Spatial modelling of lake water quality state

Incorporating monitoring data for the period 2016 to 2020

February 2022

Prepared By:

Ton Snelder

Caroline Fraser

Amy Whitehead (NIWA)

For any information regarding this report please contact:

Ton Snelder

Phone: 0275758888
Email: ton@lwp.nz

LWP Client Report Number:	2021-15
Report Date:	February 2022
LWP Project:	2021-15

Quality Assurance Statement

Version	Reviewed By	
Draft October 2021		
Final February 2022	Scott Larned (NIWA)	

Table of Contents

Figures

Tables

Executive Summary

Measurements of current lake attribute states comprising statistics calculated from observations of chlorophyll *a* (CHLA, total phytoplankton biomass), total nitrogen (TN), total phosphorus (TP), ammoniacal nitrogen (NH4N), Secchi depth (SECCHI), *Escherichia coli* (ECOLI) and two trophic level indices (TLI3 and TLI4) were obtained for between 55 to 124 lakes, depending on the variable, from regional council state-of-the-environment monitoring programmes, for the period 2016 to 2020. For each variable, the attributes 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 attribute states of all monitored lakes were calculated from the time series of these variables representing the 5-year period ended 2020. The relevant attributes for lakes are defined in the National policy statement for freshwater management. The attribute states were calculated for each lake, then combined with environmental data describing the lakes and their catchments to make spatial predictions for all 3,821 lakes in New Zealand that are larger than 1 hectare.

Good to very good performance was achieved for six of the lake spatial models (TLI3, TLI4, SECCHI_Median, TN_Median, TP_Median and ECOLI_Median); the spatial models for CHLA_Median, CHLA_AnnMax, NH4N_Median and ECOLI_Q95 had satisfactory performance. The models for NH4N_adj_Median, NH4N_adj_AnnMax, ECOLI_G260 and ECOLI_G540 had unsatisfactory performance indicating the predictions from these models should be used very cautiously.

The predicted values for all 14 attributes had similar spatial patterns, with high values of CHLA, ECOLI, TN, TP, NH4N, TLI3, and TLI4 and low values of SECCHI, in low-elevation lakes on the coasts of the North and South Island. Predicted values of CHLA, TN, TP, NH4N, and TLI3 were also high in inland areas of both islands that are dominated by agricultural land use (e.g., Southland, parts of Otago, Hawkes Bay, Bay of Plenty, Waikato, Northland. Predicted values of CHLA, ECOLI, TN, TP, NH4N, TLI3, and TLI4 were generally low and Secchi depth was high in inland areas of the South Island.

The predictions are uncertain at the lake-scale and actual data should be used in preference to the modelled predictions when evaluating individual lakes. However, the broader-scale predictions will be useful for strategic purposes such as identifying high-priority areas for interventions.

1 Introduction

Lake water quality across New Zealand was characterised in terms of attributes as defined by the National policy statement for freshwater management (NPS-FM , NZ Government, 2020) by a recent national analysis of state and trends at monitored lakes (Whitehead *et al.*, 2021). The lakes are monitored as part of the State of Environment (SOE) programmes operated by regional councils. The datasets used for these analyses contained quarterly or monthly measurements of physical, chemical, and biological variables over time periods that started as early as 1990 and extended to 2020. For each variable, the attributes comprised at least one statistic describing an aspect of the distribution of the observed values (e.g., median value, 95^{th} percentile value, annual maximum value). The attribute states of all monitored lakes were calculated from the time series of these variables representing the 5-year period ended 2020 (Whitehead *et al.*, 2021).

The objective of this study was to estimate and map the current state of 14 attributes in all of the large (> 1 ha) lakes 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.

This study used the lake water quality monitoring data from the Whitehead *et al.* (2021) study. Those data were used to develop spatial models that predict attribute state in all large lakes. The benefit of spatial modelling is that it provides a large-scale assessment of lake attribute state that is more representative than assessments based on aggregating raw monitoring lake data. The latter approach can lead to conclusions about water quality patterns that are biased by the non-random locations of monitored lakes (Larned *et al.*, 2014).

This report is a companion to the primary output for the project, which is a supplementary file containing the outputs (predictions of attribute state for all large lakes) from the spatial models. This report provides a detailed description of the methods used to define spatial models of lake water quality and to produce predictions for unmonitored lakes. The methods used to prepare the water quality data, make assessments of the representativeness of the monitored lakes, and undertake the spatial modelling are described. The results section of the report provides national maps of predicted lake attribute states. Measures of model performance and the important relationships between predictors and attribute state are described. A short discussion is provided with a minimal interpretation of the results.

2 Data

2.1 Lake State Data

We used the SOE data for lakes analysed by Whitehead *et al.* (2021) for the current study. That report has detailed methods for obtaining and grooming the data. The lake SOE data analysed by Whitehead *et al.* (2021) included lake attributes that are measures of physical, chemical, microbiological and biological conditions. Each attribute is a statistic that is derived from a variable. For example, lake attributes include the median values of the variables total nitrogen (TN) and total phosphorus (TP) and the median and annual maximum values of the variables chlorophyll a (CHLA) and ammoniacal nitrogen adjusted for pH (NH4N_adj) (Table 1).

This study included modelling of the current state of 14 attributes (Table 1). The variables NO3N and DRP were not included in this study because the number of lake sites for which data were available was small (<35 sites) and spatial coverage was poor. Therefore, these variables are poorly represented at the national scale and predictions from models developed with these data would have low accuracy.

This study used observed values (i.e., derived from monitoring data) of attribute states for the five-year period from 2016 to 2020. Two inclusion rules were applied to ensure that the data were representative of each lake and variable, following the same approach as Snelder *et al.* (2016) and Whitehead *et al.* (2021). First, if the sampling frequency was monthly, at least 30 samples over the five-year period (i.e., 50% of sample occasions) were required for the site and variable combination to be included. Second, if the sampling frequency was quarterly, at least 10 samples over the five-year period (i.e., 50% of sample occasions) were required for the site and variable combination to be included.

A summary of the number of lakes per attribute included in this study is in Table 1. The two inclusion rules were more lenient than the rules used in Whitehead *et al.* (2021), which required lake × attribute combinations in the state analyses to have measurements for at least 80% of the years (four out of five years) and at least 80% of the seasons in the period (either 48 of 60 months, or 16 of 20 quarters). The modified inclusion rules in the current study increased the number of lakes for which attribute state was assessed compared to the Whitehead *et al.* (2021) report (Table 1). Table 1 also provides the numbers of lakes used in the previous lake spatial modelling study (Fraser and Snelder, 2019), which used similarly lenient filtering rules, for comparison.

Table 1. Lake attributes included in this study. Note that the combination of the variable and the statistic defines an attribute. NM: not modelled

Attribute type	**Variable**	**Abbreviation**	**Statistic**	**Units**	**Number of lakes**		
					This study	**State and trends[1]**	**Previous spatial modelling[2]**
Physical	Secchi depth	SECCHI	Median	m	75	70 (85)	61
Chemical	Total nitrogen	TN	Median	g m^{-3}	124	83 (101)	104
	Total phosphorus	TP	Median	g m^{-3}	124	83 (101)	97
	Ammoniacal nitrogen adjusted for pH	NH4N_adj	Median, Ann. Max	g m^{-3}	80	69 (82)	NM
	Ammoniacal nitrogen	NH4N	Median	g m^{-3}	97	81 (96)	64
Phytoplankton	Chlorophyll *a*	CHLA	Median, Ann. Max	g m^{-3}	124	83 (103)	92
Microbiological	*Escherichia coli*	ECOLI	Median, Q95	MPN 100ml^{-1}	55	26 (40)	NM
			G260, G540	% exceedance			NM
Water quality index	Trophic Level Index 3	TLI3	Median	unitless	124	83 (101)	99
	Trophic Level Index 4	TLI4	Median	unitless	75	69 (84)	NM

Note: (1) The numbers of lakes included in the companion state and trends report (Whitehead *et al.*, 2021).

(2) The numbers of lakes included in the Fraser and Snelder (2019) study.

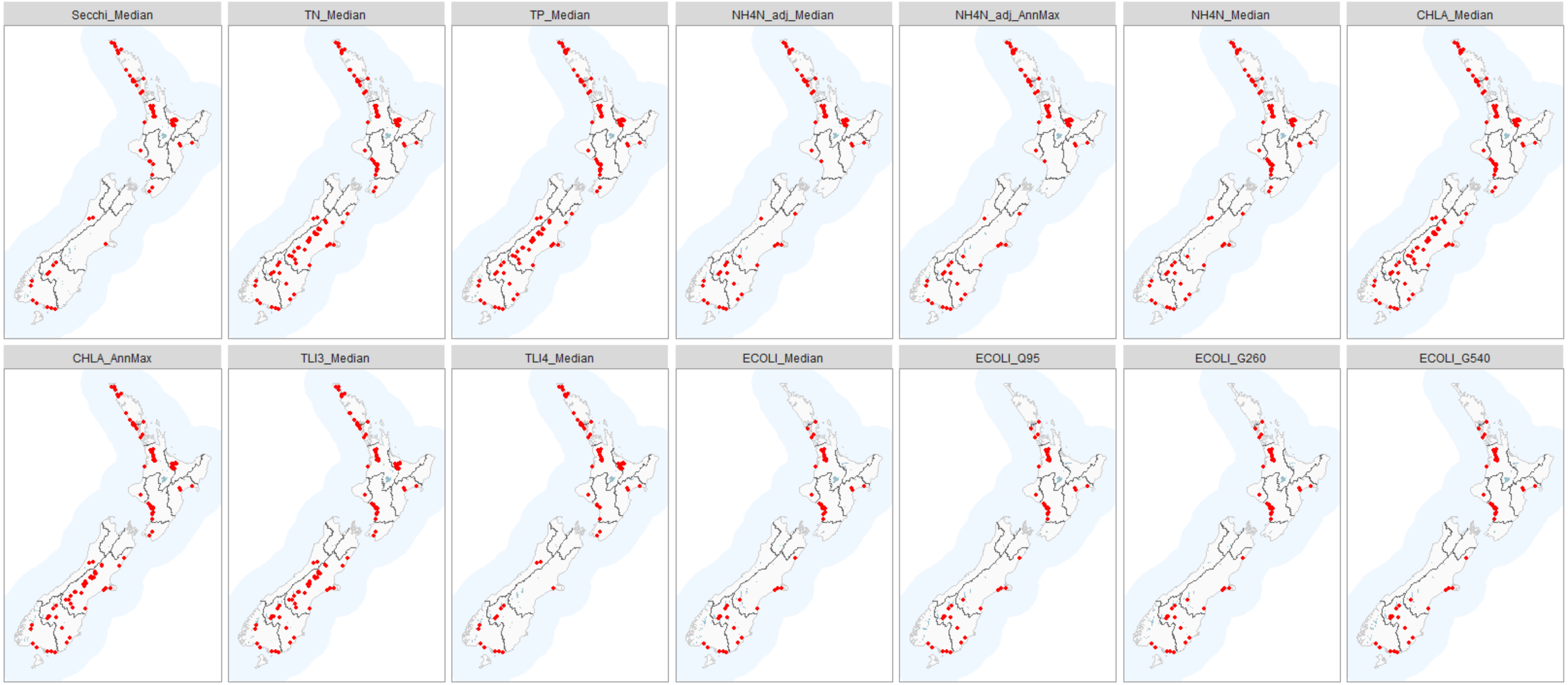

Figure 1. Locations of monitored lakes used in this study for each of the 14 attributes. The locations shown on each panel correspond to the lakes that were included in this study for each attribute listed in Table 1.

LWP

2.2 Lake predictor data

The FENZ database provides characteristics of all 3821 lakes in New Zealand that are greater than one hectare in area. Details of these variables and their derivation are provided by Snelder *et al.* (2006). Characteristics include descriptors of climatic, geological, topographic, bathymetric, land cover, and hydrological conditions in New Zealand lakes and their catchments.

In this study, land cover data variables for each lake were extracted from the Land Cover Database Version 5.0 (LCDB5)[1]. LCDB5 differentiates 34 categories based on analysis of satellite imagery from the 2018–2019 summer. We collapsed these into six super categories as shown in Table 2. In addition, for each lake catchment, we obtained five predictors representing land use intensity. The land use intensity predictors were derived in four steps from animals in the four pastoral stock type categories that are periodically surveyed on all livestock farms as part of the agricultural production census (APC). First, we intersected animal numbers for the 2017 census year that was provided by Statistics New Zealand at the mesh block level[2] with a coverage describing the catchments of all 3821 lakes. Second, the animal numbers within each lake catchment were expressed as a standardised stock unit (SU), which is a commonly used measure of metabolic demand by livestock in New Zealand (Parker, 1998). Third, the numbers of animals were converted to total stocking density (SU ha^{-1}) by multiplying each stock type by the stock unit equivalent shown in Table 3 (Snelder *et al.*, 2021) and dividing by the catchment area. Fourth, we evaluated the proportion of the total SUs associated with each stock type in the catchment as the stocking density of animal of each type shown in Table 3 divided by the total stocking density.

[1] https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/
[2] https://datafinder.stats.govt.nz/layer/105176-meshblock-2021-generalised/

Table 2. Predictor variables used in the spatial models of lake attribute state.

Predictor	Abbreviation	Description	Unit
Lake	lkArea	Lake surface area	m^2
	lkDistCoast	Straight line distance to coast	km
	lkDepth	Estimated average lake depth	m
	lkElev	Lake elevation	m ASL
Catchment topography	catSlope	Catchment average slope	Degrees
	catArea	Catchment area	m^2
	catElev	Catchment elevation	m ASL
Climate and flow	lkDecSolRad	Lake summer (December) solar radiation	W m^{-2}
	lkJuneSolRad	Lake winter (June) solar radiation	W m^{-2}
	lkDecTemp	Lake average summer (December) air temperature	Degrees
	lkJunTemp	Lake average winter (June) air temperature	Degrees
	lkFetch	Lake wind fetch	m
	lkSumWind	Lake summer (December) wind speed	m s^{-1}
	lkWinWind	Lake winter (June) wind speed	m s^{-1}
	catSumTemp	Catchment average summer (December) air temperature	Degrees
	catWinTemp	Catchment average winter (June) air temperature	Degrees
	catFlow	Catchment average annual discharge	m^3 yr^{-1}
Geology	catPhos	Catchment average phosphorus	Ordinal*
	catCalc	Catchment average calcium	Ordinal*
	catHard	Catchment average induration or hardness value	Ordinal*
	catPsize	Catchment average particle size	Ordinal*
	catPeat	Proportion of catchment occupied by peat	Proportion
	catAlluv	Proportion of catchment occupied by alluvium	Proportion
Land cover	usIntensiveAg	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)	
	usPastoralLight	Proportion of catchment in low producing grassland (LCDB5 class 41)	Proportion
	usNativeForest	Proportion of catchment in native forest (LCDB5 class 69)	Proportion
	usUrban	Proportion of catchment in built-up areas, urban parkland, surface mines, dumps and transport infrastructure (LCDB5 classes 1, 2, 6, 5)	Proportion
	usExoticForest	Proportion of catchment in exotic forest (LCDB3 class 71)	Proportion
	usBare	Proportion of catchment occupied in bare or lightly-vegetated cover (LCDB5 classes 10, 12, 14, 15, 16)	Proportion
Land use intensity	TotalSUDensity	Catchment density of total stock units (SU)	SU/ha
	usBeef	Proportion of total stock units attributable to beef cows in catchment	Proportion
	usDairy	Proportion of total stock units attributable to dairy cows in catchment	Proportion
	usDeer	Proportion of total stock units attributable to deer in catchment	Proportion
	usSheep	Proportion of total stock units attributable to sheep in catchment	Proportion

*Geological variables are based on regolith, using averages of ordinal values assigned to LRI top-rock categories by (Leathwick *et al.*, 2003). The variables catHard and catPsize characterise physical regolith conditions; and catPhos and catCalc characterises regolith fertility.

Table 3. Stock unit equivalent values assumed for sheep, beef, dairy and deer livestock classes in 2017. The values in this table are the 2017 values used by the study of Snelder et al. *(2021).*

Stock type	Stock unit equivalents per animal
Sheep	1.35
Beef	6.9
Dairy	8
Deer	2.3

3 Modelling Methods

3.1 Random forest models

We fitted the observed values of the 14 attributes associated with the monitored lakes to the predictors using random forest (RF) models (Breiman, 2001; Cutler *et al.*, 2007). A RF model is an ensemble of individual classification and regression trees (CART). In a regression context, CART partitions observations (in this case the individual water quality variables) 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). RF 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 RF 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 RF model produces a limiting value of the generalization error (i.e., the model maximises its prediction accuracy for previously unseen data; Breiman, 2001). The generalization 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 predictors 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, RF models cannot be expressed as equations. However, the relationships between predictor and response represented by RF models can be represented by importance measures and partial dependence plots (Breiman, 2001; Cutler *et al.*, 2007). During the fitting process, RF 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 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 predictor 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).

RF models include any of the original set of predictors that are chosen during the model fitting process. However, marginally important predictors may be redundant (i.e., their removal does not affect model performance) and their inclusion complicates 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 mean square 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 successive model. The final, 'reduced' model is defined by the "one standard error rule" 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) (Breiman *et al.*, 1984). Importance levels for predictors were not recalculated at each reduction step to avoid over-fitting (Svetnik *et al.*, 2004).

Although RF models do not depend on distributional assumptions, transformation of the response variable to an approximately symmetric distribution can improve model performance. We investigated transformations of the modelled water quality (i.e., response) variables on the model performance. Where performance was improved, we made predictions using these models.

All calculations were performed in the R statistical computing environment (R Development Core Team 2009) using the *randomForest* package and other specialised packages.

3.2 Model performance

Model performance was assessed by comparing observations with independent predictions (i.e., lakes that were not used in fitting the model), which were obtained from the OOB observations. We summarised the model performance using five statistics: regression R^2; Nash-Sutcliffe efficiency (NSE); percent bias (PBIAS); the relative root mean square deviation (RSR); and the 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 indicates the proportion of the total variance explained by the model, but is not a complete description of model performance (Piñeiro *et al.*, 2008).

NSE indicates how closely the observations coincide with predictions (Nash and Sutcliffe, 1970). NSE values range from $-\infty$ to 1. A NSE of 1 corresponds to a perfect match between predictions and the observations. An NSE of 0 indicates the model is only as accurate as the mean of the observed data and values less than 0 indicate the model predictions are less accurate than using the mean of the observed data.

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 4.

Table 4: 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 ≤ 0.65	15 ≤ |PBIAS| < 20
Satisfactory	$0.30 < R^2 \leq 0.60$	0.35 < NSE ≤ 0.50	20 ≤ |PBIAS| < 30
Unsatisfactory	$R^2 < 0.30$	NSE ≤ 0.35	|PBIAS| ≥ 30

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

3.3 Representativeness of monitored lakes used in RF models

A graphical comparison was used to gauge how well the monitored lakes used to fit the RF models represented environmental variation at the national scale. Here, representativeness refers to the degree to which the distribution of monitored lakes over the range of an environmental predictor variable matches the distribution of all lakes over the range of the same environmental variable. Poor representativeness can reduce the reliability of the model predictions because certain sets of environmental conditions are not represented in the fitting data.

Histograms of the proportions of monitored lake numbers over the ranges of the most important predictor variables in the RF models (i.e., the predictors with the greatest explanatory power) were visually compared with histograms of the proportions of all lakes over the same predictor variables. Note that representativeness of monitored lakes is different from model bias, which is defined in Section 3.2.

3.4 Model predictions

Predictions are made with RF models by “running” new cases down every tree in the fitted forest and averaging the predictions made by each tree (Cutler *et al.*, 2007). Some of the models in this study were fitted to $\log_{10}$-transformed data and when the model predictions were back-transformed, we corrected for retransformation bias using the smearing estimate (Duan, 1983; Equation 1, but using base 10, not base *e*). The back-transformed predictions were used to produce national maps depicting the variation in each attribute.

4 Results

4.1 Model performance

The performance of the models of all attributes, apart from TLI3, TLI4, *E. coli* G260 and *E. coli* G540, was improved by $\log_{10}$-transformation of the values of the water quality variables (the model responses). The raw variable distributions were strongly right-skewed and the transformations made these more symmetric. The distributions of the measured values of the attributes *E. coli* G260 and *E. coli* G540 could not be transformed to approximate normality because for 51% and 67% of lakes respectively, these values were zero.

The RF models of SECCHI_Median, TP_Median TLI3, TLI4, and ECOLI_Median had good performance and the RF model for TN_Median, had very good performance, as indicated by R^2 and NSE and the criteria of Moriasi *et al.* (2015) (Table 5, Figure 2). The models for CHLA_Median, CHLA_AnnMax, NH4N_Median and ECOLI_Q95 had satisfactory performance. The models for NH4N_adj_Median, NH4N_adj_AnnMax, ECOLI_G260 and ECOLI_G540 had unsatisfactory performance based on the criteria of Moriasi *et al.* (2015). All 14 models had very low bias (PBIAS; Table 5). RMSD values provide an indication of the magnitude of the characteristic error in the transformed units of each variable.

Table 5. Performance of the models representing the 14 attributes. Performance was determined using independent predictions (i.e., lakes that were not used in fitting the models) generated from the out-of-bag observations. R^2 = coefficient of determination of observation versus predictions, NSE = Nash-Sutcliffe efficiency, PBIAS = percent bias, RMSD = root mean square deviation. RMSD units are the $\log_{10}$-transformed original units for all variables except TLI3 and TLI4, which were not transformed. The colours indicate the performance ratings shown in Table 4.

Attribute	N	R^2	NSE	PBIAS	RMSD	Rating
Secchi_Median	75	0.64	0.61	-0.08	0.31	Good
TN_Median	124	0.78	0.77	3.66	0.23	Very good
TP_Median	124	0.62	0.62	0.44	0.35	Good
NH4N_adj_Median	80	0.31	0.31	0.74	0.35	Unsatisfactory
NH4N_adj_AnnMax	80	0.29	0.29	0.71	0.55	Unsatisfactory
NH4N_Median	97	0.51	0.51	0.76	0.34	Satisfactory
CHLA_Median	124	0.45	0.45	-0.82	0.41	Satisfactory
CHLA_AnnMax	124	0.58	0.57	-0.42	0.45	Satisfactory
TLI3	124	0.65	0.65	0.01	0.83	Good
TLI4	75	0.66	0.66	0.46	0.75	Good
ECOLI_Median	55	0.65	0.64	-0.21	0.34	Good
ECOLI_Q95	55	0.38	0.38	0.09	0.53	Satisfactory
ECOLI_G260	55	0.31	0.29	7.52	0.07	Unsatisfactory
ECOLI_G540	55	0.30	0.30	3.50	0.04	Unsatisfactory

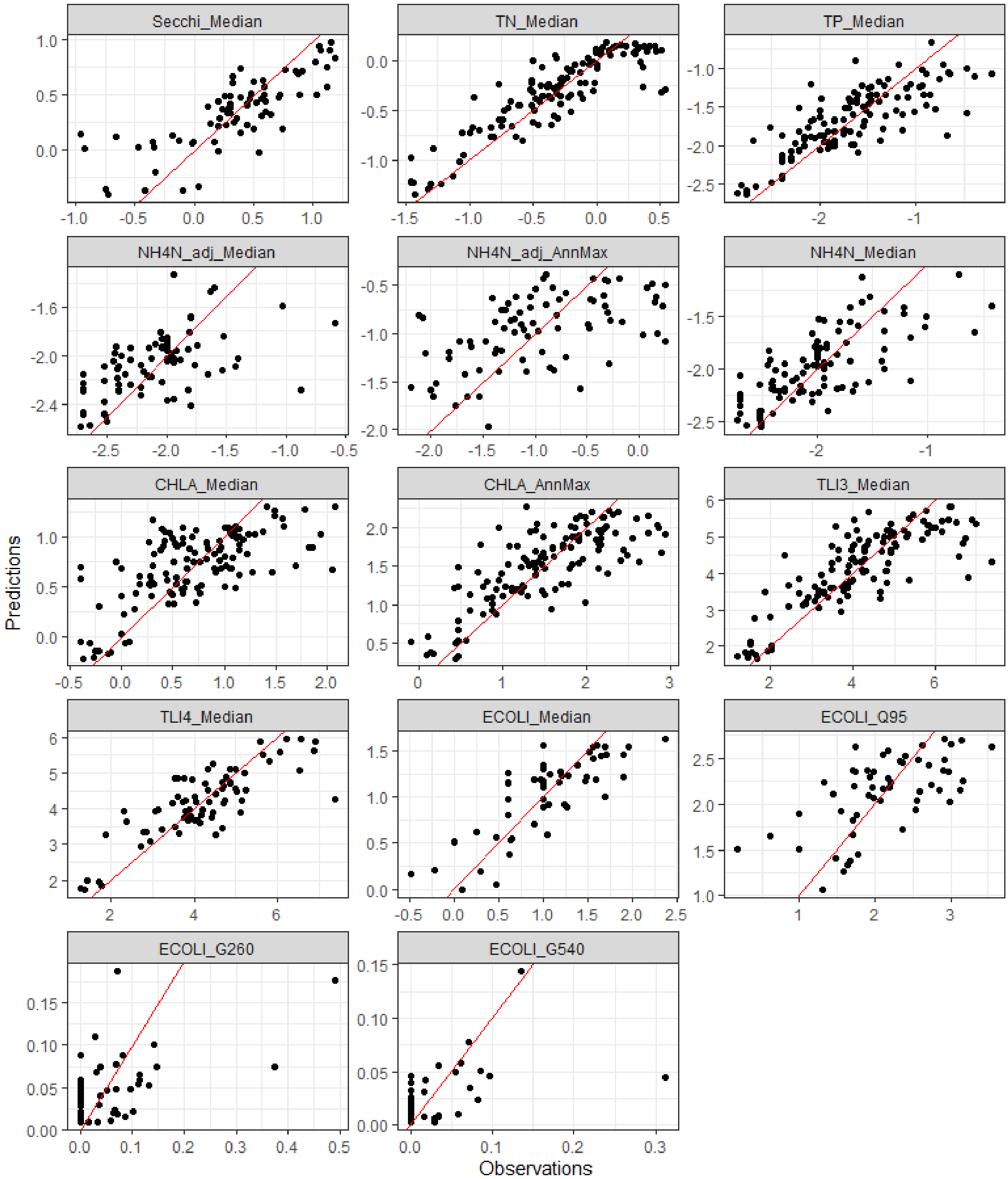

Figure 2. Comparison of observed attribute states versus values predicted by the RF models. Note that the observed values are plotted on the Y-axis and predicted values on the X-axis, following Piñeiro et al. (2008). Red line is one-to-one. Units for the attributes are the log10 of the original units for all attributes except TLI3, TLI4, ECOLI_G260 and ECOLI_G540, which are untransformed units.

4.2 Modelled relationships

Each of the reduced RF models retained only a subset of the original set of predictors (Table 6). The retained predictors (26 to 3 per model) reflected associations between water quality and lake and catchment elevation, geological and climatic factors (Table 6).

The lake attribute states had logical relationships with many of the individual predictor variables included in the reduced RF models (Figure 3). Nutrient concentrations and chlorophyll *a* increased and Secchi depth decreased with increasing catchment area occupied by intensive agricultural land cover (lcdb5_usIntensiveAg) and total stock unit density (TotalSUDensity). Nutrient concentrations and chlorophyll *a* decreased and Secchi depth increased with increasing lake elevation (lkElev). This is consistent with an observed gradient in trophic conditions for lakes that is associated with altitude and associated climatic conditions (Sorrell *et al.*, 2006). TLI3, TLI4, TN and TP decreased with lake area (lkArea), which may reflect the generally lower trophic status of larger lakes rather than the effect of area on stratification and wind mixing on lakes.

Table 6. Predictors retained by the reduced RF models of the 14 lake attributes in overall order of importance. The values indicate the rank importance of the predictor for the individual models. NA indicates that the predictor was not included in the reduced model. Predictors are defined in Table 2.

Predictor	Attribute													
	Secchi_Median	TN_Median	TP_Median	NH4N_adj_Median	NH4N_adj_AnnMax	NH4N_Median	CHLA_Median	CHLA_AnnMax	TLI3_Median	TLI4_Median	ECOLI_Median	ECOLI_Q95	ECOLI_G260	ECOLI_G540
lcdb5_usIntensiveAg	2	1	1	1	1	1	1	1	1	1	3	NA	1	1
lkElev	1	2	5	3	6	9	8	4	3	2	1	1	4	2
lcdb5_usBare	NA	3	4	12	3	13	9	5	4	NA	NA	3	6	3
TotalSUDensity	8	4	2	6	5	3	2	2	2	5	NA	NA	NA	NA
lkArea	3	5	10	NA	7	NA	3	3	5	NA	NA	NA	2	NA
catPsize	NA	8	NA	4	12	4	NA	NA	14	NA	2	2	5	NA
lkDepth	4	7	9	NA	9	NA	7	6	9	4	NA	NA	NA	NA
catArea	9	6	6	NA	NA	11	5	7	6	NA	NA	NA	NA	NA
catHard	7	NA	NA	2	10	2	NA	NA	18	NA	NA	NA	3	NA
lkWinWind	5	11	NA	9	NA	NA	6	12	13	3	NA	NA	NA	NA
catCalc	NA	NA	3	NA	4	NA	NA	8	8	NA	NA	NA	NA	NA
lkDecTemp	10	NA	NA	10	8	7	4	10	11	NA	NA	NA	NA	NA
catGlacial	NA	9	13	NA	NA	NA	12	9	7	6	NA	NA	NA	NA
lkJuneSolRad	6	NA	8	NA	13	8	NA	11	12	NA	NA	NA	NA	NA
lcdb5_usNativeForest	NA	10	NA	7	2	6	NA	13	21	NA	NA	NA	NA	NA
Prop_Dairy	NA	NA	NA	5	11	10	NA	NA	20	NA	NA	NA	NA	NA
catPhos	NA	NA	7	NA	NA	12	10	NA	16	NA	NA	NA	NA	NA
lkSumWind	11	12	NA	8	NA	NA	NA	NA	15	NA	NA	NA	NA	NA
lcdb5_usPastoralLight	NA	NA	NA	NA	NA	NA	11	NA	10	NA	NA	NA	NA	NA
catAlluv	NA	13	NA	11	NA	5	NA	NA	22	NA	NA	NA	NA	NA
lcdb5_usExoticForest	12	NA	12	13	NA	NA	NA	NA	23	NA	NA	NA	NA	NA
Prop_Sheep	NA	NA	11	NA	NA	NA	NA	NA	17	NA	NA	NA	NA	NA
lcdb5_usUrban	NA	NA	NA	NA	NA	NA	13	NA	24	NA	NA	NA	NA	NA
Prop_Deer	13	NA	NA	NA	NA	NA	NA	NA	25	NA	NA	NA	NA	NA
Prop_Beef	NA	NA	NA	NA	NA	NA	NA	NA	19	NA	NA	NA	NA	NA
catPeat	NA	NA	NA	NA	NA	NA	NA	NA	26	NA	NA	NA	NA	NA

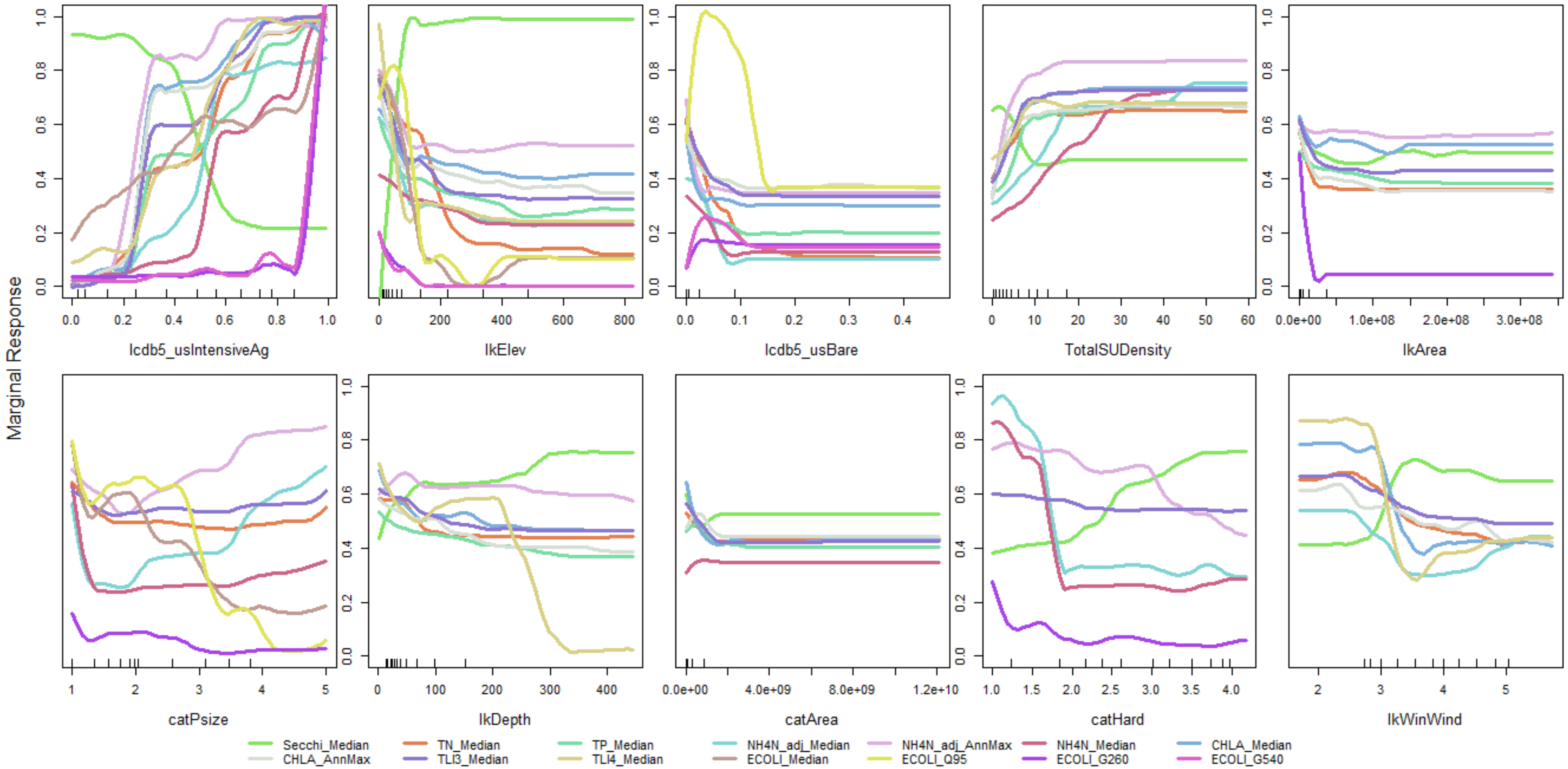

Figure 3. Partial plots for the ten most important predictor variables in the RF models of the 14 lake attributes. Each panel corresponds to one predictor. The Y-axis is the standardised value of the marginal response for each of the 14 modelled attributes. In each case, the original marginal responses over all ten predictors were standardised to have a range between zero and one. Plot amplitude (the range of the marginal response on the Y-axis) is directly related to a predictor's importance; amplitude is large for predictor variables with high importance.

4.3 Monitored lake representativeness

The distributions of monitored lakes across the environmental gradients retained in the reduced RF models were generally consistent with the distribution of all lakes nationally across the same gradients (Figure 4). For some environmental gradients, there was moderate over- and under-representation. Monitored lakes (represented by the blue histograms in Figure 4) were over-represented in environments characterised by low elevations (lkElev), and catchments with high alluvium (catAlluv) (Figure 4). Monitored lakes were under-represented in lakes with low area (lkArea) and shallow depths (lkDepth). They were also under-represented in lake catchments with low summer temperature (lkDecTemp). For example, there were no lakes in our dataset with values of lkDecTemp < 10 ^{o}C, however, 17% of lakes nationally have values of lkDecTemp in this category.

Figure 4. Histograms comparing the distributions of the 20 most important predictors used to build the RF models for all lakes and the monitored lakes. The national pool of lakes is represented by the red histograms and the monitored lakes used for RF models are represented by the blue histograms. Similarities in the distributions shown in the two histograms in each panel provide an indication of the degree to which environmental variation across the monitored lakes represent environmental variation across all lakes in New Zealand; complete representativeness would be indicated by exact matches between the histograms. The figure shows the 20 most important predictors (defined in Table 2) retained in the reduced RF models.

4.4 Model predictions

Predictions of the current state of all 14 attributes are shown in Figure 5 to Figure 18 for the 3802 lakes that had complete data in the FENZ dataset. The mapped predictions for all attributes had similar spatial patterns, with high values of CHLA, ECOLI, NH4N, TN, TP, TLI3 and TLI4 and and low values of SECCHI, in low-elevation areas on the coasts of the North and South Island, apart from areas with little or no pastoral land cover (e.g., Fiordland). Values of CHLA, NH4N, TN, TP and TLI3 were also high and values of SECCHI were low further inland in areas of both islands that are dominated by agricultural land use such as Southland, parts of Otago, Hawkes Bay, Bay of Plenty, Waikato and Northland (Figure 5 to Figure 18). Values of CHLA, ECOLI, NH4N, TN, TP and TLI3 were generally low and SECCHI high in inland areas of the South Island. Full tables of predictions are provided in a supplementary file.

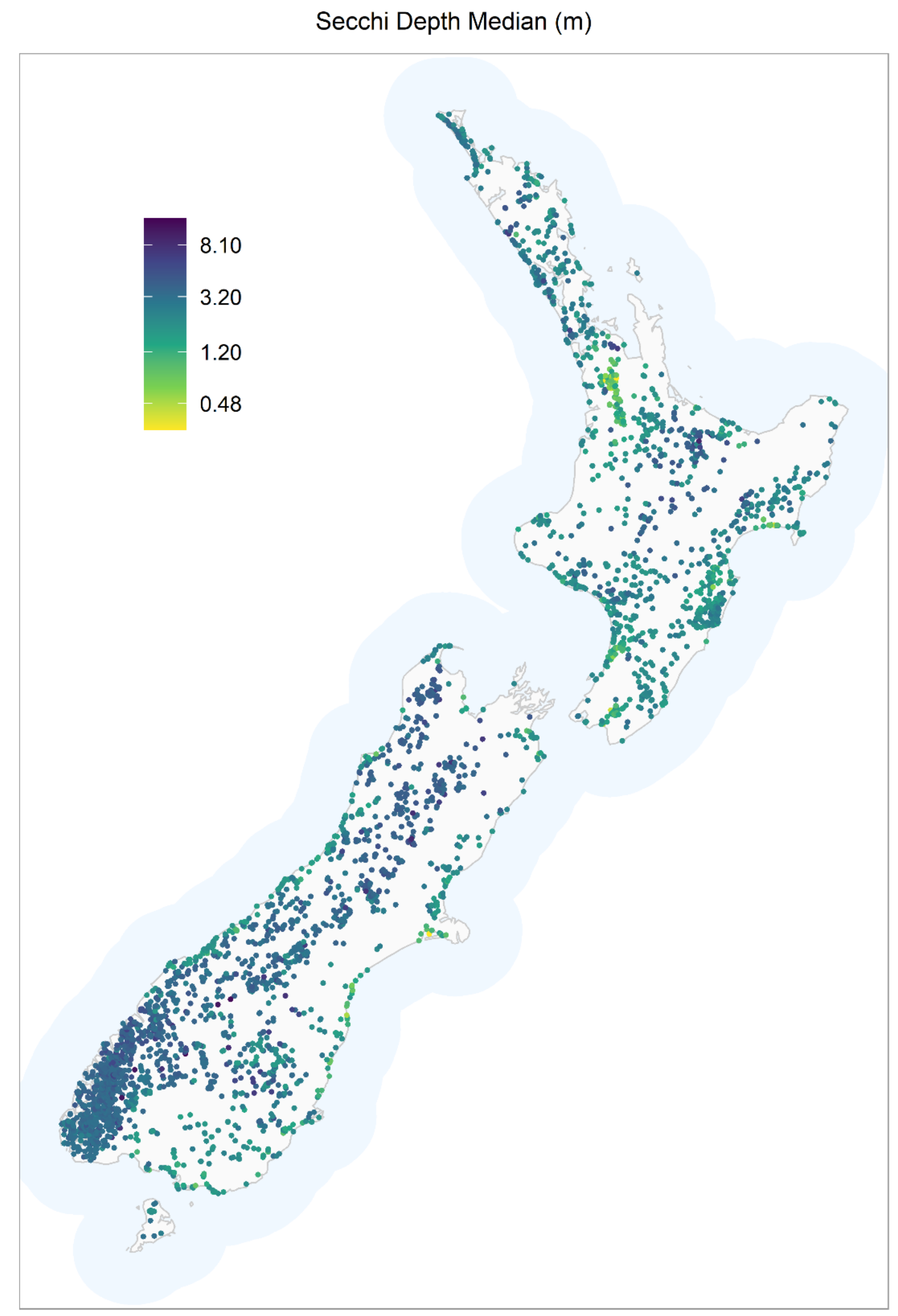

Figure 5. Predicted median Secchi depth for New Zealand lakes. The lakes are indicated by points located at the lake centre.

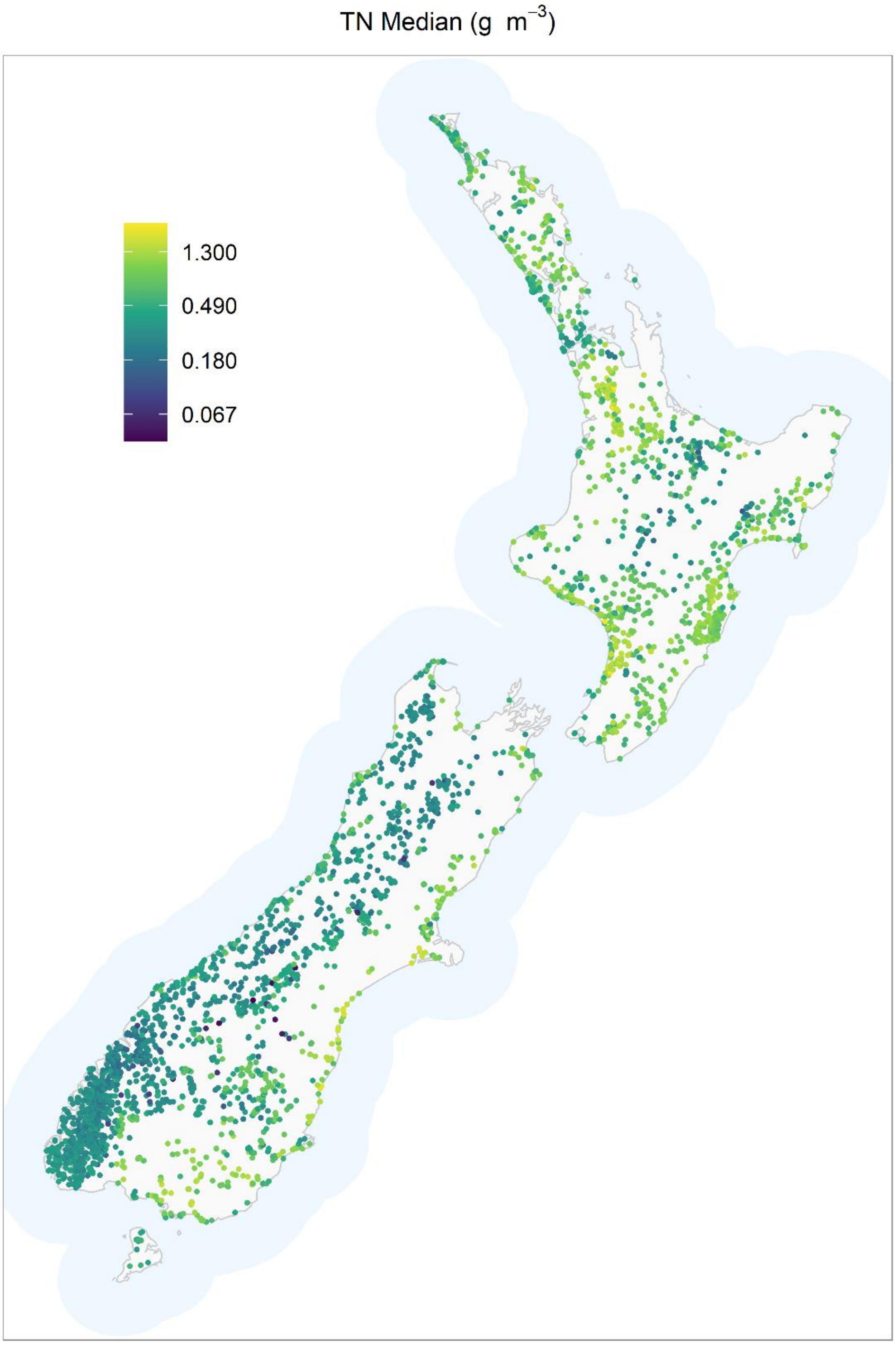

Figure 6. Predicted median total nitrogen for New Zealand lakes. The lakes are indicated by points located at the lake centre.

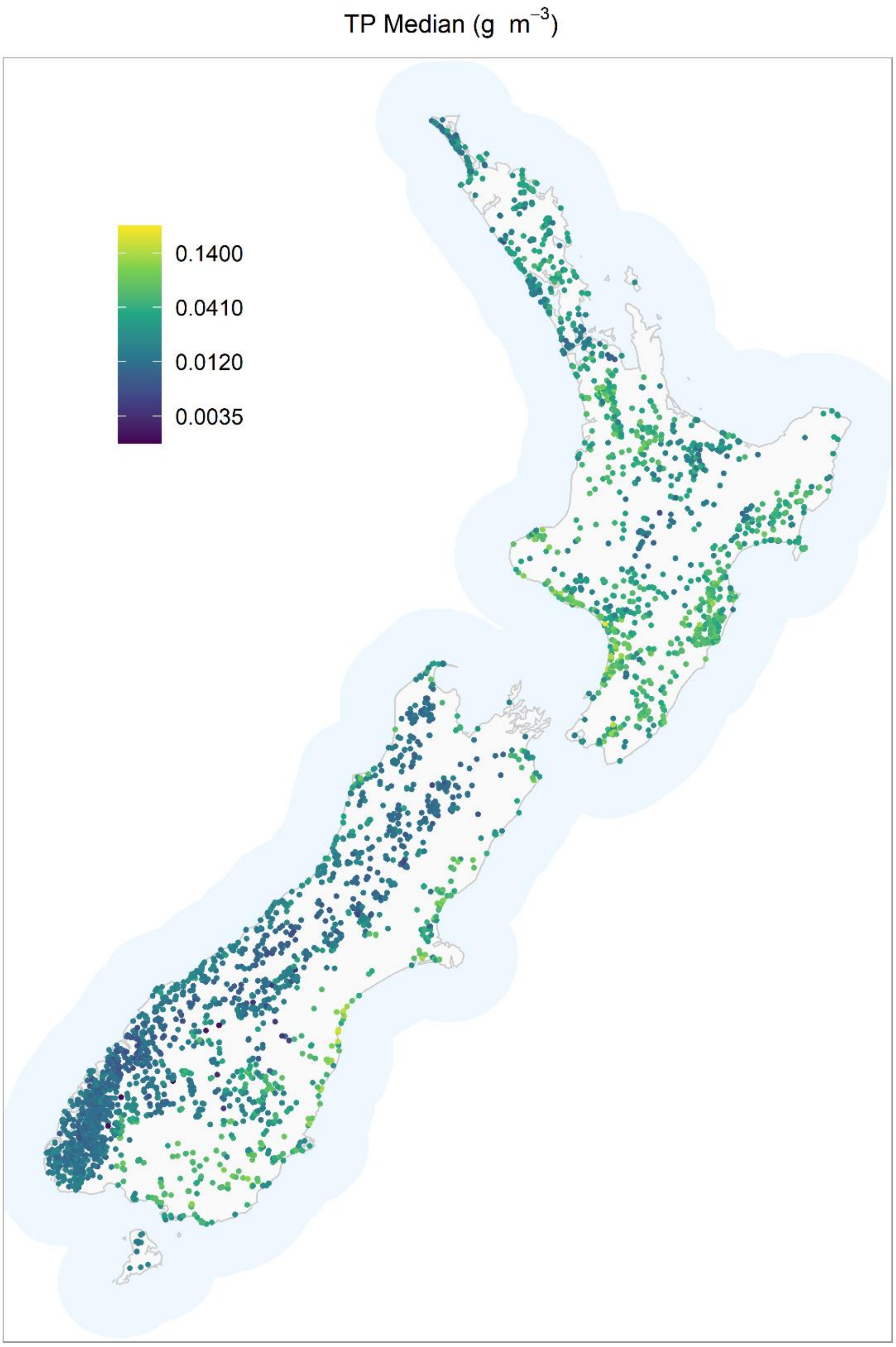

Figure 7. Predicted median total phosphorus for New Zealand lakes. The lakes are indicated by points located at the lake centre.

NH$_4$N (adjusted) Median (g m^{-3})

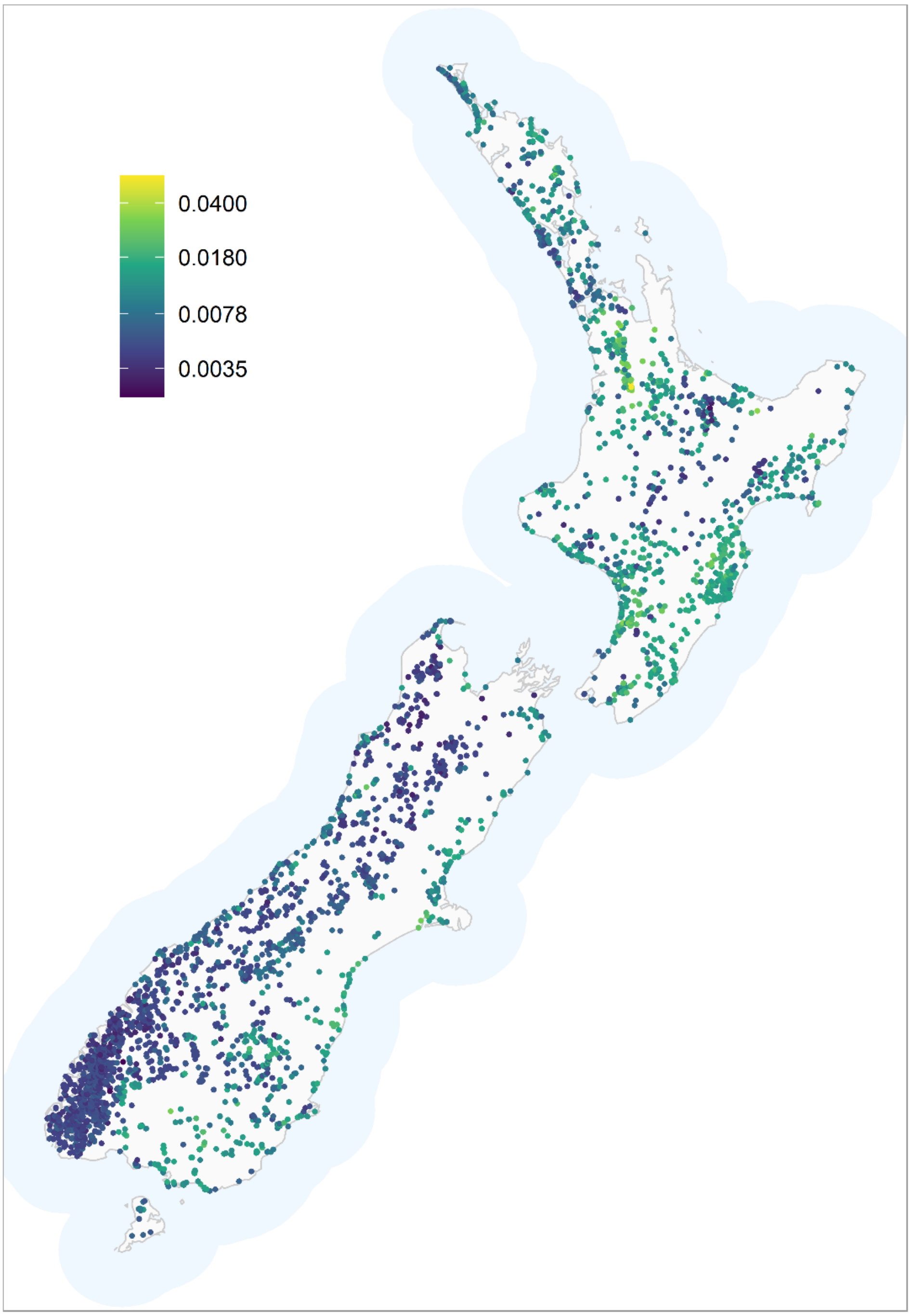

Figure 8. Predicted median ammoniacal nitrogen (adjusted) for New Zealand lakes. The lakes are indicated by points located at the lake centre.

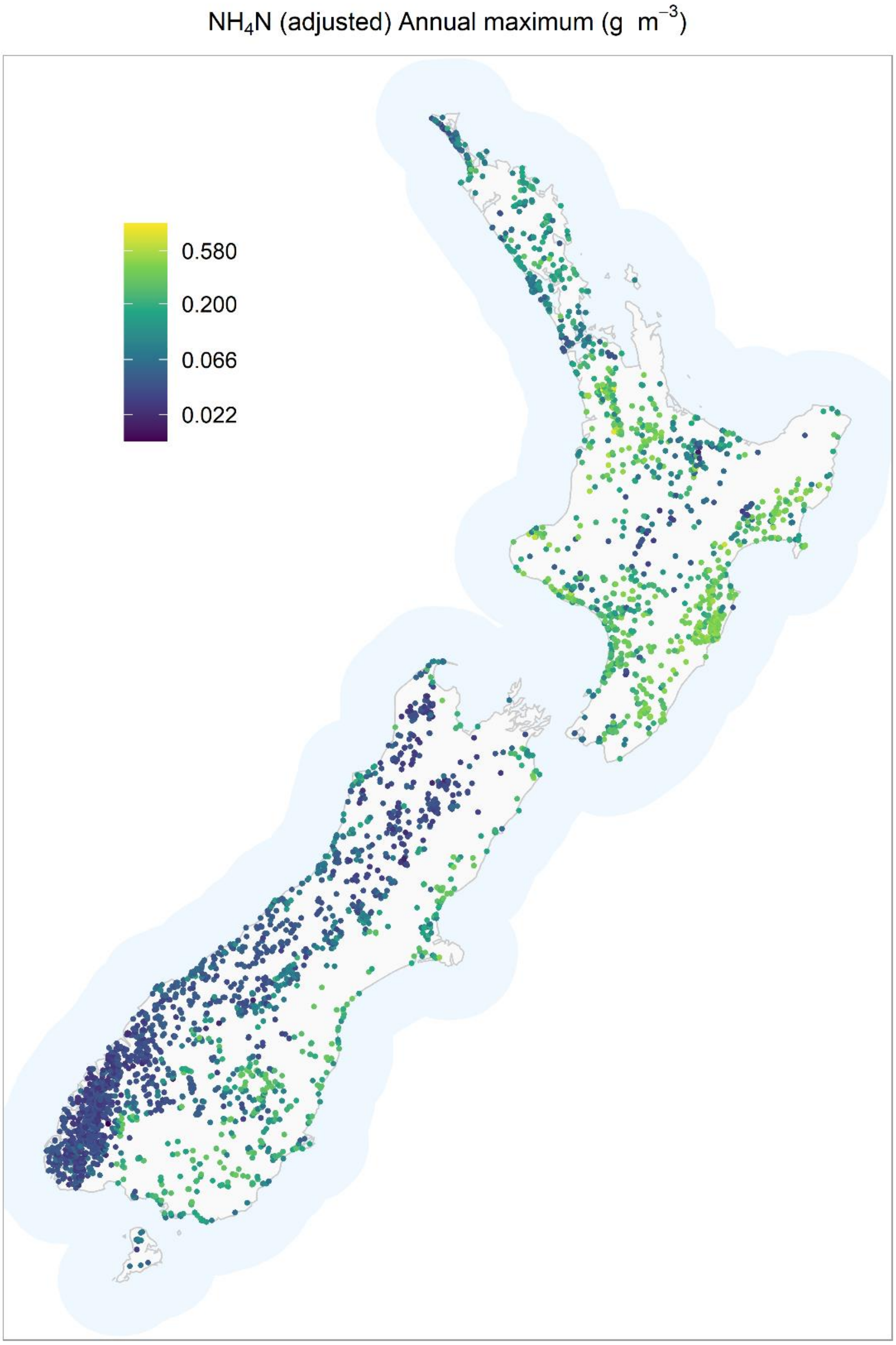

Figure 9. Predicted annual maximum ammoniacal nitrogen (adjusted) for New Zealand lakes. The lakes are indicated by points located at the lake centre.

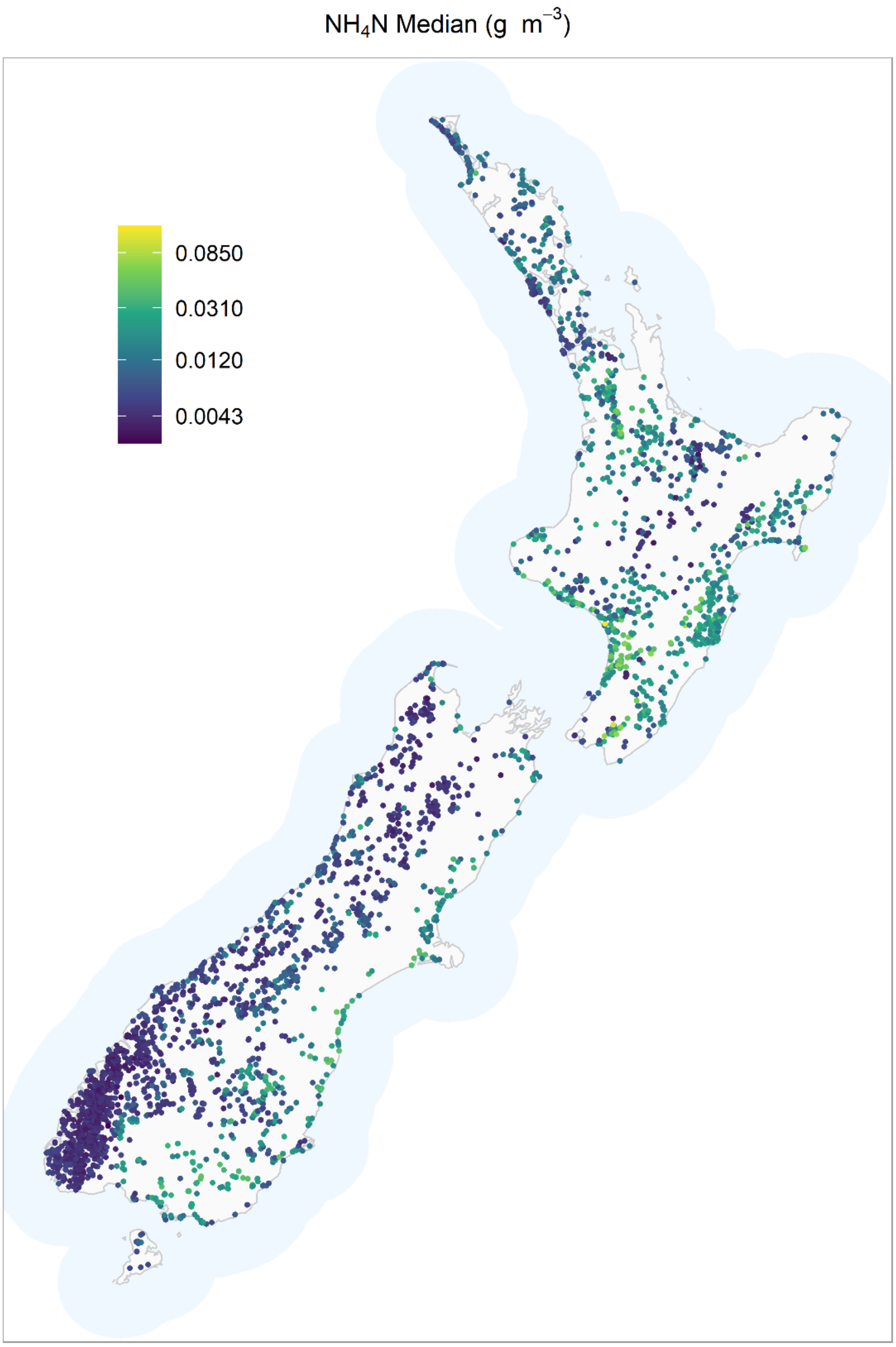

Figure 10. Predicted median ammoniacal nitrogen (not adjusted) for New Zealand lakes. The lakes are indicated by points located at the lake centre.

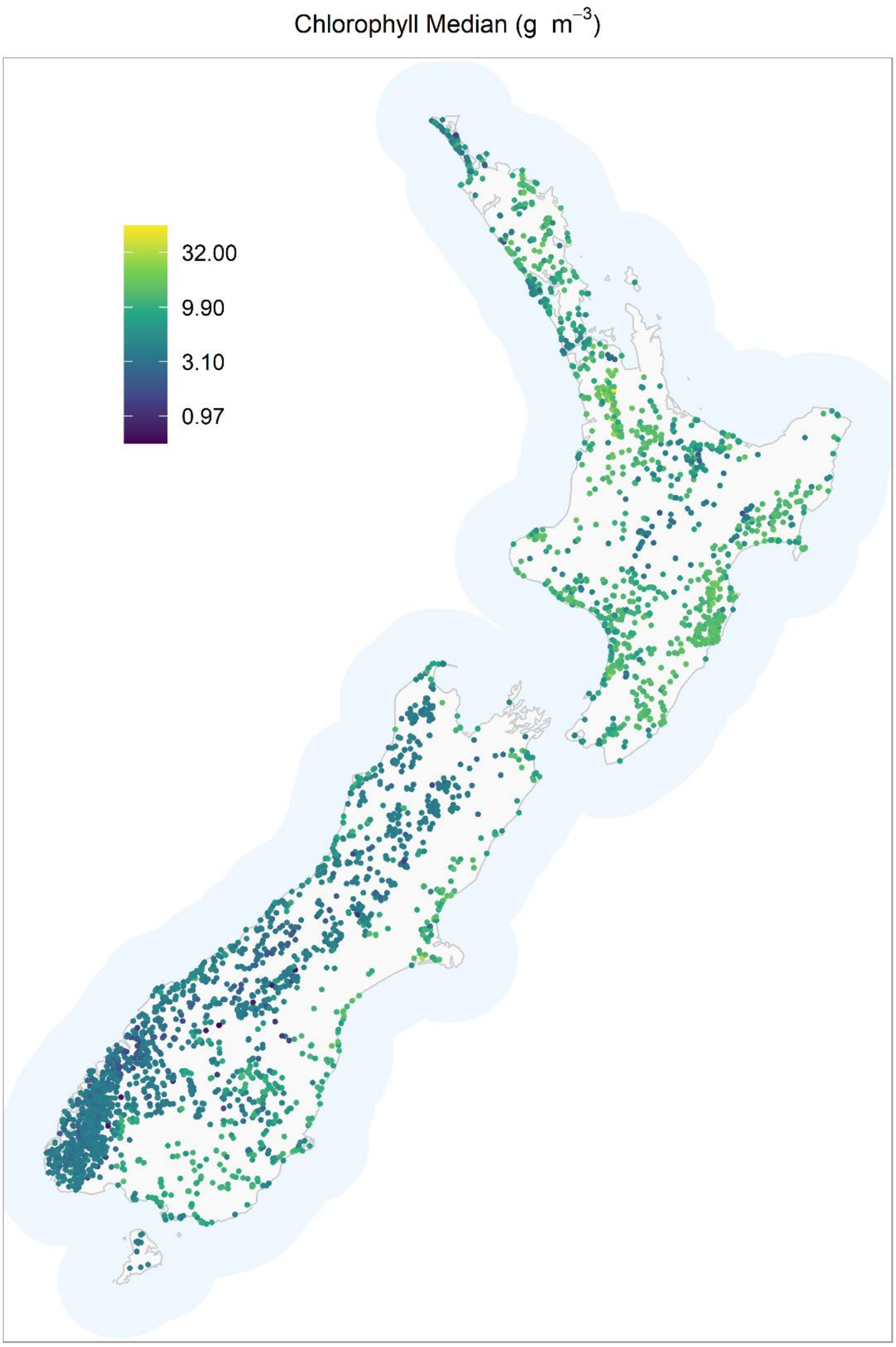

Figure 11. Predicted median chlorophyll a for New Zealand lakes. The lakes are indicated by points located at the lake centre.

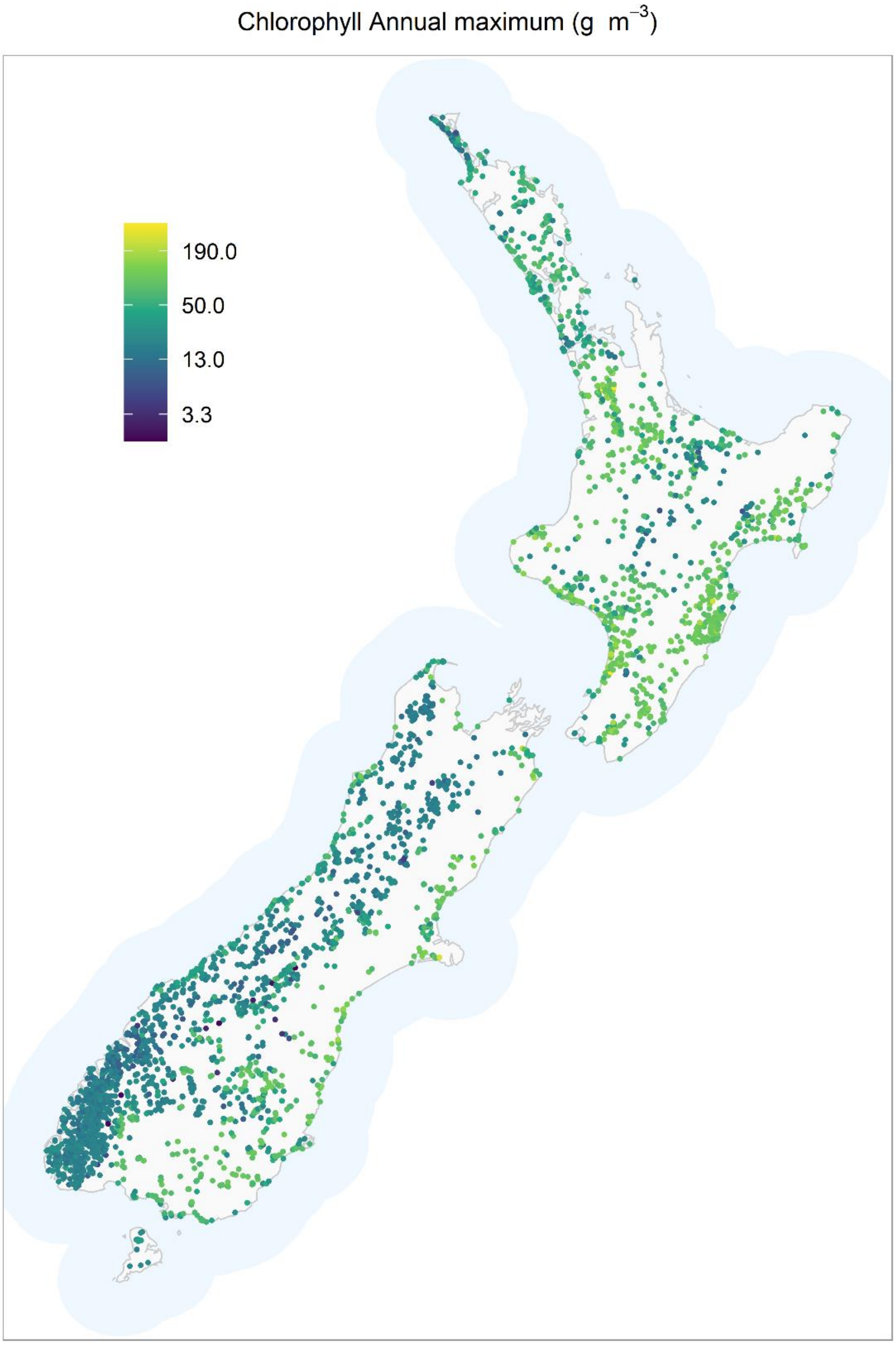

Figure 12. Predicted annual maximum chlorophyll a for New Zealand lakes. The lakes are indicated by points located at the lake centre.

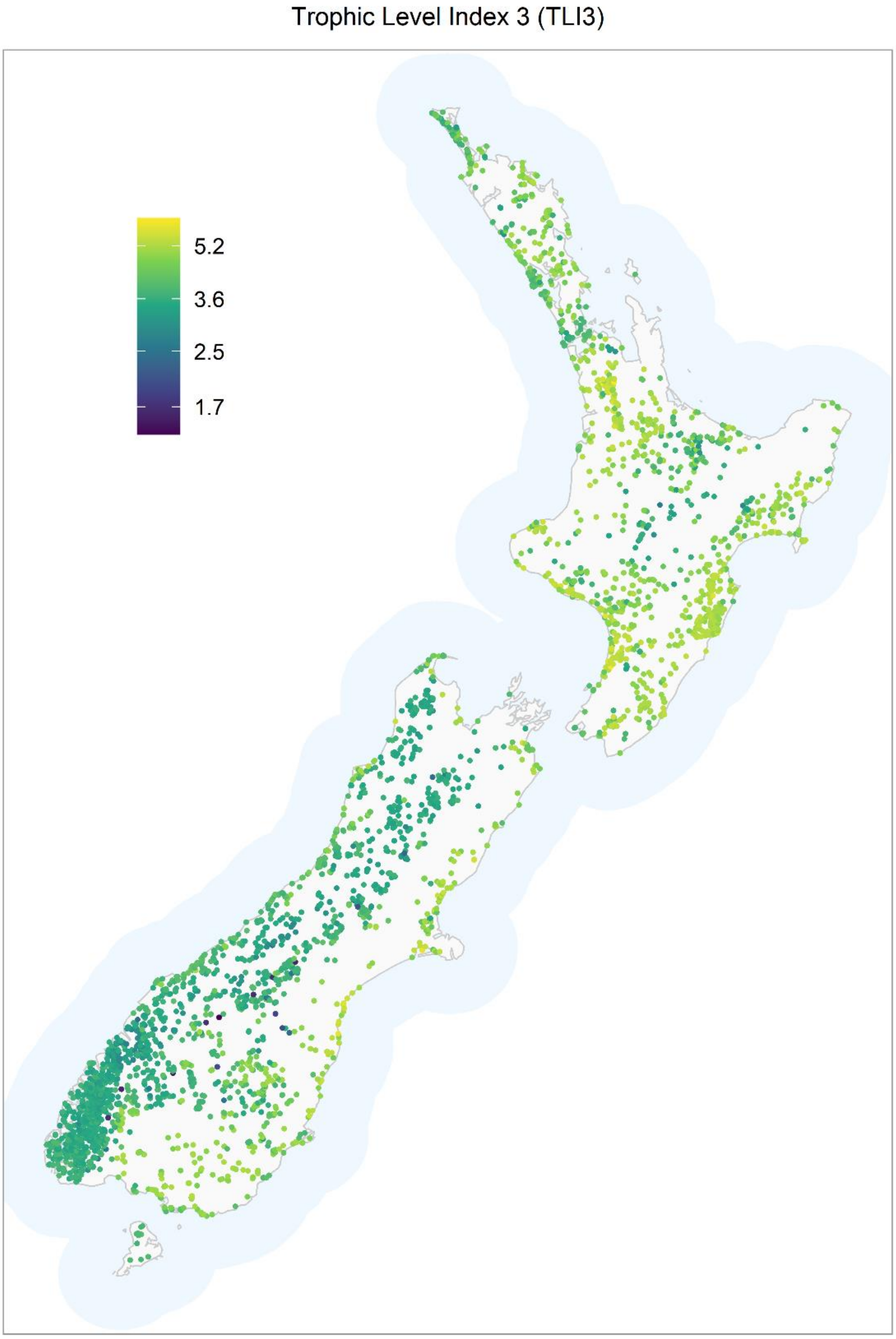

Figure 13. Predicted Trophic Level Index 3 for New Zealand lakes. The lakes are indicated by points located at the lake centre.

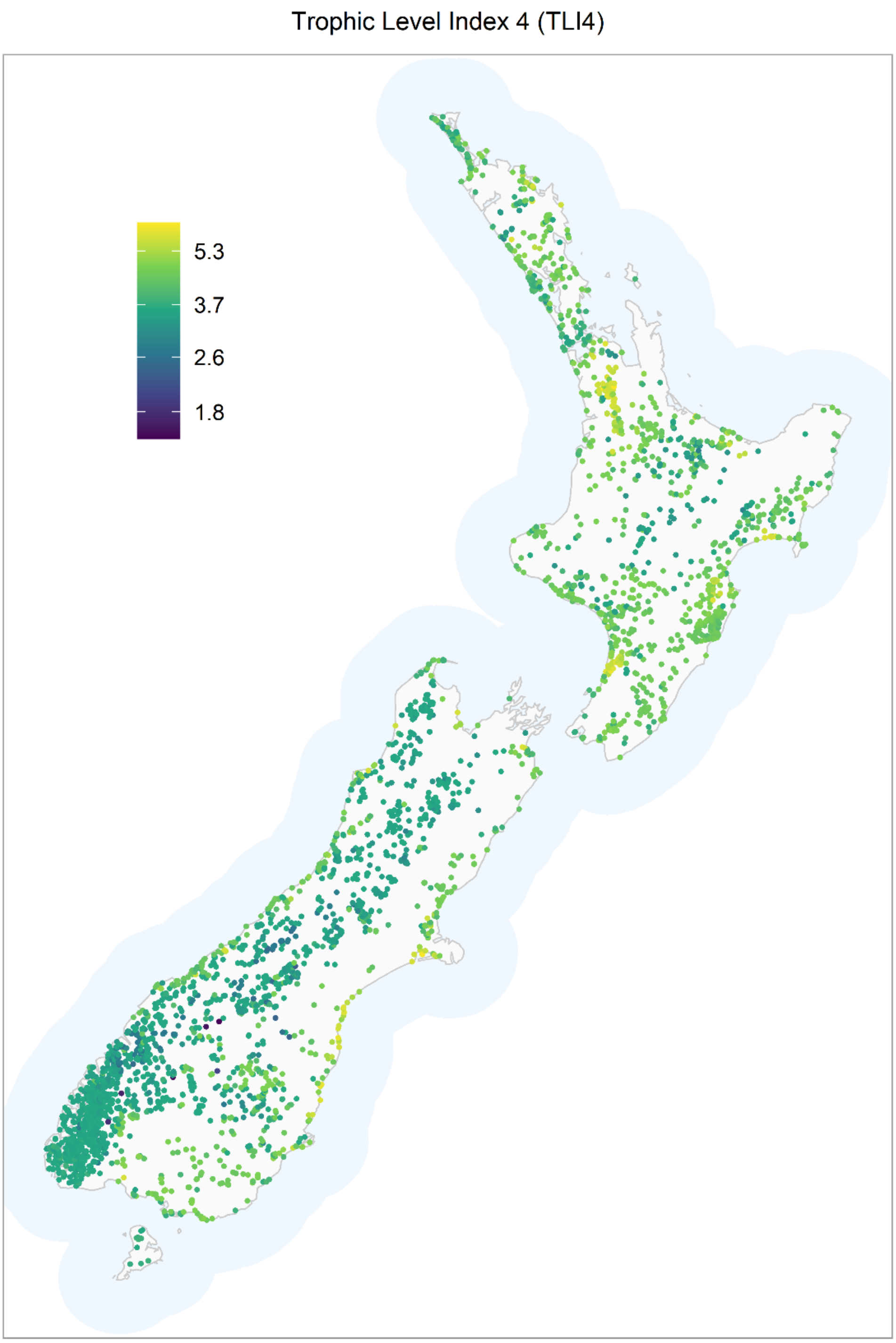

Figure 14. Predicted Trophic Level Index 4 for New Zealand lakes. The lakes are indicated by points located at the lake centre.

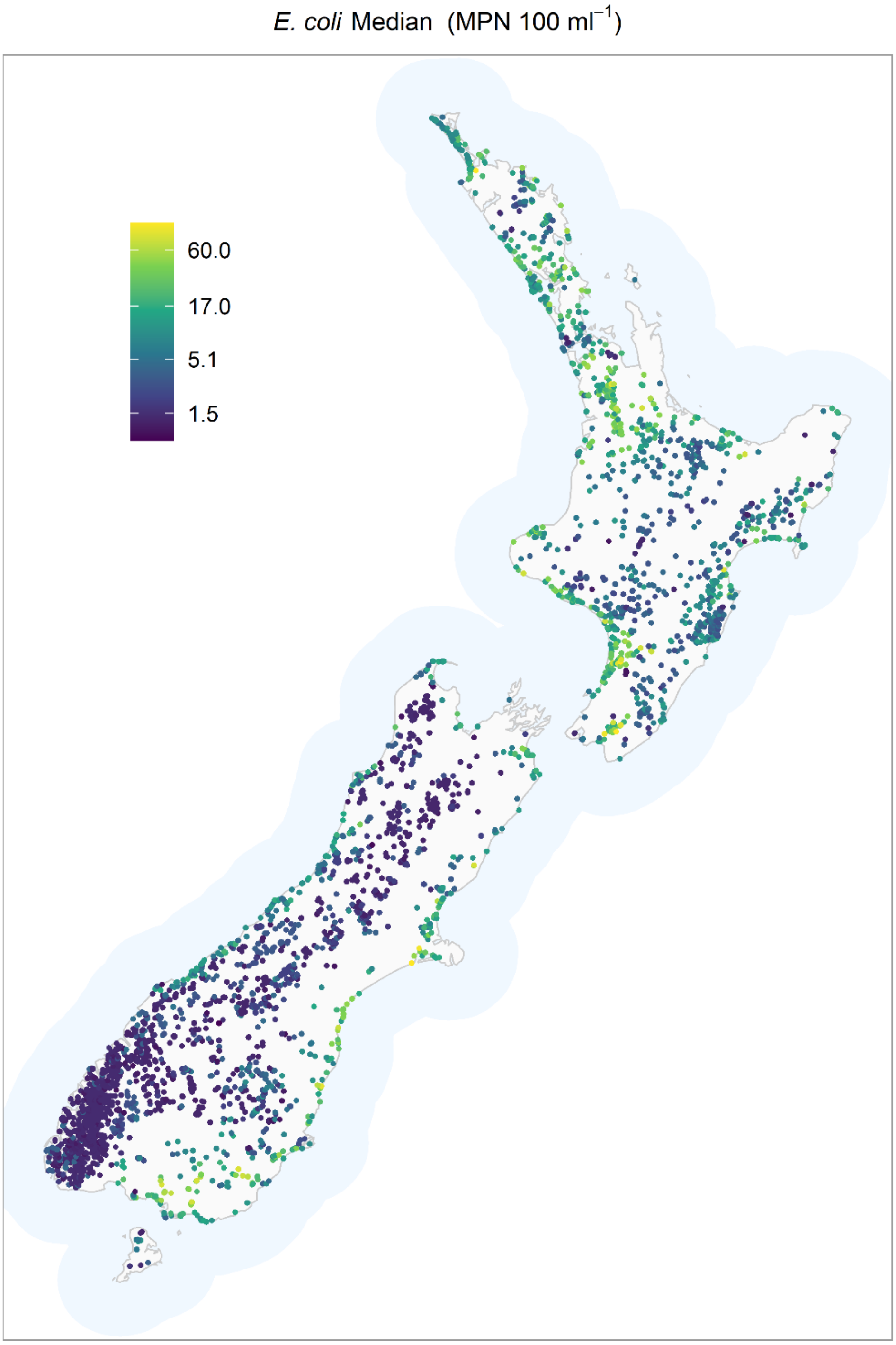

Figure 15. Predicted median Escherichia coli for New Zealand lakes. The lakes are indicated by points located at the lake centre.

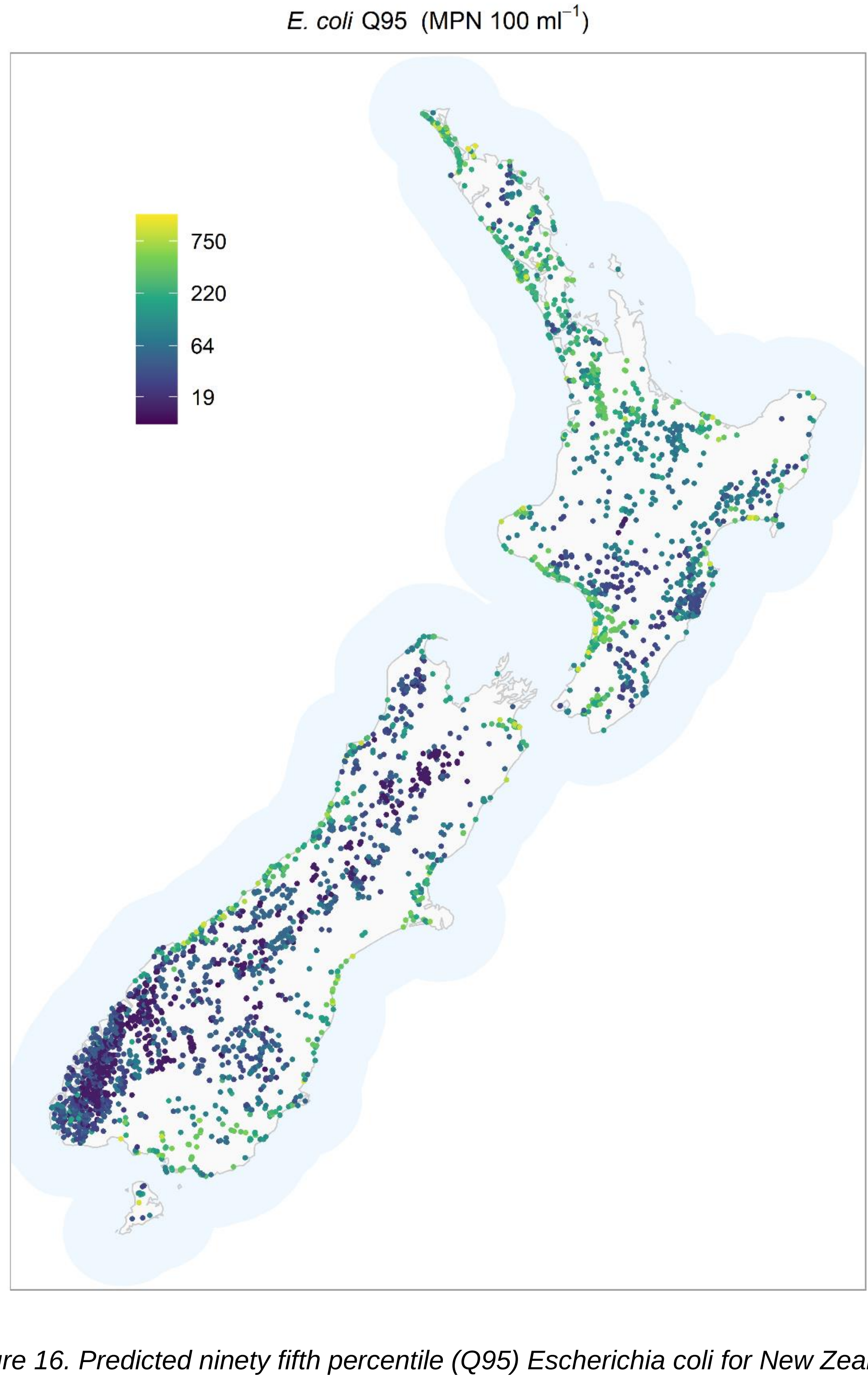

Figure 16. Predicted ninety fifth percentile (Q95) Escherichia coli for New Zealand lakes. The lakes are indicated by points located at the lake centre.

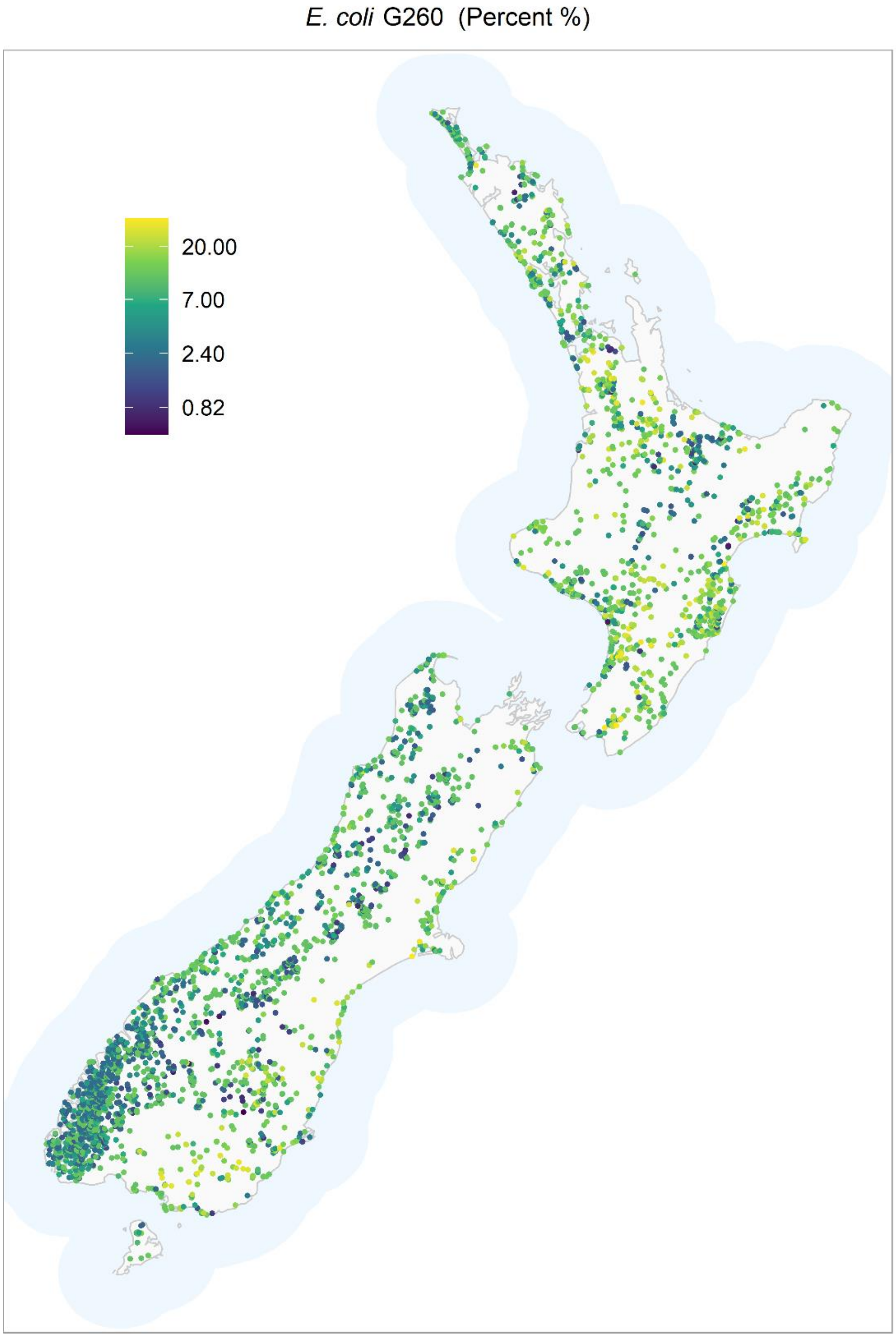

Figure 17. Predicted proportion of samples exceeding 260 E. coli 100 ml^{-1} (G260) Escherichia coli for New Zealand lakes. The lakes are indicated by points located at the lake centre.

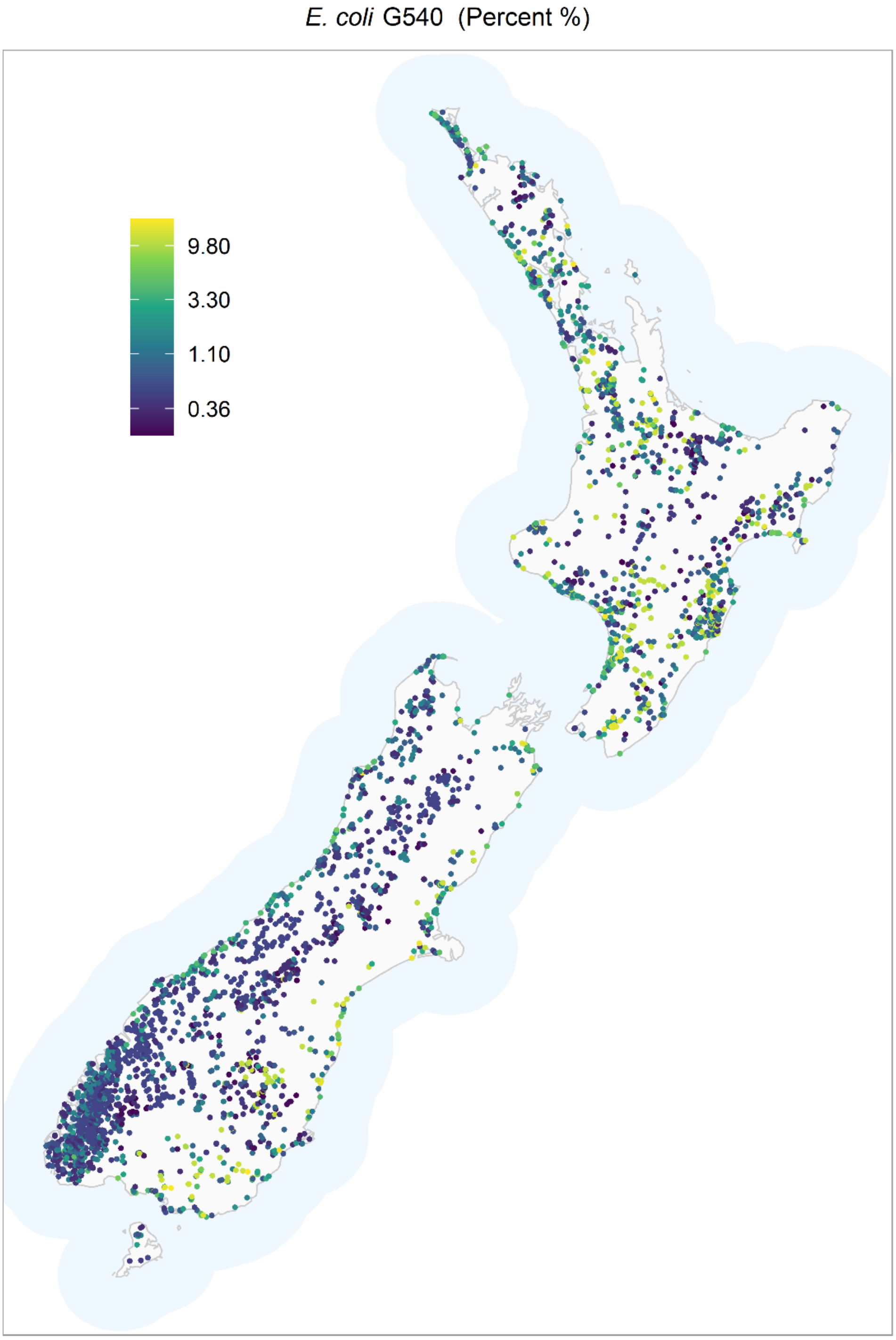

Figure 18. Predicted proportion of samples exceeding 540 E. coli 100 ml^{-1} (G540) Escherichia coli for New Zealand lakes. The lakes are indicated by points located at the lake centre.

5 Discussion

5.1 Representativeness and modelled relationships

The lake dataset was small (≤120 sites) and had a restricted geographic coverage (Figure 1). There were no or very limited data available for the Hawkes Bay, Taranaki and Gisborne regions in the North Island, and the top and west coast of the South Island. Monitored lakes were slightly over-representative of low elevations and lakes in regions with warmer climates and were under-representative of lakes in regions with colder climates (Figure 4). We note that Figure 4 only considers the representativeness of the samples in one-dimension (i.e., with respect to the variable shown on the x-axis), whereas the true representativeness of the sample needs to be considered within the multi-dimensional space defined by all the predictors. More complex methodologies exist to determine the reliability of the model predictions by considering the degree to which predictions are based on interpolation or extrapolation (Booker and Whitehead, 2018). Generally, the smaller the training set size the greater degree to which model predictions are based on extrapolation and the lower the overall prediction reliability. However, conducting this type of analysis was beyond the scope of the current project.

5.2 Comparison with previous study

The lake attributes represented in this study update previous modelling work carried out by Snelder *et al.* (2016), for the period 2009-2013 and Fraser and Snelder (2019), (2013-2017). The current study used the same methodology as the previous two studies, so the main difference is related to the change in time-period, which led to differences in the number of lakes included in the spatial models (Table 1).

In general, the spatial models produced by this study were similar to those of both Snelder *et al.* (2016) and Fraser and Snelder (2019), in terms of model performance, predictor importance levels and predicted patterns. Therefore, the broad scale conclusions for this study are similar to the previous work. The models produced in this study included predictors that described catchment land cover based on LCDB5, whereas land cover was not included in the models produced by the previous studies. In addition, the models produced in this study included predictor variables representing catchment land use intensity based on the APC stocking data.

The inclusion of predictors representing land cover and land use intensity in the current study's models potentially improves the reliability of the predictions compared to the previous study. This is because the previous study's most important predictors (catElev, lkElev and catWinTemp) probably represented correlative relationships with actual drivers such as the supply of contaminants to the lakes from their catchments. Those correlative relationships mean that model predictions may be unrealistic in situations where the relationship between catElev, lkElev and catWinTemp and the actual causative variables (catchment contaminant loads) is significantly different to the fitting dataset. The most obvious situations where this is likely are lakes at low elevations whose catchments are largely unmodified, and lakes with cold climates but low elevation. In the previous lake modelling report, we indicated that the lake water quality predictions were likely to be less reliable in geographic regions that have low elevation lakes combined with lake catchments that have relatively unmodified catchment land cover, such as the West Coast of the South Island, Fiordland and Stewart Island (Fraser and Snelder, 2019). The inclusion of catchment land cover and land use intensity as important predictors in this study is therefore likely to represent an advance in predicting water quality in unmonitored New Zealand lakes.

5.3 Model uncertainty

In this study, we modelled broad-scale patterns in current attribute state using catchment characteristics and lake-scale descriptors as predictors. Because the processes that determining lake attribute state are complex, some unexplained variation in our models is to be expected. Predictions made for individual lakes are associated with uncertainties that are characterised by model RMSD (Table 5). However, the level of model bias for each attribute was low, which indicates that the predicted patterns reflect broad scale relative differences between lakes.

For all models except TLI3, TLI4, *E. coli* G260 and *E. coli* G540, the 95% confidence intervals for the predicted attribute state for individual lakes can be obtained using the Equation 1. Equation 1 accounts for the $\log_{10}$ transformation of the response variables prior to model fitting, which means the prediction uncertainty (RMSD) values have been reported in the $\log_{10}$ transformed space.

$$95\%\ CI = 10^{[\log_{10}(x) \pm 1.96 \times RMSD]} \qquad (1)$$

where x is the estimated value in the original units, RMSD is the model error and 1.96 is the standard normal deviate or Z-score for probability (0.025 ≤ Z ≥ 0.975). The prediction confidence intervals for the $\log_{10}$-transformed attributes, when expressed in the original units of the attribute, are asymmetric and their values vary in proportion to the predicted value. For example, if we let x be a predicted value for SECCHI 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.

For TLI3, TLI4, *E. coli* G260 and *E. coli* G540, the 95% confidence intervals for the predictions can be obtained using the Equation 2.

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

RF model performance differed between modelled attributes and this variation may be attributable to differences in the biophysical processes that control different aspects of water quality in lakes. Some biophysical processes may be poorly represented by our catchment-averaged spatial predictor variables. For example, concentrations of TN and TP in lakes 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 RF models. The absence of predictors that account for these and other processes means that some level of unexplained variation is inevitable.

We used the criteria of Moriasi *et al.* (2015) to categorise the performance of the 14 models as ‘very good’, ‘good’, ‘satisfactory’ and ‘unsatisfactory’ (Table 5). These criteria are subjective and are used as an indication of the quality of the predictions. The actual acceptability of the predictions depends on their use and needs to be considered in the context of each application. The models for NH4N_adj_Median, NH4N_adj_AnnMax, ECOLI_G260 and ECOLI_G540 had ‘unsatisfactory’ performance indicating the predictions from these models should be used very cautiously.

Acknowledgements

We thank the many regional council staff who collected the data and information used by this study. Thanks to Isaac Bain (Ministry for the Environment), Lance Morell and Sonja Miller (Statistics NZ) 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. Finally, we thank Scott Larned for reviewing the draft report.

References

Booker, D. J., & Whitehead, A. L. (2018).Inside or outside: Quantifying extrapolation across river networks.Water Resources Research, 54. https://doi.org/10.1029/2018WR023378

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

Breiman, L., J.H. Friedman, R. Olshen, and C.J. Stone, 1984. Classification and Regression Trees. Wadsworth, Belmont, California.

Cutler, D.R., J.T.C. Edwards, K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, and J.J. Lawler, 2007. Random Forests for Classification in Ecology. Ecology 88:2783–2792.

Duan, N., 1983. Smearing Estimate: A Nonparametric Retransformation Method. Journal of the American Statistical Association 78:605–610.

Fraser, C. and T. Snelder, 2019. Spatial Modelling of Lake Water Quality State. Incorporating Monitoring Data for the Period 2013 to 2017. LWP Client Report, LWP Ltd, Christchurch, New Zealand.

Hastie, T., R. Tibshirani, and J.H. Friedman, 2001. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York.

Larned, S., M. Unwin, and J. Sykes, 2014. Recommendations for New Sites to Improve Representativeness in the New Zealand River Environmental Monitoring Network. NIWA Client Report, NIWA, Christchurch, New Zealand.

Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith, 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE 50:885–900.

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

Nash, J.E. and J.V. Sutcliffe, 1970. River Flow Forecasting through Conceptual Models Part I—A Discussion of Principles. Journal of Hydrology 10:282–290.

NZ Government, 2020. National Policy Statement for Freshwater Management 2020.

Parker, W.J., 1998. Standardisation between Livestock Classes: The Use and Misuse of the Stock Unit System. Proceedings of the Conference New Zealand Grassland Association., pp. 243–248.

Piñeiro, G., S. Perelman, J. Guerschman, and J. Paruelo, 2008. How to Evaluate Models: Observed vs. Predicted or Predicted vs. Observed? Ecological Modelling 216:316–322.

Snelder, T., C. Fraser, S. Larned, R. Monaghan, S. de Malmanche, and A. Whitehead, 2021. Attribution of River Water-Quality Trends to Agricultural Land Use and Climate Variability in New Zealand. Marine and Freshwater Research.

Snelder, T., J.R. Leathwick, and D. Kelly, 2006. Definition of a Multivariate Classification of New Zealand Lakes. Christchurch.

Snelder, T., S. Woods, and J. Atalah, 2016. Strategic Assessment of New Zealand's Freshwaters for Recreational Use: A Human Health Perspective Escherichia Coli in Rivers and Planktonic Cyanobacteria in Lakes. LWP Client Report, LWP Ltd, Christchurch, New Zealand.

Sorrell, B., M. Unwin, K. Dey, and H. Hurren, 2006. A Snapshot of Lake Water Quality. NIWA, Christchurch.

Svetnik, V., A. Liaw, C. Tong, and T. Wang, 2004. Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules. T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. P. Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, M. Vardi, and G. Weikum (Editors). Springer, Cagliari, Italy, pp. 334–343.

Whitehead, A., C. Fraser, T. Snelder, and R. White, 2021. Water Quality State and Trends in New Zealand Lakes. Analyses of National Lakes Data Ending in 2020. NIWA Client Report, NIWA, Christchurch, New Zealand.