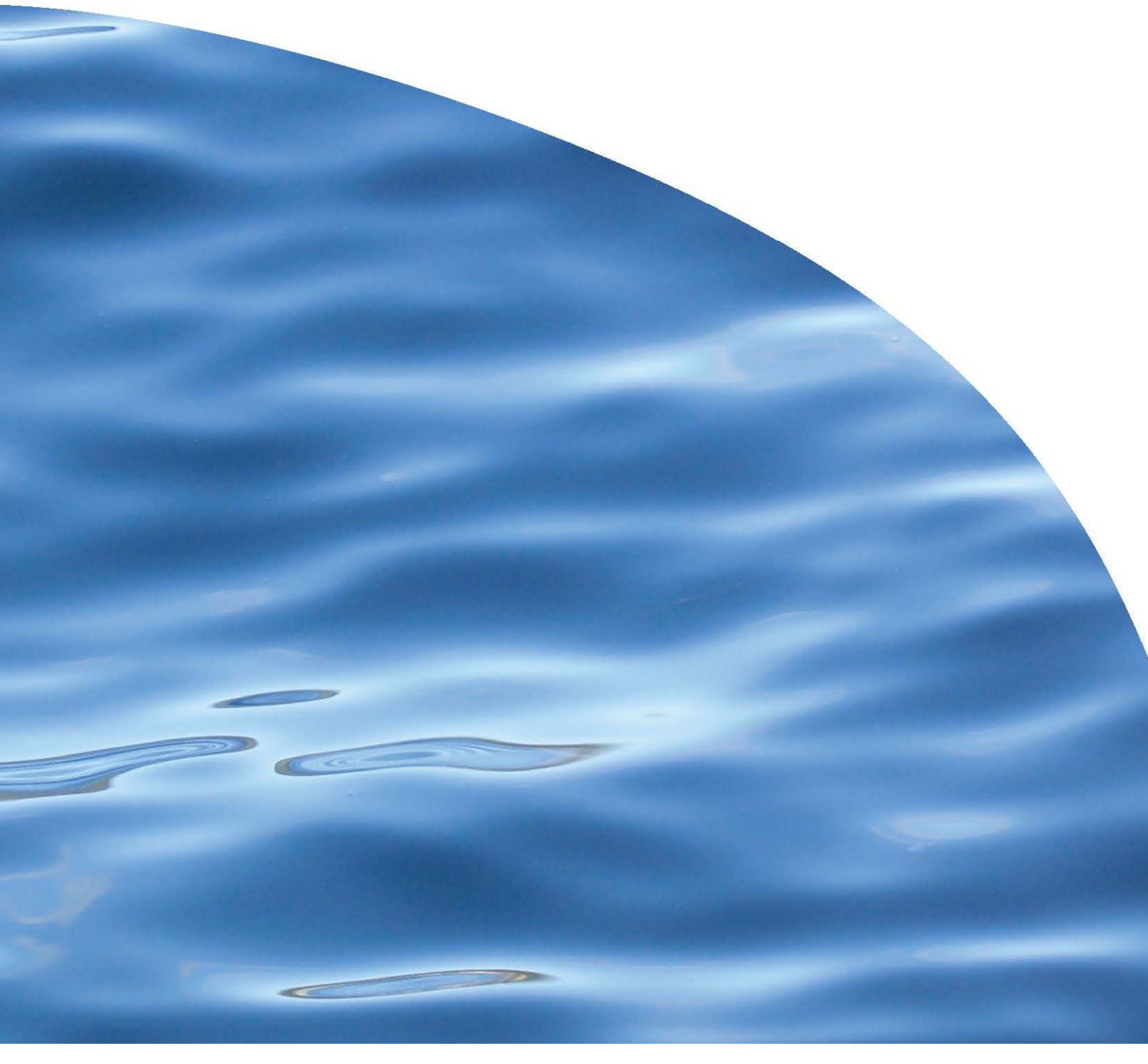
CAWTHRON INSTITUTE

REPORT NO. 2301

PREDICTIVE MODELS OF BENTHIC MACROINVERTEBRATE METRICS



PREDICTIVE MODELS OF BENTHIC MACROINVERTEBRATE METRICS

JOANNE CLAPCOTT[1], ERIC GOODWIN[1], TON SNELDER[2]

[1] CAWTHRON INSTITUTE
[2] AQUALINC

Prepared for Ministry for the Environment.

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

REVIEWED BY:
Joe Hay

APPROVED FOR RELEASE BY:
Rowan Strickland

ISSUE DATE: 27 February 2013

RECOMMENDED CITATION: Clapcott J, Goodwin E, Snelder T 2013. Predictive Models of Benthic Macroinvertebrate Metrics. Prepared for Ministry for the Environment. Cawthron Report No. 2301. 35 p. plus appendices.

EXECUTIVE SUMMARY

Ministry for the Environment are interested in using predictive models to estimate aspects of ecological integrity for the river network of New Zealand. This report provides a comparative analysis of models used to estimate indicators of benthic macroinvertebrate health for non-monitored sites, based on available landuse and environmental data. Two machine learning model approaches are applied, boosted regression trees (BRT) and random forest (RF) models, to predict the contemporary status of the Macroinvertebrate Community Index (MCI) and Ephemeroptera, Plecoptera, or Trichoptera (EPT) richness metrics. These two model approaches are also compared to a linear regression model (ANCOVA) in estimating the expected reference (historical, unimpacted) status of the macroinvertebrate metrics.

Models were developed for response metrics using a national compilation of stream monitoring data collected during 2007-2011 from 1033 unique stream segments. Predictor variables included land cover derived from the Land Cover Database (LCDB3), a measure of surface water allocation pressure, and environmental descriptors from the Freshwater Ecosystem of New Zealand (FENZ) database.

Results indicated no consistently large difference in model performance between BRT and RF models when predicting contemporary status of macroinvertebrate metrics, although BRT predictions were less biased. Similarly, there was no significant difference in reference model performance between BRT, RF or linear models for MCI. The performance of all reference models for EPT richness was relatively weak. Based on our comparative assessment of predictive model approaches we recommend a BRT approach for predicting macroinvertebrate metrics.

We applied a BRT approach and predicted contemporary and reference status for all stream segments of the river network for four macroinvertebrate metrics. Model output is summarised by stream classes including FENZ and River Environment Classification (REC) groups.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

1. INTRODUCTION

As part of the National Environmental Monitoring and Reporting (NEMaR) project Ministry for the Environment are interested in using predictive models to inform values of ecological integrity for the river network of New Zealand. Previous research has demonstrated a strong link between land use, environmental variability and indicators of stream ecological integrity at sites monitored as part of the State of the Environment network (Unwin *et al.* 2010; Clapcott *et al.* 2011). The relationships between these predictor and response variables can be used to inform indicator values for non-monitored sites. However, there are different modelling approaches available and to date limited information on which modelling approach is likely to provide the best predictions of indicators of stream ecological integrity.

This document reports on a comparative analysis of models used to predict indicators of benthic macroinvertebrate health. The Ministry for the Environment's guidelines for the project outputs emphasise generating useful datasets which can be used for subsequent analysis. To inform which model provides the most informative predictions to populate such a dataset, this report contains discussion on the reliability, accuracy and robustness of comparative models, and a brief recommendation of which of these models is likely to provide the best prediction of i) current state for macroinvertebrate metrics, and ii) reference state for macroinvertebrate metrics. Hence this report:

- describes a Random Forest (RF) methodology used to model data
- describes a Boosted Regression Tree (BRT) methodology used to model data
- describes a linear model approach to model data
- briefly summarises the results of these analyses when applied to two metrics of macroinvertebrate health and compares predictive model performance
- discusses the strengths and weaknesses of the resulting models.

The most robust modelling approach is then applied to two additional metrics of macroinvertebrate health. In total, this provides national predictions for:

- Macroinvertebrate Community index (MCI, or hbMCI): an index reflecting environmental pollution (hb more specifically referring to the variant of MCI adapted for hard-bottomed streams, and in fact used throughout this study)
- EPT richness: the number of taxa present belonging to orders Ephemeroptera, Plecoptera, or Trichoptera
- %EPT richness: the percentage of taxa belonging to orders Ephemeroptera, Plecoptera, or Trichoptera
- Taxon richness: the number of taxa present.

1.1. Predictive model approaches

1.1.1. Machine learning techniques

Machine learning (ML) is a rapidly growing area of predictive modelling that is concerned with identifying structure in complex, often nonlinear, data and generating accurate predictive models (Olden *et al.* 2008). A number of ML techniques have been promoted in ecology as powerful alternatives to traditional (linear regression) modelling approaches. These include approaches that attempt to model the relationship(s) between a set of inputs and known outputs, such as artificial neural networks, classification and regression trees, fuzzy logic, and genetic algorithms and programming. In contrast, traditional modelling approaches include techniques, such as general linear or additive models, with stricter statistical assumptions and requirements.

Machine learning approaches often exhibit greater power for resolving complex (non-linear, non-monotonic, multimodal) ecological relationships, as they are not restricted by the data assumptions of conventional, parametric approaches (Olden *et al.* 2008). Further advantages of regression tree techniques, including both boosted regression trees and random forest models, include an ability to accommodate different types of predictor variables and missing values, immunity to the effects of extreme outliers and the inclusion of irrelevant predictors, and a facility for fitting interactions between predictors (Friedman & Meulman 2003).

1.1.2. Random forest models

We used Random Forest (RF) models to relate invertebrate indices to predictor variables. For a detailed description of RF models see Breiman 2001 and Cutler *et al.* 2007. Briefly, an RF model comprises an ensemble of many individual Classification and Regression Trees (CART, Breiman *et al.* 1984) that can be used in a regression mode to partition the observations into groups, which minimise the squared error.These groups are derived by a series of binary rules (splits) constructed from the predictor variables (here the environmental variables). Single-tree CART models have two desirable features for modelling complex relationships: they are free from distributional assumptions, and they automatically fit non-linear relationships and high order interactions. However, CART models have two limitations: they do not produce an optimal tree structure and they are sensitive to small changes in input data (Hastie *et al.* 2009).

The limitations in CART models can be reduced by using RF models (Breiman 2001), in which a final prediction is based on the average of all the individual predictions obtained from the trees in the ensemble (the forest). An important feature of RF models is that each tree is 'grown' (its split rules are determined) with a bootstrap sample of the training data. In addition, at each node only small, random samples of the available predictors are used to define the split. The introduction of these two

stochastic aspects, combined with averaging individual predictions from the ensemble of trees, increases the prediction accuracy of RF models, while retaining the desirable features of CART.

The predictive accuracy (or internal cross validation) of the RF model can be reported as an R^2 value, not to be confused with the percentage variance explained (R^2) of a traditional modelling approach (*e.g.* linear modelling). The R^2 value of an RF model represents an estimate of future predictive performance (whereas the R^2 of a linear model represents how much deviance in the response variable is explained by the predictor variable(s)).

The structures of RF models can be examined using importance measures and partial dependence plots. Importance measures indicate the contribution of the predictors to model accuracy (Breiman 2001). Partial dependence plots show the marginal contribution of a predictor to the response (*i.e.* the response as a function of the predictor when the other predictors are held at their mean value) (Friedman & Meulman 2003). These plots are not a perfect representation of the effects of each predictor, particularly if there are interactions or predictors are strongly correlated, but they provide useful information for interpretation (Friedman & Meulman 2003).

1.1.3. Boosted regression trees

We used Boosted Regression Tree (BRT) models to relate invertebrate indices to predictor variables. The boosting approach used in BRT has its origins within ML, but subsequent developments in the statistical community reinterpret it as an advanced form of regression (Friedman *et al.* 2000).The model development for BRT analysis is discussed in detail in the literature (Friedman 2001; Elith *et al.* 2008; Hastie *et al.* 2009). Briefly, the BRT method combines additive regression modelling with boosting techniques, and provides an estimate from numerous (often thousands of) models. Results include 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 data held out from the fitting process (the 'holdout data').

Whereas in RF model development predictor variables are randomly selected to split (partition) the observations into separate groups, in BRT the variable on which to base the split is selected so as to maximise the difference between the two groups, or minimise the residual deviance.What's more, each subsequent tree in a BRT model focusses on explaining hitherto-unexplained deviance. That is, each subsequent tree being added models the residuals from the existing BRT forest.

The internal cross validation coefficient (CV) provides a measure of predictive accuracy comparable to the RF model predictive accuracy (R^2). Here we report it (CV) as an R^2 to compare to RF R^2.

From BRT models one can generate representations of the non-parametric relationships (*e.g.* linear, curvilinear, multimodal fitted functions) between response and predictor variables, comparably to the partial dependence plots that can be generated for RF models.

2. METHODS

2.1. Raw data

Benthic invertebrate metric data were compiled by Ministry for the Environment and consisted of data from regional council and unitary authorities, collected predominantly from State of the Environment river monitoring sites during 1998 to 2011. Metrics included:

- Macroinvertebrate Community Index (MCI): an index of organic enrichment widely applied as a measure of stream health
- EPT richness: the number of taxa present belonging to orders Ephemeroptera, Plecoptera, or Trichoptera
- %EPT richness: the percentage of taxa belonging to orders Ephemeroptera, Plecoptera, or Trichoptera
- Taxon richness: the number of taxa present.

2.2. Working dataset

We restricted our analysis to using the median of site values from the last five years (2007-2011), so that responses were comparable to potential land cover predictors derived from the most recent satellite imagery (2007-2008; Land Cover Database 3). 'Site' was defined by stream segment (NZReach), being a section of river between tributary confluences. All data for any given site were combined to calculate median values, except when there were two obvious upstream and downstream locations in a segment, potentially indicating monitoring above and below a point source input, in which case only values from the upstream (pre-impact) location were used. The working dataset included 1033 sites, from all regions (Figure 1). The sites were not evenly distributed over classes defined by the REC or FENZ stream classifications (Table 1, Table 2) and were predominantly located in lowland areas. This affects the predictive accuracy to reaches in environments that have not been sampled, regardless of the modelling method employed. However, methods which treat the predictor variables themselves (RF and BRT) are more likely to be able to extrapolate into these regions than the ANCOVA method which aggregates the environmental predictors into classes. The method will not be able to make estimates for classes not represented in the training data.

Table 1. Number of sites in the working dataset located in River Environment Classification Climate/Source-of-Flow (REC CSOF) groupings. Stream lengths exclude first order streams.

Climate/Source-of-Flow	N	% N	Stream length (km)	% total length
Cool-dry/Hill	51	4.9	18,210	9.1
Cool-dry/Low-elevation	130	12.6	25,111	12.6
Cool-dry/Lake	1	0.1	492	0.2
Cool-dry/Mountain	5	0.5	2,317	1.2
Cool-wet/Hill	191	18.5	33,264	16.7
Cool-wet/Low-elevation	136	13.2	16,011	8.0
Cool-wet/Lake	3	0.3	1,714	0.9
Cool-wet/Mountain	17	1.6	15,447	7.7
Cool-wet/Glacial-mountain	1	0.1	97	0.0
Cool-extremely-wet/Hill	44	4.3	17,228	8.6
Cool-extremely-wet/Low-elevation	34	3.3	5,896	3.0
Cool-extremely-wet/Lake	9	0.9	1,801	0.9
Cool-extremely-wet/Mountain	7	0.7	12,593	6.3
Warm-dry/Low-elevation	53	5.1	9,993	5.0
Warm-dry/Lake	1	0.1	112	0.1
Warm-wet/Hill	5	0.5	620	0.3
Warm-wet/Low-elevation	324	31.4	30,783	15.4
Warm-wet/Lake	2	0.2	368	0.2
Warm-extremely-wet/Hill	2	0.2	418	0.2
Warm-extremely-wet/Low-elevation	17	1.6	1,492	0.7
Other	0	0.0	5,793	2.9

Table 2. Number of sites in the working dataset located in Freshwater Ecosystems of New Zealand (FENZ) geo-database C20 and C100 level groupings.

C20	N	% N	Stream length (km)	% total length
A	215	20.8	86,314	19.9
B	3	0.3	1,458	0.3
C	730	70.7	176,549	48.2
D	11	1.1	18,592	4.9
E	2	0.2	1,124	0.3
G	66	6.4	41,692	10.7
H	6	0.6	28,884	6.2
Other	0	0.0	37,224	9.5

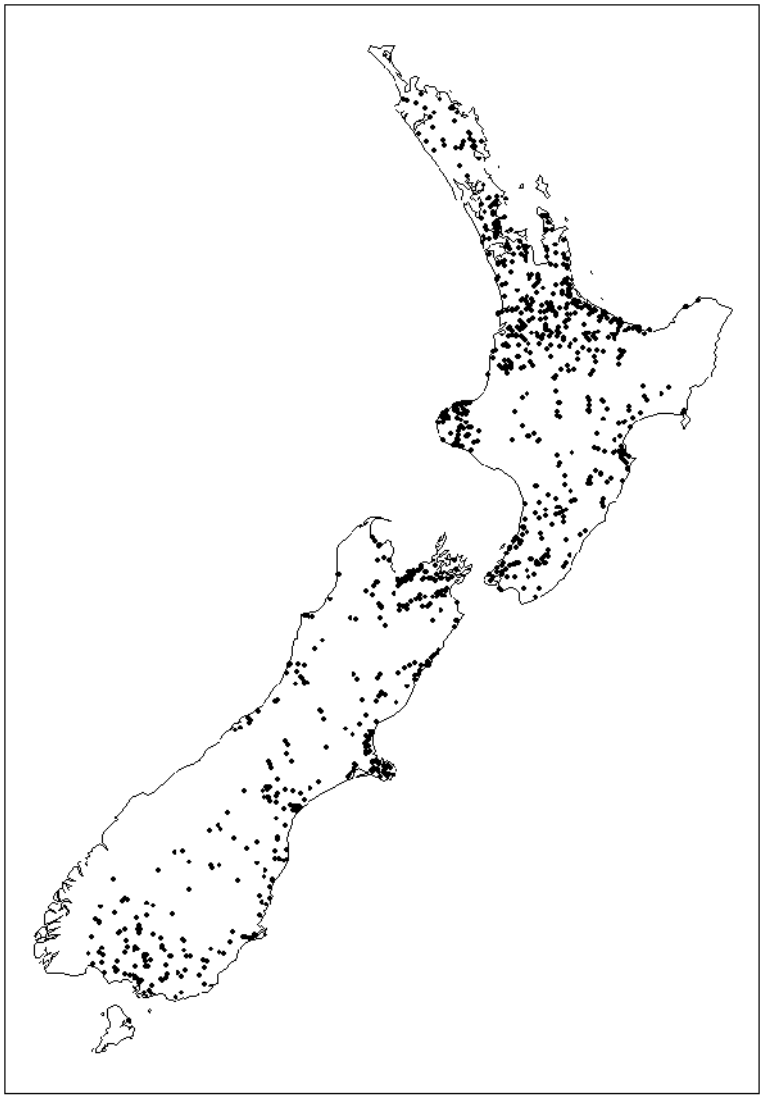


Figure 1. Distribution of working dataset sites, N = 1033.

2.3. Predictor variables

Potential predictors of macroinvertebrate metrics included measures of land cover and other environmental descriptors. We merged the set of 33 land cover descriptors from LCDB3 into six predictor variables that represented broad land cover categories as follows:

- **Native vegetation** (including Broadleaved Indigenous Hardwoods, Indigenous Forest, Alpine Grass/Herbfield, Fernland, Manuka and/or Kanuka, Sub Alpine Shrubland)
- **Exotic vegetation** (Forest – Harvested, Deciduous Hardwoods, Exotic Forest, Gorse and/or Broom, Mixed Exotic Shrubland)

- **Pastoral heavy** (Short-rotation Cropland, Orchard Vineyard & other Perennial Crops, High Producing Grassland)
- **Pastoral light** (Low Producing Grassland, Tall Tussock Grassland, Depleted Grassland)
- **Urban** (Built-up Area (Settlement), Urban Parkland/Open Space, Transport Infrastructure, Surface Mines and Dumps)
- **Bare ground** (Sand and Gravel, Landslide, Permanent Snow and Ice, Gravel and Rock)
- **Wetland** (Herbaceous Freshwater Vegetation, Herbaceous Saline Vegetation, Flaxland, Mangrove).

Environmental descriptors were accessed from the Freshwater Ecosystems of New Zealand (FENZ) database. We selected variables previously demonstrated to have informative relationships with the distribution of freshwater invertebrates (Leathwick *et al.* 2011). Environmental predictor variables included measures of:

- Geology and topography
- Slope
- Flow and flow influencing factors
- Geology.

The total predictor dataset was reduced to 23 predictors (Table 3) during initial model building, where the relative importance of predictors and meaningfulness of response relationships with invertebrate metrics was assessed. Nonsense responses and low contributing predictors were excluded from the predictor set. Transforms detailed in Table 3 were established during the development of the FENZ database (Leathwick *et al.* 2010), to distribute data more evenly across the range. While this is not necessary for the modelling methods employed in this study, it facilitates inspection of response curves, as the detail of fitted functions is revealed by expanding dense areas.

In addition, we included a measure of surface water allocation pressure, as an estimate of the pressure anthropogenic water use has on river flows. This predictor variable is described by Clapcott and Goodwin (2010).

Table 3. Description of the land-use pressure gradients and environmental variables, including the mean and range of values, used in this study.

Variable	Description	Mean (range)
NativeVeg	Native vegetation cover in the catchment (%)	34.4 (0, 100)
PastoralHeavy	Pastoral heavy cover in the catchment (%)	42.3 (0, 100)
PastoralLight	Pastoral light cover in the catchment (%)	7.1 (0, 92)
Urban	Urban impervious cover in the catchment (%)	3.2 (0, 99)
BareGround	Bare ground in the catchment (%)	1.0 (0, 40)
Surface Water Allocation	Low flow remaining after the upstream daily surface water allocation is deducted (proportion)	0.1 (0, 1)
SEGFLOWSTA	Annual low flow/annual mean flow (ratio)	0.2 (0, 0.5)
SEGLFLOW4T	Mean annual 7-day low flow (m3/s), fourth-root transformed	1.1 (1, 4.1)
SEGJANAIRT	Segment summer air temperature (°C)	17 (12.6, 19.6)
SEGMINTNOR	Segment winter air temperature (°C), normalised with respect to SEGJANAIRT	0.5 (-4.2, 3.5)
SEGRIPSHAD	Segment riparian shade (proportional)	0.3 (0, 0.8)
SEGSLOPESQ	Segment slope (°), square-root transformed	1.3 (1, 3.9)
LOCHAB	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 (2.3, 4.8)
LOCSED	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	3.6 (1, 5.9)
USDAYSRAIN	Days/year with rainfall in the catchment greater than 25 mm	14.2 (1.9, 71.4)
USAVGTNORM	Average air temperature (°C) in the catchment, normalised with respect to SEGJANAIRT	-0.5 (-6.0, 1.6)
USAVGSLOPE	Average slope in the catchment (°)	12.9 (0.0, 32.0)
USHARDNESS	Average hardness of rocks in the catchment, 1 = very low to 5 = very high	2.9 (1, 5.0)
USCALCIUM	Average calcium concentration of rocks in the catchment, 1 = very low to 4 = very high	1.6 (1.0, 4.0)
USPHOSPHOR	Average phosphorus concentration of rocks in the catchment, 1 = very low to 5 = very high	2.4 (1.0, 5)
USLAKEPC	Area of lake in upstream catchment (%)	0.0 (0.0, 0.1)
USPEATPC	Area of peat in upstream catchment (%)	0.0 (0.0, 1.0)
USGLACIER	Area of glacier in upstream catchment (%)	0.0 (0.0, 0.1)

3. COMPARISON BETWEEN RANDOM FOREST AND BOOSTED REGRESSION TREE MODELS FOR PREDICTING CURRENT METRIC VALUES

The relationship between Macroinvertebrate Community Index (MCI) and the land cover and environmental variability predictor variables was modelled using both random forest (RF) models and boosted regression tree (BRT) approaches. The aim of this investigation was to compare the models and determine the most suitable approach to use for predicting current MCI in New Zealand streams and rivers. The comparison was repeated on a second metric of stream health, being the number of macroinvertebrate taxa belonging to the orders Ephemeroptera, Plecoptera and Trichoptera (EPT richness).

Predictive models were first developed using all of the training dataset (n = 1033) to compare model diagnostics and the relative importance of predictor variables. The output from these models was used to predict metric values for all stream segments and calculate summaries for two stream classifications: River Environment Classification at the Climate/Source of Flow level (REC CSOF), and the Freshwater Ecosystems of New Zealand C20 level (FENZ C20).

Secondly, 'holdout models' were developed by fitting the BRT and RF models using a random sample of 80% of the training dataset (n = 826). The holdout models were then used to predict metric values for the remaining 20% of sites (n = 217). This allowed us to independently test the predictive performance of the models and assess model consistency and bias, where inconsistency manifests as a deviation from unity slope of a regression line between observed and predicted values, and bias manifests as a vertical offset of the regression line from a 1:1 line. We assessed model performance with:

- the correlation between predicted and observed values (R) for the 20% held out sites
- the Nash-Sutcliffe efficiency (NSE) statistic which indicates how well the plot of observed versus predicted values fits the 1:1 line, where values greater than 0 are satisfactory but values greater than 0.5 indicate good model performance (Nash & Sutcliffe 1970)
- root mean squared deviation (RMSD) is an estimate of model inaccuracy (departure between observed and predicted values), where smaller values indicate lower inaccuracy than large values (Pineiro *et al.* 2008)
- bias (Bias) which measures the average tendency of the predicted values to be larger or smaller than the observed, where positive values indicate model underestimation and negative values indicate overestimation bias

- a test whether the slope (Pslope) and intercept (Pinter) of regressions of the observed versus predicted values differed significantly from 1 and 0, respectively.

3.1. Macroinvertebrate Community Index

3.1.1. Model performance

The BRT model had an internal cross validation R^2 of 63.7% which indicates very good predictive performance. Similarly, the internal cross validation statistic of the RF model indicated very good predictive accuracy (R^2 = 63.2%).

The BRT model predicted a mean MCI value of 112 and the RF model 109.5 at the national scale. A comparison of mean MCI values for REC CSOF classes suggested BRT predictions were on average 2.5 MCI units higher than RF model predictions ($t_{(22)} = 3.52$, $p < 0.01$). Similarly, a comparison of mean MCI values for FENZ C20 classes suggested BRT predictions were on average 2.9 MCI units higher than RF model predictions ($t_{(20)} = 5.60$, $p < 0.01$).

The relationships between predicted and observed MCI values (from the hold-out models) showed similar performance for the RF and BRT models in terms of R and NSE, both suggesting excellent model performance (Figure 2). However, the RF model did have a slope (Pslope < 0.001) and intercept (Pinter = 0.001) significantly different to one and zero, respectively, compared with no significant difference for slope (Pslope = 0.069) and intercept (Pinter = 0.103) for the BRT model. This indicates that the RF model is biased and likely to overestimate low values and underestimate high values.

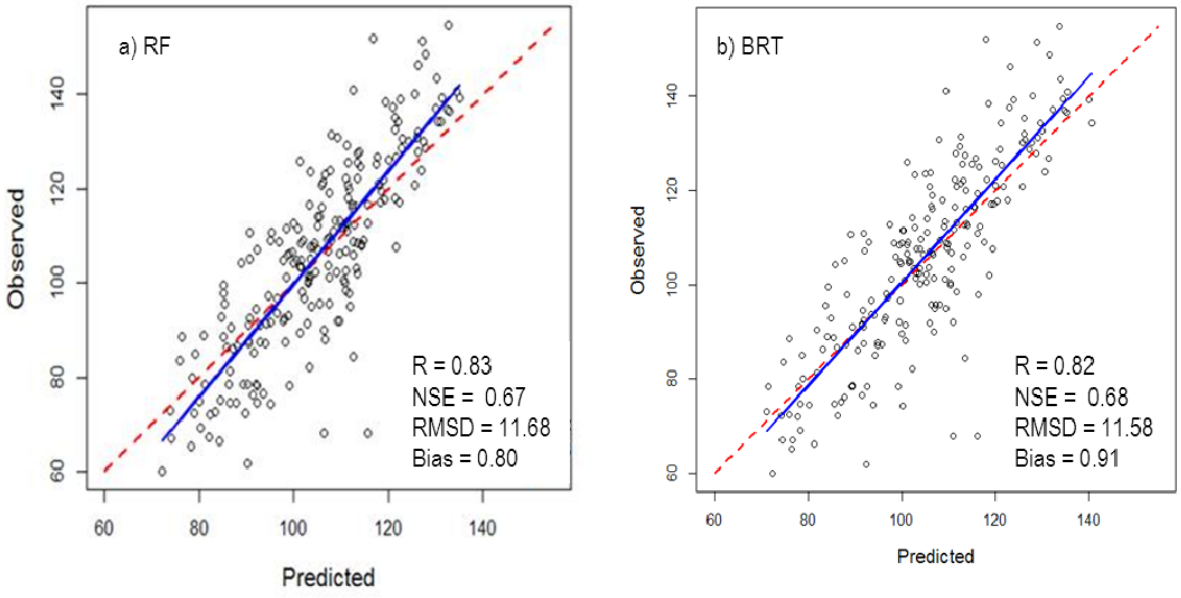


Figure 2. Scatter plots of observed versus predicted values from a) Random Forest (RF) model and b) Boosted Regression Tree (BRT) model for Macroinvertebrate Community Index (MCI). The dashed line is the 1:1 line and the blue line is the line of best fit. Model performance statistics are explained in the text.

3.1.2. Predictor variables

The relative importance of predictor variables are reported differently in the BRT and RF model approaches. In BRT the relative importance of a predictor variable is gained from the number of times that variable is selected for tree splitting during model development, weighted by the squared improvement to the model as a result of each split, and averaged over all trees. The relative influence (or contribution) of each variable is scaled so that the sum adds to 100 (i.e. scaled to a percentage contribution), with higher numbers indicating stronger influence on the response (Elith *et al.* 2008).

In RF models there are two importance measures, accuracy importance and node purity. The former reports the contribution of a specific predictor variable to model accuracy. It is computed by randomly permuting the values of the variable for the out of bag (OOB, internally held out) observations and predictions are then obtained from the tree for these modified data. The difference between the prediction accuracy (R^2) for the modified and original OOB data, divided by the standard error, measures the importance of each variable (Unwin *et al.* 2010). Node purity is more consistent with the BRT measure of relative importance and is the total decrease in node deviance from splitting on the variable, averaged over all trees (Unwin *et al.* 2010). Here we scale Node Purity scores so that they sum to 100 to make values more readily comparable to BRT output.

Both model approaches identified the same four variables as most important for explaining deviance in MCI data, including native vegetation, % heavy pasture, flow stability and summer temperature (Table 4).

Table 4. Comparison of predictor variable importance in predictive models of Macroinvertebrate Community Index (MCI) using Boosted Regression Tree (BRT) and Random Forest (RF) model approaches. The four most important variables in both models are highlighted in bold text.

Predictor	RFM	BRT
% Native vegetation	**22.45**	**29.52**
% Pastoral heavy	**11.82**	**9.82**
% Pastoral light	0.86	0.41
% Urban	3.85	6.12
% Bare ground	0.34	0.33
Surface Water Allocation	1.57	0.89
Segment flow stability	**6.83**	**7.19**
Segment low flow	3.02	2.06
Segment summer temperature	**5.48**	**8.68**
Segment winter temperature normalised	2.51	3.08
Segment shade	3.22	3.12
Segment slope	4.00	3.37
Segment habitat	5.56	4.78
Segment particle size	6.85	2.66
Catchment rain days > 25mm	4.74	5.31
Catchment average temperature	2.51	2.20
Catchment slope	5.89	3.80
Catchment hardness	2.96	2.42
Catchment calcium	2.98	2.23
Catchment phosphorus	2.18	1.84
% lake	0.11	0.08
% peat	0.28	0.10
% glacier	0.00	0.00

The RF and BRT models exhibited very similarly shaped relationships between environmental predictor variables and MCI (Figure 3). The partial dependence plots show the difference from the mean MCI value (103) in response to a change in environmental predictor value. In general, a monotonic decline in MCI was associated with decreasing native vegetation cover and increasing heavy pasture and urban development in the catchment. Low MCI values were also associated with higher temperatures, low rain days and low flow stability.

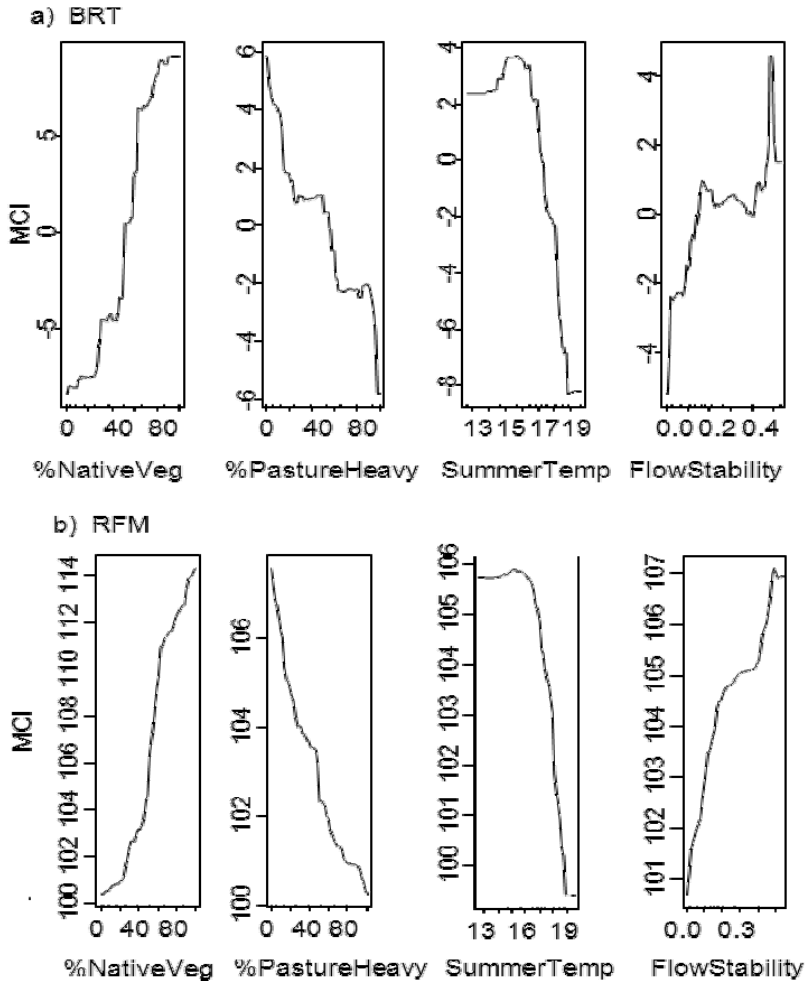


Figure 3. The relationships between Macroinvertebrate Community Index (MCI) and environmental predictors in order of importance from a) Boosted Regression Tree (BRT) and b) Random Forest (RF) models. Note the scales on the y-axes are not the same; the BRT scale provides the marginal contribution of each predictor to the mean MCI value (103), whereas the RF model scale is the absolute value of the prediction when other variables were held at their mean.

3.2. EPT richness

3.2.1. Model performance

The BRT model had an internal cross validation R^2 of 56.5% which indicates very good predictive performance. Similarly, the internal cross validation statistic of the RF model indicated very good predictive accuracy (R^2 = 56.9%).

The BRT models predicted a mean EPT richness value of 9.0 and the RF model 9.5 at the national scale. A comparison of mean EPT richness values for REC CSOF classes suggested BRT predictions were on average 0.5 units lower than RF model predictions ($t_{(22)}$ = 2.48, p = 0.02). Similarly, a comparison of mean EPT richness

values for FENZ C20 classes suggested BRT predictions were on average 1.6 units lower than RF model predictions ($t_{(20)} = 7.70$, $p < 0.01$).

The relationships between predicted and observed EPT values (from the hold-out models) showed similar performance for the RF and BRT models in terms of R and NSE, both suggesting excellent model performance (Figure 4). However, the RF model did have a slope (Pslope < 0.001) and intercept (Pinter = 0.001) significantly different to one and zero respectively, compared with no significant difference for slope (Pslope = 0.437) and intercept (Pinter = 0.737) for the BRT model. This indicates that the RF model is biased and likely to overestimate low values and underestimate high values.

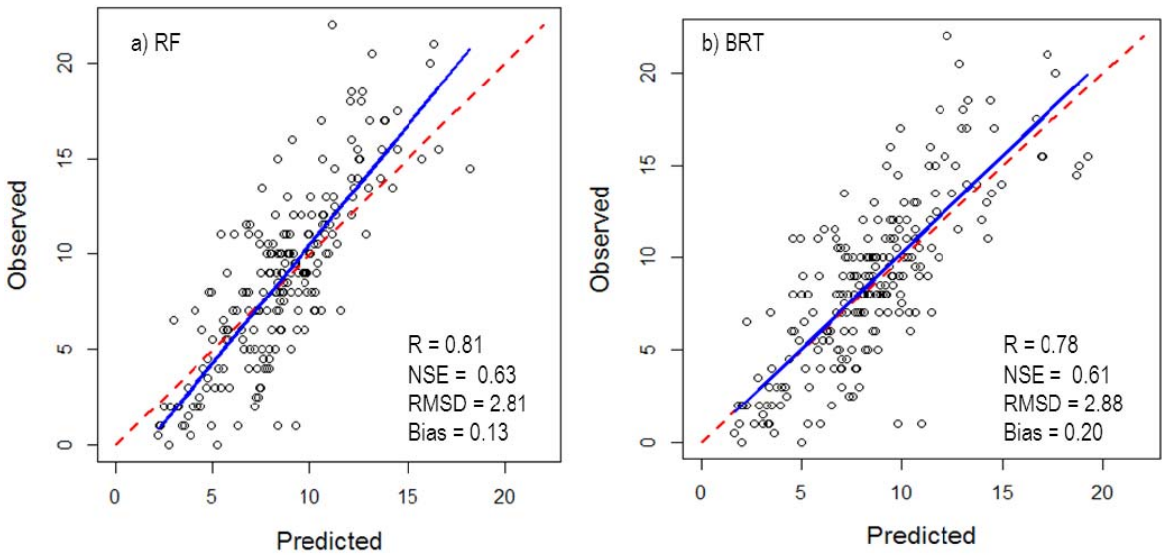


Figure 4. Correlations between observed and predicted values from a) Random Forest (RF) model and b) Boosted Regression Tree (BRT) model for EPT richness. The dashed line is the 1:1 line and the blue line is the line of best fit. Model performance statistics are explained in the text.

3.2.2. Predictor variables

Both BRT and RF model approaches identified the same three variables as most important for explaining deviance in EPT richness data, including native vegetation cover, flow stability and particle size (Table 5).

Table 5. Comparison of predictor variable importance in predictive models of EPT richness using Boosted Regression Tree (BRT) and Random Forest (RF) model approaches. The three most important variables in both models are highlighted in bold text.

Predictor	RF	BRT
% Native vegetation	**20.10**	**26.60**
% Pastoral heavy	4.94	3.67
% Pastoral light	1.58	2.82
% Urban	4.32	6.01
% Bare ground	1.01	1.03
Surface Water Allocation	2.04	1.43
Segment flow stability	**7.72**	**7.90**
Segment low flow	4.11	4.23
Segment summer temperature	4.01	5.05
Segment winter temperature normalised	4.22	1.65
Segment shade	4.52	3.33
Segment slope	3.60	1.83
Segment habitat	3.31	2.90
Segment particle size	**12.05**	**7.80**
Catchment rain days > 25mm	4.59	5.18
Catchment average temperature	4.22	4.46
Catchment slope	4.44	2.95
Catchment hardness	3.05	3.83
Catchment calcium	3.34	3.30
Catchment phosphorus	3.46	3.05
% lake	0.13	0.00
% peat	0.67	0.89
% glacier	0.00	0.10

The RF and BRT models exhibited very similarly shaped relationships between individual environmental predictor variables and EPT. Mean EPT values (16.7) increased in association with high native vegetation cover, high flow stability and greater substrate particle size (Figure 5).

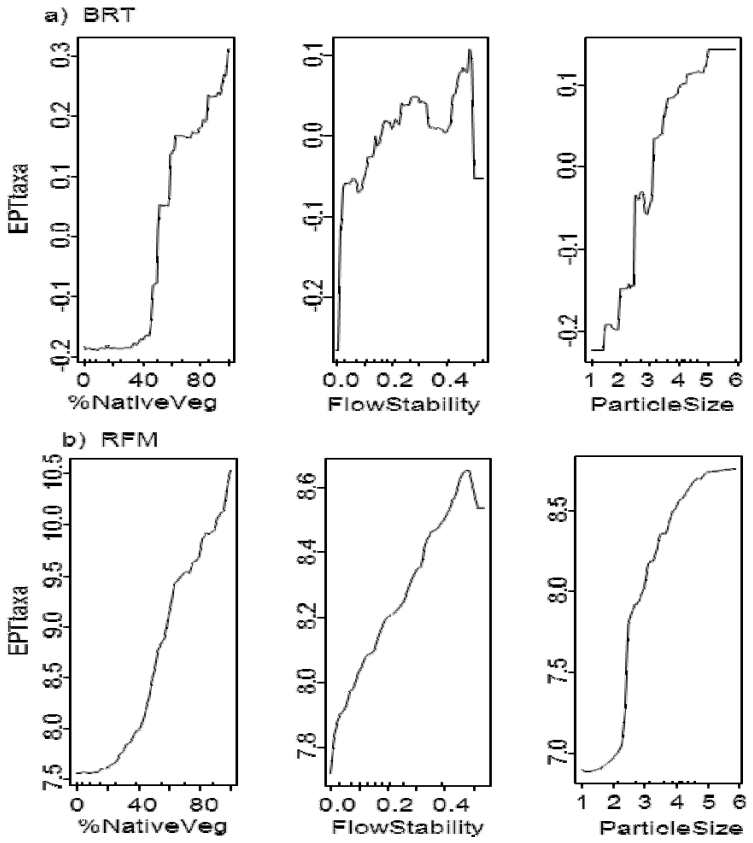


Figure 5. The relationships between EPT richness and environmental predictors ordered by importance for a) Boosted Regression Tree (BRT) and b) Random Forest (RF) models. Note the scales on the y-axes are not the same; the BRT scale provides the marginal contribution of each predictor to the mean value, whereas the RF model scale is the absolute value of the prediction when other variables are at the mean.

3.3. Model comparison summary

The BRT and RF models developed to predict MCI and EPT richness result in very similar output. The model diagnostics are similar, as are the order of importance in explanatory variables. However, the cross validation analyses indicated that the RF models were slightly biased (where bias is a consistent offset from a 1:1 relationship) and inconsistent (where this is a slope different to the 1:1 relationship). Both models suggest meaningful and logical relationships between predictor and response variables. The stochastic error associated with model approaches is likely to be greater than any difference in the predictive accuracy of models and for this reason the performance of the models is very similar. However, the bias and inconsistency of the RF models means they are likely to under estimate low values and overestimate high values (Table 6) and on this basis we suggest the BRT model is used to predict contemporary MCI.

Table 6. Comparison of current model performances of Random Forest (RF) and Boosted Regression Tree (BRT) models for Macroinvertebrate Community Index (MCI) and EPT richness. * Indicates slope significantly different from one; + indicates intercept significantly different from zero. The 95% confidence interval was estimated as 1.96 times the RMSD from the correlation with hold out data.

Model	% deviance explained	Predictive accuracy (internal model cross validation)	Correlation with hold out data (independent model cross validation)	NSE	95% CI	Bias	National mean
MCI RF	na	63.2	0.83	0.67	22.9	0.80*+	109.5
MCI BRT	63.9	63.7	0.82	0.68	22.7	0.91	112.0
EPT richness RF	na	56.9	0.81	0.63	5.5	0.13*+	9.5
EPT richness BRT	52.9	56.5	0.78	0.61	5.6	0.20	9.0

4. PREDICTING REFERENCE CONDITIONS

We examined three model approaches for predicting macroinvertebrate metrics:

1. One-step reset land use approach using BRT and RFM
2. Two-step offset land use approach using BRT (Clapcott *et al.* 2011)
3. Linear models with data grouped by REC and FENZ classes.

To validate and compare model approaches we used a selection of the working dataset identified as potential reference sites, according to a set of nominated rules based on land cover. Two land-use rule sets (Refset1 and Refset2) were investigated as outlined in Table 7. The 'relaxing' of land-use constraints doubled the number of sites in the reference site dataset and improved the geographical coverage of sites (Figure 6) without significantly changing the mean values of response variables, so we proceeded using Refset2.

Table 7. Description of land-use rules and summary data for two potential reference site groupings.

	Refset1	**Refset2**
Land-use rules		
Native vegetation	≥ 90%	≥ 85%
Pastoral heavy	0%	≤ 5%
Pastoral light	≤ 10%	≤ 15%
Urban	0%	0%
Surface water allocation	0%	0%
No. sites	27	63
No. REC CSOF classes	7	8
No. FENZ C100 classes	10	11
MCI mean (min, max)	130.4 (103.8, 155.9)	130.3 (103.8, 155.9)
EPT mean (min, max)	15.1 (7, 23)	14.1 (7, 23)

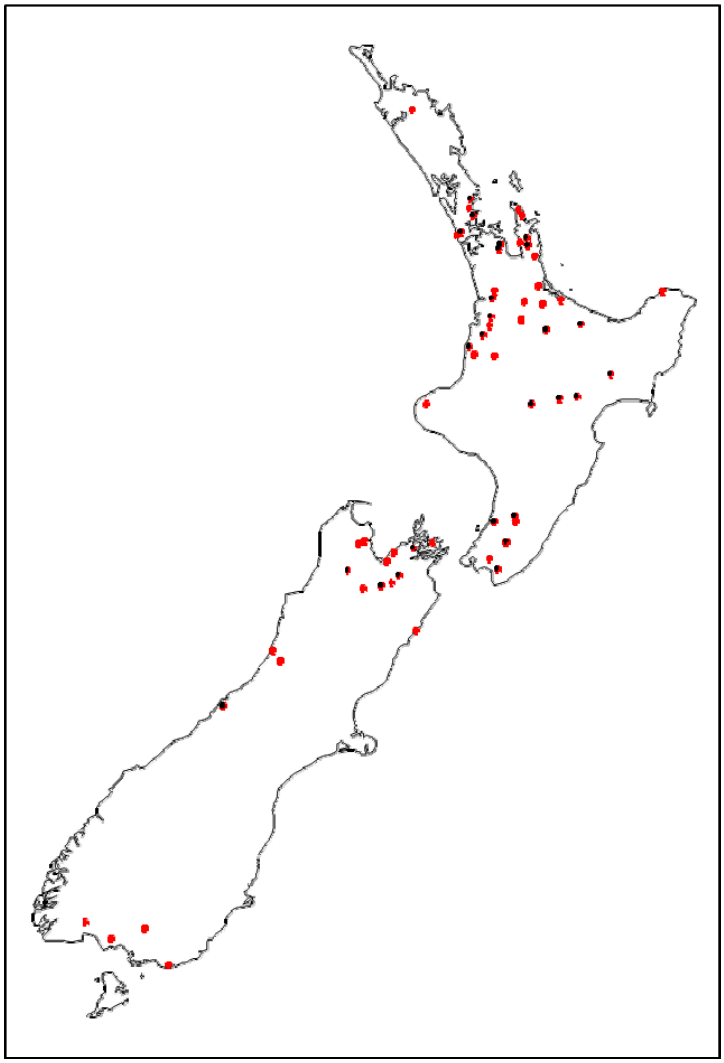


Figure 6. Distribution of land-use defined reference sites. Blue dots are Refset1 (N = 27) and red dots are Refset2 (N = 63).

4.1. Reset land use

The models developed in section 3 were employed here: the 23 variables describing land cover, water allocation pressure, and environmental descriptors were used to predict MCI and EPT richness metrics. However, in this instance the value of the land-use pressure predictors were reset to reference conditions at all sites, *i.e.* native vegetation set to 100%, and urban, pastoral land cover and surface water allocation were all set to zero. The models were then used to predict MCI and EPT richness values for all sites. Hence the model is used to predict metric values as a product of environmental variability in the absence of anthropogenic pressure.

4.1.1. Macroinvertebrate Community Index

The BRT model predicted a mean reference MCI value of 126.7 and the RF model 123.3 at the national scale. A comparison of mean MCI values for REC CSOF classes suggested BRT predictions were on average 2.5 MCI units higher than RF model predictions ($t_{(22)} = 5.31$, $p < 0.01$). Similarly, a comparison of mean MCI values for FENZ C20 classes suggested BRT predictions were on average 1.9 MCI units higher than RF model predictions ($t_{(20)} = 6.54$, $p < 0.01$).

We made an independent test of the performance of the models when predicting reference MCI values. First, a random selection of half of the Refset2 data (n = 30) were excluded from a model building dataset and the RF and BRT models were refitted. This second model was then used to predict MCI values for the 30 reference sites that were held out of the model building selection from Refset2. There was a strong relationship between predicted and observed MCI values for both the RF model (R = 0.75) and the BRT model (R = 0.65) (Figure 7). The relationships between predicted and observed MCI values suggested good model performance in terms of both consistency and bias for the RF model, with only a slight improvement observed for the BRT model. Both model approaches tended to underestimate MCI values at reference sites.

4.1.2. EPT richness

The BRT model predicted a mean reference EPT richness value of 12.8 and the RF model 12.9 at the national scale. A comparison of mean values for REC CSOF classes suggested there was no significant difference between BRT and RF model predictions ($t_{(22)} = 0.31$, $p = 0.76$). In comparison, the mean EPT richness values for FENZ classes suggested BRT reference predictions were on average 0.7 units lower than RF reference predictions ($t_{(20)} = 2.82$, $p = 0.01$).

We made independent tests for the performance of the models for predicting the reference EPT richness in the same manner as for the MCI models. Following refitting of the model excluding half of the Refset2 data, there was substantial difference in the correlation between observed and predicted EPT richness for both the RF model (R = 0.46) and the BRT model (R = 0.36) (Figure 7). Both model approaches resulted in poor predictive performance of reference values (NSE ≤ 0); the bias was low but the RMSD was high for both models indicating wide deviation from the 1:1 line.

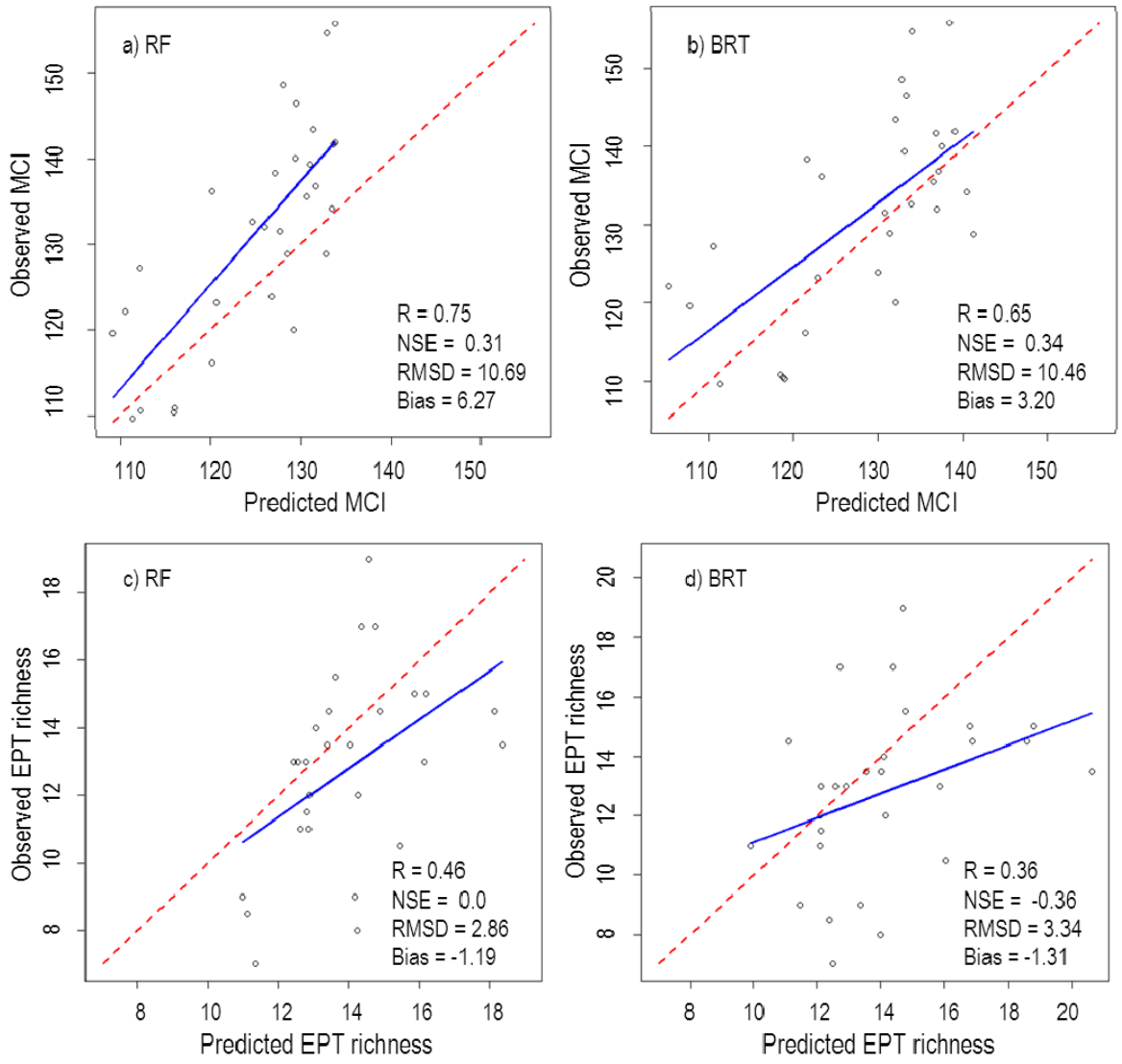


Figure 7. Correlations between observed values held out from half of the Refset2 dataset and predictions made for these sites for Macroinvertebrate Community Index (MCI) a) Random Forest (RF) and b) Boosted Regression Tree (BRT) models, and for EPT richness c) RF and d) BRT models.

4.2. Offset land use

This model uses a two-step approach but theoretically has the same aim as the previous model, to remove the effect of land use pressures and allow natural variability to explain variation in metric values. However, in a two-step approach the metric is first modelled as a response to current land use pressure (*e.g.* MCI in response to five variables — Step 1) then the output from this model is used as a fixed offset in a second model to explain the remaining deviance in metric data (Step 2). The Step 2 model is then applied to a starting value, in this case the mean value

from Refset2, resulting in a range of reference values as a product of natural variability alone (*i.e.* MCI in response to 18 environmental variables).

In contrast to the one-step model, the two-step model allows the user to define what level of land-use is acceptable at a reference site and this information is used to inform the starting reference value. In practise there is little difference between the approaches. A limitation of both methods is that neither allows for the factoring out of the collinearity between land cover and environmental variables, *e.g.* low native vegetation cover is associated with unstable flows, lower slopes and warmer temperatures.

The BRT offset model (land-use pressure plus environmental variability) for MCI had very good predictive accuracy (cross validation R^2 = 66.3%). The mean predicted reference MCI value was 132.7, and mean reference predictions for stream class were consistently greater those observed for the BRT reset model.

From the refitted model excluding half of the Refset2 data, there was a reasonable correlation between predicted reference values and observed values (R = 0.53, Figure 8). The BRT offset model was unbiased and consistent (slope and intercept not significantly different to zero and one). The BRT offset model did not perform as well as the BRT and RF land use reset models.

The BRT offset model for EPT richness had good prediction accuracy (cross validation R2 = 58.4%). The mean predicted reference EPT richness value was 14.1. However, there was a poor correlation between predicted reference values and observed values from the refitted model excluding half of the Refset2 data (R = 0.35, NSE = -3.1, Figure 8). The BRT offset model significantly overestimated EPT richness values and had poor model performance compared to both the BRT and RF land use reset models.

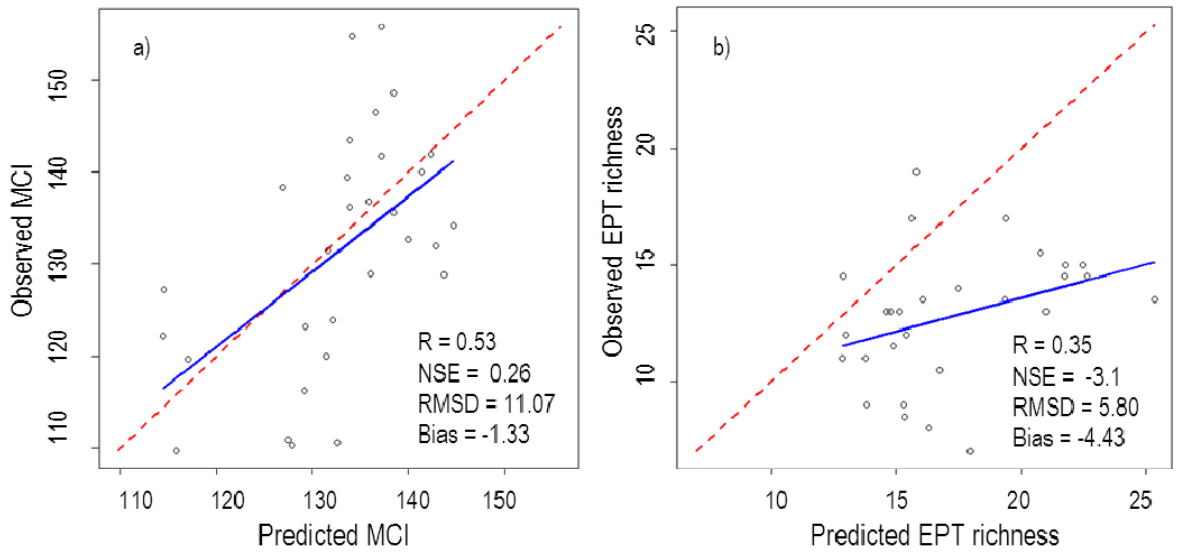


Figure 8. Correlations between measured observed values from Refset2 and predicted reference values from the Boosted Regression Tree (BRT) offset models for a) Macroinvertebrate Community Index (MCI) and b) EPT richness.

4.3. Reference predictions by Class

As an alternative to the use of continuous gradients of environmental variability to predict reference we used the approach of Dodds and Oakes (2004). This approach uses Analysis of Covariance (ANCOVA), which includes categorical and continuous variables in a linear regression model. ANCOVA tests for significant differences in the response metrics among sites, grouped by the categorical variable, while accounting for the variation due to the continuous variable(s). In our analysis the categorical variables were stream class (from REC or FENZ) and the continuous variables were land cover and water allocation pressure gradients.

When the categorical variable is significant, the intercept of the regression is the estimated value of the indicator in the absence of anthropogenic influence, or a reference value for each class represented in the model. The approach reduces the problem that land cover may be correlated with the environmental variables, because the relationship between land cover and the indices is defined for a group for which environment is considered homogeneous (*i.e.* a REC or FENZ class). This means the method may produce estimates of reference values that are more accurate. However, the method can only be applied to classes for which there is sufficient representation to define the regression relationships and does not allow estimates of reference values in non-represented classes.

We developed a Class model for MCI using both REC CSOF (n = 20 groups) and FENZ C20 (n = 7 groups) classifications. First we tested the response of MCI to the most explanatory land cover variable (*i.e*. native vegetation cover) in an ANCOVA

model with Class as a covariate. Then we tested the response of MCI to multiple land cover variables using a stepwise reduction procedure to only include those land cover variables that were informative in the ANCOVA model.

The linear model predicting hbMCI (MCI customised for hard-bottomed streams, in fact used throughout this study) as a product of native vegetation cover alone showed no interactions among REC CSOF classes (*i.e.* all classes showed similar slopes for the relationships with native vegetation) (Figure 9). The ANCOVA model adjusted R^2 was 50% ($p < 0.001$). All land use pressures except surface water allocation were retained in the multi-predictor model and the interaction between class and urban cover was also retained, but other interactions were dropped by model selection. The multi-predictor ANCOVA model had an adjusted R^2 of 56.4% ($p < 0.001$).

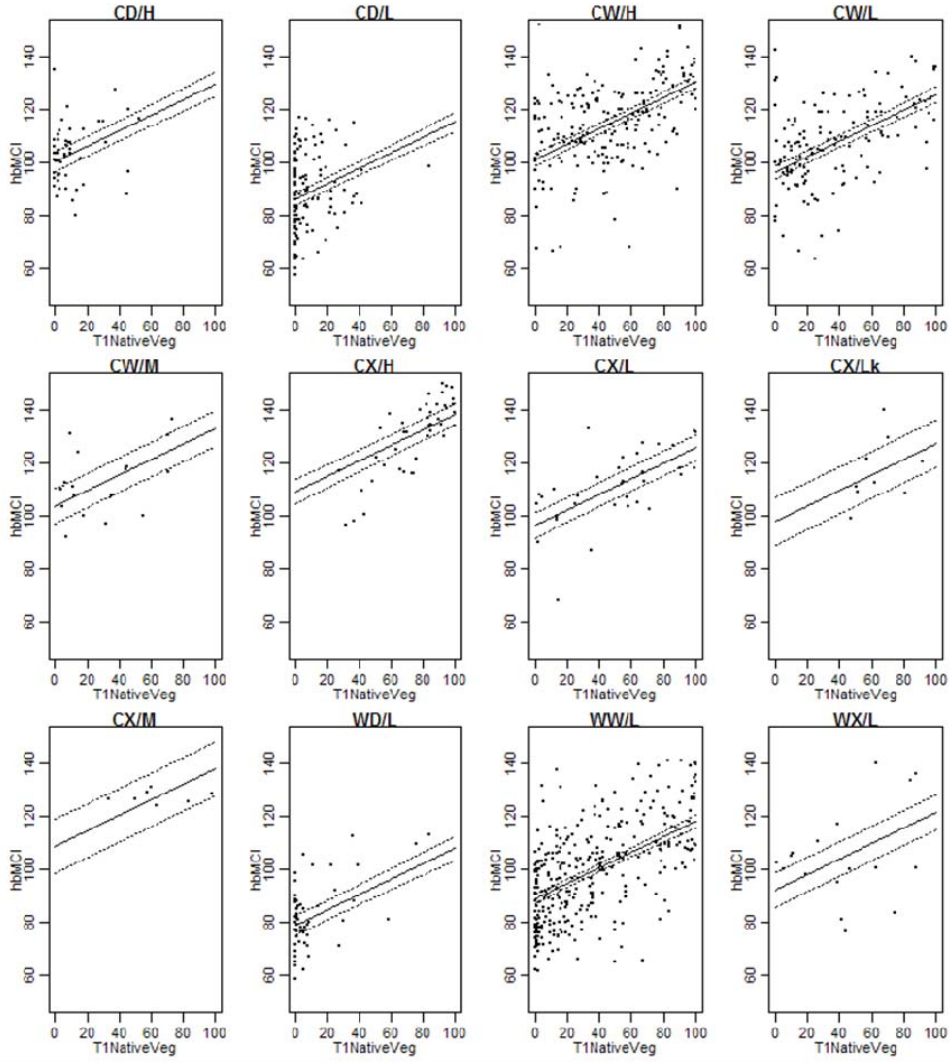


Figure 9. Relationship between Macroinvertebrate Community Index (MCI) and native vegetation cover by River Ecosystem Classification Climate/Source-of-Flow (REC CSOF) classification showing 95% confidence intervals of the mean. Twelve out of 20 groups are shown, which illustrate where there are sufficient data to inform meaningful predictions of reference values.

The linear model predicting MCI as a product of native vegetation cover showed no interactions among FENZ C20 classes (Figure 10). The ANCOVA model adjusted R^2 was 45% ($p < 0.001$). Native vegetation, pastoral heavy and urban covers were retained in the multi-predictor model, but interactions were dropped by model selection. The multi-predictor ANCOVA model adjusted R^2 was 52% ($p < 0.001$).

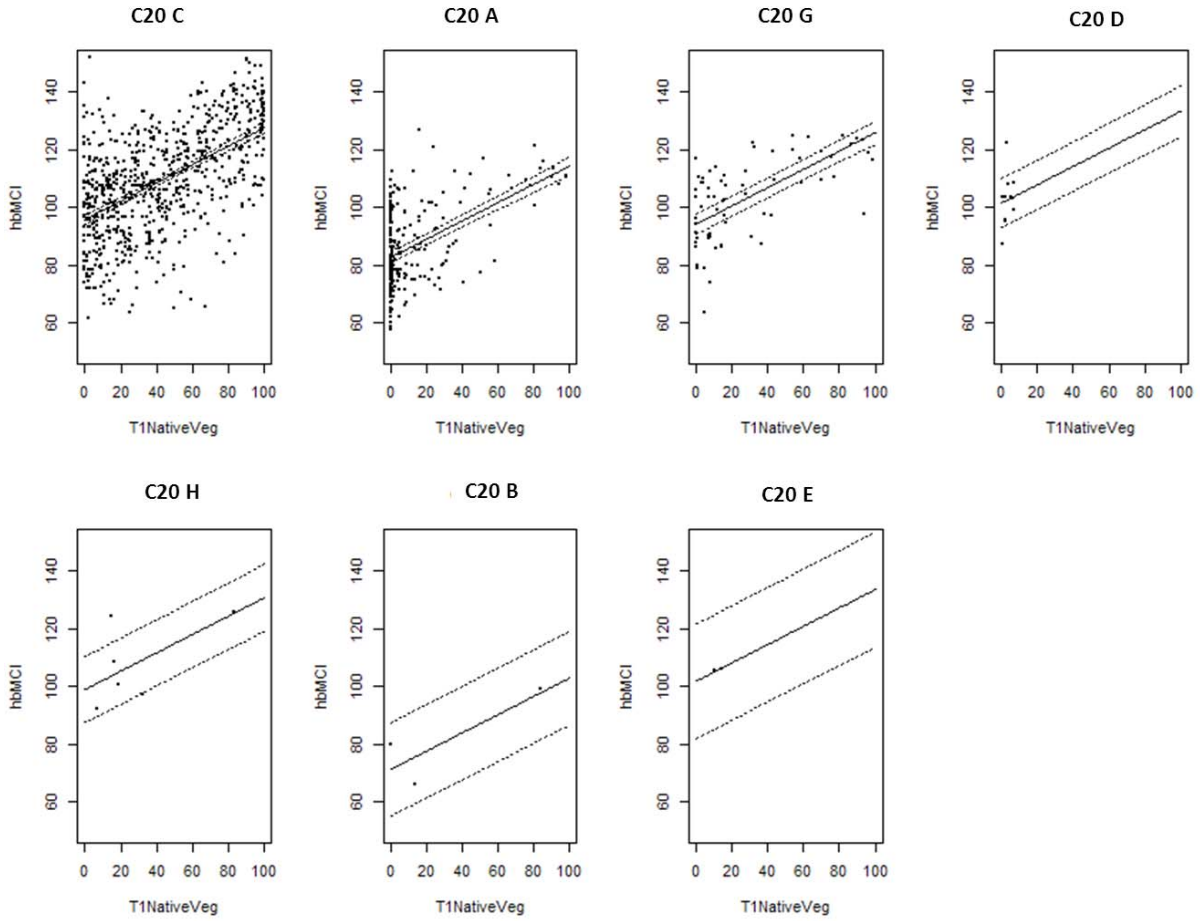


Figure 10. Relationship between Macroinvertebrate Community Index (MCI) and native vegetation cover by Freshwater Ecosystem of NZ (FENZ) C20 classification showing 95% confidence intervals of the mean. Seven out of seven groups are illustrated but only three groups have sufficient data to inform useful predictions of reference.

We made an independent test of the performance of the ANCOVA models when predicting reference MCI and EPT values. First, a random selection of half of the Refset2 data (n = 30) were excluded from a model building dataset and the models were refitted. We then compared the intercepts of the fitted models to the 30 reference sites that were held out of Refset2. The correlations between measured reference values and predicted reference values from the multi-predictor ANCOVA by REC CSOF (R = 0.74) and for the FEWNZ C20 (R = 0.35) clearly demonstrate the categorical nature of the classification analysis; all sites within a class are assigned a

similar reference value (Figure 11). However, the model performance statistics (at least for the REC CSOF model) were comparable to both RF and BRT model approaches (Figure 7, Figure 8), indicating that predictions from the ANCOVA model are as accurate as those of these other two model approaches for classes represented in the existing dataset.

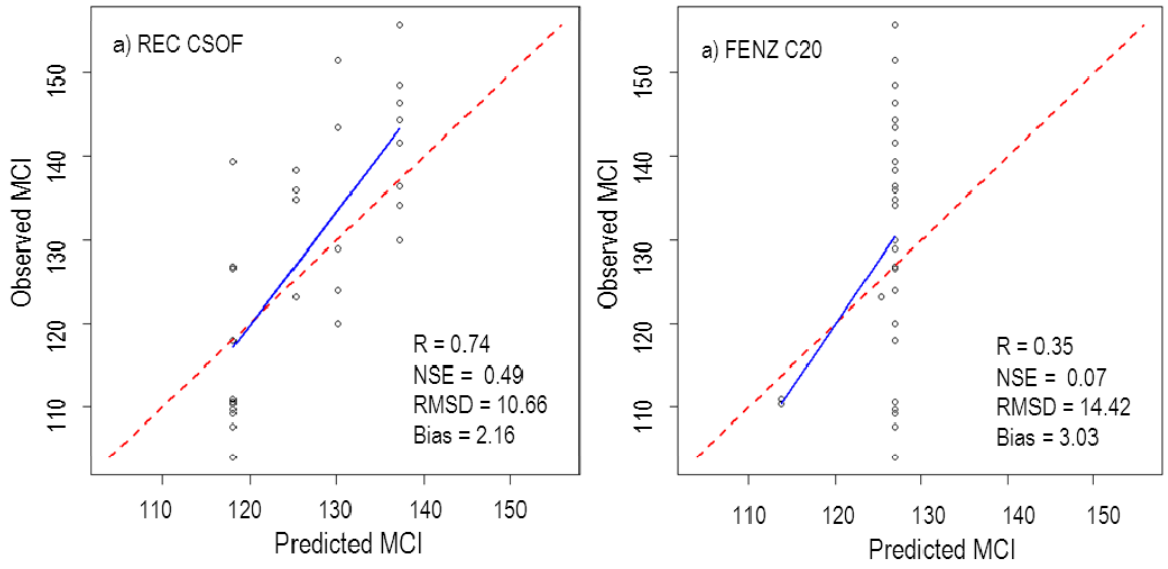


Figure 11. Correlations between measured observed values from Refset2 and predicted reference values from the multi-predictor models for Macroinvertebrate Community Index (MCI) by a) River Ecosystem Classification Climate/Source-of-Flow and b) Freshwater Ecosystem of NZ (FENZ) C20 groups.

4.4. Model comparison summary

Comparison of reference models based on predictive performance suggests little difference in the RF, BRT and ANCOVA model based on REC classes for MCI (Table 8). All MCI models performed well. However, model performance was poorer for EPT richness.

Table 8. Comparison of reference model performances for Macroinvertebrate Community Index (MCI) and EPT richness. * Indicates slope significantly different from one; + indicates intercept significantly different from zero. The 95% confidence interval was estimated as 1.96 times the RMSD from the correlation with hold out data.

Model	% deviance explained	Predictive accuracy (internal model cross validation)	Correlation with hold out data (independent model cross validation)	NSE	95% CI	Bias	National mean
MCI							
Reset land use BRT	63.9	63.7	0.65	0.34	20.5	3.21	126.7
Reset land use RF	na	63.2	0.75	0.31	20.9	6.27	123.3
Offset land use BRT	66.0	66.3	0.53	0.26	21.7	-1.33	132.7
Class by REC	56.4	na	0.74	0.49	20.9	2.16	na
Class by FENZ	52.0	na	0.35	0.07	28.3	3.03	na
EPT richness							
Reset land use BRT	52.9	56.5	0.36	-0.36	6.5	-1.31*+	12.8
Reset land use RF	na	56.9	0.46	0.0	5.6	-1.19	12.9
Offset land use BRT	58.3	58.4	0.35	-0.31	11.4	-4.43*+	14.1

The main difference between RF and BRT and ANCOVA models for predicting reference site metrics is the ability of the former two to predict to stream classes not represented in the working data set. As noted in section 2.2 the working data set is unevenly distributed amongst stream classes. However, sites are described by environmental predictors that cover a wide range of conditions (Figure 12). As such, we can have confidence in the prediction to other stream classes, if they fall within the range of environmental variability observed in the working data set. In contrast, whilst the ANCOVA models reported no interaction among classes (they all had a similar slope in response to land cover pressure), there is no way of predicting the intercept or reference value for classes not represented it the working data set.

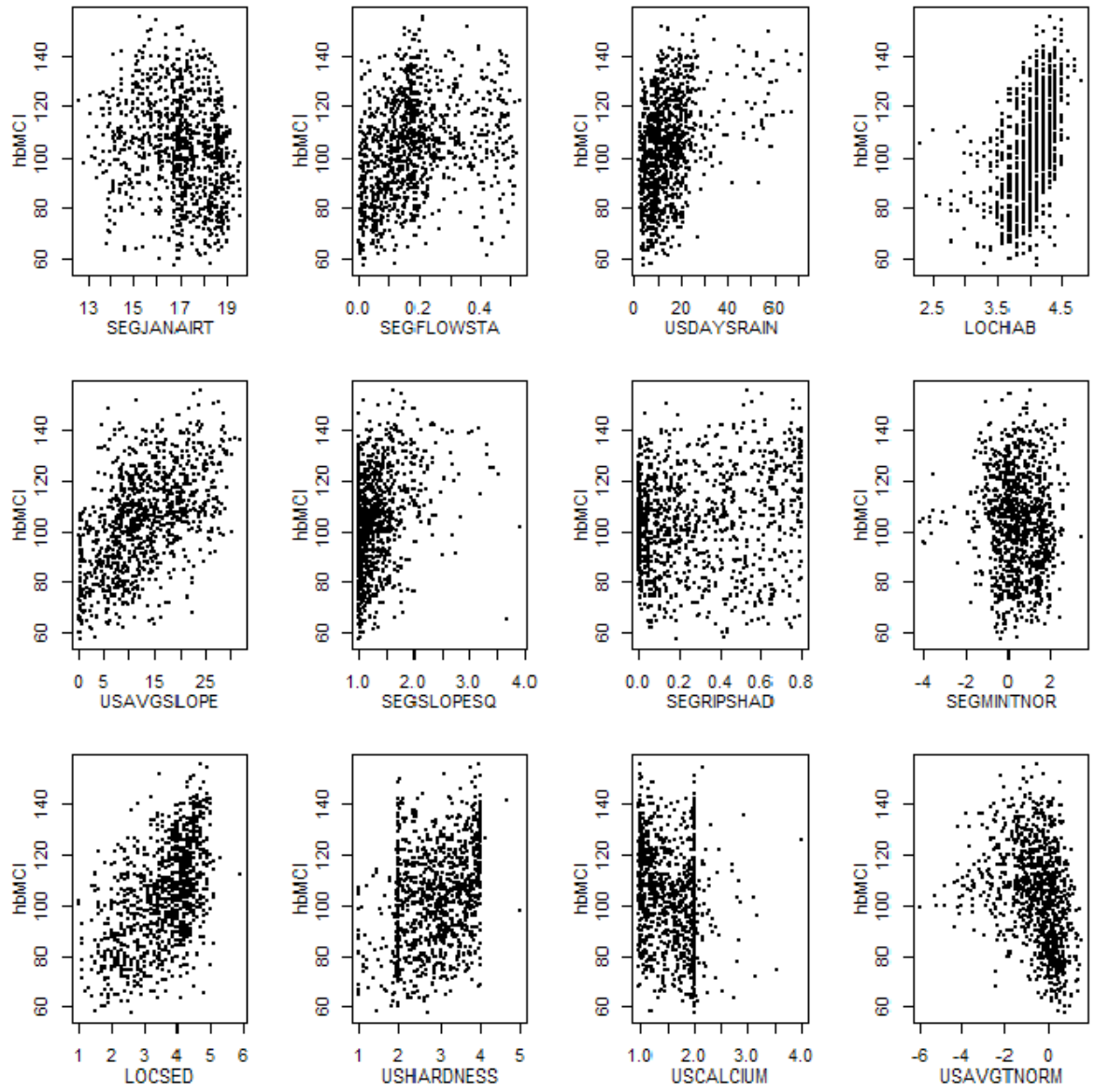


Figure 12. Scatterplots of the distribution of sites within the working data set across gradients of environmental descriptors.

5. MODEL RECOMMENDATION

The 'best' predictive models will explain the highest percentage of deviance in the metric data and have good predictive performance for independent values (*i.e.* hold-out data). The predictive performance (*i.e.* R and NSE for independent data) for the BRT and RF models was not significantly different for predicting contemporary MCI and EPT richness values. However, BRT models were unbiased and consistent whereas tests of the independent predictions made using the RF models indicated they were biased and inconsistent. In practice, the inherent stochastic nature of these models means that any model error is likely to be greater than the difference in predictive output.

For predicting **contemporary** MCI and EPT the BRT model is best because it is consistent and unbiased. We acknowledge that this is probably a very marginal difference to the RF model because of the limited dataset.

NSE is the best overall indicator of predictive performance because it combines uncertainty, bias and inconsistency in a single measure of performance. On that basis the reset land use BRT model is the best for predicting **reference** MCI, while EPT predictions were all poor — the NSE value of zero or less indicates that predicting all sites to have the mean value would be a better model. This is based on our best assessment of performance of these models for predicting reference values for new sites. The test itself is uncertain as it is based on few values. Until there are more data we are not able to improve this result.

An alternative test of model performance could be to perform a 'leave one out' cross validation (Snelder *et al.* 2011), whereby the models are replicated numerous times excluding a single site (in contrast to the single 'hold out' cross validation we performed). This alternative approach may provide a 'tighter' estimate of confidence intervals for national predictions, as the RMSD is derived from the same models that are used to make national estimates. As currently provided, the confidence intervals may be viewed as conservative; they may overestimate uncertainty. However, the internal BRT cross validation statistics report similar error to that observed by our external cross validations, indicating that model replication may not necessarily reduce uncertainty.

The ultimate aim of having predictions of both contemporary and reference status is the calculation of O/E (observed/expected) values to indicate the degree to which current conditions have deviated from natural conditions. Applying a one-tailed test based on our current conservative assessment of model performance, we can be 95% confident that predicted current conditions are poorer than predicted reference conditions when values differ by 36.3 units for MCI (6110 or 1.1% of stream segments nationally) and 12.1 units for EPT richness (143 or 0.3% of stream segments nationally). We can be 90% confident that predicted current conditions are poorer than

predicted reference conditions when values differ by 28.3 units for MCI (81,629 or 14.2% of stream segments nationally) and 7.9 units for EPT richness (17,168 or 3% of stream segments nationally).

National predictions can be summarised by stream classifications to inform expected current and reference values for any modelled metric. An alternative reference condition at the class level can be obtained by averaging current metric predictions for only those sites that fit a predetermined set of land cover rules, *e.g.* mean value at sites with > 90% native vegetation cover and 0% other land-use pressures (as shown in the tables in appendix 1). This alternative reference benchmark would reflect best obtainable current condition and be based on a model with strong predictive performance. Class summaries for all BRT model predictions are in Appendices.

6. OTHER METRICS

The following sections summarise BRT model statistics for two further invertebrate metrics: Taxa richness and %EPT richness. We did not independently test model performance (*i.e.* external cross validation) for these two metrics, nor base the modelling method selection on them.These two facts account for the simplicity of this section. Class summaries are provided in Appendices.

6.1. Taxa richness

The BRT model for Taxa richness had an internal cross validation R^2 of 44.9% which indicates acceptable predictive performance. In general, Taxa richness increased with increasing native vegetation cover, lower catchment slope, lower phosphorous geology upstream, and lower segment slope (Figure 13). The BRT model predicted a mean current Taxa richness value of 38.5 and a mean reference Taxa richness value of 45.7.

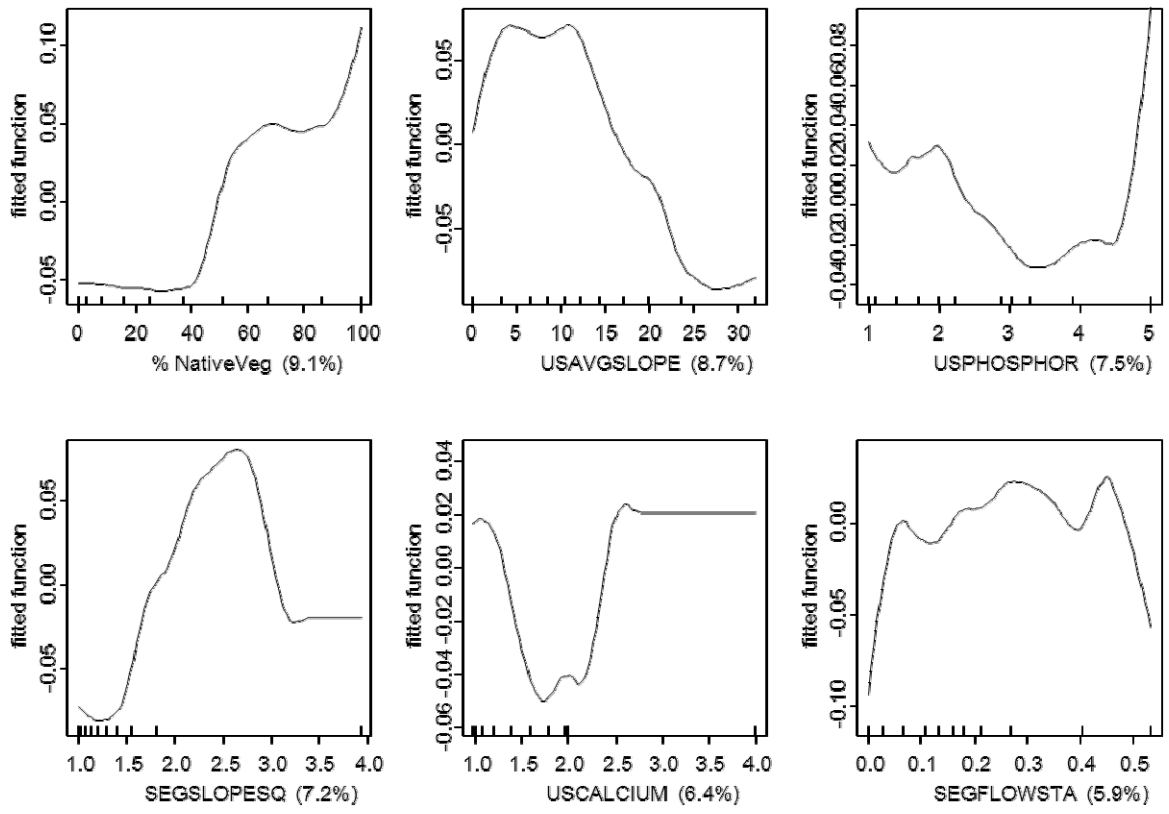


Figure 13. The relationships between Taxa richness and the top six model predictors ordered by relative contribution to predictive performance.

6.2. %EPT richness

The BRT model for %EPT richness had an internal cross validation R^2 of 60.2% which suggested very good predictive performance. In general, %EPT richness increased with greater substrate size, decreasing heavy pasture cover, decreasing urban cover, and lower summer temperatures (Figure 13). The BRT model predicted a mean current %EPT richness value of 47.5 and a mean reference %EPT richness value of 54.4.

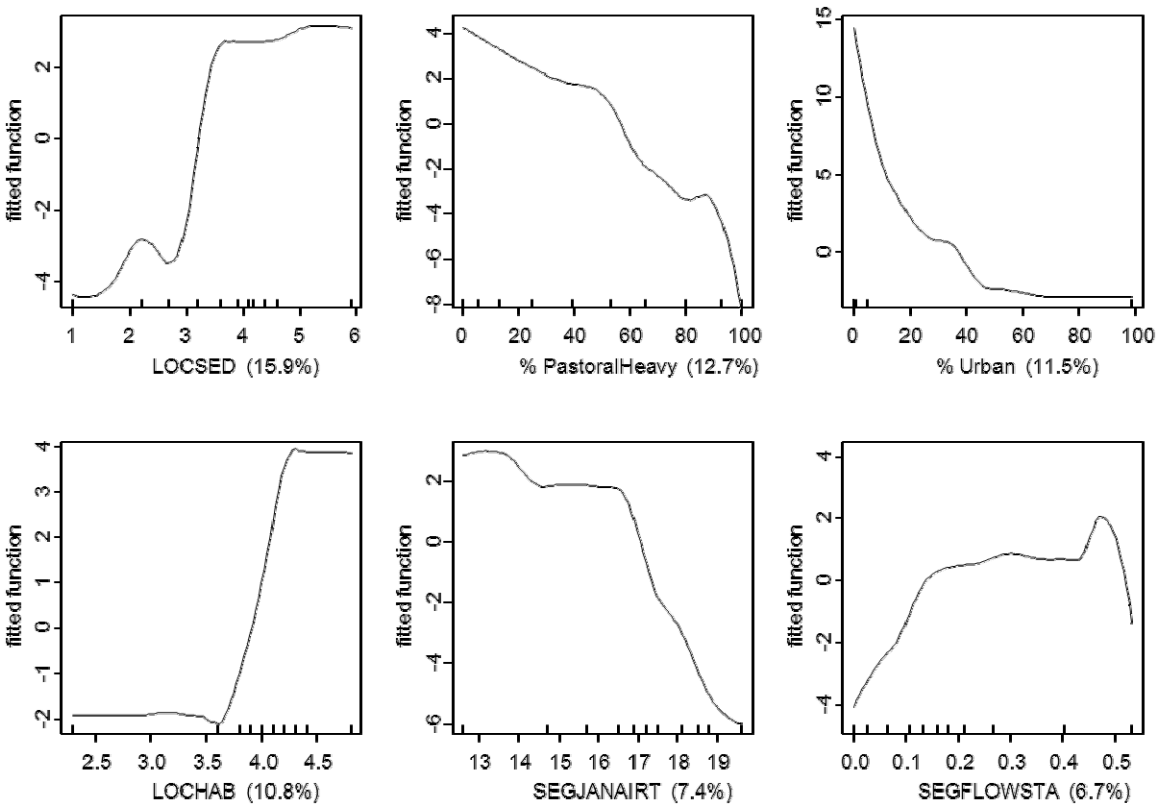


Figure 14. The relationships between % EPT richness and the top six model predictors ordered by relative contribution to predictive performance.

7. ACKNOWLEDGMENTS

The code used to develop predictive models has been greatly progressed by the previous contributions of several people; in particular we thank John Leathwick and Doug Booker.

8. REFERENCES

Breiman L, Friedman JH, Olshen R, Stone CJ 1984. Classification and Regression Trees. Wadsworth, Belmont, California.

Breiman L 2001. Random forests. Machine Learning 45: 5-32.

Clapcott J, Young R, Goodwin E, Leathwick J, Kelly D 2011. Relationships between multiple land-use pressures and individual and combined indicators of stream ecological integrity. In DOC Research and Development Series. Department of Conservation, Wellington. pp 57.

Clapcott JE, Goodwin EO 2010. The response of indicators of river integrity to multiple land-use stressors: Further development towards a multi-metric index of ecological integrity. Prepared for Department of Conservation. Cawthron Report No. 1859. 21 p.

Cutler DR, Edwards TC, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ 2007. Random forests for classification in ecology. Ecology 88: 2783-2792.

Dodds WK, Oakes RM 2004. A technique for establishing reference nutrient concentrations across watersheds affected by humans. Limnology and Oceanography: Methods 2: 333-341.

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

Friedman JH, Hastie T, Tibshirani R 2000. Additive logistic regression: a statistical view of boosting. Annals of Statistics 28: 337-407.

Friedman JH 2001. Greedy function approximation: a gradient boosting machine. Annals of Statistics 29: 1189-1232.

Friedman JH, Meulman JJ 2003. Multiple additive regression trees with application in epidemiology. Statistics in Medicine 22: 1365-1381.

Hastie T, Tibshirani R, Friedman J 2009. Elements of statistical learning: Data mining, inference and prediction. Springer, New York. 745 p.

Leathwick JR, West D, Gerbeaux P, D Kelly H, Robertson H, Brown D Chadderton WL, Ausseil A-G 2010.Freshwater Ecosystems of New Zealand (FENZ) Geodatabase. Version one - August 2010 User Guide. Deparment of Conservation.

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

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

Olden JD, Lawler JJ, Poff NL 2008. Machine learning without tears: A primer for ecologists. The Quarterly Review of Biology 83 (2): 171-193.

Pineiro G, Perelman S, Guerschman J, Paruelo J 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecological Modelling 216: 316-322.

Snelder TH, Lamouroux N, Pella H 2011. Empirical modelling of large scale patterns in river bed surface grain size. Geomorphology 127 (3-4): 189-197.

Unwin M, Snelder T, Booker D, Ballantine D, Lessard J 2010. Modelling water quality in New Zealand rivers from catchment-scale physical, hydrological and land-cover descriptors using random forest models. Prepared for NIWA Client Report: CHC2010-037.

9. APPENDICES

Appendix 1. Current (O = observed) and reference (E = expected) macroinvertebrate metric values predicted by Boosted Regression Tree (BRT) models and summarised by stream classes. Expected values are based on predictions made using the reset land cover BRT model. An alternative reference value is calculated as the mean of Observed values at sites with > 90% native vegetation and 0% other land-use pressures.

Table A1.1. Summary of mean values for River Environment Classification Climate/Source-of-Flow (REC CSOF) groups of current (MCIO) and reference (MCIE) values for Macroinvertebrate Community Index (MCI). *Sites with > 90% native vegetation and 0% other land-use pressures.

	N	MCIO (mean)	N	MCIO (mean of reference sites*)	N	MCIE (mean)
CD/H	50190	106.4	366	128.8	50190	124.7
CD/L	62303	92.7	206	122.6	62303	119.9
CD/Lk	2935	94.2	1	117	2935	115.6
CD/M	5927	114.0	8	127	5927	127.7
CW/GM	145	109.9	0	na	145	127.3
CW/H	89673	120.0	24737	134.1	89673	130.0
CW/L	45990	113.0	11068	129.8	45990	126.3
CW/Lk	5113	104.9	41	123.5	5113	120.6
CW/M	45149	118.2	2949	134.8	45149	129.8
CX/GM	15391	119.3	21	138	15391	135.1
CX/H	51938	135.6	26791	139.7	51938	138.5
CX/L	17537	127.1	8438	136.3	17537	133.3
CX/Lk	4276	120.9	123	129.9	4276	130.7
CX/M	46414	127.4	3815	138.8	46414	137.6
WD/H	1	124.7	0	na	1	127.0
WD/L	30508	86.9	289	113.6	30508	114.1
WD/Lk	582	92.5	0	na	582	112.5
WW/H	1845	124.5	990	132.7	1845	129.8
WW/L	94270	100.5	8509	127.2	94270	118.0
WW/Lk	807	94.7	11	120.6	807	112.9
WX/H	1129	122.8	408	133.9	1129	129.5
WX/L	4024	111.9	718	133.9	4024	126.7
WX/Lk	9	120.5	0	na	9	128.8
National	576156	112.0	89489	134.9	576156	126.7

Table A1.2. Summary of mean values for River Environment Classification Climate/Source-of-Flow (REC CSOF) groups of current (EPTO) and reference (EPTE) values for EPT richness. *Sites with > 90% native vegetation and 0% other land-use pressures.

	N	EPTO (mean)	N	EPTO (mean of reference sites*)	N	EPTE (mean)
CD/H	50190	8.6	366	12.6	50190	13.4
CD/L	62303	6.2	206	10.9	62303	11.4
CD/Lk	2935	5.3	1	8.4	2935	9.3
CD/M	5927	9.0	8	13.5	5927	13.0
CW/GM	145	7.0	0	0	145	12.0
CW/H	89673	11.3	24737	14.2	89673	14.5
CW/L	45990	10.3	11068	14.0	45990	13.7
CW/Lk	5113	7.7	41	13.0	5113	12.1
CW/M	45149	9.2	2949	13.5	45149	13.1
CX/GM	15391	7.6	21	10.4	15391	11.7
CX/H	51938	12.0	26791	13.4	51938	13.7
CX/L	17537	11.7	8438	13.5	17537	13.2
CX/Lk	4276	9.1	123	11.9	4276	12.5
CX/M	46414	9.3	3815	13.0	46414	13.1
WD/H	1	14.2	0	0	1	15.6
WD/L	30508	4.2	289	9.3	30508	8.8
WD/Lk	582	4.8	0	0	582	8.4
WW/H	1845	13.5	990	14.9	1845	15.1
WW/L	94270	7.7	8509	13.7	94270	12.0
WW/Lk	807	5.5	11	12.8	807	9.3
WX/H	1129	13.6	408	16.1	1129	15.6
WX/L	4024	10.5	718	14.6	4024	14.3
WX/Lk	9	10.7	2	12.0	9	13.2
National	576156	9.0	89489	13.8	576156	12.8

Table A1.3. Summary of mean values for River Environment Classification Climate/Source-of-Flow (REC CSOF) groups of current (nTaxaO) and reference (nTaxaE) values for Taxa richness. *Sites with > 90% native vegetation and 0% other land-use pressures.

	N	nTaxaO (mean)	N	nTaxaO (mean of reference sites*)	N	nTaxaE (mean)
CD/H	50190	38.2	366	44.2	50190	47.7
CD/L	62303	36.1	206	42.0	62303	43.9
CD/Lk	2935	32.8	1	40.7	2935	41.5
CD/M	5927	34.2	8	42.5	5927	44.3
CW/GM	145	25.7	0	0	145	41.2
CW/H	89673	43.2	24737	46.7	89673	49.7
CW/L	45990	44.2	11068	48.6	45990	49.7
CW/Lk	5113	35.2	41	45.2	5113	46.2
CW/M	45149	33.1	2949	42.4	45149	43.5
CX/GM	15391	26.3	21	30.3	15391	37.0
CX/H	51938	37.0	26791	39.6	51938	40.9
CX/L	17537	38.7	8438	40.2	17537	40.6
CX/Lk	4276	32.0	123	38.1	4276	41.2
CX/M	46414	30	3815	38.4	46414	38.7
WD/H	1	54.2	0	0	1	60.4
WD/L	30508	38.6	289	42.7	30508	45.2
WD/Lk	582	34.7	0	0	582	41.5
WW/H	1845	49.4	990	50.8	1845	52.3
WW/L	94270	43.0	8509	50.5	94270	49.8
WW/Lk	807	34.4	11	53.2	807	41.0
WX/H	1129	48.5	408	52.2	1129	52.7
WX/L	4024	44.3	718	47.0	4024	49.4
WX/Lk	9	40.5	2	41.1	9	43.7
National	576156	38.5	89489	44.1	576156	45.7

Table A1.4. Summary of mean values for River Environment Classification Climate/Source-of-Flow (REC CSOF) groups of current (%EPTO) and reference (%EPTE) values for %EPT richness. *Sites with > 90% native vegetation and 0% other land-use pressures.

	N	%EPTO (mean)	N	%EPTO (mean of reference sites*)	N	%EPTE (mean)
CD/H	50190	46.7	366	59.5	50190	54.6
CD/L	62303	35.8	206	51.1	62303	46.2
CD/Lk	2935	38.1	1	42.3	2935	42.8
CD/M	5927	53.7	8	60.9	5927	60.2
CW/GM	145	52.5	0	0	145	59.5
CW/H	89673	53.2	24737	61.2	89673	58.2
CW/L	45990	47.3	11068	57.4	45990	53.7
CW/Lk	5113	45.5	41	53.2	5113	50.1
CW/M	45149	56.4	2949	63.7	45149	61.8
CX/GM	15391	57.4	21	65.4	15391	63.3
CX/H	51938	63.2	26791	65.3	51938	64.6
CX/L	17537	56.1	8438	61.8	17537	58.6
CX/Lk	4276	56.2	123	58.3	4276	59.7
CX/M	46414	61.1	3815	65.7	46414	65.5
WD/H	1	49.3	0	0	1	49.9
WD/L	30508	23.2	289	39.6	30508	36.5
WD/Lk	582	30.1	0	0	582	36.0
WW/H	1845	55.2	990	59.9	1845	57.8
WW/L	94270	35.2	8509	55.2	94270	45.3
WW/Lk	807	33.7	11	45.6	807	40.4
WX/H	1129	54.9	408	60.9	1129	58.9
WX/L	4024	46.2	718	60.7	4024	54.7
WX/Lk	9	46.5	2	50.2	9	49.0
National	576156	47.5	89489	61.6	576156	54.4

Table A1.5. Summary of mean values for Freshwater Ecosystems of New Zealand (FENZ) C20 groups of current (MCIO) and reference (MCIE) values for Macroinvertebrate Community Index (MCI). *Sites with > 90% native vegetation and 0% other land-use pressures.

	N	MCIO (mean)	N	MCIO (mean of reference sites*)	N	MCIE (mean)
A	109791	88.9	1121	120.1	109791	115.4
B	1817	87.2	76	111.8	1817	110.1
C	250837	117.1	63479	135.0	250837	128.6
D	22874	105.5	103	132.3	22874	123.6
E	1083	107.0	0	Na	1083	124.2
F	659	94.1	0	Na	659	116.1
G	57478	111.4	9354	129.2	57478	125.3
H	45621	121.8	4811	135.6	45621	132.0
I	3095	125.9	133	136.5	3095	134.6
J	27968	135.7	9318	141.4	27968	140.2
K	599	119.1	0	Na	599	135.6
L	1294	119.2	0	Na	1294	135.8
M	141	112.9	0	Na	141	129.7
N	23335	120.1	608	132.8	23335	133.1
O	7481	128.4	315	138.6	7481	139.3
P	5944	119.9	25	134.5	5944	135.1
Q	2476	116.8	6	127.8	2476	133.4
R	64	119.5	1	131.8	64	135.6
S	3118	118.2	3	134.9	3118	135.2
T	1546	117.8	0	na	1546	134.8
National	576156	112.0	89489	134.9	576156	126.7

Table A1.6. Summary of mean values for Freshwater Ecosystems of New Zealand (FENZ) C20 groups of current (EPTO) and reference (EPTE) values for EPT richness. *Sites with > 90% native vegetation and 0% other land-use pressures.

	N	EPTO (mean)	N	EPTO (mean of reference sites*)	N	EPTE (mean)
A	109791	4.8	1121	11.3	109791	9.9
B	1817	2.5	76	4.6	1817	4.4
C	250837	10.6	63479	14.0	250837	13.9
D	22874	7.7	103	10.6	22874	11.8
E	1083	6.8	NA	0	1083	11.8
F	659	4.6	NA	0	659	7.9
G	57478	10.4	9354	14.7	57478	14.2
H	45621	9.5	4811	13.1	45621	13.3
I	3095	7.8	133	11.4	3095	11.7
J	27968	10.2	9318	12.2	27968	12.6
K	599	7.3	NA	0	599	11.7
L	1294	7.4	NA	0	1294	11.8
M	141	6.4	NA	0	141	10.6
N	23335	9.2	608	13.1	23335	13.3
O	7481	9.1	315	12.9	7481	12.6
P	5944	8.7	25	12.0	5944	12.9
Q	2476	8.8	6	14.1	2476	13.3
R	64	8.2	1	12.8	64	11.7
S	3118	7.7	3	10.3	3118	11.2
T	1546	8.6	NA	0	1546	11.8
National	576156	9.0	89489	13.8	576156	12.8

Table A1.7. Summary of mean values for Freshwater Ecosystems of New Zealand (FENZ) C20 groups of current (nTaxaO) and reference (nTaxaE) values for Taxa richness. *Sites with > 90% native vegetation and 0% other land-use pressures.

	N	nTaxaO (mean)	N	nTaxaO (mean of reference sites*)	N	nTaxaE (mean)
A	109791	36.7	1121	45.5	109791	44.0
B	1817	30.2	76	26.0	1817	31.6
C	250837	42.9	63479	45.4	250837	49.1
D	22874	34.9	103	37.2	22874	44.4
E	1083	27.1	NA	0	1083	41.4
F	659	29.1	NA	0	659	36.7
G	57478	41.0	9354	47.9	57478	48.2
H	45621	33.0	4811	40.4	45621	42.3
I	3095	25.2	133	32.3	3095	36.7
J	27968	30.4	9318	33.8	27968	35.8
K	599	24.9	NA	0	599	36.4
L	1294	25.4	NA	0	1294	36.3
M	141	24.3	NA	0	141	37.8
N	23335	31.6	608	40.2	23335	41.3
O	7481	29.4	315	36.9	7481	36.6
P	5944	30.1	25	36.6	5944	40.1
Q	2476	30.2	6	46.5	2476	41.7
R	64	27.6	1	34.4	64	36.4
S	3118	27.2	3	32.3	3118	36.3
T	1546	30.4	NA	0	1546	38.7
National	576156	38.5	89489	44.1	576156	45.7

Table A1.8. Summary of mean values for Freshwater Ecosystems of New Zealand (FENZ) C20 groups of current (%EPTO) and reference (%EPTE) values for %EPT richness. *Sites with > 90% native vegetation and 0% other land-use pressures.

	N	%EPTO (mean)	N	%EPTO (mean of reference sites*)	N	%EPTE (mean)
A	109791	27.6	1121	43.4	109791	39.1
B	1817	19.4	76	34.6	1817	30.1
C	250837	50.2	63479	61.3	250837	56.4
D	22874	45.9	103	60.0	22874	52.6
E	1083	50.1	NA	0	1083	55.3
F	659	34.9	NA	0	659	39.9
G	57478	50.2	9354	59.1	57478	56.6
H	45621	58.3	4811	64.4	45621	63.1
I	3095	59.6	133	64.4	3095	62.9
J	27968	64.5	9318	66.9	27968	66.5
K	599	57.5	NA	0	599	63.9
L	1294	57.9	NA	0	1294	64.1
M	141	52.9	NA	0	141	59.3
N	23335	57.8	608	63.0	23335	63.4
O	7481	61.2	315	65.6	7481	65.9
P	5944	57.7	25	64.4	5944	63.6
Q	2476	57.1	6	61.0	2476	63.2
R	64	57.7	1	64.5	64	63.6
S	3118	56.7	3	63.7	3118	62.8
T	1546	56.1	NA	0	1546	62.1
National	576156	47.5	89489	61.6	576156	54.4

Appendix 2. Box plots (median, 25th and 75th percentiles, 95th percentiles and outliers showing minimum and maximum values) of reference predictions for Macroinvertebrate Community Index (MCI) by classification group. Plots also show number of sites for each group.

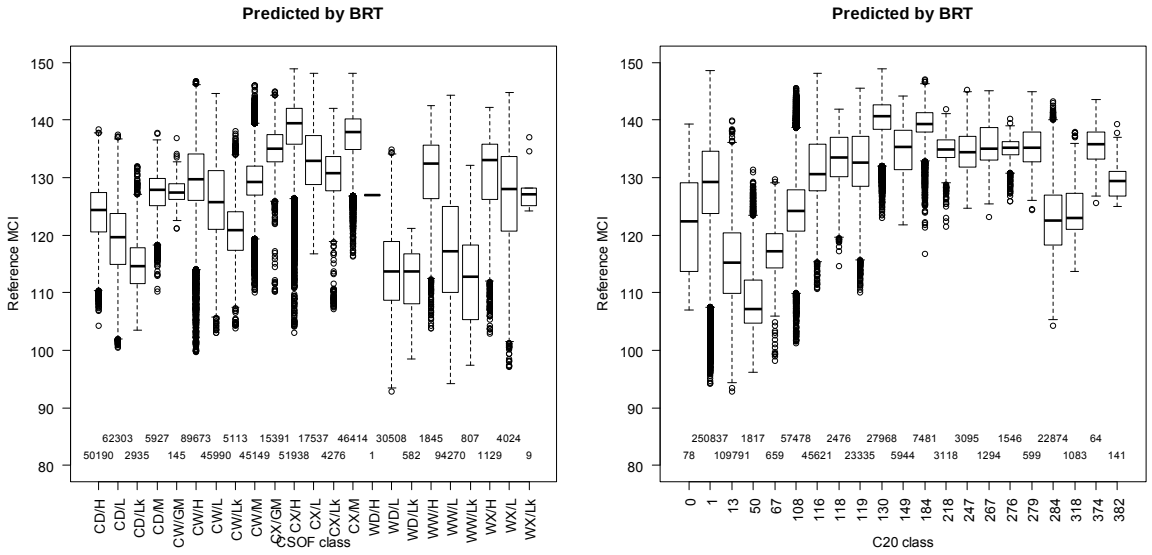


Figure A2.1. Box plots of reference Macroinvertebrate Community Index (MCI) grouped by River Environment Classification Climate/Source-of-Flow (REC CSOF) and Freshwater Ecosystems of New Zealand (FENZ) C20 predicted by Boosted Regression Tree (BRT) model.

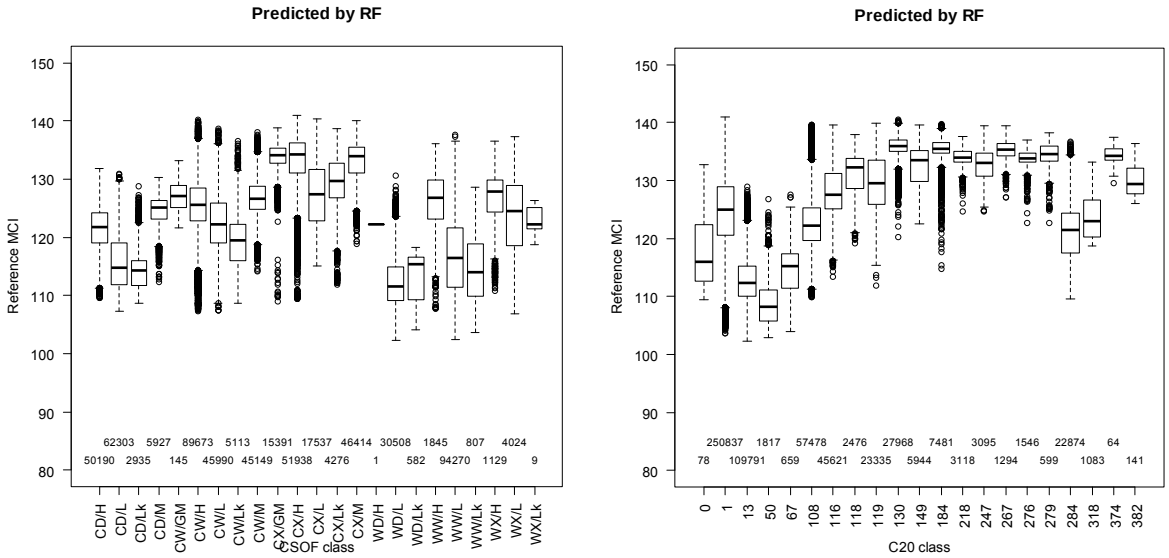


Figure A2.2. Box plots of reference Macroinvertebrate Community Index (MCI) grouped by River Environment Classification Climate/Source-of-Flow (REC CSOF) and Freshwater Ecosystems of New Zealand (FENZ) C20 predicted by Random Forest (RF) model.

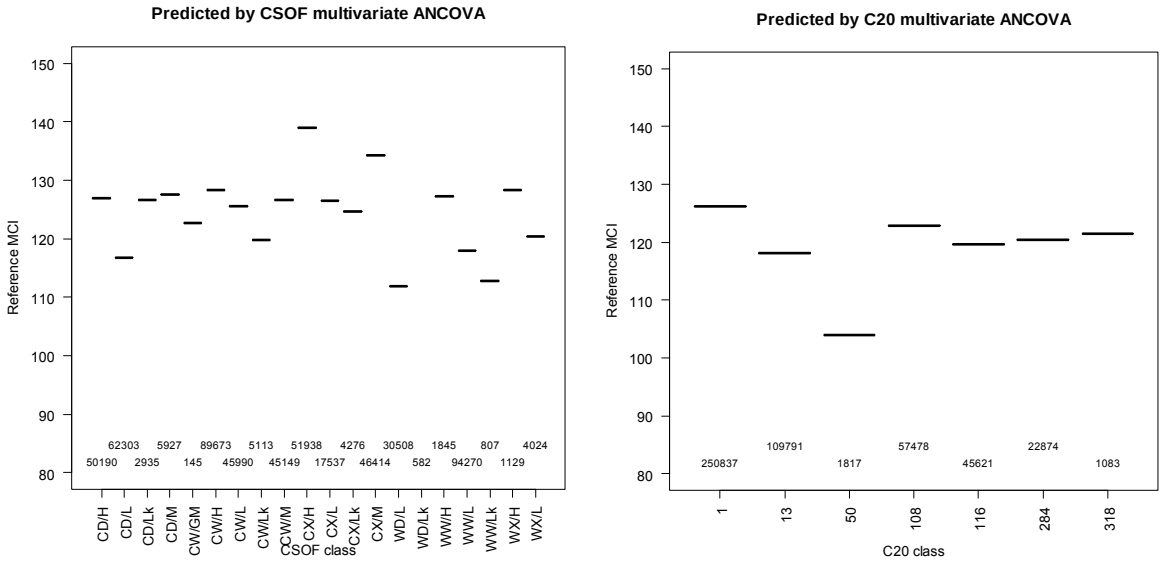


Figure A2.3. Box plots of reference Macroinvertebrate Community Index (MCI) grouped by River Environment Classification Climate/Source-of-Flow (REC CSOF) and Freshwater Ecosystems of New Zealand (FENZ) C20 predicted by linear models.