

A multimetric approach for predicting the ecological integrity of New Zealand streams

J.E. Clapcott^{(1),*}, E.O. Goodwin⁽¹⁾, R.G. Young⁽¹⁾, D.J. Kelly^{(1),(2)}

Received June 25, 2014

Revised August 18, 2014

Accepted August 19, 2014

ABSTRACT

Key-words:
multimetric index, ecological integrity, stream health, boosted regression trees, New Zealand

Integrating multiple measures of stream health into a combined metric can provide a holistic assessment of the ecological integrity of a stream. The aim of this study was to develop a multimetric index (MMI) of stream integrity based on predictive modelling of national data sets of water quality, macroinvertebrates, fish and ecosystem process metrics. We used a boosted regression tree approach to calculate an observed/expected score for each metric prior to combining metrics in a MMI based on data availability and the strength of predictive models. The resulting MMI provides a geographically meaningful prediction of the ecological integrity of rivers in New Zealand, but identifies limitations in data and approach, providing focus for ongoing research.

RÉSUMÉ

Une approche multimétrique pour prédire l'intégrité écologique des cours d'eau en Nouvelle-Zélande

Mots-clés :
indice multimétrique, intégrité écologique, santé des cours d'eau, arbres de décision stimulés, Nouvelle-Zélande

L'intégration de plusieurs mesures de la santé des cours d'eau dans une métrique combinée peut fournir une évaluation globale de l'intégrité écologique d'un cours d'eau. Le but de cette étude était de développer un indice multimétrique (MMI) de l'intégrité des rivières basé sur la modélisation prédictive des ensembles de données nationales de métrologie de la qualité de l'eau, des macro-invertébrés, des poissons et des processus de l'écosystème. Nous avons utilisé une approche par l'arbre de décision stimulé pour calculer un score observé / prévu pour chaque métrique avant de combiner les paramètres dans une MMI basée sur la disponibilité des données et la force de modèles prédictifs. Le MMI résultant fournit une prédiction géographique significative de l'intégrité écologique des rivières en Nouvelle-Zélande, mais identifie les limites des données et de l'approche, et donne une orientation à la recherche en cours.

INTRODUCTION

Conservation planning at a national scale requires knowledge of the ecological integrity (EI) of streams. Ecological integrity is the degree to which the physical, chemical and biological components (including composition, structure and process) of a stream and their relationships are present, functioning and maintained close to a reference condition reflecting negligible or minimal anthropogenic impacts (Schallenberg et al., 2011). A single indicator

(1) Cawthron Institute, Private Bag 2, Nelson 7042, New Zealand

(2) Department of Conservation, Private Bag 4715, Christchurch 8140, New Zealand

* Corresponding author: joanne.clapcott@cawthron.org.nz

which combines multiple stream health metrics, or a multimetric index (MMI), can provide a measure of EI. A single indicator provides a transparent, objective and reproducible assessment of environmental health (USEPA, 2008). Although it has been argued that information is lost rather than gained by multimetric application (Baker and King, 2010), we suggest that it is the way in which information is combined that becomes crucial in MMI development.

Karr (1981) was the first to suggest the combination of fish metrics to create an index of biological integrity (IBI). Subsequently single groups of organisms have been the focus of a MMI approach. Multi-metric indexes have been developed using fish data (Joy and Death, 2004; Pont *et al.*, 2006; Stoddard *et al.*, 2008), macroinvertebrates (Collier, 2008; Maxted *et al.*, 2000; Vlek *et al.*, 2004), diatoms (Cao *et al.*, 2007) and periphyton (Hill *et al.*, 2000). However, recent comparative studies of multiple indicators have shown how different groups of organisms provide complementary information of ecological condition. For example, in a parallel investigation of fish, macroinvertebrates and diatom assemblages, Carlisle *et al.* (2008) showed how a single group evaluation indicated impaired conditions on average much less often than when several groups were used. Similarly, a recent New Zealand study illustrated how different water quality, macroinvertebrate and fish indicators varied in their responses to land-use stressors (Clapcott *et al.*, 2012). Comparable studies suggest that the assessment of multiple groups of organisms has the potential to provide a more robust evaluation of ecological integrity than the assessment of a single group (Brown *et al.*, 2009; Waite, 2014).

Our goal was to combine multiple indicators of stream health into a MMI to provide a more holistic measure of EI than the use of individual indicators. To address this goal we: modelled reference condition for each candidate metric; calculated an observed/expected (O/E) score for each candidate metric; determined the weighting functions of candidate scores for aggregation in the MMI, and characterised the ecological response curve of an MMI to disturbance gradients.

MATERIALS AND METHODS

We analysed metrics calculated using a range of regional and national datasets as described in Clapcott *et al.* (2012). Metrics included water quality metrics (water visual clarity – Clarity, nitrate-N and nitrite-N concentration – NO_x), macroinvertebrate metrics (Macroinvertebrate Community Index – MCI (average sensitivity score per taxon (Stark, 1985)), number of species producing once in a life cycle – Cycle), fish metrics (index of biological integrity – F-IBI (Joy and Death, 2004), percent of native fish species – Native) and ecosystem process metrics (ecosystem respiration – ER, gross primary production – GPP, cellulose decomposition potential – Cotton, $\delta^{15}\text{N}$ of primary consumers – D15N). The quantity and spatial distribution of datasets varied substantially for each group of data (Figure 1, Table I). The inclusion of candidate metrics in our analysis was determined to meet a balance between conceptual inclusiveness by measuring EI components of pristineness, diversity, nativeness and resilience (Schallenberg *et al.*, 2011) and management focussed indicators (*i.e.* measures that have been and can be widely adopted and communicated).

We used a boosted regression tree (BRT) approach (Elith *et al.*, 2008), to model the response of candidate metrics to land use and environmental gradients at a national scale as described in Clapcott *et al.* (2011). A two-step model was developed where in the first step we examined the metric response to three land-use gradients: percentage of native vegetation cover remaining (VegR), percentage of impervious cover (IC), and the log-transformed predicted stream N concentration (LogN; based on land-use intensity whereby estimated nitrogen load is standardised by stream flow (Woods *et al.*, 2006)). Fitted values from the first step were then used as a fixed offset in a second step where we examined the degree to which remaining variation could be explained by the addition of 21 environmental descriptors. Additional environmental descriptors included measures of upstream and segment-scale climatic, topographical, geological and morphological characteristics (<http://www.doc.govt.nz/conservation/land-and-freshwater/freshwater/freshwater-ecosystems-of-new-zealand/>). The output from

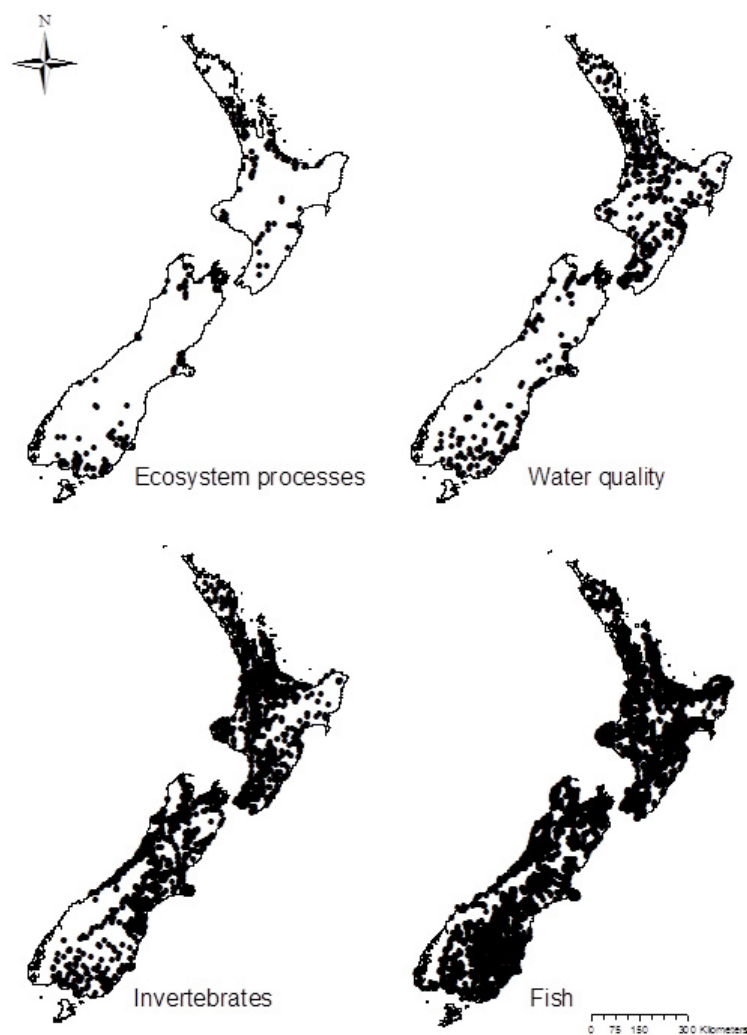


Figure 1

Map of New Zealand showing location of sites for water quality, fish, macroinvertebrate and ecosystem process sample data.

the resultant BRT model was used to predict the “observed” contemporary value for all stream segments in New Zealand. Model validation showed that this two-step approach had improved predictive accuracy compared to a simple “all-in” model, we suspect due to the fixed partitioning of variance to land-use effects.

To predict “expected” reference values, first we calculated the average metric value at sites with >95% native vegetation cover and 0% urban cover (N ranged from 5–40 and represented approximately 10% of our training data with good geographical representation, Clapcott *et al.*, 2011). We then used the fitted functions from the second step to estimate how these reference values will vary spatially on a national scale due to environmental variability in the absence of land-use effects. We note that when using BRT, predictions are constrained to the training data range, unlike random forests which allow extrapolation beyond (Breiman, 2001); as such reference predictions are likely to reflect best attainable condition rather than historical condition (*sensu* Stoddard *et al.*, 2006).

We divided observed values by expected values to produce an O/E score; for metrics that increased in response to increasing land use the inverse of the O/E score was calculated (e.g. $1 / (O/E)$). O/E scores were not constrained by 0 and 1 when a non-Gaussian model was used (Table I; see Clapcott *et al.*, 2012 for more details). O/E scores were weighted prior to aggregation based on two factors: power of the initial BRT model (average of the

Table 1 Mean and range in modelled-expected and modelled-observed values for 10 candidate metrics of ecological integrity compared to sample data, including boosted regression tree (BRT) model diagnostics used in weighting metrics during MMI development (from Clapcott et al., 2012). See Materials and Methods for a description of metrics. Predicted reference = the low land-use value used in the calculation of modelled-expected values.

Metric	No. sites in sample of dataset	Proportion of stream types* represented by sample data	BRT % deviance explained	BRT cross validation coefficient	Sample data mean (min, max)	Sample reference value	Modelled-expected mean (min, max)	Modelled-observed mean (min, max)	O/E scores mean (min, max)
NOx (mg L⁻¹)	360	0.74	56.96	68.4	060 (0.01, 8.45)	0.07	0.18 (0.09, 0.32)	0.66 (0.11, 8.56)	0.64 (0.03, 1.02)
Clarity (m)	525	0.78	29.94	53.9	2.08 (0.10, 10.43)	4.8	4.62 (2.06, 11.8)	2.85 (0.55, 12.5)	0.58 (0.24, 1.16)
MCI	2666	0.92	35.6	60.1	115.9 (20, 180)	138.4	136.6 (97.5, 161.9)	117.9 (55.5, 162.1)	0.86 (0.48, 1.00)
Cycle (%)	2626	0.92	31.64	56.7	82.5 (0, 100)	93.5	92.5 (62.5, 100)	83.05 (16.8, 100)	0.90 (0.24, 1.00)
FIBI	6858	0.95	11.32	33.8	34.3 (0, 60)	37.8	35.5 (0, 60)	29.8 (0, 60)	0.81 (<0.1, >10)
Native (%)	6858	0.95	10.03	31.8	75.3 (0, 100)	83.8	92 (0, 100)	79 (0, 100)	16.40 (<0.1, >10)
ER (g O m⁻² d⁻¹)	156	0.65	10.27	28.2	9.94 (0.34, 55.78)	6.50	6.26 (5.07, 8.28)	9.10 (4.59, 24.83)	0.92 (0.32, 1.11)
GPP (g O m⁻² d⁻¹)	156	0.65	0.25	18.1	4.60 (0.08, 35.99)	3.04	2.27 (1.55, 5.13)	2.66 (1.2, 39.49)	0.97 (0.11, 1.37)
D15N (‰)	73	0.31	48.19	72.9	6.75 (1.03, 20.99)	5.22	5.18 (5.09, 5.31)	5.62 (2.25, 19.0)	1.04 (0.28, 2.27)
Cotton (kddd⁻¹) × 100	108	0.55	12.91	38.4	0.38 (0.02, 2.58)	0.16	0.17 (0.12, 0.20)	0.27 (0.11, 1.53)	0.74 (0.13, 1.10)

* as defined by Freshwater Environments of New Zealand 100-level classification (Leathwick et al., 2011)

ability to explain residual deviance and potential predictive accuracy) and the spatial representation of the training data. For example, the FIBIOE score was multiplied by the average of the deviance explained (0.113, Table 1) and cross validation coefficient (0.338) from the BRT model, and then multiplied by national representation of stream types (0.95), so that FIBIOE contribution = $FIBIOE \times (0.113 + 0.338)/2 \times 0.95 = 0.214$. Ten weighted O/E scores were aggregated first into hierarchical groups of water quality, invertebrate, fish and process components. Then the averages of components were aggregated into MMI scores which were scaled to a range between 0 and 1 by multiplying by the theoretical maximum sum of weighting factors (*i.e.* 2.56). When databases increase in size, the appropriateness of these weightings could be reassessed. Similarly, new indices could be used interchangeably in the creation of a MMI score or together with current indices to strengthen the inference of change.

RESULTS

There was a wide range in modelled-expected values, modelled-observed values and O/E scores for component metrics (Table 1). The ecological response curves for the O/E scores of most component metrics suggested a general decrease in stream condition in response to an increase in land use pressure (Figure 2).

Values for the MMI ranged from 0.18 to 1.99, with values beyond 1 due to the contribution from non-linear models of fish and some ecosystem process metrics. The geographical distribution of the summary metric for ecological integrity showed meaningful patterns across New Zealand (Figure 3). There were high values in the Conservation estate and low values in areas of intense urban and agricultural development.

The shape of relationships between predicted MMI scores and land-use gradients were similar to existing conceptual ecological response curves for pressure estimates (*i.e.* Leathwick and Julian, 2007). The BRT model predicted a rapid decrease in EI before 5% and after 95% vegetation removal; otherwise a generally linear decline in EI is apparent in response to native vegetation removal (Figure 4). A rapid decline in EI is predicted as impervious cover increases to 5%, followed by a continued but slow decline (Figure 4). The predicted MMI response to nitrogen concentration had a distinct non-monotonic shape (Figure 4).

DISCUSSION

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

The form of the MMI response to the three land-use gradients appears most similar to the metrics with greatest influence, *i.e.* water quality and invertebrates. However, the influence of other metrics is also apparent, especially fish. It appears that weaker models such as those for fish introduce noise into the MMI predictions, both in the form of the response but also in geographical patterns. This noise could be addressed by further down weighting the contribution of fish to the MMI until more robust models are developed (*e.g.* through reanalysis with new pressures or new fish metrics). Alternatively, metrics with weak models 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 to improve 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. O/E scores for the MMI represent ecological integrity; therefore close to 1 represent high EI. Based on a qualitative evaluation of the MMI, results suggest 36% of stream segments in New Zealand have high EI (>0.9); roughly equivalent to the proportion of land in the Conservation estate and hence

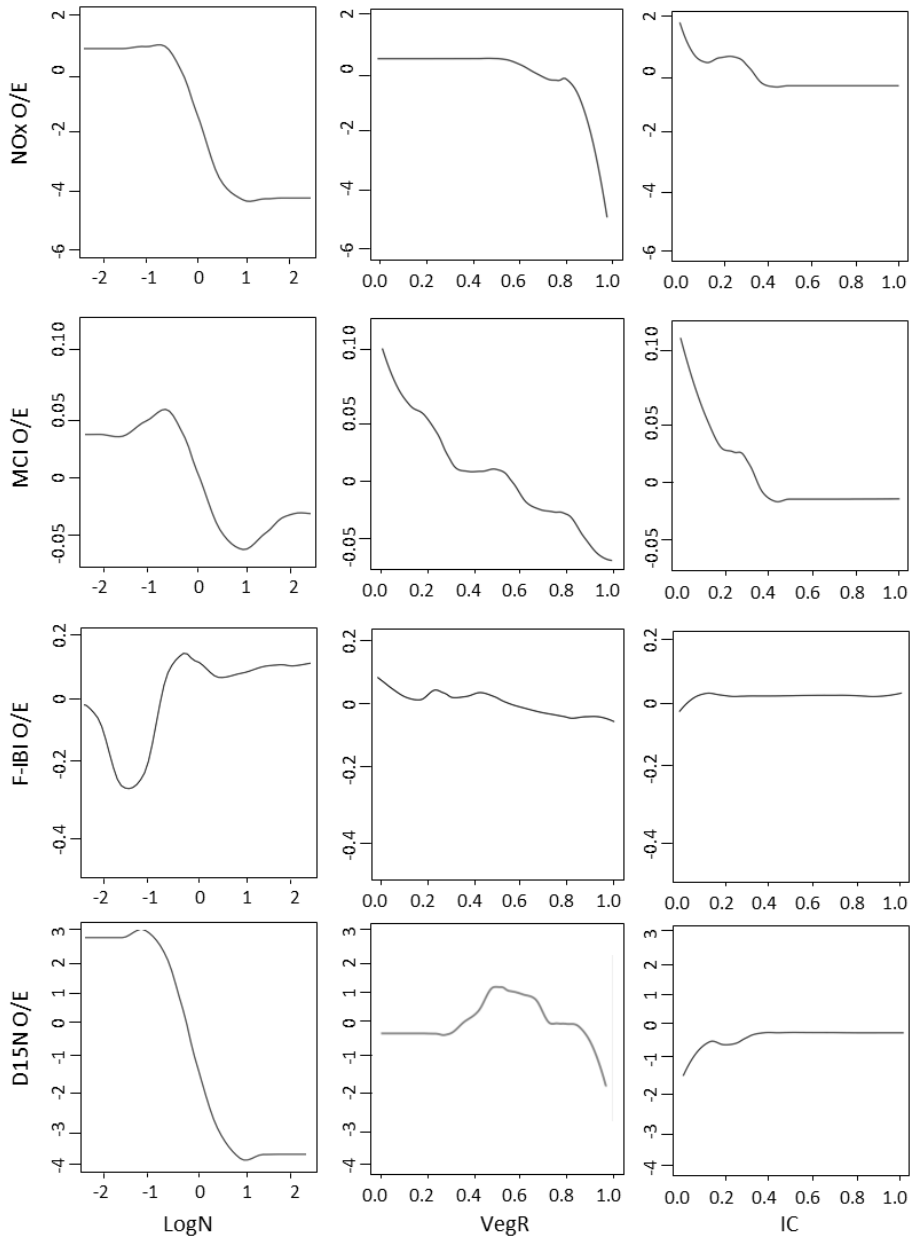


Figure 2

Ecological response curves (BRT fitted functions) of the O/E scores of selected component metrics (NO_x, MCI, F-IBI, D15N) against land-use pressure estimates (log nitrogen concentration [LogN], % native vegetation loss [VegR], % impervious cover [IC]). Functions are plotted with a loess-smoothing span of 0.3.

probably demonstrating the importance of native vegetation cover as a predictor of stream integrity. However, we do not develop a method of assigning qualitative scores in this study and suggest 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 threshold for change at 0.32 g·m⁻³ nitrogen concentration.

Commonly, the development of a multimetric index involves scoring component metrics prior to aggregation. This makes a 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.* 6, 3, 1 (Fore *et al.*, 1996) or 1 to 10 (Hughes *et al.*, 1998). Scores usually result

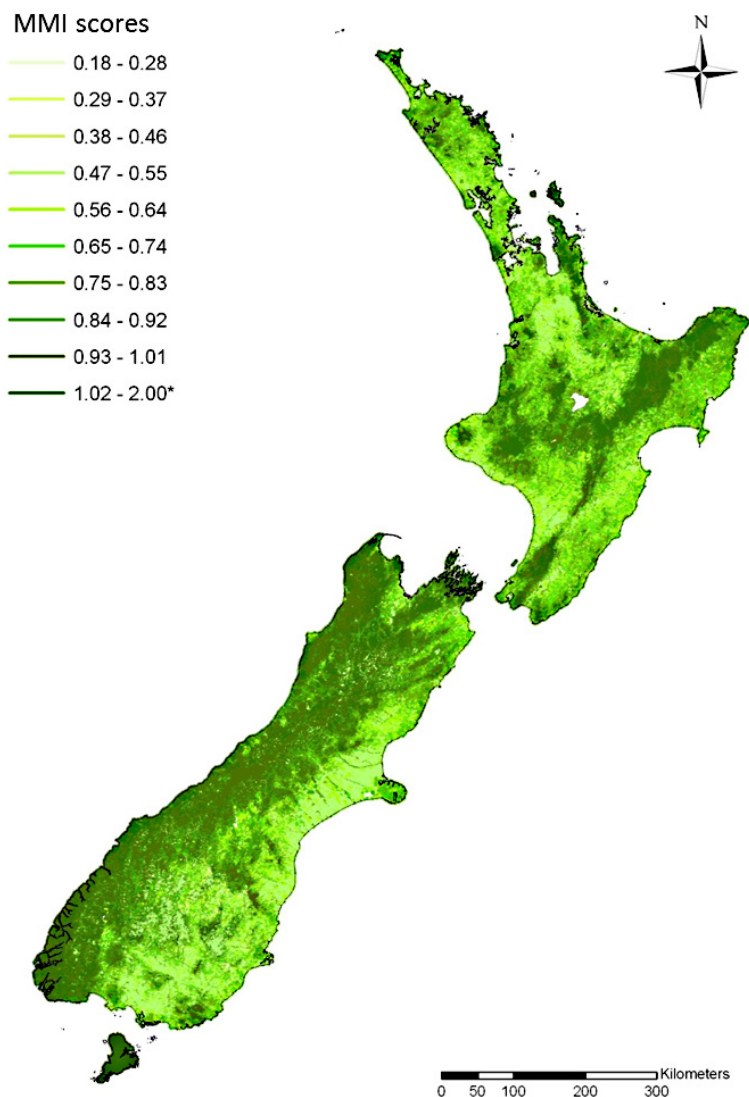


Figure 3

Geographical variation in predicted scores for the multimetric index of ecological integrity (MMI). Scores are divided by equal intervals except for the final group (*), which includes a greater range in values.

in categorical groupings that make trends in EI readily communicable e.g. Good, Moderate, Poor (European Union, 2000). This approach could also be further examined in future studies. The multimetric index (MMI) developed in this study represents four categories of stream health indicators (water quality, invertebrates, fish and ecosystem processes) as well as two components of the intrinsic worth of stream integrity (nativeness and pristineness). Whilst 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, Acuna *et al.* (2007) showed that functionally, streams were more resilient to flood disturbances during summer because they were efficiently processing carbon entering the stream at this time. Clearly there is a need to re-examine diversity metrics or examine alternative resilience metrics to include such measures in a holistic measure of EI.

Predictions of MMI and each of the component metrics were made at a national scale assuming the model predictions are applicable to all stream types. A strength of our BRT modelling

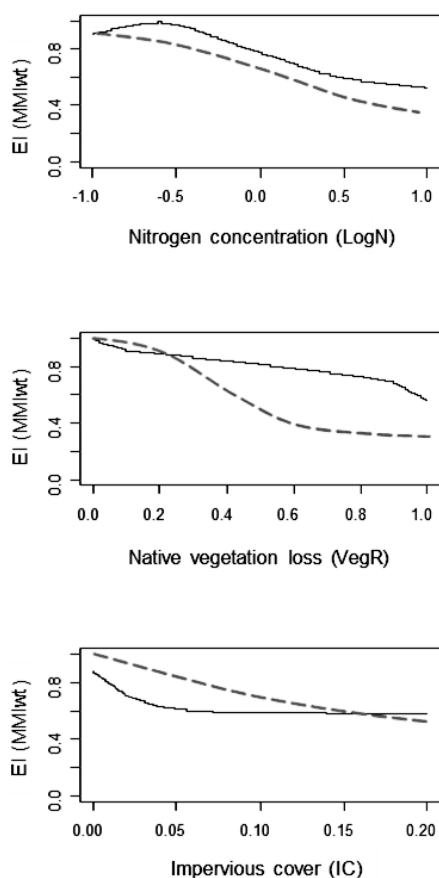


Figure 4

Ecological response curves (BRT fitted functions) of the multimetric index of ecological integrity (MMI) across the gradient in land-use pressure where there is confidence in the model predictions. Functions are plotted with a loess-smoothing span of 0.3. Dotted lines are conceptual ecological response curves from Leathwick and Julian (2007).

approach is that no *a priori* definition of typology is required. However, we only have confidence in the predictions for streams that fall within the range of environmental types described by the training data. Any stream types outside this sample group may not respond to pressures in the same way. The MMI 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).

All of the predictor 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.

CONCLUSIONS

Predictive modelling of component metrics of a multimetric index of ecological integrity was most informative when there were strong, monotonic relationships between the components of the MMI and vegetation removal, predicted stream *N* concentration and impervious cover

land-use gradients. 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 investigate alternative fish indices, to ensure fish are represented in a holistic measure of EI. Based on a qualitative evaluation of the MMI, results suggest 36% of streams have “high” 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. In summary, the boosted regression tree modelling provides a good approach for predicting expected and observed values for many component measures of ecological integrity. Verifying reference offsets and testing predictions with temporal and spatial measurements would strengthen models. The MMI metric in its current form provides a geographically meaningful prediction of the ecological integrity of rivers in New Zealand, at a broad spatial scale.

ACKNOWLEDGEMENTS

This research was funded by the Ministry of Science and Innovation through a Cross Departmental Research Pool project administered by the Department of Conservation. We thank John Leathwick for assistance with BRT analysis and Cawthron Institute for further funding for the preparation of the manuscript and two anonymous reviews for constructive feedback and editorial advice.

REFERENCES

- Acuña V., Giorgi A., Munõz I., Sabater F. and Sabater S., 2007. Meteorological and riparian influences on organic matter dynamics in a forested Mediterranean stream. *J. N. Am. Benthol. Soc.*, 26, 54–69.
- Baker M.E. and King R.S., 2010. A new method for detecting and interpreting biodiversity and ecological community thresholds. *Method Ecol. Evol.*, 1, 25–37.
- Breiman L., 2001. Random forests. *Mach. Learn.*, 45, 5–32.
- Brown L.R., Cuffney T.F., Coles J.F., Fitzpatrick F., McMahon G., Steuer J., Bell A.H. and May J.T., 2009. Urban streams across the USA: lessons learned from studies in 9 metropolitan areas. *J. N. Am. Benthol. Soc.*, 28, 1051–1069.
- Cao Y., Hawkins C.P., Olson J. and Kosterman M.A., 2007. Modelling natural environmental gradients improves the accuracy and precision of diatom-based indicators. *J. N. Am. Benthol. Soc.*, 26, 566–585.
- Carlisle D.M., Hawkins C.P., Meador M.R., Potapova M. and Falcone J., 2008. Biological assessments of Appalachian streams based on predictive models for fish, macroinvertebrates, and diatom assemblages. *J. N. Am. Benthol. Soc.*, 27, 16–37.
- Clapcott J., Young R., Goodwin E., Leathwick J. and Kelly D., 2011. Relationships between multiple land-use pressures and individual and combined indicators of stream ecological integrity. *DOC Res. Dev.*, 326, 57.
- Clapcott J.E., Collier K.J., Death R.G., Goodwin E.O., Harding J.S., Kelly D.J., Leathwick J.R. and Young R.G., 2012. Quantifying the relationships between land-use gradients and structural and functional indicators of stream ecological integrity. *Freshwater Biol.*, 57, 74–902.
- Collier K.J., 2008. Average score per metric: an alternative metric aggregation method for assessing wadeable stream health. *New Zeal. J. Mar. Fresh.*, 42, 367–378.
- Elith J., Leathwick J.R. and Hastie T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.*, 77, 802–813.
- European Union, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for community action in the field of water policy, 73 p.
- Fore L.S., Karr J.R. and Wisseman R.W., 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. *J. N. Am. Benthol. Soc.*, 15, 212–231.
- Hill B.H., Herlihy A.T., Kaufman P.R., Stevenson R.J., McCormick F.H. and Johnson C.B., 2000. Use of periphyton assemblage data as an index of biotic integrity. *J. N. Am. Benthol. Soc.*, 19, 50–67.

- Hughes R.M., Kaufmann P.R., Herlihy A.T., Kincaid T.M., Reynolds L. and Larsen D.P., 1998. A process for developing and evaluating indices of fish assemblage integrity. *Can. J. Fish. Aquat. Sci.*, 55, 1618–1631.
- Joy M.K. and Death R.G., 2004. Application of the index of biotic integrity methodology to New Zealand freshwater fish communities. *Environ. Manage.*, 34, 415–428.
- Karr J.R., 1981. Assessment of biotic integrity using fish communities. *Fisheries*, 6, 21–27.
- Leathwick J. and Julian K., 2007. *Production of pressure estimates for New Zealand river catchments*. NIWA client report: HAM2007-027, National Institute of Water & Atmospheric Research Ltd, Hamilton.
- Leathwick J.R., Snelder T., Chadderton W.L., Elith J., Julian K. and Ferrier S., 2011. Use of generalised dissimilarity modelling to improve the biological discrimination of river and stream classifications. *Freshwater Biol.*, 56, 21–38.
- Maxted J.R., Barbour M.T., Gerritsen J., Poretti V., Primrose N., Silvia A., Penrose D. and Renfrow R., 2000. Assessment framework for mid-Atlantic coastal plain streams using benthic macroinvertebrates. *J. N. Am. Benthol. Soc.*, 19, 128–144.
- Pont D., Hugueny B., Beier U., Goffaux D., Melcher A., Noble R., Rogers C., Roset N. and Schmutz S., 2006. Assessing river biotic condition at a continental scale: a European approach using functional metrics and fish assemblages. *J. Appl. Ecol.*, 43, 70–80.
- Schallenberg M., Kelly D., Clapcott J., Death R., MacNeil C., Young R., Sorrell B. and Scarsbrook M., 2011. Approaches to assessing ecological integrity of New Zealand freshwaters. *Sci.Cons.*, 307, 84.
- Stark J.D., 1985. *A macroinvertebrate community index of water quality for stony streams*. Water and Soil Miscellaneous Publication National Water and Soil Conservation Authority, Wellington, 53 p.
- Stoddard J.L., Larsen D.P., Hawkins C.P., Johnson R.K. and Norris R.H., 2006. Setting expectations for the ecological condition of streams: The concept of reference condition. *Ecol. Appl.*, 16, 1267–1276.
- Stoddard J.L., Herlihy A.T., Peck D.V., Hughes R.M., Whittier T.R. and Tarquinio E., 2008. A process for creating multimetric indices for large-scale aquatic surveys. *J. N. Am. Benthol. Soc.*, 27, 878–891.
- Uehlinger U., 2000. Resistance and resilience of metabolism in a flood prone river system. *Freshwater Biol.*, 45, 319–332.
- Vlek H.E., Verdonschot P.F.M. and Nijboer R.C., 2004. Towards a multimetric index for the assessment of Dutch streams using benthic invertebrates. *Hydrobiologia*, 516, 173–189.
- Waite I., 2014. Agricultural disturbance response models for invertebrate and algal metrics from streams at two spatial scales within the U.S. *Hydrobiologia*, 726, 285–303.
- Woods R., Elliot S., Shankar U., Bidwell V., Harris S., Wheeler D., Clothier B., Green S., Hewitt A., Gibb R. and Parfitt R., 2006. *The CLUES project: Predicting the effects of land-use on water quality – Stage II*. National Institute of Water & Atmospheric Research Ltd, Hamilton, 109 p.