirical results suggest amendments should be made which identify a much more rapid response to impervious cover and an almost linear rather than sigmoidal response to native vegetation loss. Furthermore, the potential for other pressures to interact with the three studied should be noted and be the focus of future studies.

4.3   Limitations of the study

Data limitations

One source of uncertainty in quantifying the relationship between indicators of EI and land use comes from the temporal discrepancy between land-use data (1996–2000; LCDB2) and indicator data for water quality (1996–2002), fish (1980–2004), macroinvertebrates (1987–2005) and ecosystem processes (1994–2008). It is likely that some of the noise observed in metric response is due to this disparity. In particular, a recent report of the temporal change in fish communities suggests a significant difference between metric patterns from the 1970s to 1990s (Joy 2009). These data are combined in the current analyses. Similarly, it has been suggested that some in-stream measures are more indicative of previous land-use history than contemporary patterns (Harding et al. 1998; McTammany et al. 2007), especially if the land use has resulted in a change in the stream habitat template. Hence, it is important to bear in mind that some of the noise in current models may be due to using data from multiple years.

Most environmental variables used in this study were estimated, rather than measured, for each site. There is a clear need for model validation and this could be achieved by measuring local site characteristics at the same time that response variables are collected. It is important that broad-scale assessments based on model predictions are used primarily as a tool to screen large numbers of basins and identify general relationships, but not for direct management decisions at a local scale (Carlisle et al. 2009).

The response variables included data from sites that spanned the full range of land-use pressure for *IC* and *VegL* (i.e. 0–100%), although 94% of the data have values of less than 20% *IC*. For the *LogN* pressure gradient, 96.8% of the data are within 0.1 and 10 g/m^3 predicted nitrogen loads. Hence, less confidence can be placed in predicting a response outside of these ranges, as is further illustrated by the rug plots and the 95% confidence bands in the figures. To improve the applicability of model predictions, especially for sites with high *IC*, for example, would require expansion of indicator datasets. However, it is consistently shown in the literature that a threshold for the influence of impervious cover on stream systems is less than 20%—typically around 10% or less (Wang et al. 2001; Gergel et al. 2002; Anderson & Cabana 2006). The response

functions calculated in this study also suggest that most of the ecological response occurs at less than 10% impervious cover. Therefore, it is likely that the ecological response curves observed in this study already provide the most meaningful information in terms of the response of stream integrity to the three pressures of interest.

Holistic representation

MMIwt in this study represents four categories of stream health indicators (fish, invertebrates, water quality and ecosystem processes) as well as two components of the intrinsic worth of stream integrity (nativeness and pristineness). While none of the MMI subindices represent the 'diversity' or 'resilience' components of EI, as recognised by Schallenberg et al. (2011), some of the measures have been discussed as providing an indirect assessment of these qualities. The metabolic status of streams has been suggested as a surrogate measure for resilience, with sustained metabolism indicative of ecosystem efficiency (Uehlinger 2000). For example, Acuña et al. (2007) showed that, functionally, streams were more resilient to flood disturbances during summer than at other times of the year because they were efficiently processing carbon entering the stream at this time. Neither of the taxonomic richness metrics performed well in the first part of this study where ecological response curves were empirically characterised and, hence, they were not investigated as potential MMI components. Clearly, there is a need to re-examine diversity metrics or examine alternative diversity metrics to include such measures in a holistic measure of EI.

Data gaps

Predictions of *MMIwt* and each of the component metrics were made at a national scale under the assumption that the model predictions are applicable to all stream types. However, we can have confidence only in the predictions for those streams that form part of the sample data on which the models were based. Any stream types outside this sample group may not respond to pressures in the same way. Output from Model C strongly suggests that different stream types respond differently to the three pressures investigated. Therefore, there is a strong argument for caution in extrapolating any of the modelled values to sites outside the currently sampled database, as indicated in Table 2. The *MMIwt* metric provides some correction for the current limitation in sample representation by down-weighting those metrics with limited data coverage (i.e. ecosystem process components).

Similarly, the degree of confidence in model predictions will vary at different levels of land-use pressure. As illustrated in Figs 3, 4, 5 and 6, at certain levels of pressure, the 95% confidence intervals are greater than the size of the response. At these levels, we cannot be confident in the model output to predict an ecological response. Conversely, we can have strong confidence that model predictions are accurate for sites that have between 0 and 100% vegetation removal, 0 and 20% impervious cover and from –1 to 1 log-nitrogen load. At a national scale, 88% of mapped stream segments have pressure values within the 95% range of model confidence.

Model validation

All of the predictor and pressure variables (apart from native vegetation removal and impervious cover) used in the analytical modelling in this study were themselves the result of extrapolations and predictive models based on climate, topography, geology and land cover. There is a clear need to validate the models in this study by measuring predictor and response variables, at the same time. Examining temporal datasets to test the precision and sensitivity of predictions over time could further refine models. In the interim, the database produced in this study provides an invaluable resource for broad-scale planning, contributing to the initial identification of areas to focus restoration, conservation or resource development.

4.4 Future research opportunities identifying land-use thresholds

The United States Environmental Protection Agency (USEPA) is actively advocating the use of a Biological Condition Gradient (BCG) as a model to allow for the assessment of stream biological condition independent of methodology (i.e. using a range of contrasting field methods and metrics) as well as a means to better communicate outcomes to the public (Davies & Jackson 2006). In New Zealand, relatively comparable methods are used in the assessment of ecological condition, hence there is limited need to develop a framework for comparing variable methods. However, several lessons can be gained from the BCG approach. Firstly, there is a universal need for communication with the public, and a categorical ranking as proposed in the BCG provides a means to achieve this. Secondly, the results from this study support the framework for assigning scores in the BCG. For example, in the BCG, 'condition' is assigned based on the status of multiple in-stream attributes such as taxonomic composition, native taxa, organism condition and ecosystem function, as well as scale-dependent factors such as connectivity (Fig. 29). The results from this study suggest that different indicators of EI (water quality, biotic endpoints and ecosystem functions) are potentially sensitive to different levels of land-use stress, after which stage a relative loss in integrity is apparent. These 'thresholds' could be used to determine ranking classifications. For example, a change in rank from 1 to 2 could signify a loss in water quality, whereas a change in rank from 3 to 4 could signify a loss in biodiversity, although this assumes that changes always occur in the same order. To explore this framework further, a formal hypothesis-testing analytical approach would be required to identify such thresholds. Recent research suggests a quantile piecewise regression approach may provide the means to do this (Brenden et al. 2008).
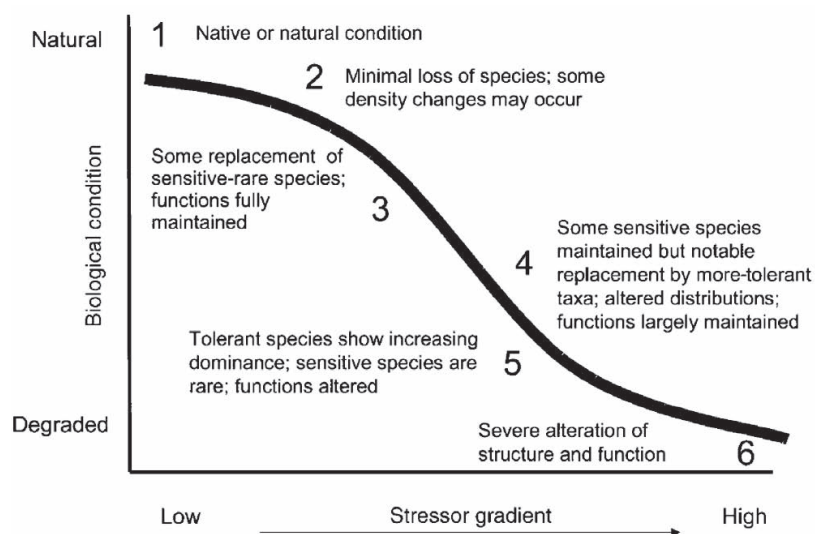

Figure 29. A conceptual model depicting stages of change in biological conditions in response to an increasing stressor gradient. From Davies & Jackson (2006).

5. Conclusions

The statistical analyses used in this study showed that metrics of water quality, invertebrates, fish and ecosystem processes can all be used to assess components of the EI of rivers. Within each metric group, at least one indicator responded significantly to land-use pressures in terms of the proportion (%) of deviance explained and the cross-validation coefficients. For all indicators, a large proportion of the deviance that was not explained by land use was explained by environmental variables. Stream classification data also improved the predictive capability of models for some indicators, but was not as informative as the inclusion of the underlying environmental variables used to construct that classification. The ability of FENZ classes to improve the predictive capability of ecological responses suggests that FENZ provides a convincing summary of environmental variables.

The BRT analyses illustrated how most indicators show a monotonic response to land-use pressures, whereas some fish and invertebrate metrics showed a non-monotonic response. All indicators illustrated potential land-use thresholds where a change in the form of the response occurred. The threshold triggers could be used to develop standards for regional planning and to identify areas where conservation efforts could be most effective at a landscape scale.

Predictive modelling of the component metrics of an MMI of EI was most informative when there were strong, monotonic relationships between the components of the MMI and vegetation removal, nitrogen loading and impervious cover. Fish metrics appeared to introduce significant noise to the MMI on a national scale. This suggests the need to further develop the fish metrics or to investigate alternative fish indices, to ensure fish are represented in a holistic measure of EI. Based on a qualitative evaluation of the MMI, the results suggest that 36% of streams have 'high' EI and less than 1% of streams have 'poor' EI. However, we suggest ordinal ratings are not as useful as assessments based on the quantitative analysis of ecological response curves and recommend further analysis of land-use thresholds.

The form of the ecological response curves between EI and land-use pressures was similar to those previously predicted by Leathwick & Julian (2007). Our results suggest some refinements based on a rapid decline in EI at less than 10% impervious cover rather than 20%, a steep decline in EI between –0.5 and 0.5 log-nitrogen concentration rather than between –1 and 2 log-nitrogen concentration, and an almost linear decline in EI in response to native vegetation removal (except for rapid declines at very high and very low values) rather than a gradual sigmoid curve with thresholds at 20% and 60% vegetation loss.

In summary, BRT modelling provides a good approach for predicting expected and observed values for many component measures of EI. Verifying reference offsets and testing predictions with temporal and spatial measurements would strengthen models. There is an apparent need to further develop fish indices because they appear to contribute noise to the MMI. The MMI metric in its current form provides a geographically meaningful prediction of the EI of rivers in New Zealand, at a broad spatial scale.

6. Acknowledgements

7. References

Acuña, V.; Giorgi, A.; Munõz, I.; Sabater, F.; Sabater, S. 2007: Meterological and riparian influences on organic matter dynamics in a forested Mediterranean stream. *Journal of the North American Benthological Society 26*: 54–69.

Adams, S.M.; Hill, W.R.; Peterson, M.J.; Ryon, M.G.; Smith, J.G.; Stewart, A.J. 2002: Assessing recovery in a stream ecosystem: applying multiple chemical and biological endpoints. *Ecological Applications 12*: 1510–1527.

Allan, J.D. 2004: Landscapes and riverscapes: the influence of land use on stream ecosystems. *Annual Review of Ecology, Evolution and Systematics 35*: 257–284.

Allan, J.D.; Erickson, D.L.; Fay, J. 1997: The influence of catchment land use on stream integrity across multiple spatial scales. *Freshwater Biology 37*: 149–161.

Anderson, C.; Cabana, G. 2005: δ^{15}N in riverine food webs: effects of N inputs from agricultural watersheds. *Canadian Journal of Fisheries and Aquatic Sciences 62*: 333–340.

Anderson, C.; Cabana, G. 2006: Does δ^{15}N in river food webs reflect the intensity and origin of N loads from the watershed? *Science of the Total Environment 367*: 968–978.

Angermeir, P.L.; Karr, J.R. 1994: Biological integrity versus biological diversity as policy directives: protecting biotic resources. *Bioscience 44*: 690–697.

Angradi, T.R.; Pearson, M.S.; Jicha, T.M.; Taylor, D.L.; Bolgrien, D.W.; Moffett, M.F.; Blocksom, K.A.; Hill, B.H. 2009: Using stressor gradients to determine reference expectations for great river fish assemblages. *Ecological Applications 9*: 748–64.

ANZECC 2000: Australian and New Zealand guidelines for fresh and marine water quality. Volume 1, The guidelines. Australian and New Zealand Environment and Conservation Council, Agriculture and Resource Management Council of Australia and New Zealand.

Ausseil, A.-G.; Chadderton, W.L.; Gerbeaux, P.; Stephens, T.; Brown, D.; Leathwick, J. 2008: Wetland ecosystems of national importance for biodiversity: criteria, methods and candidate list of nationally important inland wetlands. Landcare Research Contract Report LC0708/158. 174 p.

Barbour, M.T.; Swietlik, W.F.; Jackson, S.K.; Courtemanch, D.L.; Davies, S.P.; Yoder, C.O. 2000: Measuring the attainment of biological integrity in the USA: a critical element of ecological integrity. *Hydrobiologia 422/423*: 453–464.

Boothroyd, I.K.D.; Harding, J.S.; Death, R.G. 2002: Guide to the selection of reference sites for environmental monitoring of New Zealand rivers and streams. New Zealand Macroinvertebrate Working Group Report No. 2. Ministry for the Environment, Wellington.

Brenden, T.O.; Wang, L.; Su, Z. 2008: Quantitative identification of disturbance thresholds in support of aquatic resource management. *Environmental Management 42*: 821–832.

Bunn, S.E.; Davies, P.M. 2000: Biological processes in running waters and their implications for the assessment of ecological integrity. *Hydrobiologia 422/423*: 61–70.

Burcher, C.L.; Valett, H.M.; Benfield, E.F. 2007: The land-cover cascade: relationships coupling land and water. *Ecology 88*: 228–242.

Cao, Y.; Hawkins, C.P.; Olson, J.; Kosterman, M.A. 2007: Modelling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *Journal of the North American Benthological Society 26*: 566–585.

Carlisle, D.M.; Falcone, J.; Meador, M.R. 2009: Predicting the biological condition of streams: use of geospatial indicators of natural and anthropogenic characteristics of watersheds. *Environmental Monitoring and Assessment 151*: 143–160.

Carlisle, D.M.; Hawkins, C.P.; Meador, M.R.; Potapova, M.; Falcone, J. 2008: Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrates, and diatom assemblages. *Journal of the North American Benthological Society 27*: 16–37.

Chadderton, W.L.; Brown, D.J.; Stephens, R.T. 2004: Identifying freshwater ecosystems of national importance for biodiversity: criteria, methods, and candidate list of nationally important rivers. Discussion document, December 2004. Department of Conservation, Wellington.

Clapcott, J.E.; Barmuta, L.A. 2010: Forest clearance increases metabolism and organic matter processes in small headwater streams. *Journal of the North American Benthological Society 29*: 546–561.

Clapcott, J.; Young, R.; Goodwin, E.; Leathwick, J. 2009: Exploring the response of functional indicators of stream ecological integrity to land-use pressure gradients. *Cawthron Report No. 1611*. Prepared for the Department of Conservation. Cawthron, Nelson.

Clapcott, J.E.; Young, R.G.; Goodwin, E.O.; Leathwick, J.R. 2010: Exploring the response of functional indicators of stream health to land-use gradients. *Freshwater Biology 55*: 2181–2199.

Collier, K.J. 2008: Average score per metric: an alternative metric aggregation method for assessing wadeable stream health. *New Zealand Journal of Marine and Freshwater Research 42*: 367–378.

Davies, S.P.; Jackson, S.K. 2006: The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. *Ecological Applications 16*: 1251–1266.

Death, R.G.; Collier, K.J. 2009: Measuring stream macroinvertebrate responses to gradients of vegetation cover: when is enough enough? *Freshwater Biology 55:* 1447–1464.

Diebel, M.W.; Zanden, M.J.V. 2009: Nitrogen stable isotopes in streams: effects of agricultural sources and transformations. *Ecological Applications 19*: 1127–1134.

Doledec, S.; Phillips, N.R.; Scarsbrook, M.R.; Riley, R.H.; Townsend, C.R. 2006: Comparison of structural and functional approaches to determining land use effects on grassland stream invertebrate communities. *Journal of the North American Benthological Society 25*: 44–60.

Donohue, I.; McGarrigle, M.L.; Mills, P. 2006: Linking catchment characteristics and water chemistry with the ecological status of Irish rivers. *Water Research 40*: 91–98.

Downes, B.J.; Barmuta, L.A.; Fairweather, P.G.; Faith, D.P.; Keough, M.J.; Lake, P.S.; Mapstone, B.D.; Quinn, G.P. 2002: Monitoring ecological impacts: concepts and practice in flowing waters. Cambridge University Press, New York.

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

Fellows, C.S.; Clapcott, J.E.; Udy, J.W.; Bunn, S.E.; Harch, B.D.; Smith, M.J.; Davies, P.M. 2006: Benthic metabolism as an indicator of stream ecosystem health. *Hydrobiologia 572*: 71–87.

Flinders, C.A.; Horwitz, R.J.; Belton, T. 2008: Relationship of fish and macroinvertebrate communities in the mid-Atlantic uplands: implications for integrated assessments. *Ecological Indicators 8*: 588–598.

Fore, L.S.; Karr, J.R.; Wisseman, R.W. 1996: Assessing invertebrate responses to human activities: evaluating alternative approaches. *Journal of the North American Benthological Society 15*: 212–231.

Friedman, J.H. 2001: Greedy function approximation: a gradient boosting machine. *Annals of Statistics 29*: 1189–1232.

Gergel, S.E.; Turner, M.G.; Miller, J.R.; Melack, J.M.; Stanley, E.H. 2002: Landscape indicators of human impacts to riverine systems. *Aquatic Sciences 64*: 118–128.

Gessner, M.O.; Chauvet, E. 2002: A case for using litter breakdown to assess functional stream integrity. *Ecological Applications 12*: 498–510.

Griffith, M.B.; Hill, B.H.; McCormick, F.H.; Kaufman, P.R.; Herlihy, A.T.; Selle, A.R. 2005: Comparative application of indices of biotic integrity based on periphyton, macroinvertebrates, and fish to southern Rocky Mountain streams. *Ecological Indicators 5*: 117–136.

Gücker, B.; Brauns, M.; Pusch, M.T. 2006: Effects of wastewater treatment plant discharge on ecosystem structure and function of lowland streams. *Journal of the North American Benthological Society 25*: 313–329.

Harding, J.S.; Benfield, E.F.; Bolstad, P.V.; Helfman, G.S.; Jones, E.B.D. 1998: Stream biodiversity: the ghost of land use past. *Proceedings of the National Academy of Sciences of the United States of America 95*: 14843–14847.

Harding, J.S.; Hawke, D.J.; Holdaway, R.N.; Winterbourn, M.J. 2004: Incorporation of marine-derived nutrients from petrel breeding colonies into stream food webs. *Freshwater Biology 49*: 576–586.

Harding, J.S.; Winterbourn, M.J. 1995: Effects of contrasting land use on physico-chemical conditions and benthic assemblages of streams in a Canterbury (South Island, New Zealand) river system. *New Zealand Journal of Marine and Freshwater Research 29*: 479–492.

Harding, J.S.; Winterbourn, M.J. 1997: An ecoregion classification of the South Island, New Zealand. *Journal of Environmental Management 51*: 275–287.

Harding, J.S.; Young, R.G.; Hayes, J.W.; Shearer, K.A.; Stark, J.D. 1999: Changes in agricultural intensity and river health along a river continuum. *Freshwater Biology 42*: 345–357.

Harris, G.P. 2001: Biogeochemistry of nitrogen and phosphorus in Australian catchments, rivers and estuaries: effects of land use and flow regulation and comparisons with global patterns. *Marine and Freshwater Research 52*: 139–149.

Hastie, T.; Tibshirani, R.; Friedman, J. 2009: Elements of statistical learning: data mining, inference and prediction. Springer, New York.

Hawkins, C.P.; Norris, R.H.; Gerritsen, J.; Hughes, R.M.; Jackson, S.K.; Johnson, R.K.; Stevenson, R.J. 2000: Evaluation of the use of landscape classifications for the prediction of freshwater biota: synthesis and recommendations. *Journal of the North American Benthological Society 19*: 541–556.

Hay, J.; Hayes, J.; Young, R. 2006: Water quality guidelines to maintain trout fishery values. *Cawthron Report No. 1205*. Prepared for Horizons Regional Council. Cawthron, Nelson.

Hayes, J.W.; Leathwick, J.R.; Hanchet, S.M. 1989: Fish distribution patterns and their association with environmental factors in the Mokau River catchment, New Zealand. *New Zealand Journal of Marine and Freshwater Research 23*: 171–180.

Herlihy, A.T.; Stoddard, J.L.; Johnson, C.B. 1998: The relationship between stream chemistry and watershed land cover data in the mid-Atlantic region, U.S. *Water, Air and Soil Pollution 105*: 377–386.

Hicks, B.J.; Barrier, R.F.G. 1996: Habitat requirements of black mudfish (*Neochanna diversus*) in the Waikato region, North Island, New Zealand. *New Zealand Journal of Marine and Freshwater Research 30*: 135–150.

Hill, B.H.; Herlihy, A.T.; Kaufman, P.R.; Stevenson, R.J.; McCormick, F.H.; Johnson, C.B. 2000: Use of periphyton assemblage data as an index of biotic integrity. *Journal of the North American Benthological Society 19*: 50–67.

Hughes, R.M.; Kaufmann, P.R.; Herlihy, A.T.; Kincaid, T.M.; Reynolds, L.; Larsen, D.P. 1998: A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences 55*: 1618–1631.

Hughes, R.M.; Paulsen, S.G.; Stoddard, J.L. 2000: EMAP-Surface waters: a multiassemblage, probability survey of ecological integrity in the U.S.A. *Hydrobiologia 422/423*: 429–443.

Huryn, A.D.; Butz Huryn, V.M.; Arbuckle, C.J.; Tsomides, L. 2002: Catchment land-use, macroinvertebrates and detritus processing in headwater streams: taxonomic richness versus function. *Freshwater Biology 47*: 401–415.

Johnson, R.K.; Hering, D. 2009: Response of taxonomic groups in streams to gradients in resource and habitat characteristics. *Journal of Applied Ecology 46*: 175–186.

Joy, M.K. 2009: Temporal and land-cover trends in freshwater fish communities in New Zealand's rivers: an analysis of data from the New Zealand Freshwater Fish Database—1970–2007. Prepared for the Ministry for the Environment, Wellington.

Joy, M.K.; Death, R.G. 2004: Application of the index of biotic integrity methodology to New Zealand freshwater fish communities. *Environmental Management 34*: 415–428.

Joy, M.K.; Henderson, I.M.; Death, R.G. 2000: Diadromy and longitudinal patterns of upstream penetration of freshwater fish in Taranaki, New Zealand. *New Zealand Journal of Marine and Freshwater Research 3*: 531–543.

Karr, J.R. 1981: Assessment of biotic integrity using fish communities. *Fisheries 6*: 21–27.

Karr, J.R. 1996: Ecological integrity and ecological health are not the same. Pp. 97–109 in Schulze, P. (Ed.): Engineering within ecological constraints. National Academy Press, Washington, DC.

King, R.S.; Baker, M.E.; Whigham, D.F.; Weller, D.E.; Jordan, T.E.; Kazyak, P.F.; Hurd, M.K. 2005: Spatial considerations for linking watershed land cover to ecological indicators in streams. *Ecological Applications 15*: 137–153.

Lake, J.; McKinney, R.A.; Osterman, F.A.; Pruell, R.J.; Kiddon, J.; Libby, A.D. 2002: Stable isotopes as indicators of anthropogenic activities in small freshwater systems. *Canadian Journal of Fisheries and Aquatic Sciences 58*: 870–878.

Lake, P.S. 2000: Disturbance, patchiness, and diversity in streams. *Journal of the North American Benthological Society 19*: 573–592.

Larned, S.T.; Scarsbrook, M.; Snelder, T.; Norton, N.J. 2003: Nationwide and regional state trends in river water quality 1996–2002. NIWA Client Report CHC2003-051. NIWA, Christchurch. 112 p.

Leathwick, J.; Julian, K. 2007: Production of pressure estimates for New Zealand river catchments. NIWA Client Report HAM2007-027. Prepared for the Department of Conservation. NIWA, Hamilton.

Leathwick, J.R.; Elith, J.; Chadderton, W.L.; Rowe, D.; Hastie, T. 2008a: Dispersal, disturbance and contrasting biogeographies of New Zealand's diadromous and non-diadromous fish species. *Journal of Biogeography 35*: 1481–1497.

Leathwick, J.R.; Julian, K.; Elith, J.; Chadderton, L.; Ferrier, S.; Snelder, T.H. 2008b: A biologically-optimised environmental classification of New Zealand rivers and streams: reanalysis excluding human impact variables. NIWA Client Report HAM2008-027. Prepared for the Department of Conservation. NIWA, Hamilton.

Lenat, D.R.; Crawford, J.K. 1994: Effects of land use on water quality and aquatic biota of three North Carolina Piedmont streams. *Hydrobiologia 294*: 185–199.

Marchant, R.; Norris, R.H.; Milligan, A. 2006: Evaluation and application of methods for biological assessment of streams: summary of papers. *Hydrobiologia 572*: 1–7.

Maxted, J.R.; Barbour, M.T.; Gerritsen, J.; Poretti, V.; Primrose, N.; Silvia, A.; Penrose, D.; Renfrow, R. 2000: Assessment framework for mid-Atlantic coastal plain streams using benthic macroinvertebrates. *Journal of the North American Benthological Society 19*: 128–44.

Maxted, J.R.; Evans, B.F.; Scarsbrook, M.R. 2003: Development of standard protocols for macroinvertebrate assessment of soft-bottomed streams in New Zealand. *New Zealand Journal of Marine and Freshwater Research 37*: 793–807.

McDowall, R.M. 1993: Implications of diadromy for the structuring and modelling of riverine fish communities in New Zealand. *New Zealand Journal of Marine and Freshwater Research 27*: 453–462.

McTammany, M.E.; Benfield, E.F.; Webster, J.R. 2007: Recovery of stream ecosystem metabolism from historical agriculture. *Journal of the North American Benthological Society 26*: 532–545.

Minns, C.K. 1990: Patterns of distribution and association of freshwater fish in New Zealand. *New Zealand Journal of Marine and Freshwater Research 24*: 31–44.

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

Niyogi, D.K.; Simon, K.; Townsend, C.R. 2003: Breakdown of tussock grass in streams along a gradient of agricultural development in New Zealand. *Freshwater Biology 48*: 1698–1708.

Norris, R.H.; Linke, S.; Prosser, I.; Young, W.J.; Liston, P.; Bauer, N.; Sloane, N.; Dyer, F.; Thoms, M. 2007: Very-broad-scale assessment of human impacts on river condition. *Freshwater Biology 52*: 959–976.

O'Neill, R.V.; Hunsaker, C.T.; Jones, K.B.; Riitters, K.H.; Wickham, J.D.; Schwartz, P.M.; Goodman, I.A.; Jackson, I.A.; Baillargeon, W.S. 1997: Monitoring environmental quality at the landscape scale. *Bioscience 47*: 513–519.

Paulsen, S.G.; Mayio, A.; Peck, D.V.; Stoddard, J.L.; Tarquinio, E.; Holdsworth, S.M.; Van Sickle, J.; Yuan, L.L.; Hawkins, C.P.; Herlihy, A.T.; Kaufman, P.R.; Barbour, M.T.; Larsen, D.P.; Olsen, A.R. 2008: Condition of stream ecosystems in the US: an overview of the first national assessment. *Journal of the North American Benthological Society 27*: 812–821.

Plafkin, J.L.; Barbour, M.T.; Porter, K.D.; Gross, S.K.; Hughes, R.M. 1989: Rapid bioassessment protocols for use in streams and rivers: benthic macroinvertebrates and fish. U.S. Environmental Protection Agency, Office of Water Regulations and Standards, Washington, DC. EPA 440-4-89-001.

Poff, N.L. 1997: Landscape filters and species traits: towards a mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society 16*: 391–409.

Pont, D.; Hugueny, B.; Beier, U.; Goffaux, D.; Melcher, A.; Noble, R.; Rogers, C.; Roset, N.; Schmutz, S. 2006: Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. *Journal of Applied Ecology 43*: 70–80.

Quinn, J.M. 2000: Effects of pastoral development. Pp. 208–229 in Collier, K.J.; Winterbourn, M.J. (Eds): New Zealand stream invertebrates: ecology and implications for management. New Zealand Limnological Society, Christchurch.

Quinn, J.M.; Cooper, A.B.; Davies-Colley, R.J.; Rutherford, J.C.; Williamson, R.B. 1997: Land use effects on habitat, water quality, periphyton, and benthic invertebrates in Waikato, New Zealand, hill-country streams *New Zealand Journal of Marine and Freshwater Research 31*: 579–597.

Quinn, J.M.; Stroud, M.J. 2002: Water quality and sediment and nutrient export from New Zealand hill-land catchments of contrasting land use. *New Zealand Journal of Marine and Freshwater Research 36*: 409–429.

Raven, P.J.; Holmes, N.T.H.; Dawson, F.H.; Fox, P.J.A.; Everand, M.; Fozzard, I.R.; Rowen, K.J. 1998: River habitat quality: the physical character of rivers and streams in the UK and Isle of Man. Environment Agency, Bristol, England.

Resh, V.H. 2008: Which group is best? Attributes of different biological assemblages used in freshwater biomonitoring programs. *Environmental Monitoring and Assessment 138*: 131–138.

Resh, V.H.; Brown, A.V.; Covich, A.P.; Gurtz, M.E.; Li, H.W.; Minshall, G.W.; Reice, S.R.; Sheldon, A.L.; Wallace, J.B.; Wissmar, R.C. 1988: The role of disturbance in stream ecology. *Journal of the North American Benthological Society 7*: 433–455.

Richardson, J. 1989: The all-new freshwater fish database. *Freshwater Catch 41*: 20–21.

Ridgeway, G. 2006: Generalised boosted regression models. Documentation on the R Package 'gbm' version 1.5-7. www.i-pensieri.com/gregr/gbm.shtml.

Roberts, B.J.; Mulholland, P.J.; Hill, W.R. 2007: Multiple scales of temporal variability in ecosystem metabolism rates: results from 2 years of continuous monitoring in a forested headwater stream. *Ecosystems 10*: 588–606.

Roth, N.E.; Allan, J.D.; Erickson, D.L. 1996: Landscape influences on stream biotic integrity assessed at multiple spatial scales. *Landscape Ecology 11*: 141–156.

Scarsbrook, M. 2008: Quantifying relationships between human pressures and ecological integrity in rivers: pilot analyses. NIWA Client Report HAM2008-114. NIWA, Hamilton. 51 p.

Schallenberg, M.; Kelly, D.; Clapcott, J.; Death, R.; MacNeil, C.; Young, B.; Sorrell, B.; Scarsbrook, M. 2011: Approaches to assessing ecological integrity of New Zealand freshwaters. *Science for Conservation 307.* Department of Conservation, Wellington. 84 p.

Sheeder, S.A.; Evans, B.M. 2004: Estimating nutrient and sediment threshold criteria for biological impairment in Pennsylvania watersheds. *Journal of the American Water Resources Association 40*: 881–888.

Smith, M.J.; Storey, A.W. (Ed.) 2001: Design and implementation of baseline monitoring (DIBM3): developing an Ecosystem Health Monitoring Program for rivers and streams in southeast Queensland. Report to the South East Queensland Regional Water Quality Management Strategy, Brisbane.

Snelder, T.; Biggs, B.; Weatherhead, M. 2004: New Zealand River Environment Classification user guide. Ministry for the Environment, Wellington.

Snyder, C.D.; Young, J.A.; Villella, R.; Lemarié, D.P. 2003: Influences of upland and riparian land use patterns on stream biotic integrity. *Landscape Ecology 18*: 647–664.

Soranno, P.A.; Cheruvelil, K.S.; Stevenson, R.J.; Rollins, S.L.; Holden, S.W.; Heaton, S.; Torng, E. 2008: A framework for developing ecosystem-specific nutrient criteria: Integrating biological thresholds with predictive modelling. *Limnology and Oceanography 53*: 773–787.

Stark, J.D. 1998: SQMCI: a biotic index for freshwater macroinvertebrate coded abundance data. *New Zealand Journal of Marine and Freshwater Research 32*: 55–66.

Stark, J.D.; Boothroyd, I.K.G.; Harding, J.S.; Maxted, J.R.; Scarsbrook, M.R. 2001: Protocols for sampling macroinvertebrates in wadeable streams. *New Zealand Working Group Report No.1.* Prepared for the Ministry for the Environment, Wellington.

Stoddard, J.L.; Herlihy, A.T.; Peck, D.V.; Hughes, R.M.; Whittier, T.R.; Tarquinio, E. 2008: A process for creating multimetric indices for large-scale aquatic surveys. *Journal of the North American Benthological Society 27*: 878–891.

Strayer, D.L.; Beighley, R.E.; Thompson, L.G.; Brooks, D.; Nilsson, C.; Pinay, G.; Naiman, R.J. 2003: Effects of land cover on stream ecosystems: roles of empirical models and scaling issues. *Ecosystems 6*: 407–423.

Stribling, J.B.; Jessup, B.K.; Feldman, D.L. 2008: Precision of benthic macroinvertebrate indicators fo stream condition in Montana. *Journal of the North American Benthological Society 27*: 58–67.

Suding, K.N.; Hobbs, R.J. 2009: Threshold models in restoration and conservation: a developing framework. *Trends in Ecology and Evolution 24*: 271–279.

Townsend, C.R.; Arbuckle, C.J.; Crowl, T.A.; Scarsbrook, M.R. 1997a: The relationship between land-use and physiochemistry, food resources and macroinvertebrate communities in tributaries of the Taieri River, New Zealand: a hierarchically-scaled approach. *Freshwater Biology 37*: 177–191.

Townsend, C.R.; Scarsbrook, M.R.; Doledec, S. 1997b: The intermediate disturbance hypothesis, refugia, and biodiversity in streams. *Limnology and Oceanography 42*: 938–949.

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

Uehlinger, U. 2000: Resistance and resilience of metabolism in a flood prone river system. *Freshwater Biology 45*: 319–332.

Van Sickle, J.; Baker, J.; Herlihy, A.; Bayley, P.; Gregory, S.; Haggerty, P.; Ashkenas, L.; Li, J. 2004: Projecting the biological condition of streams under alternative scenarios of human land use. *Ecological Applications 14*: 368–380.

Van Sickle, J.; Johnson, C.B. 2008: Parametric distance weighting of landscape influence on streams. *Landscape Ecology 23*: 427–438.

Vlek, H.E.; Verdonschot, P.F.M.; Nijboer, R.C. 2004: Towards a multimetric index for the assessment of Dutch streams using benthic invertebrates. *Hydrobiologia 516*: 173–189.

Wang, L.; Lyons, J.; Kanehl, P. 2001: Impacts of urbanisation on stream habitat and fish across multiple spatial scales. *Environmental Management 28*: 255–266.

Weigel, B.M.; Wang, L.; Rasmussen, P.W.; Butcher, J.T.; Stewart, P.M.; Simon, T.P.; Wiley, M.J. 2003: Relative influence of variables at multiple spatial scales on stream macroinvertebrates in the Northern Lakes and Forest ecoregion, U.S.A. *Freshwater Biology 48*: 1440–1461.

Weijters, M.J.; Janse, J.H.; Alkemade, R.; Verhoeven, J.T.A. 2009: Quantifying the effect of catchment land use and water nutrient concentrations on freshwater river and stream biodiversity. *Aquatic Conservation: Marine and Freshwater Ecosystems 19*: 104–112.

Whittier, T.R.; Hughes, R.M.; Stoddard, J.L.; Lomnicky, G.A.; Peck, D.V.; Herlihy, A.T. 2007: A structured approach for developing indices of biotic integrity: three examples from streams and rivers in the western USA. *Transactions of the American Fisheries Society 136*: 718–735.

Young, R.G.; Clapcott, J.E. 2010: Temporal variability in ecosystem metabolism of rivers in the Manawatu-Whanganui Region—updated. *Cawthron Report No. 1791*. Prepared for Horizons Regional Council. Cawthron, Nelson.

Young, R.G.; Collier, K.J. 2009: Contrasting responses to catchment modification among a range of functional and structural indicators of river ecosystem health. *Freshwater Biology 54*: 2155–2170.

Young, R.G.; Huryn, A.D. 1999: Effects of land use on stream metabolism and organic matter turnover. *Ecological Applications 9*: 1359–1376.

Young, R.G.; Matthaei, C.D.; Townsend, C.R. 2008: Organic matter breakdown and ecosystem metabolism: functional indicators for assessing river ecosystem health. *Journal of the North American Benthological Society 27*: 605–625.

Appendix 1

Boosted regression tree model parameters

Model A   Predictors = *VegL*, *LogN*, *IC*

INDICATOR*	N	SCALE	MEAN	LEARNING RATE	TREE COMPLEXITY	FOLDS	STEP SIZE	TREES	MODEL ERROR
NOx (ppm)	360	×0	7.583	0.005	3	5	50	1050	Poisson
DRP (ppm)	578	×100	2.048	0.0025	3	5	50	1150	Poisson
Clarity (m)	525	×10	20.808	0.0025	3	5	50	1200	Poisson
*MC*I	2666	na	115.810	0.005	3	10	50	1150	Gaussian
ITR	2626	na	16.998	0.02	3	10	50	1050	Gaussian
EPT (%)	2626	na	50.506	0.015	3	10	50	1450	Gaussian
Cycle (%)	2626	na	82.516	0.005	3	10	50	1200	Gaussian
F-IBI	6858	na	34.259	0.025	3	10	50	1000	Gaussian
FTR	6858	×5	13.117	0.02	3	10	50	1000	Poisson
Native (%)	6858	arc sin	66.109	0.015	3	10	50	1750	Gaussian
GPP ($gO_2\,m^{-2}\,d^{-1}$)	156	na	4.603	0.002	3	5	50	500	Poisson
ER ($gO_2\,m^{-2}\,d^{-1}$)	156	na	9.936	0.001	3	5	50	1400	Poisson
Del15N (‰)	73	×3	20.274	0.02	3	5	50	900	Poisson
Cotton (*k*/dd)	108	×5000	19.074	0.0025	3	5	50	1200	Poisson

*   See Table 2 for ecological indicator names in full.

Model B   Predictors = Model A output as a fixed offset plus *SegHabitat*, *SegSubstrate*, *SegLowFlow*, *SegFlowStability*, *SegSumT*, *SegTSeas*, *SegShade*, *SegSlope*, *USRainDays*, *USAvgT*, *USSlope*, *USCalcium*, *USHardness*, *USPhosphorus*, *USPeat*, *USLake*, *USGlacier*, *DSDist2Coast*, *DSDist2Lake*, *DSAvgSlope*, *DSMaxSlope*

INDICATOR*	N	SCALE	MEAN	LEARNING RATE	TREE COMPLEXITY	FOLDS	STEP SIZE	TREES	MODEL ERROR
NOx (ppm)	360	×10	7.583	0.0015	7	5	50	1150	Poisson
DRP (ppm)	578	×100	2.048	0.005	7	5	50	1450	Poisson
Clarity (m)	525	×10	20.808	0.005	7	5	50	1150	Poisson
MCI	2666	na	115.810	0.005	7	10	50	1850	Gaussian
ITR	2626	na	16.998	0.01	7	10	50	1250	Gaussian
EPT (%)	2626	na	50.505	0.005	7	10	50	950	Gaussian
Cycle (%)	2626	na	82.516	0.005	7	10	50	1550	Gaussian
F-IBI	6858	na	34.259	0.025	7	10	50	2100	Gaussian
FTR	6858	×5	13.117	0.05	7	10	50	1200	Poisson
Native (%)	6858	arc sin	66.109	0.05	7	10	50	1150	Gaussian
GPP ($gO_2\,m^{-2}\,d^{-1}$)	156	na	4.603	0.001	7	5	50	1050	Poisson
ER ($gO_2\,m^{-2}\,d^{-1}$)	156	na	9.936	0.0005	7	5	50	1000	Poisson
Del15N (‰)	73	×3	20.274	0.00025	7	5	50	1900	Poisson
Cotton (*k*/dd)	108	×5000	19.074	0.0005	7	5	50	1500	Poisson

*   See Table 2 for ecological indicator names in full.

Model C   Predictors = *VegL*, *LogN*, *IC* plus *Group*

INDICATOR*	N	SCALE	MEAN	LEARNING RATE	TREE COMPLEXITY	FOLDS	STEP SIZE	TREES	MODEL ERROR
NOx (ppm)	360	×100	7.583	0.005	3	7	50	1300	Poisson
DRP (ppm)	578	×250	5.157	0.0025	3	7	50	1450	Poisson
Clarity (m)	525	×200	41.608	0.003	3	7	50	1300	Poisson
MCI	2666	na	115.810	0.005	3	10	50	1200	Gaussian
ITR	2626	na	16.998	0.01	3	10	50	1000	Gaussian
EPT (%)	2626	na	50.505	0.0075	3	10	50	1050	Gaussian
Cycle (%)	2626	na	82.516	0.005	3	10	50	950	Gaussian
F-IBI	6860	na	34.255	0.025	3	10	50	1400	Gaussian
Native (%)	6860	arc sin	66.097	0.015	3	10	50	1350	Gaussian
FTR	6860	×5	13.117	0.025	3	10	50	1250	Poisson
GPP ($gO_2 m^{-2} d^{-1}$)	156	na	4.603	0.001	5	5	50	500	Poisson
ER ($gO_2 m^{-2} d^{-1}$)	156	na	9.936	0.001	5	5	50	950	Poisson
Del15N (‰)	73	×3	20.274	0.005	5	5	50	900	Poisson
Cotton (*k*/dd)	108	×5000	19.074	0.002	5	5	50	850	Poisson

* See Table 2 for ecological indicator names in full.

Appendix 2

Relative contributions of environmental predictor variables

Table A2.1. Relative contributions (%) of environmental predictor variables for the boosted regression models (Model B) that increase the deviance explained in the indicator data for ecological integrity after the influence of three human land-use pressures has been fixed. See Table 2 for ecological indicator names in full. '--' = not included in the model.

	FISH				INVERTEBRATES			WATER QUALITY			PROCESSES			
	F-IBI	FTR	NATIVE	CYCLE	EPT	ITR	MCI	CLARITY	DRP	NOX	GPP	ER	DEL15N	COTTON
N	6858	6858	6858	2626	2626	2626	2666	525	578	360	156	156	73	108
% deviance increase*	41.89	34.90	37.50	10.76	17.78	30.77	12.90	24.87	38.32	18.59	25.67	23.53	37.95	41.70
SegHabitat	1.02	2.32	1.19	3.29	1.51	1.72	3.48	7.14	3.19	1.50	0.58	1.37	1.54	6.38
SegSubstrate	2.63	4.09	2.53	5.67	3.94	5.37	4.16	1.97	10.09	2.52	1.59	4.26	3.92	6.43
SegLowFlow	1.10	2.48	1.84	3.36	1.78	1.81	3.01	9.83	2.98	1.47	47.93	13.09	0.31	5.07
SegFlowStability	3.98	5.20	6.67	8.73	11.46	12.34	6.66	6.72	21.69	5.01	3.92	12.89	3.66	3.36
SegSumT	7.85	4.84	8.87	8.75	14.42	8.36	15.37	4.18	5.68	3.05	0.50	5.24	1.65	18.33
SegTSeas	6.16	4.29	6.90	7.33	9.90	9.32	8.64	9.66	5.68	25.57	1.48	3.95	31.00	10.85
SegShade	3.38	3.89	3.85	5.88	3.31	5.03	5.64	3.69	2.99	1.69	8.55	6.72	7.78	12.00
SegSlope	3.31	4.29	4.30	10.36	6.92	5.79	6.81	3.17	9.21	4.97	2.20	6.79	31.29	2.25
USRainDays	15.68	7.15	7.57	11.50	11.62	12.72	10.20	8.54	3.15	8.72	2.35	3.25	4.62	3.12
USAvgT	3.17	4.37	5.47	6.61	5.38	6.13	4.47	9.84	5.63	15.38	1.72	8.39	2.61	8.18
USSlope	5.33	7.31	5.20	6.46	4.81	5.11	6.27	5.34	5.37	15.38	2.00	3.97	1.43	3.70
USCalcium	3.81	4.42	3.45	4.20	5.06	5.21	4.10	8.72	2.52	2.91	6.45	3.20	2.68	3.09
USHardness	3.25	3.55	2.61	6.07	8.24	4.63	6.04	8.00	5.32	4.14	14.12	22.41	6.69	15.03
USPhosphorus	3.17	3.99	3.23	5.99	7.26	13.64	6.49	8.86	11.99	7.43	3.22	4.20	0.82	2.21
USPeat	0.20	0.64	0.51	2.21	3.13	0.69	2.43	0.52	4.18	0.25	3.39	0.27	0.00	0.00
USLake	0.21	0.43	0.39	3.59	1.10	0.36	6.20	3.48	0.33	<0.01	0.00	0.00	0.00	0.00
USGlacier	0.41	0.14	0.17	0	0.15	1.78	0.04	0.33	0.00	0.00	0.00	0.00	0.00	0.00
DSDist2Coast	22.42	11.15	18.32	--	--	--	--	--	--	--	--	--	--	--
DSDist2Lake	2.15	0.86	4.10	--	--	--	--	--	--	--	--	--	--	--
DSAvgSlope	5.38	5.46	6.29	--	--	--	--	--	--	--	--	--	--	--
DSMaxSlope	5.41	19.11	6.56	--	--	--	--	--	--	--	--	--	--	--

* Model B minus Model A.