Spatial modelling of river water-quality state

Incorporating monitoring data from 2016 to 2020

Prepared for Ministry for the Environment

February 2022

Climate, Freshwater & Ocean Science

Prepared by:
Amy Whitehead
Caroline Fraser (LWP)
Ton Snelder (LWP)

For any information regarding this report please contact:
Amy Whitehead
Quantitative Freshwater Ecologist

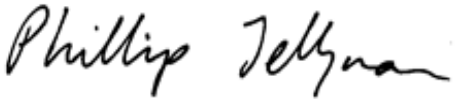
+64-3-343 7854
amy.whitehead@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 8602
Riccarton
Christchurch 8011

Phone +64 3 348 8987

NIWA CLIENT REPORT No: 2021303CH
Report date: October 2021
NIWA Project: MFE21502

Revision	Description	Date
Version 1.0	Final Report	18 October 2021
Version 2.0	Updated Final Report	10 February 2022

Quality Assurance Statement		
	Reviewed by:	Scott Larned
	Formatting checked by:	Rachel Wright
	Approved for release by:	Phil Jellyman

Contents

Tables

Figures

Executive summary

This report provides model-based predictions of current attribute states for each of ~590,000 unique river segments that comprise New Zealand's national river network, using statistics calculated from observations of water quality state for the period 2016–2020. Three comparable reports were produced in 2010, 2016 and 2019 using data for the periods 2003–2007, 2009–2013 and 2013–2017, respectively. This report is the second in a series of 2021 reports prepared for the Ministry for the Environment on national-scale state and trends in river freshwater quality. The first report provided site-specific river water quality state and trends for several hundred river monitoring sites operated by Regional Councils and NIWA. The river water quality data acquired and processed for the first report were used in the current report.

The predicted water quality attribute states in the current report were generated using random forest models. The random forest empirical modelling method predicts the values of response variables using a suite of predictors and a dataset of observations (the 'training data'). Random forest models are an advanced form of regression-tree models. Single regression trees do not identify the optimum tree structure (i.e., the most accurate predictions) and they are sensitive to small changes in the observational data. To overcome these problems, random forest modelling employs an ensemble of trees (a forest) and makes predictions based on the average of all trees. Random forest models have several additional properties that make them suitable for use in situations where the observational data are heterogeneous and the predictors are inter-correlated (as is often the case in water quality analyses); they require no assumptions about data distributions to be met, they are minimally affected by multi-collinearity among predictor variables, and they cannot be over-fitted.

Random forest models were developed for 16 attribute states based on ten water quality variables: visual clarity (CLAR), turbidity (TURB), ammoniacal nitrogen (NH4N), pH-adjusted ammoniacal nitrogen (NH4N_pH), nitrate + nitrite-nitrogen (NNN), total nitrogen (TN), dissolved reactive phosphorus (DRP), total phosphorus (TP), *Escherichia coli* (ECOLI), and the macroinvertebrate community index (MCI). For each variable, the attribute states comprised at least one statistic describing an aspect of the distribution of the observed values (e.g., median value, 95th percentile value, annual maximum value). The attributes are defined by the National Policy Statement for Freshwater Management (NPS-FM). The predictors consisted of 37 variables for which georeferenced data are stored in the River Environment Classification geodatabase (REC2.4). These predictors were selected to represent climatic, geological, topographic, land cover, and hydrological conditions in New Zealand rivers and their catchments.

The observational data used in the random forest models consisted of site attribute states from monthly and quarterly measurements (and annual invertebrates for MCI scores) for the period 2016–2020. These data came from 715–973 monitoring sites (depending on the attribute state). The sites are reasonably well-distributed across the North and South Islands, with some gaps in inaccessible areas. To assess the degree to which the monitoring sites used for observational data represent the range of environmental conditions present in New Zealand, we compared histograms of the distributions of predictor values for the monitoring sites, with the distributions of the same predictors for all river segments in New Zealand. The monitoring sites were reasonably representative, with moderate over-representation of low-elevation, low-gradient catchments with large proportions of intensive agricultural land cover.

The random forest models performed well in predicting attribute states, based on the amount of variation in the observational data explained, the congruence between observed and predicted values, low model bias (tendency to over- or under-estimate), and low prediction uncertainty.

The four most important predictors for models of nutrients (NNN, NH4N, TN, DRP, TP), CLAR, TURB and ECOLI were the upstream stock density, catchment elevation, variability in upstream rainfall and the proportion of urban land cover in catchments. For the MCI model, the most important predictors were the proportion of upstream native forest, local and upstream catchment elevation and the proportion of upstream intensive agriculture. Collectively, the models suggest that attribute state is most severely compromised in low-elevation, low-gradient land under intensive land use.

National maps of the predicted attribute states for nutrients, TURB and ECOLI have relatively high values in low-elevation areas on the east coasts of the North and South Island, and in the inland Waikato, Wairarapa Valley, Rangitikei-Manawatu coastal plain, Taranaki Ring Plain, and Auckland Region. Predicted nutrient, TURB and ECOLI attribute states are generally low in major mountain ranges, in large areas of the Department of Conservation estate and in other native forest-dominated areas. Predicted DRP and TP attribute states appear to be elevated in rivers draining phosphorus-rich tertiary mudstone and volcanic ash on the North Island, suggesting that parent geology affects large scale patterns in river DRP and TP. However, the effects of geology are likely to be inter-correlated with land use and topography. Geographic patterns in predicted CLAR and MCI scores are generally the reverse of the patterns for chemical and microbial attribute states, with high values in mountain ranges and Department of Conservation estate, and low values in areas dominated by intensive agriculture and urban land cover.

1 Introduction

River water quality across New Zealand was characterised in terms of attributes as defined by the National Policy Statement-for Freshwater Management (NPS-FM; New Zealand Government 2020) by a recent national analysis of state and trends at monitored rivers (Whitehead et al. 2021). The sites are monitored as part of state-of-environment (SoE) programmes operated by Regional Councils and unitary authorities and the National River Water Quality Network (NRWQN) operated by NIWA. The datasets used for these analyses contained quarterly, monthly or annual measurements of physical, chemical, and biological variables over time periods that started as early as 1990 and extended to 2020.

The objective of this study was to estimate and map the current state of 16 water quality attribute states in all river segments in New Zealand. The resulting large dataset of estimates can then be used in a wide range of applications, including identifying environmental drivers of water quality variation and setting water-quality reference and baseline levels.

In this study, we developed random forest models to relate spatial variation in the same 16 attribute states used in Whitehead et al. (2021) to a large suite of environmental predictors. The predictors represent climatic, geological, topographic, land cover, and hydrological conditions in New Zealand rivers and their catchments. The random forest models were then used to predict attribute states at un-monitored locations and to produce spatially continuous maps of predicted attribute states that illustrate patterns in water quality at the national scale.

One of the primary purposes of predictive random forest models is to provide large-scale water-quality assessments that are more representative of the true patterns of current attribute states than assessments based on aggregated data from multiple monitoring sites. The latter approach can lead to conclusions about water quality patterns that are biased by the non-random locations of monitoring sites. Previous analyses have shown that the aggregate network of river water quality monitoring sites in New Zealand is over-represented by sites in catchments dominated by pastoral land cover and under-represented by sites in catchments dominated by native forest (e.g., Larned and Unwin 2012; Whitehead 2019). This non-representative distribution of sites can produce biased results when multiple environmental classes are merged. Using random forest models to predict current attribute state in all river reaches across New Zealand can reduce this problem.

In this report, we provide detailed methods for using random forest models to predict current river attribute state across the heterogeneous New Zealand environment. The methodological steps include preparation of the attribute state data, selection of environmental predictors, assessment of site representativeness, the random forest modelling process, and assessments of model performance. In the results section, we present maps of national predictions of current attribute state, identify important predictors and quantify model performance. In the discussion section, we compare the current random forest models with previous models of river attribute state and comment briefly on uncertainty in random forest models and alternative modelling methods. We also provided MfE with the model outputs for every river segment in New Zealand in the supplementary file “RiverRF_WQModel_Predictions_220204.csv”.

2 Data

2.1 River attribute state data

The monitoring sites and data used in Whitehead et al. (2021) to analyse attribute state were used as input for the random forest models. The water quality data consisted of measurements of 16 physical, chemical, microbiological and invertebrate attribute states from river monitoring sites in council SoE networks and the NRWQN sites (Table 2-1). Detailed methods for processing the water quality observations are given in Whitehead et al. (2021). The state dataset consisted of data for each of the 16 attribute states, for the 2016–2020 period, at sites for which measurements were available in at least 90% of the sampling intervals in that period (i.e., at least 54 of 60 months or 18 of 20 quarters). For the annually sampled Macroinvertebrate Community Index (MCI), which is generally less variable than physical or chemical attribute states, we required that data were available for at least four of the five years. In addition, we shifted the 2016–2020 time period by six months (1 July 2015 – 30 June 2020) to align with water years, in order to prevent splitting summer samples into two calendar years, and only included samples if they occurred from December to March to align with the NPS-FM MCI attribute (New Zealand Government 2020). In the final dataset used for random forest modelling, 1389 sites met the inclusion criteria for at least one of the 16 attribute states (Table 2-1).

The geographic distribution of river monitoring sites used for modelling is shown in Figure 2-1. The sites are reasonably well-distributed, although there are gaps in the central North Island and west coast of the South Island. There is a high degree of overlap among the sites used for physical, chemical and microbiological water quality monitoring, as some or all of the corresponding variables are measured at each site in council SoE programmes. There is less overlap among sites used for invertebrate monitoring; several councils operate separate programmes for monitoring physical-chemical water quality and invertebrates, with variable levels of site overlap between programmes.

Table 2-1: River attributes, measurement units and site numbers used to develop random forest models. A random forest model was developed for each combination variable and statistic (referred to as an attribute state). Sites = the number of monitoring sites with observations available for each variable.

Attribute type	Variable	Abbreviation	Statistic	Units	Sites
Physical	Visual clarity	CLAR	Median	m	715
	Turbidity	TURB	Median	NTU	834
Chemical	Ammoniacal nitrogen	NH4N	Median	mg l^{-1}	973
	Ammoniacal nitrogen adjusted for pH	NH4N_pH	Median, AnnMax	mg l^{-1}	885
	Nitrate + nitrite-nitrogen	NNN	Median, Q95	mg l^{-1}	946
	Total nitrogen (unfiltered)	TN	Median	mg l^{-1}	938
	Dissolved reactive phosphorus	DRP	Median, Q95	mg l^{-1}	973
	Total phosphorus (unfiltered)	TP	Median	mg l^{-1}	901
Microbiological	*Escherichia coli*	ECOLI	Median, Q95 G260, G540	cfu 100 ml^{-1} % exceedances	967
Invertebrate	Macroinvertebrate Community Index	MCI	Median	unitless	863

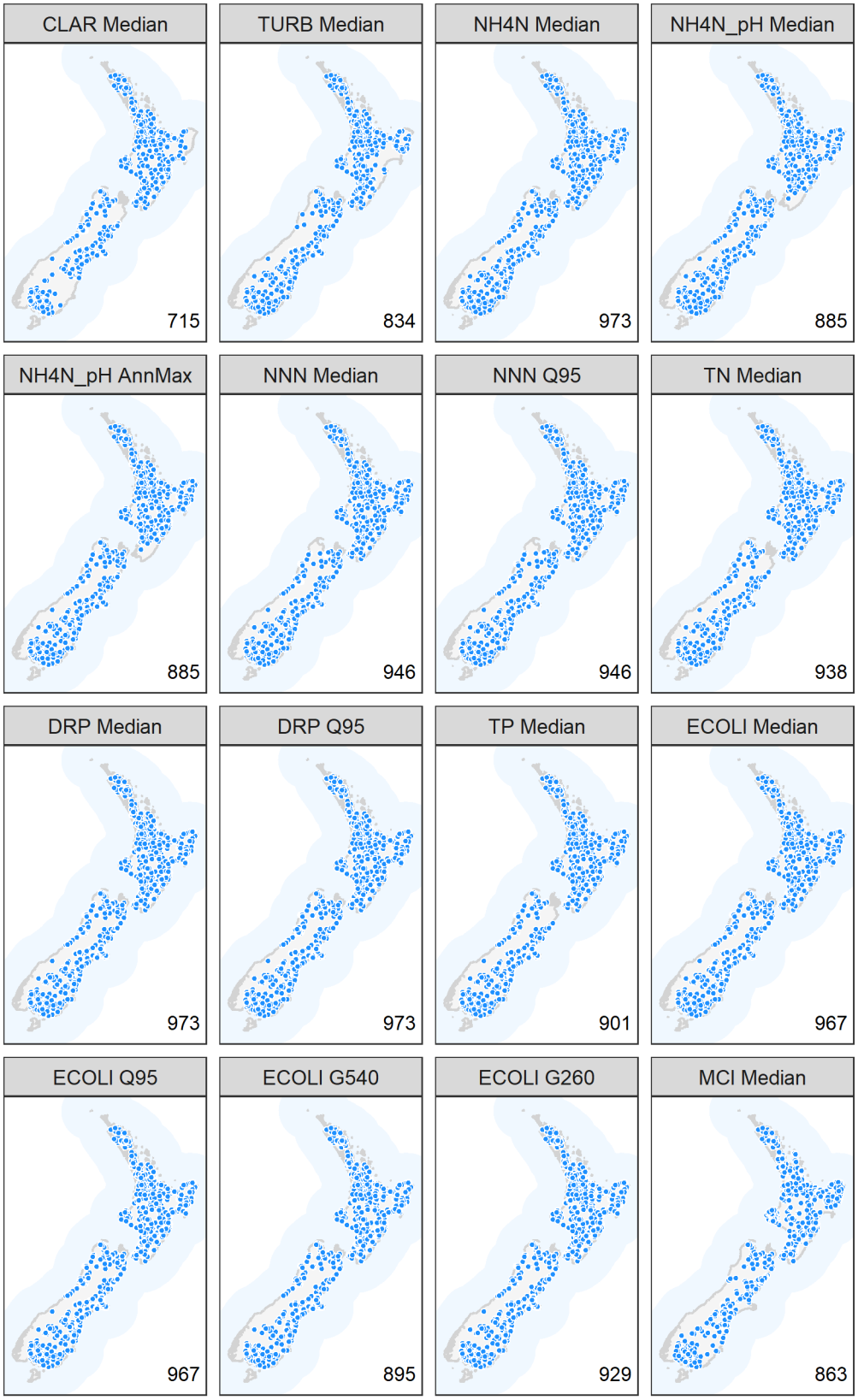


Figure 2-1: Locations of river monitoring sites used for modelling the current state of the 16 attributes. The number in the lower right of each panel corresponds to the number of sites included in each attribute state model after data transformation (Table 2-1).

2.2 Predictor data

We used the digital river network (REC2.4) to provide the spatial framework for the random forest models of river attribute state. The river network and catchment boundaries were derived from a digital elevation model (DEM) with a spatial resolution of 50 m. The digital network represents New Zealand's rivers as ~590,000 segments (bounded by upstream and downstream confluences) and their corresponding catchments. Each segment in the digital network has a unique identifier, the nzsegment number. The links between each nzsegment and its catchment, between adjacent nzsegments and between adjacent catchments facilitate analyses of upstream-downstream connectivity and the accumulation of catchment characteristics in the downstream direction.

In addition to the digital river network, REC2.4 is associated with spatial data layers describing the climate, topography, geology, vegetation, infrastructure and hydrology of New Zealand[1]. These spatial data are used to link each nzsegment to many attributes that describe the environmental characteristics of the segment and its catchment. Catchment land cover was derived from the national Land Cover Database 5 (LCDB5) which differentiates 34 categories based on analysis of satellite imagery from the 2018–2019 summer[2]. Descriptions of catchment regolith were derived from the Land Resources Inventory (LRI) including interpretations of the LRI categories made by Leathwick et al. (2003). Additional variables for each segment have been derived from national-scale hydrological modelling (e.g., Booker and Snelder 2012). In addition, estimates of the mean land use intensity associated with four different stock types (beef, dairy, deer, sheep) were derived from the 2017 agricultural production census[3] (Snelder et al. 2022).

We selected 37 network attributes from REC2.4 (Table 2-2) for predictors in spatial models of the sixteen attribute states listed in Table 2-1. The predictors were selected based on their predicted mechanistic or correlative relationships with water quality, and on previous experience with national scale modelling of water quality (e.g., Larned et al. 2016; Unwin et al. 2010; Whitehead 2019) and invertebrate communities (Clapcott et al. 2013; Leathwick et al. 2011).

[1] https://www.niwa.co.nz/freshwater-and-estuaries/management-tools/river-environment-classification-0
[2] https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/
[3] https://statisticsnz.shinyapps.io/livestock_numbers/

Table 2-2: Predictors used in random forest models of river attribute states.

Predictor class	Predictor description	Abbreviation	Unit
Geography & topography	Catchment area	usArea	km^2
	Segment mean elevation	segElev	m ASL
	Percentage of catchment occupied by lakes	usLakePerc	%
	Mean catchment elevation	usElev	m ASL
	Mean catchment slope	usSlope	degrees
	Distance to the coast	DistToCoast	km
	Mean segment slope	SegSlope	degrees
	Segment sinuosity (segment length divided by the straight-line distance between endpoints)	Sinuosity	unitless
	Distance to furthest headwater segment	DistToHead	km
Climate & flow	Mean segment June air temperature	segTmin	degrees C
	Mean segment January air temperature.	segTwarm	degrees C
	Mean catchment June air temperature	usTmin	degrees C
	Mean catchment January air temperature	usTwarm	degrees C
	Mean annual catchment rainfall	usRain	mm
	Mean catchment coefficient of variation of annual rainfall	usRainvar	mm/yr
	Mean catchment rain days > 10 mm	usRainDays10	days/mo
	Mean catchment rain days > 200 mm	usRainDays20	days/mo
	Mean catchment rain days > 100 mm	usRainDays100	days/mo
	Mean annual catchment potential evapotranspiration	usPET	mm/yr
	Estimated mean flow	MeanFlow	m^3/s
Geology	Mean catchment induration (hardness) of regolith	usHard	Ordinal
	Mean catchment phosphorous content of regolith	usPhos	Ordinal
	Mean catchment particle size of regolith	usPsize	Ordinal
	Mean catchment calcium content of regolith	usCalc	Ordinal
Land cover	Proportion of catchment occupied by combination of high producing exotic grassland, short-rotation cropland, orchard, vineyard and other perennial crops (LCDB5 classes 40, 30, 33)	usIntensiveAg	%
	Proportion of catchment in low producing grassland (LCDB5 class 41)	usPastoralLight	%
	Proportion of catchment in native forest (LCDB5 class 69)	usNativeForest	%
	Proportion of catchment in built-up areas, urban parkland, surface mines, dumps and transport infrastructure (LCDB5 classes 1, 2, 6, 5)	usUrban	%
	Proportion of catchment in scrub and shrub cover (LCDB5 classes 50, 51, 52, 54, 55, 56, 58)	usScrub	%
	Proportion of catchment occupied by lake and pond, river and estuarine open water (LCDB5 classes 20, 21, 22)	usWetland	%
	Proportion of catchment in exotic forest (LCDB3 class 71)	usExoticForest	%
	Proportion of catchment occupied in bare or lightly-vegetated cover (LCDB5 classes 10, 12, 14, 15, 16)	usBare	%
Land use intensity	Catchment density of total stock units (SU) on pastoral land	usTotalStock	SU/ha
	Proportion of total stock units attributable to beef cattle in catchment	usBeef	Proportion
	Proportion of total stock units attributable to dairy cows in catchment	usDairy	Proportion
	Proportion of total stock units attributable to deer in catchment	usDeer	Proportion
	Proportion of total stock units attributable to sheep in catchment	usSheep	Proportion

3 Modelling methods

3.1 Random forest models

We modelled each attribute state as a function of the predictors using random forest models (Breiman 1984, 2001; Cutler et al. 2007), with most attribute states log-transformed (i.e., the $\log_{10}$ of the untransformed raw data). The exceptions were ECOLI exceedances (G260, G540) which were logit-transformed and MCI where we used the untransformed raw data. A random forest model is an ensemble of individual classification and regression trees (CART). In a regression context, CART partitions observations (in this case the individual attribute states) into groups that minimise the sum of squares of the response (i.e., assembles groups that minimise differences between observations) based on a series of binary rules or splits that are constructed from the predictor variables. CART models have several desirable features including requiring no distributional assumptions and the ability to automatically fit non-linear relationships and high order interactions. However, single regression trees have the limitations of not searching for optimal tree structures, and of being sensitive to small changes in input data (Hastie et al. 2001). Random forest models reduce these limitations by using an ensemble of trees (a forest) and making predictions based on the average of all trees (Breiman 2001). An important feature of random forest models is that each tree is grown with a bootstrap sample of the fitting data (i.e., the observation dataset). In addition, a random subset of the predictors is made available at each node to define the split. Introducing these random components and then averaging over the forest increases prediction accuracy while retaining the desirable features of CART.

A random forest model produces a limiting value of the generalisation error (i.e., the model maximises its prediction accuracy for previously unseen data; Breiman 2001). The generalisation error converges asymptotically as the number of trees increases, so the model cannot be over-fitted. The number of trees needs to be set high enough to ensure an appropriate level of convergence, and this value depends on the number of variables that can be used at each split. We used default options that included making one third of the total number of predictors available for each split, and 500 trees per forest. Some studies report that model performance is improved by including more than ~50 trees per forest, but that there is little improvement associated with increasing the number of trees beyond 500 (Cutler et al. 2007).

Unlike linear models, random forest models cannot be expressed as equations. However, the relationships between predictor and response variables represented by random forest models can be represented by importance measures and partial dependence plots (Breiman 2001; Cutler et al. 2007). During the fitting process, random forest model predictions are made for each tree for observations that were excluded from the bootstrap sample; these excluded observations are known as out-of-bag (OOB) observations. To assess the importance of a specific predictor, the values of the response are randomly permuted for the OOB observations, and predictions are obtained from the tree for these modified data. The importance of the predictor is indicated by the degree to which prediction accuracy decreases when the response variable is randomly permuted. Importance is defined in this study as the loss in model performance (i.e., the increase in the mean square error; MSE) when predictions are made based on the permuted OOB observations compared to those based on the original observations. The differences in MSE between trees fitted with the original and permuted observations are averaged over all trees and normalized by the standard deviation of the differences (Cutler et al. 2007).

A partial dependence plot is a graphical representation of the marginal effect of a predictor on the response, when the values of all other predictors are held constant. The benefit of holding the other predictors constant (generally at their respective mean values) is that the partial dependence plot effectively ignores their influence on the response. Partial dependence plots do not perfectly represent the effects of each predictor, particularly if predictors are highly correlated or strongly interacting, but they do provide an approximation of the modelled predictor-response relationships that are useful for model interpretation (Cutler et al. 2007).

Random forest models can include any of the original set of predictors that are chosen during the model fitting process. Inclusion of marginally important and correlated predictors does not degrade the performance of the random forest models. However, these predictors may be redundant (i.e., their removal does not affect model performance) and their inclusion can complicate model interpretation. We used a backward elimination procedure to remove redundant predictors from the initial 'saturated' models (i.e., models that included any of the original predictor variables). The procedure first assesses the model error (MSE) using a 10-fold cross validation process. The predictions made to the hold out observations during cross validation are used to estimate the MSE and its standard error. The model's least important predictors are then removed in order, with the MSE and its standard error being assessed for each for each successive model. The final, 'reduced' model is defined as the model with the fewest predictors whose error is within one standard error of the best model (i.e., the model with the lowest cross validated MSE). This is equivalent to the "one standard error rule" used for cross validation of classification trees (Breiman 1984).

An alternative approach is to choose the model with the smallest error. We used the former procedure as it retains fewer predictors than the latter procedure, while achieving an error rate that is not different, within sampling error, from the "best solution". Importance levels for predictors were not recalculated at each reduction step to avoid over-fitting (Svetnik et al. 2004).

We note that, because fitting a random forest model involves randomly selecting observations and predictors throughout the fitting process, successive models fitted to the same data set will exhibit subtle differences in structure and diagnostics such as total explained deviance, MSE, partial dependence plots, and the order of predictor importance. In the current study, the variability in model error between individual fits of the model for each attribute state were within the reported model performance (see Section 3.2).

All calculations were performed in the R statistical computing environment (Version 4.0.5; R Core Team 2021) using the *randomForest* package (Liaw and Wiener 2002) and other specialised packages.

3.2 Model performance

Model performance was assessed by comparing observations with independent predictions (i.e., sites that were not used in fitting the model), which were obtained from the out-of-bag (OOB) samples. We summarised the models using four statistics; regression R^2**,** Nash-Sutcliffe Efficiencies (NSE), percent bias (PBIAS) and root mean square deviation (RMSD).

The regression R^2 value is the coefficient of determination derived from a regression of the observations against the predictions. The R^2 value shows the proportion of the total variance explained by the regression model (Piñeiro et al. 2008). However, the regression R^2 is not a complete description of model performance.

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

Bias measures the average tendency of the predicted values to be larger or smaller than the observed values. Optimal bias is zero, positive values indicate underestimation bias and negative values indicate overestimation bias (Piñeiro et al. 2008). PBIAS is computed as the sum of the differences between the observations and predictions divided by the sum of the observations (Moriasi et al. 2007). Model predictions were evaluated to be very good, good, satisfactory or unsatisfactory, following the criteria proposed by Moriasi et al (2015), outlined in Table 3-1.

The root mean square deviation (RMSD) is a measure of the characteristic model statistical error or uncertainty. RMSD is mean deviation of predicted values with respect to the observed values (distinct from the standard error of the regression model). RMSD can be used to evaluate the confidence intervals of the predictions.

Table 3-1: Performance ratings for statistics used in this study. From Moriasi et al (2015).

Performance Rating	R^2	NSE	PBIAS
Very good	$R^2 \geq 0.70$	NSE > 0.65	|PBIAS| <15
Good	$0.60 < R^2 \leq 0.70$	$0.50 < NSE \leq 0.65$	$15 \leq \|PBIAS\| < 20$
Satisfactory	$0.30 < R^2 \leq 0.60$	$0.35 < NSE \leq 0.50$	$20 \leq \|PBIAS\| < 30$
Unsatisfactory	$R^2 < 0.30$	$NSE \leq 0.35$	$\|PBIAS\| \geq 30$

3.3 Representativeness of monitoring sites used in random forest models

A graphic comparison was used to gauge how well the monitoring sites used to fit the random forest models represented environmental variation at the national scale. Here, representativeness refers to the degree to which the distribution of monitoring sites over the range of an environmental predictor matches the distribution of all network segments over the range of the same environmental predictor. Poor representativeness can reduce accuracy in model predictions because certain combinations of environmental conditions are not represented in the fitting data.

Histograms of the proportion of monitoring site numbers across the environmental gradient of the 12 most important predictors in the random forest models (i.e., the predictors with the greatest explanatory power) were visually compared with histograms of the proportion of all network segments over the same predictor variables. Two sets of comparable histograms were derived. The first represented data from all sites that monitored at least one water quality attribute state, excluding MCI (976 sites). The second set of comparable histograms represented the 863 invertebrate monitoring sites that were used for modelling MCI scores. Separate histograms were constructed due to the limited overlap in physical-chemical water quality and invertebrate monitoring sites, as noted in Section 2.1. Note that representativeness of monitoring sites is different from model bias, which is defined in Section 3.2. Model bias is a measure of systematic error in model predictions (i.e., over- or under-estimation).

3.4 Model predictions

Predictions are made with random forest models by "running" new cases down every tree in the fitted forest and averaging the predictions made by each tree (Cutler et al. 2007). The models in this study were fitted to $\log_{10}$-transformed variables (except for MCI which used non-transformed data and ECOLI exceedances, which were logit-transformed). When the predictions made by models fitted to $\log_{10}$-transformed variables are back-transformed, the model error term no longer has a mean of zero. Ignoring this results in retransformation bias (i.e., predictions that systematically underestimate the response). We corrected the retransformation bias using the smearing estimate (*S*) developed by Duan (1983):

$$S = \frac{1}{n}\sum_{i=1}^{n} 10^{\hat{\varepsilon}_i} \quad (1)$$

where $\hat{\varepsilon}$ are the residuals of a random forest model. The predictions were back-transformed by raising them to the power of 10, then corrected for retransformation bias by multiplying by *S.* Predictions of ECOLI exceedance values (G260, G540) were back-transformed using the inverse-logit function. The back-transformed and corrected predictions for all river segments in New Zealand were projected on a national map for each attribute state.

4 Results

4.1 Model performance

The random forest models for nine of the 16 attribute states were considered to be very good (NNN Q95, TN Median, ECOLI Median and MCI Median) or good (NNN Median, DRP Median, TP Median, ECOLI Q95 and ECOLI G260) as indicated by R^2 and NSE and the criteria of Moriasi *et al.* (2015) (Table 4-1, Figure 4-1). The remaining seven models (CLAR Median, TURB Median, NH4N Median, NH4N_pH Median, NH4N_pH AnnMax, DRP Q95, ECOLI G540) had satisfactory performance. All 16 models had very low bias (PBIAS; Table 4-1) as indicated by the close match between the line representing the regression of the observed versus predicted values (red dashed line in Figure 4-1) and the one-to-one line (blue solid line in Figure 4-1). The close match between the regression and one to one line also indicates that the models are consistent (i.e., that low or high values are not under or over-estimated). RMSD values provide an indication of the magnitude of the characteristic error in the transformed units of each variable. Note that the RMSD value for MCI is higher than the other variables because the MCI data were not transformed.

Table 4-1: Performance of the 16 attribute state models. Performance was determined using independent predictions (i.e., sites that were not used in fitting the models) generated from the out-of-bag (OOB) observations. R^2 = coefficient of determination, NSE = Nash-Sutcliffe efficiency, PBIAS = percent bias, RMSD = root mean square deviation). Units for RMSD are the $\log_{10}$ or logit transformed units of the attribute state except for MCI, which were not transformed. The colours indicate the performance ratings indicated in Table 3-1.

Attribute state	N	R^2	NSE	PBIAS	RMSD	Rating
CLAR Median	715	0.58	0.58	1.28	0.22	Satisfactory
TURB Median	832	0.60	0.59	-0.44	0.29	Satisfactory
NH4N Median	970	0.52	0.52	0.17	0.34	Satisfactory
NH4N_pH Median	883	0.48	0.48	0.25	0.34	Satisfactory
NH4N_pH AnnMax	883	0.48	0.48	0.34	0.44	Satisfactory
NNN Median	943	0.65	0.65	-0.79	0.43	Good
NNN Q95	943	0.75	0.75	-0.16	0.28	Very good
TN Median	935	0.76	0.76	1.16	0.23	Very good
DRP Median	970	0.62	0.62	0.09	0.3	Good
DRP Q95	970	0.59	0.59	0.32	0.33	Satisfactory
TP Median	898	0.67	0.67	0.26	0.26	Good
ECOLI Median	964	0.70	0.70	-0.02	0.33	Very good
ECOLI Q95	964	0.64	0.64	-0.04	0.39	Good
ECOLI G540	892	0.58	0.58	0.89	0.87	Satisfactory
ECOLI G260	926	0.63	0.63	1.43	0.98	Good
MCI Median	863	0.71	0.70	0.09	10.34	Very good

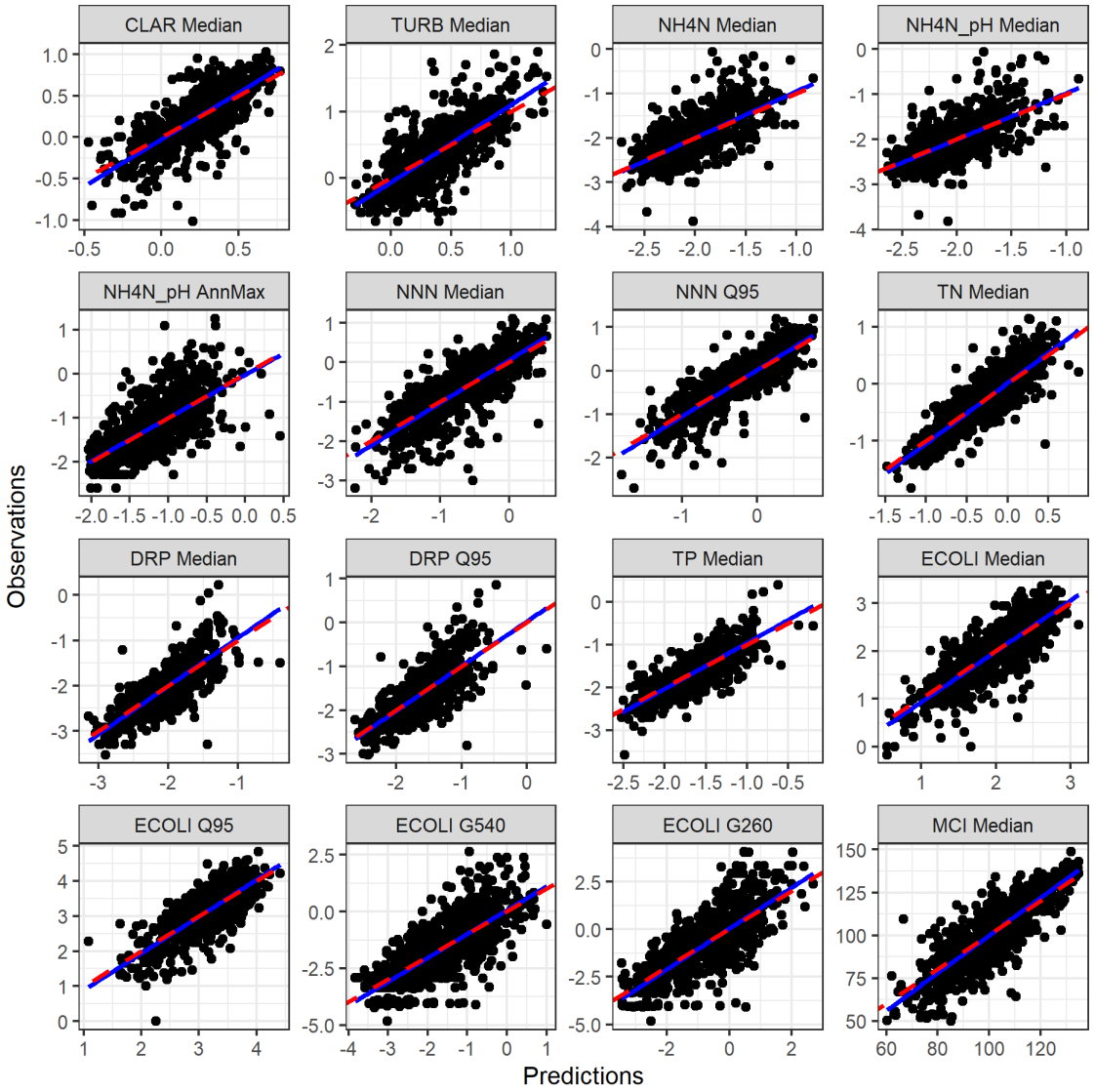


Figure 4-1: Comparison of observed attribute state versus values predicted by each of the random forest models. Note that the observed values are plotted on the Y-axis and predicted values on the X-axis, following Piñeiro et al. (2008). Blue solid line: best fit linear regression of the observed and predicted values. Red dashed line: one-to-one line. Units are the $\log_{10}$ transformed units of all water quality variables, except for MCI (non-transformed values) and ECOLI G260 and ECOLI G540 (logit-transformed values).

4.2 Monitoring site representativeness

The distributions of river water quality and MCI monitoring sites across the environmental gradients defined by 12 predictors were generally consistent with the distribution of all segments in the river network across the same gradients (Figure 4-2 and Figure 4-3). The predictors shown in the histograms were those subsequently found to be most important in the random forest models (Table 4-2).

There were several cases of moderate over- and under-representation of monitoring sites compared to the river network. Water-quality sites were over-represented in environments characterised by low catchment elevations (usElev), low catchment slopes (usSlope) and warmer winter temperatures (segTmin, usTmin; Figure 4-2). Water quality sites were under-represented in catchments with low proportions of high intensity agriculture and low stocking rates (usIntensiveAg, usStockDensity).

Invertebrate monitoring sites were over-represented in warm, low elevation and low slope rivers (segTwarm, segElev, usElev, usSlope) close to the coast (distToCoast) and under-represented in steep headwater catchments with high proportion of native forest land cover or low proportion of intensive agricultural landcover (distToHead, usNativeForest, usIntensiveAg; Figure 4-3).

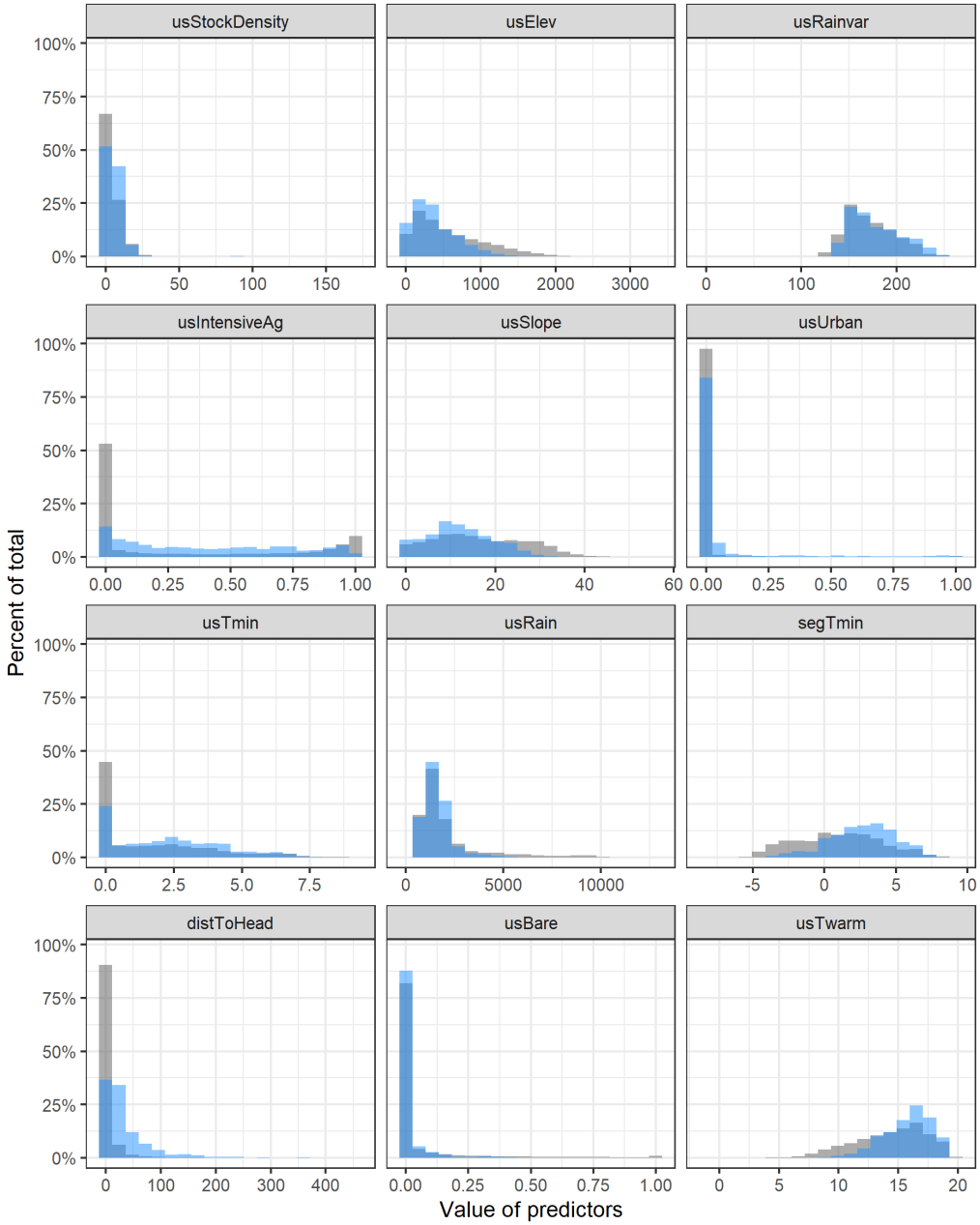


Figure 4-2: The distributions of predictors across all segments in the digital river network and at water quality sites (grey and blue histograms, respectively). Similarities in the distributions shown in the two histograms in each panel provide an indication of the degree to which environmental variation across the monitoring sites represents environmental variation across the New Zealand river network; complete representativeness would be indicated by exact matches between the histograms. These 12 predictors were the most important overall predictors in the water quality random forest models (with the exception of MCI – see Figure 4-3) and are ordered from most (top left) to least (bottom right) important (Table 4-2).

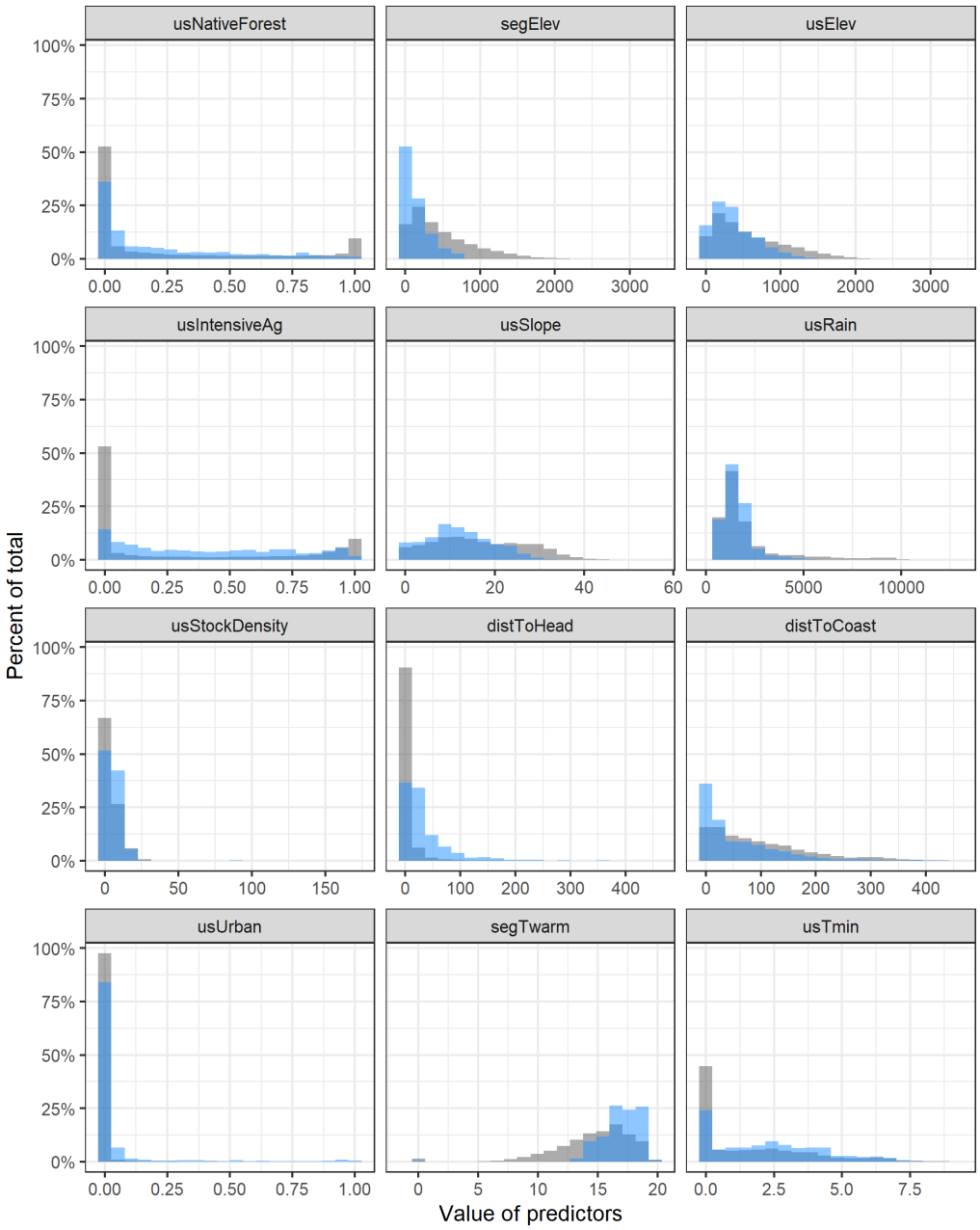


Figure 4-3: Distributions of predictors across all segments in the digital river network and at invertebrate sites (grey and blue histograms, respectively). Similarities in the distributions shown in the two histograms in each panel provide an indication of the degree to which environmental variation across the monitoring sites represents environmental variation across the New Zealand river network; complete representativeness would be indicated by exact matches between the histograms. These 12 predictors were the most important in the random forest model of MCI and are ordered from most (top left) to least (bottom right) important (Table 4-2).

4.3 Modelled relationships

The predictors with high importance in all random forest models reflected strong associations between water quality and land use intensity, landcover, climate and catchment topography. The density of stock in the upstream catchment (usStockDensity) was the most important variable when compared across all models, while usUrban and usIntensiveAg were included in the top ten most important predictors. The partial plots indicate that CLAR Median and MCI Median decreased with increasing usUrban and usIntensiveAg (ranked 4th and 7th, respectively), while all attribute states

describing nutrients, ECOLI and TURB increased (Figure 4-4 and Figure 4-5). In comparison, CLAR Median and MCI Median increased with increasing usNativeForest (the proportion of late-successional native forest in the upstream catchment, ranked 13th overall but 1st for MCI), while all attribute states describing nutrients, ECOLI and TURB decreased. Predictors of stock type intensity (usDairy and usBeef) were important predictors for NNN (Median, Q95) and TN Median but either ranked lowly (< 10th) or were not included in models for the other attribute states (Table 4-2). In contrast, usDeer and usSheep were ranked lowly or not included across all attribute states. These patterns are consistent with previous correlations between landcover and water quality state (e.g., Larned et al. 2004, 2016; Snelder et al. 2022; Whitehead 2019).

Predictors describing the slope and elevation of the upstream catchment, usElev and usSlope, ranked 2nd and 5th overall (Table 4-2). The importance values showed usElev and usSlope were both important in all models (with the exception of usElev for DRP Median). The partial plots indicated that MCI increased with increasing values of usSlope and usElev, while CLAR Median increased to approximately 550m usElev and 25^{o} usSlope before declining. The values of all other water quality variables decreased with increasing usElev and usSlope (Figure 4-4 and Figure 4-5). The distance to headwater (distToHead) was ranked 10th, with TURB Median increasing and CLAR Median and MCI Median decreasing as the distance to the top of the catchment increases.

Predictors of rainfall usRainvar (variation in mean annual catchment rainfall) and usRain (mean annual catchment rainfall) were the 3rd and 8th most important overall predictors, respectively (Table 4-2). Values of CLAR Median and MCI Median increased with increasing usRainvar and usRain, and the values of all other attribute states decreased (Figure 4-4 and Figure 4-5). These results suggest that there is a moderately strong positive association between attribute state and catchment rainfall. The mechanisms that drive this association may include solute dilution and sustained low water temperatures. Predictors of upstream and local minimum winter temperature (usTmin and segTmin) were ranked 6th and 9th, respectively. TURB Median, TP Median and all ECOLI attribute states increased with increasing minimum winter temperature, while CLAR Median decreased. The relationship with winter temperature may be associated with poorer water quality associated with lower elevation sites in lowland catchments.

The predictor usBare had relatively high importance (ranked 11th) across the random forest models (Table 4-2), indicating that the regolith of the catchment is associated with water quality state. The values of most water quality variables decreased with increasing values of usBare, with the exception of TURB Median, NH4N Median and NH4N_pH Median which increased with increasing usBare (Figure 4-4). These patterns suggest that water quality generally declines as erosion potential increases.

Table 4-2: Rank order of importance of predictors retained in the random forest models for at least one attribute state. Blank cells indicate that a predictor was not included in the reduced model. The predictors are listed in descending order of the median rank importance over all 16 models.

	CLAR	TURB	NH4N	NH4N_pH		NNN		TN	DRP		TP		ECOLI			MCI
Attribute state	**Median**	**Median**	**Median**	**Median**	**AnnMax**	**Median**	**Q95**	**Median**	**Median**	**Q95**	**Median**	**Median**	**Q95**	**G540**	**G260**	**Median**
usStockDensity	-	16	6	14	1	1	1	2	3	2	5	3	1	2	2	7
usElev	10	3	3	3	3	10	6	5	-	11	18	1	3	1	1	3
usRainvar	1	1	2	7	10	8	-	13	5	15	8	2	2	3	3	19
usIntensiveAg	17	15	15	18	5	4	2	3	18	5	7	5	4	6	4	4
usSlope	9	14	1	1	2	6	4	1	6	1	1	6	6	12	10	5
usTmin	8	4	7	4	8	11	12	15	4	6	6	14	5	7	7	12
usUrban	-	-	4	10	12	5	11	4	-	-	-	7	8	4	5	10
usRain	4	7	-	12	14	15	5	8	2	8	2	18	-	17	-	6
segTmin	2	5	5	2	4	7	8	10	11	4	9	23	11	10	12	13
distToHead	3	9	-	-	-	-	-	-	-	-	-	-	-	-	23	8
usBare	5	12	16	9	-	18	-	9	7	12	3	4	17	5	6	35
usTwarm	15	10	8	5	6	9	13	16	12	3	4	19	7	15	11	15
usNativeForest	7	2	-	-	-	13	9	12	-	14	17	12	9	16	16	1
usPET	13	13	-	-	9	-	14	17	15	10	14	11	15	11	8	20
usRainDays10	18	-	9	11	13	-	-	-	10	13	-	20	10	-	-	14
segTwarm	16	6	-	13	11	-	-	-	17	-	-	-	-	-	17	11
distToCoast	-	26	-	-	-	-	-	-	-	-	-	13	13	18	9	9
usHard	-	11	11	6	18	16	-	11	16	18	11	-	-	-	-	32
usPhos	-	24	13	8	-	-	-	-	14	16	10	10	-	-	-	31
segElev	-	-	14	-	17	-	18	-	-	-	-	9	14	-	13	2
usExoticForest	-	-	-	-	-	14	3	14	-	-	-	8	-	8	14	22
usArea	11	8	-	-	-	-	-	-	-	-	-	-	-	-	20	18
usRainDays20	6	19	10	15	-	-	-	-	9	9	-	26	-	14	21	17
usBeef	-	17	-	-	-	3	7	7	-	-	16	21	-	9	15	21
usPastoralLight	-	-	18	16	7	-	-	-	8	17	-	15	-	13	25	28
usSheep	-	18	-	-	-	12	16	-	-	-	15	24	16	-	18	16
usDairy	-	25	17	-	16	2	10	6	-	-	13	16	-	-	22	26
usPsize	-	-	12	17	-	17	-	-	1	7	12	22	-	-	26	25
usLakePerc	-	-	-	-	-	-	17	-	13	-	-	-	18	-	-	29
usScrub	-	-	-	-	-	-	15	18	-	-	-	-	-	-	-	27
usDeer	-	23	-	-	-	-	-	-	-	-	-	17	12	-	19	30
usRainDays100	14	21	-	-	-	-	-	-	-	-	-	-	-	-	-	34
usWetland	12	22	-	-	15	-	-	-	-	-	-	25	-	-	-	36
meanFlow	-	20	-	-	-	-	-	-	-	-	-	-	-	-	24	23
segSlope	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	24
usCalc	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	33
sinuosity	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	37

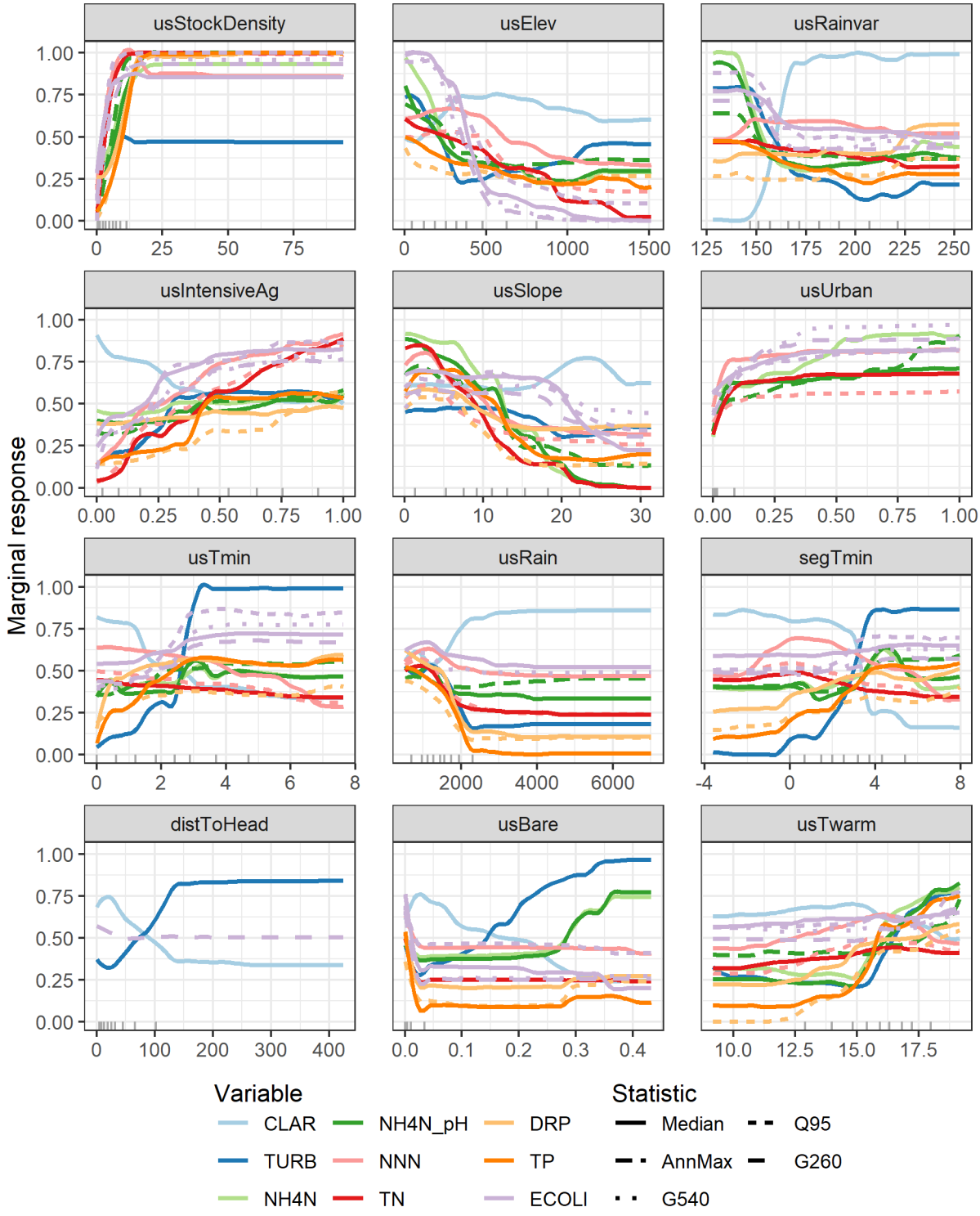


Figure 4-4: Partial plots for the 12 most important predictors in random forest models of current attribute state. Colours represent water quality variables, with the statistic indicated by line type (i.e., the combination of colour and line type represents an attribute). Each panel corresponds to one predictor, with predictors ordered by overall importance from most (top left) to least (bottom right) important. Y-axis scales represent marginal response standardised across all modelled attribute states. Plot amplitude (the range of the marginal response on the Y-axis) is directly related to a predictor's importance, with amplitude larger for predictors with higher importance. Units on X-axes are in Table 2-2.

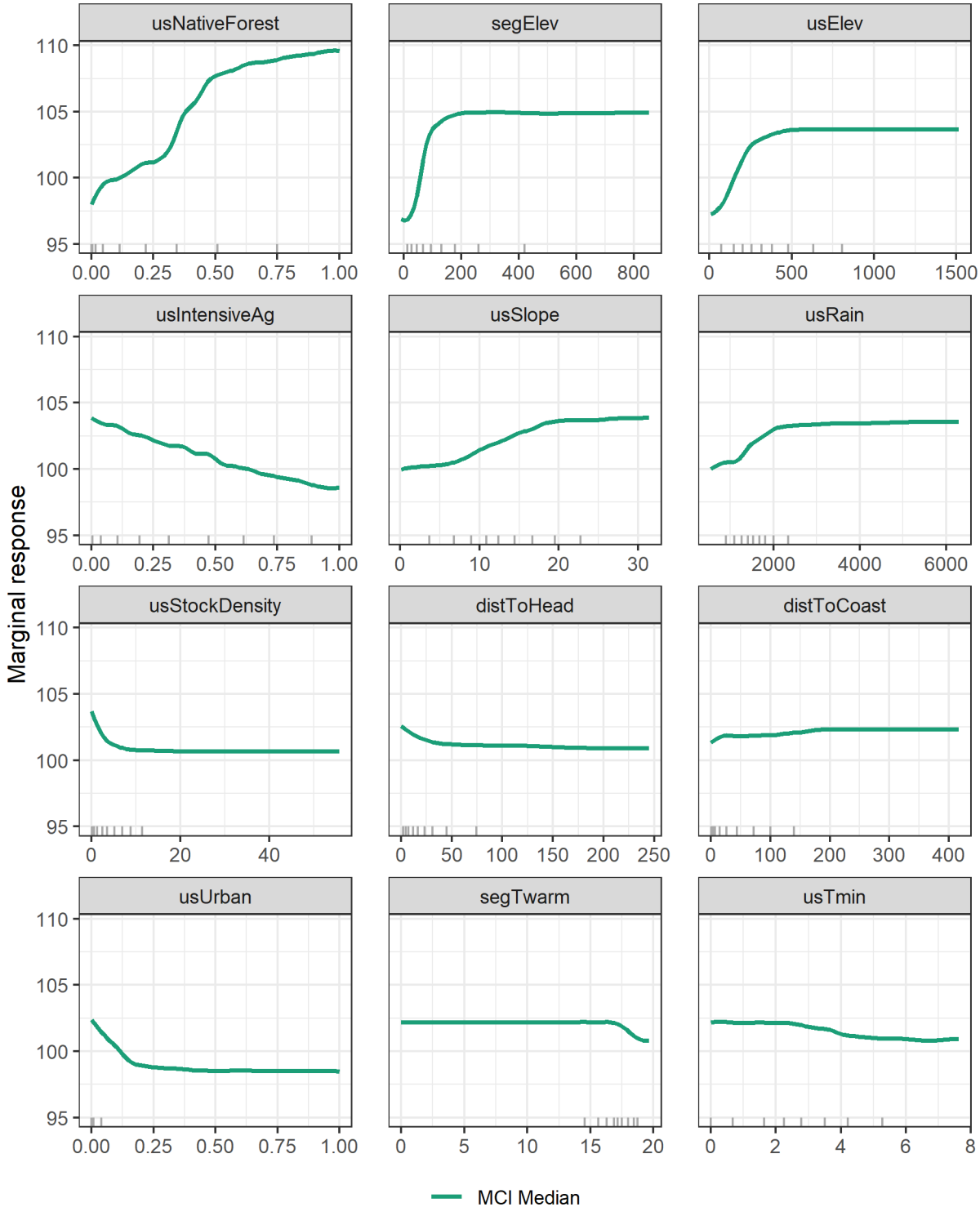


Figure 4-5: Partial plots for the 12 most important predictors for the median MCI random forest model. Each panel corresponds to one predictor, with predictors ordered from most (top left) to least (bottom right) important. The Y-axis scale represents the absolute value of the marginal response (i.e., the scale represents the marginal change in MCI values). The amplitude of each plot (i.e., the range of the marginal response shown on the y-axis) is directly related to a predictor's importance. Units on X-axes are in Table 2-2.

4.4 Model predictions

The minimum values predicted by the random forest models were always somewhat larger than the minimum of the observed values and the maximum predicted values were always somewhat smaller than the maximum observed values (Table 4-3). This is an expected outcome of random forest models, which are based on partitioning the data and predictions are derived from the means of observations that are assigned to a particular partition. As a consequence, the predictions for each attribute state were always within the range of the observations.

Table 4-3: Comparisons of the minimum and maximum observed and predicted values of water quality attribute states.

Attribute state (units)	Observed values		Predicted values	
	Minimum	Maximum	Minimum	Maximum
CLAR Median (m)	0.10	10.7	0.22	7.56
TURB Median (NTU)	0.22	79	0.39	40.15
NH4N Median (mg l^{-1})	0	0.87	0	0.27
NH4N_pH Median (mg l^{-1})	0	0.87	0	0.25
NH4N_pH AnnMax (mg l^{-1})	0	17.9	0.01	3.37
NNN Median (mg l^{-1})	0	13.05	0	7.74
NNN Q95 (mg l^{-1})	0	15.84	0.01	10.5
TN Median (mg l^{-1})	0.01	13.9	0.03	7.68
DRP Median (mg l^{-1})	0	1.66	0	0.26
DRP Q95 (mg l^{-1})	0	7.14	0	1.17
TP Median (mg l^{-1})	0	2.45	0	0.61
ECOLI Median (cfu 100 ml^{-1})	0.67	2400	1.71	1281.6
ECOLI Q95 (cfu 100 ml^{-1})	1	68150	25.38	26952.87
ECOLI G540 (% exceedances)	0.01	0.93	0.02	0.81
ECOLI G260 (% exceedances)	0.01	0.98	0.02	0.96
MCI Median	49.85	148.75	54.54	140.79

The mapped predictions for attribute states describing nutrients, ECOLI and TURB have similar coarse-scale spatial patterns, with relatively high values in low-elevation areas on the east coasts of the North and South Island, and in the inland Waikato, Wairarapa Valley, Rangitikei-Manawatu coastal plain, Taranaki Ring Plain, and Auckland Region (Figure 4-6 to Figure 4-21). In contrast, predicted nutrient and ECOLI attribute states are generally low in major mountain ranges (e.g., Southern Alps, Kahurangi, Kaimanawa, and Tararua Ranges), in large areas of the Department of Conservation estate (e.g., Fiordland, Westland, Te Urewera, Egmont, Whanganui and Tongariro National Parks), and in smaller, native forest-dominated areas of Northland and the Coromandel Peninsula.

The low elevation areas characterised by high nutrient and ECOLI attributes coincide with land used for intensive agriculture and with most of New Zealand's urban centres. High-intensity agricultural and urban land currently account for 60% of the land area below 350 m elevation (Larned et al. 2016). Within these areas, there are some finer scaled differences in predicted attribute state. The

Canterbury Plains are characterised by high TN and NNN concentrations, and intermediate TP and DRP concentrations, and the Waikato-Hauraki Plains area is characterised by high concentrations of both nitrogen and phosphorus.

Note that the maps in Figure 4-6 to Figure 4-21 consist of nzsegments of Order 3 and above, and some extensive lowland areas are dominated by low order streams (e.g., eastern Auckland, Tauranga). Steep coastal areas of the Marlborough Sounds, Fiordland, Coromandel and Banks Peninsulas and offshore islands are also dominated by low order streams. The predicted attribute state in low order streams in these areas is not shown on the maps in Figure 4-6 to Figure 4-21.

Predicted attribute states describing DRP and TP are elevated in rivers draining catchments dominated by Tertiary Mudstones (e.g., eastern Wairarapa, and the Aorangi, Puketoi and Ruahine Ranges), and in rivers draining catchments dominated by volcanic andesites, rhyolites and ignimbrites (e.g., central volcanic plateau), as indicated in Figure 4-14 and Figure 4-16. Evidence for phosphorus enrichment due to chemical weathering in these areas comes from several studies of geochemistry and river and lake chemistry (Close and Davies-Colley 1990; Eden and Parfitt 1992; McGroddy et al. 2008; Timperley 1983). The Canterbury coast, Southland and Tasman Bay also showed elevated DRP and TP attribute states that may be associated with anthropogenic sources of phosphorus, such as fertiliser.

Large-scale geographic patterns in predicted CLAR Median and MCI Median are generally the inverse of those for chemical and microbial attribute states (Figure 4-6 and Figure 4-21). Predicted CLAR Median is relatively high and predicted MCI Median scores correspond to the excellent and good ecological states (as set out in Stark and Maxted (2007)) in mountainous areas, the Department of Conservation estate and other areas dominated by native forest land cover. Predicted CLAR Median decreases and MCI Median scores correspond to the fair and poor states in low-elevation alluvial plains and other areas dominated by intensive agriculture and urban land cover. Predicted MCI Median scores are also fair to poor in some rivers in areas dominated by exotic forest and low-intensity agriculture land cover, such as Central Otago, southwest Canterbury and the Rotorua Lakes-Lake Taupo area.

Figure 4-6: Predicted median CLAR in New Zealand rivers. Map shows all nzsegments of Order 3 and higher.

Figure 4-7: Predicted median TURB in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-8: Predicted median unadjusted NH4N in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-9: Predicted median NH4N adjusted for pH in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-10: Predicted annual maximum NH4N adjusted for pH (NH4N_pH) in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-11: Predicted median NNN concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-12: Predicted 95th percentile (Q95) NNN concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-13: Predicted median TN concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-14: Predicted median DRP concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-15: Predicted 95th percentile (Q95) DRP concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-16: Predicted median TP concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-17: Predicted median ECOLI in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-18: Predicted 95th percentile (Q95) ECOLI in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-19: Predicted proportion of samples exceeding 260 *E. coli* 100 ml^{-1} (G260) ECOLI in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-20: Predicted proportion of samples exceeding 540 *E. coli* 100 ml^{-1} (G540) ECOLI in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a $\log_{10}$ colour scale.

Figure 4-21: Predicted median MCI scores in New Zealand rivers. Map shows all nzsegments of Order 3 and higher.

5 Discussion

5.1 Comparison with previous studies

The models of current river attribute state presented in this study update previous modelling work carried out by Unwin et al. (2010), Larned et al. (2016) and Whitehead (2019). The models in the previous three reports were based on data from 1996–2007, 2009–2013 and 2013–2017, respectively, while the current models are based on data from 2016–2020. The results of the current study are generally consistent with those of earlier studies, with the structures of the models (as indicated by the relative importance of predictor variables and directions of partial plots) also similar. In addition, the performance of the models in the present study (as indicated by percent variance explained) was generally comparable to model performance from earlier studies (Larned et al. 2016; Unwin et al. 2010; Whitehead 2019).

Improvements in the modelling methodology and predictor variables between the 2010 and 2016 studies (see Larned et al. 2016) increased the performance of the random forest models. In the current study, we used the same modelling procedures and most of the predictor variables as Whitehead (2019). However, we generated new landcover predictors using 2018 landcover data (LCDB5), improving on the 2008 (LCDB3) and 2012 (LCDB4) landcover data used in the previous reports. This spatial layer represents the most current landcover data available at a national scale. We also included predictors of land use intensity associated with four different stock types (beef, dairy, deer, sheep) following Snelder et al (2022).

In a previous study, Clapcott et al. (2013) fitted an random forest model to site median MCI scores and reported a cross-validated R^2 of 0.63. Clapcott et al. (2013) also reported a cross-validated R^2 of 0.64 with an alternative technique, boosted regression trees. The equivalent R^2 statistic for the MCI model in the current study was 0.71 (Table 4-1). The improvement in performance in the current study may reflect the longer-term dataset and the inclusion of the land use intensity predictors.

5.2 Model uncertainty

In this study, we modelled broad scale patterns in attribute state using catchment characteristics and segment scale descriptors as predictor variables. Because the processes determining water quality at any location are complex, some unexplained variation in our models is to be expected. Predictions made for individual locations are associated with uncertainties characterised by model RMSD (Table 4-1). However, the level of model bias for each attribute state was low, which indicates that the predicted patterns reflect broad scale relative differences between locations.

The 95% confidence intervals for values predicted by our models for individual segments can be obtained using the following equations. Equation 2 should be used for calculating the intervals for the MCI predictions. Equation 3 should be used for calculating the intervals for the water quality variables for which the variables were $\log_{10}$ transformed prior to model fitting and the prediction uncertainty (RMSD) values have been reported in the $\log_{10}$ transformed space. Equation 4 should be used for calculating the intervals for attribute states for which the variables were logit-transformed prior to model fitting (ECOLI G260, ECOLI G540) and the prediction uncertainties are reported in the logit transformed space.

$$95\%\ CI = x \pm 1.96 \times RMSD \quad (2)$$

$$95\%\ CI = 10^{[\log_{10}(x) \pm 1.96 \times RMSD]} \quad (3)$$

$$95\%\ CI = \frac{e^{[logit(x) \pm 1.96 \times RMSD]}}{(1+e^{[logit(x) \pm 1.96 \times RMSD]})}, \text{ where } logit(x) = log\left(\frac{x}{1-x}\right) \quad (4)$$

In all equations, x is the estimated value in the original units, RMSD is the model error in the transformed space and 1.96 is the standard normal deviate or Z-score for probability ($0.025 \leq Z \geq 0.975$). The prediction confidence intervals for the $\log_{10}$-transformed variables, when expressed in the original units, are asymmetric and their values vary in proportion to the predicted water quality value. For example, if we let x be a predicted value for CLAR of 0.1 m, the lower and upper 95% confidence intervals are 0.04 and 0.25 m, respectively, whereas if x is 1.0 m the lower and upper 95% confidence intervals are 0.4 and 2.5 m, respectively.

Random forest model performance differed among attribute states (Table 4-1). This variation may be attributed to differences in the biophysical processes that control the water quality variables. Some biophysical processes may be poorly represented by our catchment-averaged spatial predictors. For example, concentrations of dissolved and total nitrogen and phosphorus in rivers are influenced to differing degrees by adsorption-desorption processes, deposition and suspension, and biological assimilation, transformation and removal; these mechanisms are not explicitly represented in the random forest models. The absence of predictors that account for these and other processes means that some level of unexplained variation is inevitable.

5.3 Alternative modelling methods

The random forest method that we used to model attribute state is well-suited to data from monitoring sites that represent a wide range of environmental conditions. However, it is not the only method available. Alternative statistical models include generalised additive models (GAMs; Hastie et al. 2001), artificial neural networks (e.g., Joy and Death 2001), and boosted regression trees (e.g., Leathwick et al. 2005). We did not employ these alternatives, but it is possible that some water quality applications would be better served by models developed by one of the alternative methods. In particular, if it is important to identify areas with potentially extreme water quality values, models such as GAMS that can extrapolate beyond the range of the fitting data would be useful, although such predictions may lead to spurious results. In addition, models that incorporate biophysical processes (e.g., CLUES; Alexander et al. 2002) are available; in some circumstances, process models are better suited to inform environmental policy. We considered random forest models to be the best tool for predicting attribute state for national scale reporting for three general reasons:

1. Spatial data that correspond to land cover and other environmental characteristics are widely available in New Zealand. These data are suitable for investigating associations between water quality and environmental characteristics, and empirical models are appropriate tools for identifying those associations. In contrast, process models require measurements or estimates of catchment processes (e.g., erosion, contaminant transport and transformation) and these data are in far shorter supply. In

addition, process models are generally more time consuming and complex to calibrate than purely empirical models.

2. Random forest model predictions can be mapped at scales ranging from single network segments to the entire county. These maps provide a useful description of spatial patterns in water quality attribute state for environmental reporting purposes.

3. Among empirical modelling methods that generate associations between water quality and environmental characteristics, random forest models have several advantages: they are minimally affected by multi-collinearity among predictor variables, they cannot be over-fitted, and they are unaffected by variation in data distributions. Random forest models cannot predict beyond the range of the observations, which may limit their utility in some applications. In the present study, limiting model predictions to the range of observations was a positive attribute as it ensured that those predictions were conservative.

6 Acknowledgements

Thanks to Isaac Bain (Ministry for the Environment), Lance Morell and Sonja Miller (Stats NZ Tatauranga Aotearoa) for project support and feedback. Many thanks to the LAWA Refresh Team [Abi Loughnan (Otago Regional Council), Eric Goodwin (Cawthron), Jane Groom and Coral Grant (Auckland Council)] for data updates and advice, as well as support from Juliet Milne and Scott Larned (NIWA). Finally, many thanks to the many regional council staff who provided additional data and information about monitoring programmes.

7 Glossary of abbreviations and terms

AnnMax	The annual maximum of a given state variable
APC	Agricultural production census
Attribute state	The combination of water quality variable and statistic (e.g., 95th percentile of dissolved reactive phosphorus; DRP Q95).
CART	Classification and regression trees
CLAR	Visual clarity (m)
DEM	Digital elevation model
DRP	Dissolved reactive phosphorus (mg l^{-1})
ECOLI	*Escherichia coli* (cfu 100 ml^{-1})
G260	The proportion of ECOLI samples that exceed 260 per 100 ml
G540	The proportion of ECOLI samples that exceed 540 per 100 ml
LCDB5	Land Cover Database (Version 5.0)
LRI	Land Resources Inventory
MCI	Macroinvertebrate community index
MSE	Mean square error
NH4N	Ammoniacal nitrogen (mg l^{-1})
NH4N_pH	Ammoniacal nitrogen adjusted for pH (mg l^{-1})
NNN	Nitrate + nitrite-nitrogen (mg l^{-1})
NPS-FM	National Policy Statement for Freshwater Management
NRWQN	National River Water Quality Network
NSE	Nash-Sutcliffe Efficiencies
NTU	Nephelometric turbidity units
OOB	Out-of-bag observations
Q95	The 95th percentile of a given state variable
REC2.4	River Environmental Classification (Version 2.4)
RMSD	Root mean square deviation
RSR	Relative root mean square error
SoE	State-of-environment
TN	Total nitrogen (unfiltered; mg l^{-1})
TP	Total phosphorus (unfiltered; mg l^{-1})
TURB	Turbidity (NTU)

8 References

Breiman, L. (1984) *Classification and regression trees*. Wadsworth International Group, Belmont, Calif, 358 pp.

Breiman, L. (2001) Random Forests. *Machine Learning*, 45: 5–32.

Clapcott, J.E., Goodwin, E., Snelder, T.H. (2013) *Predictive models of benthic macroinvertebrate metrics*. Cawthron Institute Report 2301. Nelson.

Close, M.E., Davies-Colley, R.J. (1990) Baseflow water chemisty in New Zealand rivers 2. Influence of environmental factors. *New Zealand Journal of Marine and Freshwater Research*, 24: 343–356.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J. (2007) Random forests for classification in ecology. *Ecology*, 88: 2783–2792.

Eden, D., Parfitt, R. (1992) Amounts and sources of phosphate in hill country rocks, south-eastern North Island, New Zealand. *Soil Research*, 30: 357–370.

Hastie, T.J., Tibshirani, R., Friedman, J.H. (2001) *The elements of statistical learning, data mining, inference and prediction*. 2nd ed. Springer-Verlag, New York.

Larned, S.T., Scarsbrook, M.R., Snelder, T.H., Norton, N.J., Biggs, B.J.F. (2004) Water quality in low-elevation streams and rivers of New Zealand: Recent state and trends in contrasting land-cover classes. *New Zealand Journal of Marine and Freshwater Research*, 38: 347–366.

Larned, S.T., Snelder, T.H., Unwin, M.J., McBride, G.B. (2016) Water quality in New Zealand rivers: current state and trends. *New Zealand Journal of Marine and Freshwater Research*, 50: 389–417.

Larned, S.T., Unwin, M.J. (2012) *Representativeness and statistical power of the New Zealand river monitoring network. National Environmental Monitoring and Reporting: Network Design, Step 2*. NIWA Client Report CHC2012-079 prepared for Ministry for the Environment. Christchurch.

Leathwick, J.R., Overton, J.M., McLeod, M. (2003) An environmental domain classification of New Zealand and its use as a tool for biodiversity management. *Conservation Biology*, 17: 1612–1623.

Leathwick, J.R., Snelder, T.H., Chadderton, W.L., Elith, R.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.

Liaw, A., Wiener, M. (2002) Classification and Regression by randomForest. *R News*, 2: 18–22.

McGroddy, M.E., Baisden, W.T., Hedin, L.O. (2008) Stoichiometry of hydrological C, N, and P losses across climate and geology: An environmental matrix approach across New Zealand primary forests. *Global Biogeochemical Cycles*, 22: 1–14.

Moriasi, D., Gitau, M., Pai, N., Daggupati, P. (2015) Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Transactions of the ASABE*, 58: 1763–1785.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L. (2007) Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, 50: 885–900.

Nash, J.E., Sutcliffe, J.V. (1970) River flow forecasting through conceptual models: Part I — A

discussion of principles. *Journal of Hydrology*, 10: 282–290.

New Zealand Government (2020) *National Policy Statement for Freshwater Management 2020*. Ministry for the Environment, Wellington.

Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M. (2008) How to evaluate models: Observed vs. predicted or predicted vs. observed?. *Ecological Modelling*, 216: 316–322.

R Core Team (2021) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Snelder, T.H., Fraser, C., Larned, S.T., Monaghan, R., De Malmanche, S., Whitehead, A.L. (2022) Attribution of river water-quality trends to agricultural land use and climate variability in New Zealand. *Marine and Freshwater Research*, 73: 1–19.

Stark, J.D., Maxted, J.R. (2007) *A user guide for the Macroinvertebrate Community Index*. Cawthron Report 1166. Cawthron Institute, Nelson.

Svetnik, V., Liaw, A., Tong, C., Wang, T. (2004) Application of Breiman's random forest to modeling structure-activity relationships of pharmaceutical molecules. In: Roli, F., Kittler, J., and Windeatt, T. (Eds). *Multiple Classifier Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg: 334–343.

Timperley, M.H. (1983) Phosphorus in spring waters of the Taupo Volcanic Zone, North Island, New Zealand. *Chemical Geology*, 38: 287–306.

Unwin, M.J., Snelder, T.H., Booker, D.J., Ballantine, D., Lessard, J. (2010) *Predicting water quality in New Zealand rivers from catchment-scale physical, hydrological and land cover descriptors using Random Forest models*. NIWA Client Report CHC2010-037. NIWA, Christchurch.

Whitehead, A.L. (2019) *Spatial modelling of river water-quality state: incorporating monitoring data from 2013 to 2017*. NIWA Client Report 2018360CH prepared for the Ministry for the Environment. NIWA, Christchurch.

Whitehead, A.L., Fraser, C., Snelder, T.H., Walter, K., Woodward, S., Zammit, C. (2021) *Water quality state and trends in New Zealand rivers: analyses of national data ending in 2020*. NIWA Client Report 2021296CH prepared for Ministry for the Environment. NIWA, Christchurch.