



Spatial modelling of lake water quality state

Incorporating monitoring data for the period 2013 to 2017

March 2019

Prepared By:

Caroline Fraser

Ton Snelder

For any information regarding this report please contact:

Caroline Fraser

Phone: 022 049 1779

Email: caroline@lwp.co.nz

LWP Ltd
PO Box 70
Lyttelton 8092
New Zealand

LWP Client Report Number: LWP Client Report 2018-16

Report Date: March 2019

LWP Project: 2018-16

Quality Assurance Statement


Version	Reviewed By	
	Scott Larned (NIWA)	

Table of Contents

- Executive Summary v**
- 1 Introduction 6**
- 2 Data 6**
 - 2.1 Lake State Data 6
 - 2.2 Lake predictor data 8
- 3 Modelling Methods..... 9**
 - 3.1 Random forest models 9
 - 3.2 Model performance 11
 - 3.3 Representativeness of monitored lakes used in RF models 12
 - 3.4 Model predictions 12
- 4 Results 12**
 - 4.1 Model performance 12
 - 4.2 Modelled relationships 14
 - 4.3 Monitored lake representativeness..... 17
 - 4.4 Model predictions 18
- 5 Discussion..... 20**
 - 5.1 Representativeness and modelled relationships 20
 - 5.2 Comparison with previous study..... 20
 - 5.3 Model uncertainty..... 21
- Acknowledgements..... 22**
- References..... 23**
- Appendix A Exploration of lake loads as predictors for lake random forest models 25**
- Appendix B Mapped predictions from random forest models 28**
 - Figure 1. Locations of lakes reported on in this study for each of the six water quality variables..... 8
 - Figure 2. Comparison of observed water quality versus values predicted by the RF models. 14
 - Figure 3. Partial plots for the ten most important predictor variables in the RF models of lake water quality. 16
 - Figure 4. Histograms comparing the distributions of predictor variables for all lakes and the monitored lakes used to build the RF models. 18
 - Figure 5. Predicted water quality for New Zealand lakes. 19
 - Figure 6. Comparison of observed water quality versus values predicted by the RF models. 26

Figure 7. Comparison of predicted water quality for 978 lakes (those with load predictions) from RF models.	27
Figure 8. Predicted chlorophyll for New Zealand lakes.	28
Figure 9. Predicted NH ₄ N for New Zealand lakes.	29
Figure 10. Predicted SECCHI for New Zealand lakes.....	30
Figure 11. Predicted TLI3 for New Zealand lakes.....	31
Figure 12. Predicted TNI for New Zealand lakes.	32
Figure 13. Predicted TP for New Zealand lakes.	33

Table 1. Lake water quality variables included in this study. NM: not modelled.....	7
Table 2. Predictor variables used in the spatial models of lake water quality.	9
Table 3: Performance ratings for statistics used in this study, from (Moriassi et al., 2007).....	12
Table 4. Performance of the lake water quality models.	13
Table 5. Predictors retained by the reduced RF models of lake water quality variables.	15
Table 6: Number of lakes in the model training set and number of lakes that also have load predictions	25
Table 7: Performance of lake water quality models with and without nutrient loads as predictors.	26

Executive Summary

Measurements of lake water quality variables comprising chlorophyll *a* (CHLA, total phytoplankton biomass), total nitrogen (TN), total phosphorus (TP), ammoniacal nitrogen (NH₄N), Secchi depth (SECCHI), and the trophic level index (TLI₃) were obtained for 61 to 104 lakes (differed by variable) from regional council state-of-the-environment monitoring programmes, for the period 2013 to 2017. The median values of these variables were estimated 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.

Satisfactory to good performance was achieved for four of the lake spatial models (CHLA, TN, SECCHI and TLI₃); the spatial models for TP and NH₄N had a slightly poorer performance. Because of the limited number of lakes in the dataset used to develop the models, the models may have over-predicted in lowland areas with native catchments e.g., West Coast South Island. The mapped predictions for all six water-quality variables had similar spatial patterns, with high values of CHLA, TN, TP, NH₄N, and TLI₃ and low values of SECCHI, in low-elevation areas 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, TN, TP, NH₄N and TLI₃ 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 areas of most concern to target interventions.

1 Introduction

Lake water quality across New Zealand was characterised by a recent national analysis of state and trends at monitored lakes (Larned *et al.*, 2018). The lakes are monitored as part of the State of Environment (SOE) programmes operated by regional councils. The datasets underlying these analyses contained quarterly or monthly measurements of physical, chemical, and biological variables over time periods from as early as 1990 to 2017.

The objective of this study is to enable current water quality state to be estimated and mapped across all 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. In turn, these applications are needed for water management decision-making and for limit setting under the NPS-FM.

In the current report, the lake water quality monitoring data from the Larned *et al.* (2018) study were utilised. Those data were used to develop spatial models that predict water quality in all large lakes. The benefit of spatial modelling is that it provides a large-scale assessment of water quality 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 monitoring lakes (Larned and Unwin, 2012).

This report is a companion to the primary output for the project, a .csv file containing the outputs from the spatial models, which may be used by MfE in a variety of future applications. It provides a detailed description of the methods used to extract variables from available data and to produce spatial predictions for unmonitored lakes. The methods used to prepare the water quality variables data, make assessments of the representativeness of the monitored lakes, and to undertake the spatial modelling are described. The results provide national maps of predicted lake water quality. Measures of model performance and the important relationships between water quality variables and predictors 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 Larned *et al.* (2018) for the current study for spatial modelling. Detailed methods for obtaining and grooming these data are provided by (Larned *et al.*, 2018). The lake SOE data analysed by Larned *et al.*, (2018) included nine water quality variables that correspond to physical, chemical and biological conditions. Six of these variables have been modelled in the current report (Table 1). A decision was made to exclude three variables (DRP, NO₃N and ECOLI) as the number of lakes with corresponding measurements was small (<35 sites) and spatial coverage was poor. Therefore, these three variables were poorly represented at the national scale and predictions from models developed with these data would have low accuracy.

This study used water-quality data for the five-year period from 2013 to 2017. Two inclusion rules were applied to ensure that the data were representative of each lake and variable, following the approach of (Snelder *et al.*, 2016). First, at least eight samples were available for the five-year period. Second, less than 50% of the observations of each variable were censored (i.e., below analytical detection limits). A summary of the number of lakes per

variable used in this study is in Table 1. The two rules were more lenient than the inclusion rules used in Larned *et al.* (2018), which required lake x variable 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 water quality state was assessed compared to Larned *et al.* (2018) (Table 1). Table 1 also provides the numbers of lakes used in the previous lake spatial modelling study (Snelder *et al.*, 2016), for comparison.

Table 1. Lake water quality variables included in this study. NM: not modelled

Variable type	Variable	Abbreviation	Units	Number of lakes		
				This study	State and trends ¹	Previous spatial modelling ²
Physical	Secchi depth	SECCHI	m	61	52	64
Chemical	Total nitrogen	TN	g m ⁻³	104	63	82
	Total phosphorus	TP	g m ⁻³	97	63	97
	Ammoniacal nitrogen	NH4N	g m ⁻³	64	62	NM
Phytoplankton	Chlorophyll <i>a</i>	CHLA	g m ⁻³	101	63	92
Water quality index	Trophic Level Index	TLI3	unitless	99	58	76

Note: (1) The numbers of lakes included in the companion state and trends report (Larned *et al.*, 2018)

(2) The numbers of lakes included in the Snelder *et al.* (2016) study.

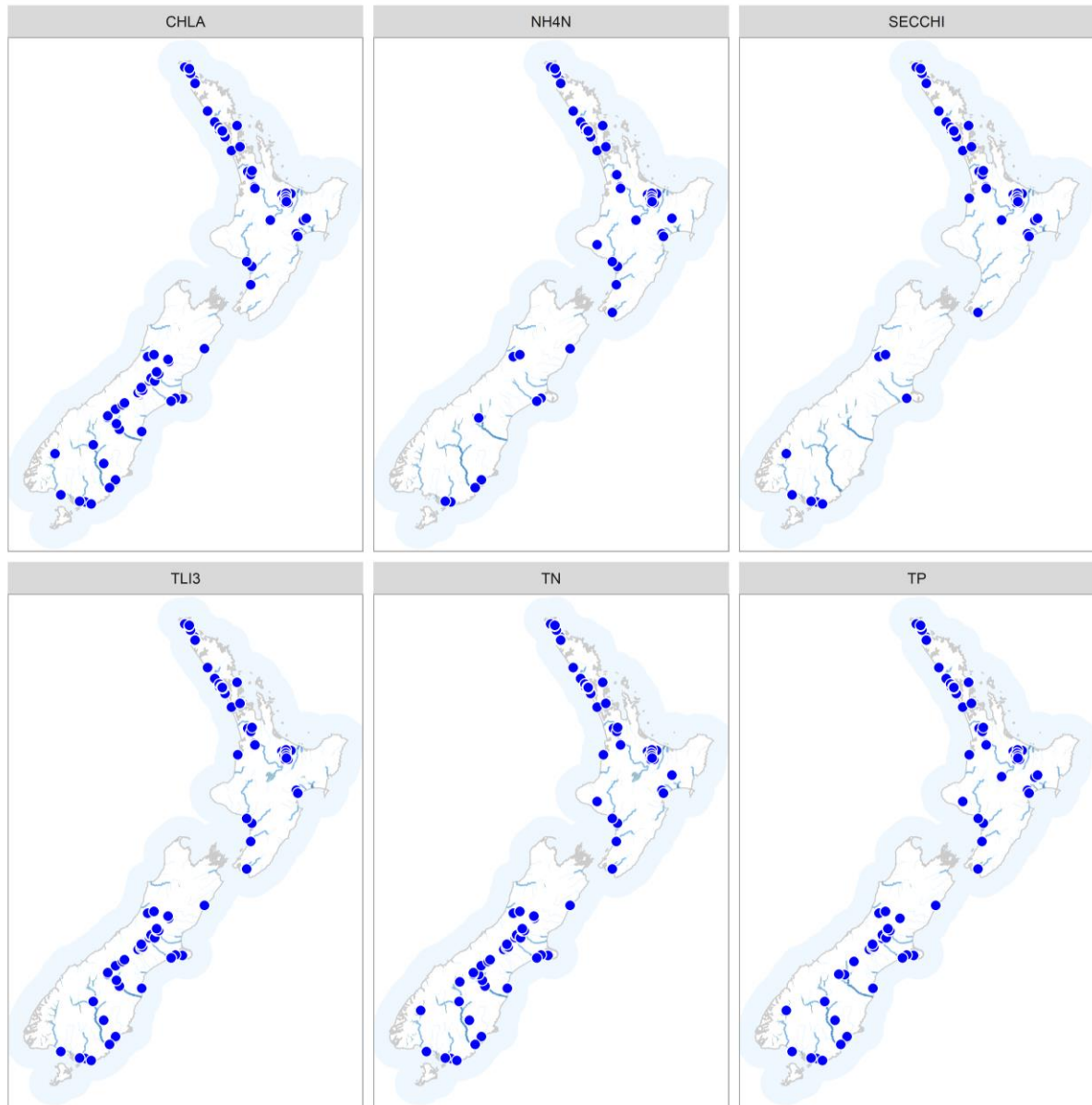


Figure 1. Locations of lakes reported on in this study for each of the six water quality variables. The locations shown on each panel correspond to the lakes that were included in this study for each variable listed in Table 1.

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.

We also explored the use of modelled nutrient loads as predictors (from Snelder *et al.*, 2017). A comparison of model results based on models with and without loads as predictors is presented in Appendix A.

Table 2. Predictor variables used in the spatial models of lake water quality.

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	catGlacial	Proportion of catchment occupied by permanent ice	Proportion
	catIndigForest	Proportion of catchment occupied by indigenous forest	Proportion
	catBare	Proportion of catchment occupied by bare ground	Proportion
	catExoticForest	Proportion of catchment occupied by exotic forest	Proportion
	catPastoral	Proportion of catchment occupied by pasture	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.

3 Modelling Methods

3.1 Random forest models

We fitted the median values of the six water quality variables associated with the monitored lakes to the predictor variables using random forest (RF) models (Breiman, 2001; Cutler *et al.*, 2007). An 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 predictor variables 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 variables that can be used at each split. We used default options that included making one third of the total number of predictor variables 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). Our models took less than a minute to fit when using the default of 500 trees per forest.

Unlike linear models, RF models cannot be expressed as equations. However, the relationships between predictor and response variables 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 variable, the values of the response variable are randomly permuted for the OOB observations, and predictions are obtained from the tree for these modified data. The importance of the predictor variable 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 variable on the response variable, when the values of all other predictor variables 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 variables. Partial dependence plots do not perfectly represent the effects of each predictor variable, particularly if predictor variables 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 predictor variables that are chosen during the model fitting process. However, marginally important predictor variables 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 predictor variables 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 predictor variables 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 predictor variables 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).

RSR is a measure of the characteristic model uncertainty. It is estimated as the mean deviation of predicted values with respect to the observed values (the root mean square deviation), divided by the standard deviation of the observations (Moriasi *et al.*, 2007).

The normalization associated with PBIAS and RSR allowed the performance of models to be compared across all of the modelled water quality variables. Model predictions were evaluated to be very good, good, satisfactory or unsatisfactory, following the criteria proposed by Moriasi *et al.*, 2007, outlined in Table 3.

Table 3: Performance ratings for statistics used in this study, from (Moriassi et al., 2007).

Performance Rating	RSR	NSE	PBIAS ¹
Very good	$RSR \leq 0.50$	$NSE > 0.75$	$ PBIAS < 25$
Good	$0.50 < RSR \leq 0.60$	$0.65 < NSE \leq 0.75$	$25 \leq PBIAS < 40$
Satisfactory	$0.60 < RSR \leq 0.70$	$0.50 < NSE \leq 0.65$	$40 \leq PBIAS < 70$
Unsatisfactory	$RSR > 0.70$	$NSE \leq 0.5$	$ PBIAS \geq 70$

Notes: (1) PBIAS shown here is only applicable for nutrient models.

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 water quality variable.

4 Results

4.1 Model performance

The performance of all models was improved by \log_{10} -transformation of the lake median 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 RF models of SECCHI, TN, CHLA, and TLI3 had satisfactory to good performance as indicated by the following statistics: $NSE > 0.5$, $RSR < 0.7$ (Table 4, Figure 2; Moriassi *et al.*, 2007). The models for TP and NH₄N had poorer performance, with NSE values of 0.48 and

0.40, respectively. All six models had very low bias (PBIAS; Table 4, Figure 2). RMSD values provide an indication of the magnitude of the characteristic error in the original units of each variable.

Table 4. Performance of the lake water quality models. 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, RSR = relative root mean square error, RMSD = root mean square deviation. RMSD units are the \log_{10} -transformed original units..

Model	N	R^2	NSE	PBIAS	RSR	RMSD
CHLA	99	0.54	0.53	0.22	0.69	0.37
NH4N	62	0.40	0.39	0.08	0.78	0.39
SECCHI	60	0.63	0.60	-6.54	0.63	0.30
TLI3	97	0.67	0.66	0.25	0.58	0.09
TN	102	0.72	0.70	-0.99	0.54	0.25
TP	95	0.48	0.46	0.62	0.73	0.34

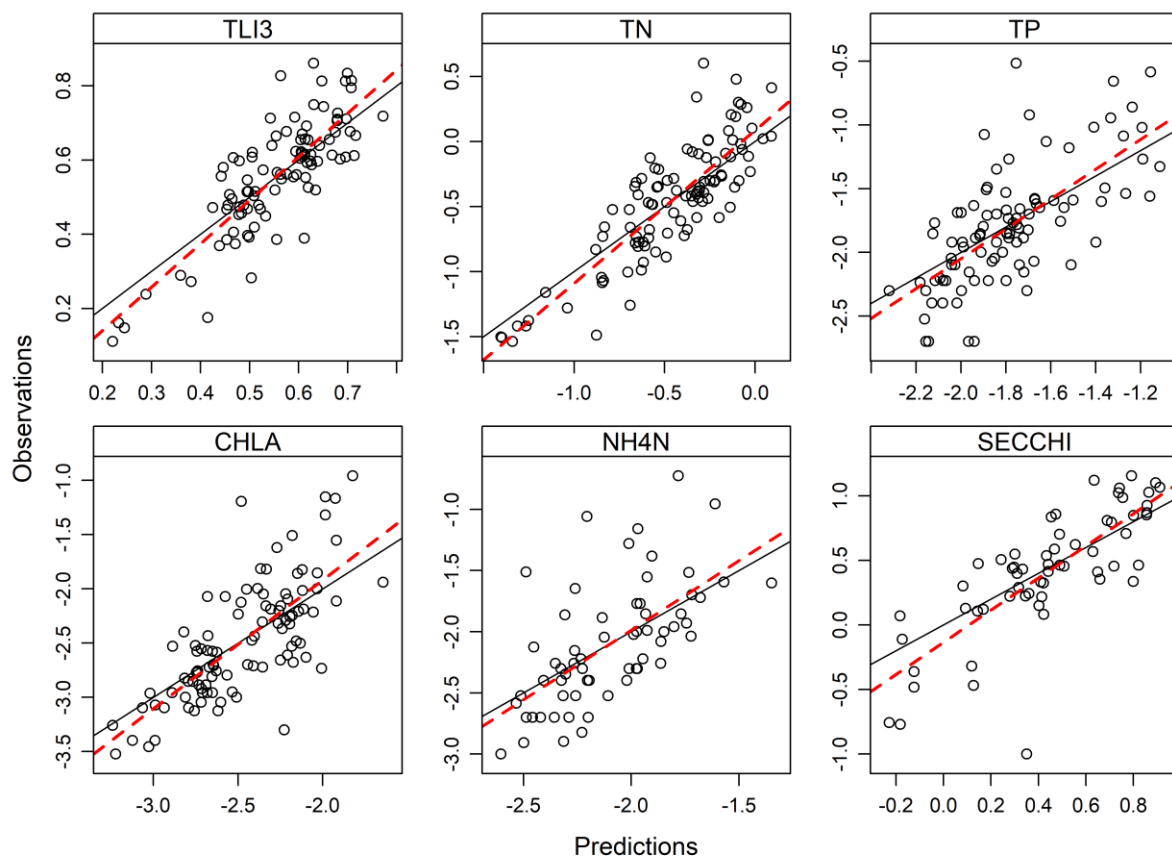


Figure 2. Comparison of observed water quality versus values predicted by the RF models. Note that the observed values are plotted on the Y-axis and predicted values on the X-axis, following Piñeiro *et al.* (2008). Red dashed line: best fit linear regression of the observed and predicted values. The solid black line is one-to-one. Units for the variables are the \log_{10} of the original units.

4.2 Modelled relationships

The reduced RF models retained only a subset of the original set of predictors (Table 5). The nineteen retained predictors (five to fourteen per model) reflected associations between water quality and lake and catchment elevation, geological and climatic factors (Table 5).

The lake water quality variables had logical relationships with many of the individual predictor variables included in the reduced RF models (Figure 3). Nutrient concentrations and chlorophyll *a* decreased and Secchi depth increased with increasing lake and catchment elevation (lkElev, catElev) and decreasing wintertime catchment air temperature (catWinTemp). This is consistent with an observed gradient in trophic conditions for lakes that is associated with altitude and climate (Sorrell *et al.*, 2006). Predictors describing catchment land cover were not retained in any of the RF models (Figure 3). However, the inclusion of elevation and catchment climate is probably partly due to these predictor's correlation with catchment land cover. TLI3, and all four nutrient variables decreased with lake fetch (lkFetch), which may reflect the generally lower trophic status of larger lakes rather the effect of wind mixing on lakes.

Table 5. Predictors retained by the reduced RF models of lake water quality variables. 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. Predictor variables are defined in Table 2.

Predictor	CHLA	NH4N	SECCHI	TLI3	TN	TP
lkElev	3	2	1	1	2	1
catElev	1	4	2	2	1	2
catWinTemp	2	10	4	3	5	7
catSlope	4	7	NA	5	4	6
lkFetch	5	13	NA	4	3	5
catHard	NA	3	NA	NA	NA	NA
catPeat	NA	NA	3	NA	NA	NA
lkSumWind	NA	6	5	8	8	NA
catPsize	NA	5	NA	NA	NA	NA
lkArea	NA	12	NA	6	7	NA
catCalc	NA	9	NA	7	12	3
catAlluv	NA	1	NA	9	11	8
lkDepth	NA	8	NA	NA	6	NA
lkDistCoast	NA	NA	NA	NA	9	NA
catPhos	NA	NA	NA	NA	NA	9
catFlow	NA	NA	NA	NA	10	NA
lkDecSolRad	NA	11	NA	NA	NA	NA
catArea	NA	14	NA	NA	14	4
catGlacial	NA	NA	NA	NA	13	NA

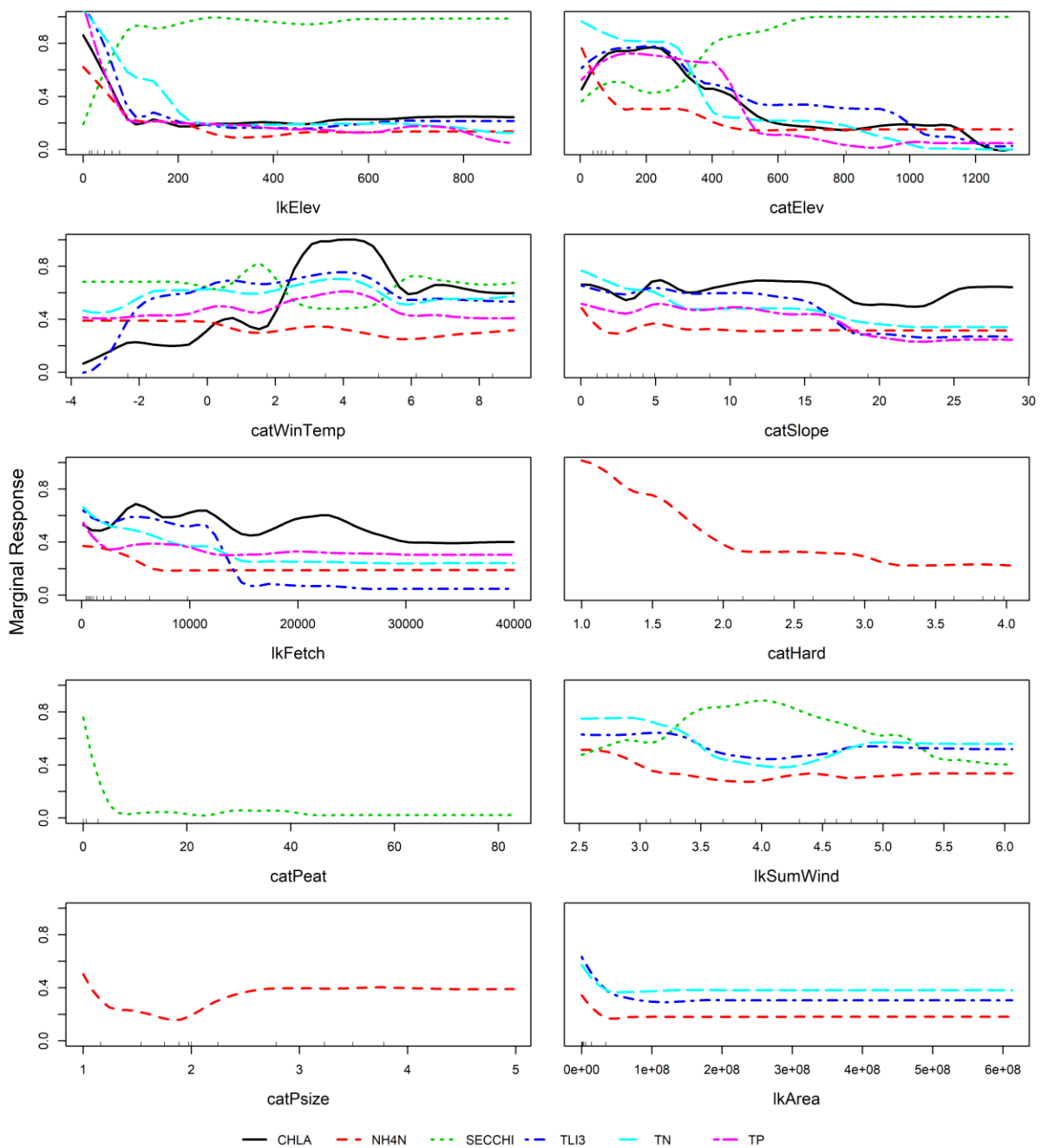


Figure 3. Partial plots for the ten most important predictor variables in the RF models of lake water quality. Each panel corresponds to one predictor. The Y-axis is the standardised value of the marginal response for each of the ten modelled variables. In each case, the original marginal responses over 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 variable'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 were slightly over-represented in environments characterised by low elevations (lkElev, catElev), and low catchment slopes (catSlope) and catchments with high alluvium (catAlluv) (Figure 4). Monitored lakes were under-represented in lakes with low fetch (lkFetch) and shallow depths (lkDepth). They were also under-represented in lake catchments with very low winter temperature (catWinTemp). For example, there were no lakes in our dataset with values of catWinTemp < -3.2 °C, however, 11% of lakes nationally have values of catWinTemp in this category. In addition, the monitored lakes were over represented in lake catchments with very high winter temperature.



Figure 4. Histograms comparing the distributions of predictor variables for all lakes and the monitored lakes used to build the RF models. The national pool of lakes is represented by the green 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 19 predictors (defined in Table 2) retained in the reduced RF models.

4.4 Model predictions

Predictions for CHLA, NH4N, SECCHI, TN, TP and TLI3 are shown in Figure 5 for the 3802 lakes that had complete data in the FENZ dataset (larger versions of the same maps are in Appendix B). The mapped predictions for all six variables had similar spatial patterns, with high values of CHLA, NH4N, TN, TP and TLI3 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). Values of CHLA, 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 the supplementary file: "LakeRF_WQModel_Predictions_7Mar19.xlsx".

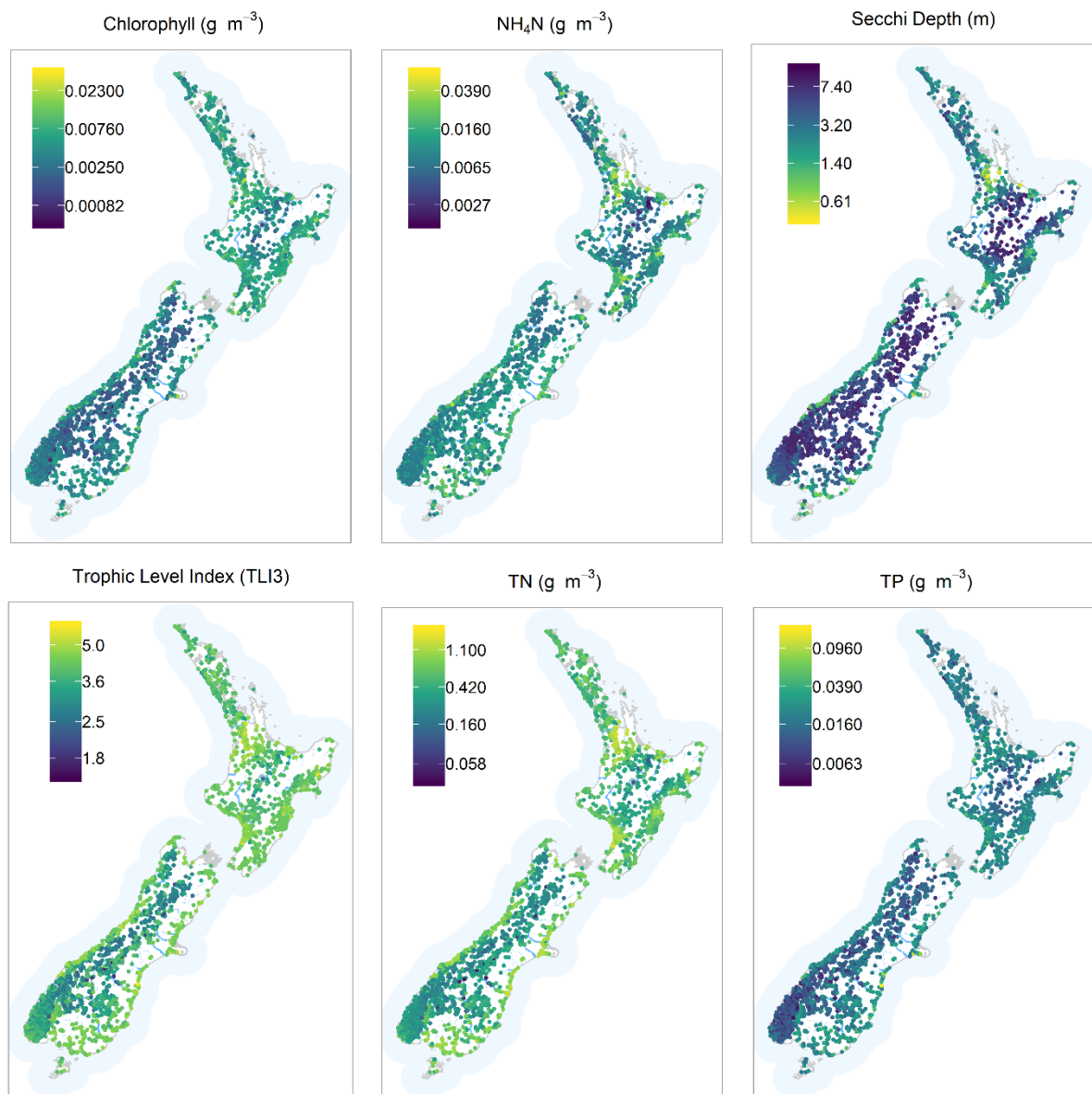


Figure 5. Predicted water quality for New Zealand lakes. The lakes are indicated by points located at the lake centre.

5 Discussion

5.1 Representativeness and modelled relationships

The lake dataset was small (<100 sites) and had a restricted geographic coverage (Figure 1). In particular, 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.

A somewhat surprising result was that the lake models included no predictors that directly described catchment land cover. It is well established that the proportion of the catchment occupied by pastoral land cover is strongly associated with magnitude of nutrient loads from agricultural source at the national scale (e.g., Larned *et al.*, 2016). However, the elevation predictors (catElev and lkElev) and the mean wintertime temperature predictor (catWinTemp) are likely included in the models partly because they are correlated with the catchment land cover. Low elevation catchments and those in warmer regions are commonly associated with greater pastoral land use intensity than catchments at higher elevations and in colder regions.

The correlative rather than causative nature of the relationships between these predictors and nutrient loads to lakes is not relevant when considering the statistical measures of predictive performance of the models. However, it does mean that the lake model predictions are unrealistic in situations where the relationship between catElev, lkElev and catWinTemp and the actual causative variables (catchment nutrient 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 (i.e., low catWinTemp) but low elevation. The model predictions are therefore 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.

5.2 Comparison with previous study

The lake water quality models represented in this study update previous modelling work carried out by Snelder *et al.* (2016), for the period 2009-2013. The same methodology was used by the two studies, so the only 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 most cases there was an increase in the number of lakes used. This study also generally had slightly higher model performance.

In general, the spatial models were very similar for both the Snelder *et al.* (2016) and current studies, in terms of model performance, predictor importance levels and prediction patterns. As such, the broad scale conclusions are the same for both studies.

5.3 Model uncertainty

In this study, we modelled broad-scale patterns in lake water quality using catchment characteristics and lake-scale descriptors as predictor variables. Because the processes determining water quality in a lake 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 4). However, the level of model bias for each water quality variable was low, which indicates that the predicted patterns reflect broad scale relative differences between lakes.

The 95% confidence intervals for median values of the water-quality variables predicted by our models 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.

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

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 ($0.025 \leq Z \leq 0.975$). The prediction confidence intervals for the \log_{10} -transformed variables, when expressed in the original units of the variables, 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 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.

RF model performance differed between modelled variables 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.

Acknowledgements

We thank Abi Loughnan of Land, Air, Water Aotearoa (LAWA) and Amy Whitehead of NIWA for assistance with for obtaining the lake water quality data. We thank Dan Elder, Annabelle Ellis and Sheree De Malmanche (Ministry for the Environment) for project support and feedback. We thank Clive Howard-Williams and 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.
- Burns, N., J. Rutherford, and J. Clayton, 1999. A Monitoring and Classification System for New Zealand Lakes and Reservoirs. *Lake and Reservoir Management* 15:255–271.
- 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.
- 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.T., T. Snelder, M. Unwin, and G.B. McBride, 2016. Water Quality in New Zealand Rivers: Current State and Trends. *New Zealand Journal of Marine and Freshwater Research* In Press. doi:<http://dx.doi.org/10.1080/00288330.2016.1150309>.
- Larned, S.T. and M. Unwin, 2012. Representativeness and Statistical Power of the New Zealand River Monitoring Network. NIWA Client Report, NIWA, Christchurch, New Zealand.
- Larned, S., A. Whitehead, C. Fraser, T. Snelder, and J. Yang, 2018. Water Quality State and Trends in New Zealand Rivers. Analyses of National-Scale Data Ending in 2017. NIWA, 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.
- 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.
- 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.H., S.T. Larned, and R.W. McDowell, 2017. Anthropogenic Increases of Catchment Nitrogen and Phosphorus Loads in New Zealand. *New Zealand Journal of Marine and Freshwater Research*:1–26.
- 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.

Appendix A Exploration of lake loads as predictors for lake random forest models

A recent study by (Snelder *et al.*, 2017) made predictions of nutrient loads for all segments in the national digital river network (REC2). By identifying those reaches that drain into lakes, it is possible to evaluate incoming lake nutrient loads using these estimates. Although these model predictions are uncertain, they have the potential to be useful predictors in the random forest modelling of lake water quality state.

In this appendix we explore the effect of including these additional predictors in the random forest models of lake water quality state. We examine the significance of any improvement in model performance (as compared to models where load estimates were not included as predictors), and compare predictions made with and without loads.

The main disadvantage of including the loads as predictors, is that there are fewer load predictions (978) than lakes in the lakes database (3802), and there are also fewer lakes in the training dataset (Table 6). This is because inflow streams and rivers for smaller lakes are often not represented by the digital river network because their catchment areas are insufficiently large to define a network segment. The numbers of lakes with estimated loads and observations of SECCHI and NH4N were too few to fit random forest models, and these variables were excluded from the following analysis.

Table 6: Number of lakes in the model training set and number of lakes that also have load predictions

Model	Number of lakes	Number of lakes with load estimates
CHLA	101	78
NH4N	64	45
SECCHI	61	46
TLI3	99	76
TN	104	81
TP	99	74

We evaluated the statistical significance of the difference in model performance by determining the 95% confidence intervals for R^2 values for the random forest models that included and excluded the estimated loads as predictors. To determine the confidence intervals of R^2 , we refitted the random forest model 500 times and used the R^2 values returned for each realisation to determine the overall mean and standard deviation of R^2 . Note, we refitted the random forest models without loads using only the subset of lakes that had load predictions (978 lakes), hence there are some differences in the model performance presented in this appendix compared with the results in the body of the report.

We calculated the 95% confidence intervals of R^2 for each model from its overall mean and standard deviation of R^2 . We inferred that there was a statistically significant difference in model performance when the 95% confidence intervals did not overlap. The results are demonstrated in Table 7 and Figure 6. . Three of the four variables (TLI3, TN and TP) had

statistically significant improvements in model performance. The increases in model performance were most notable for TP.

Table 7: Performance of lake water quality models with and without nutrient loads as predictors.

Variable	R ²		95% CI	
	Model without loads		Model including loads	
CHLA	0.50	(0.49-0.51)	0.49	(0.47-0.5)
TLI3	0.64	(0.63-0.65)	0.67	(0.66-0.67)
TN	0.69	(0.68-0.69)	0.72	(0.71-0.73)
TP	0.38	(0.36-0.39)	0.52	(0.51-0.53)

Figure 6 shows both the performance of the random forest models with and without loads as predictors, compared against observed water quality. Overall, the patterns of both models with the observations are very similar.

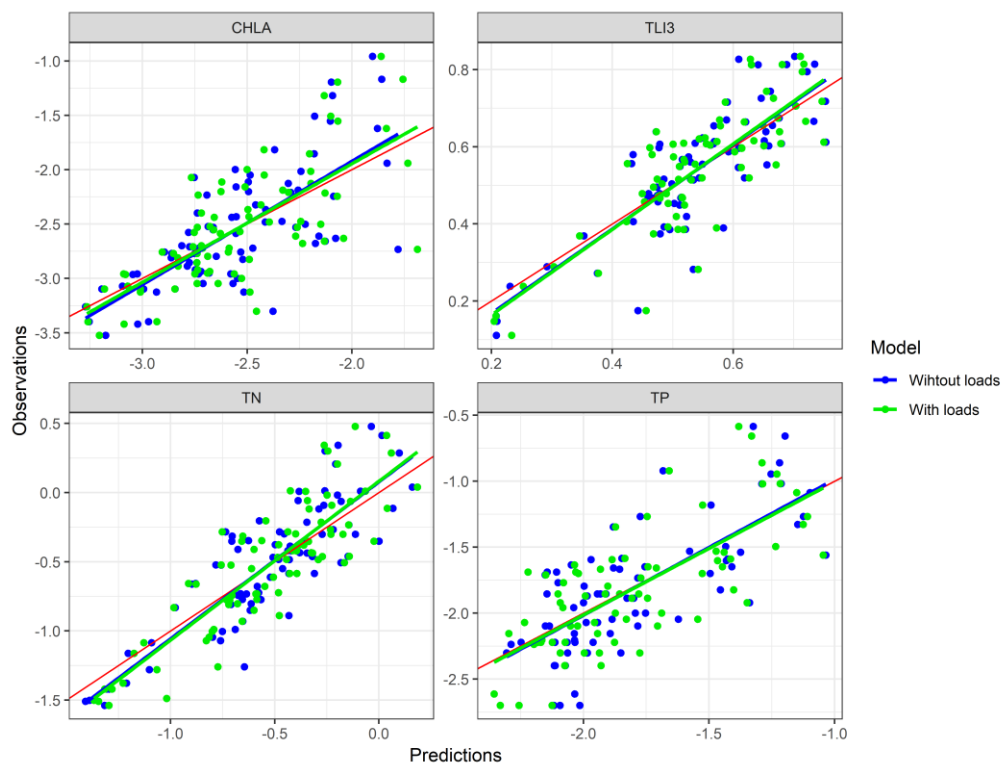


Figure 6. Comparison of observed water quality versus values predicted by the RF models. Note that the observed values are plotted on the Y-axis and predicted values on the X-axis, following Piñeiro et al. (2008). Blue and Green lines: best fit linear regression of the observed and predicted values, for the model with and without loads as predictors, respectively. The solid red line is one-to-one. Units for the variables are the log₁₀ of the original units.

Both random forest models were used to make predictions for all lakes that had load predictions (978). Figure 7 shows a comparison of the national scale predictions of water quality. TP is the only variables for which there appears to be any non-random differences in the predictions, with the RD model including loads tending to predict slightly higher concentrations at low concentrations, compared to the RF model excluding loads as predictors.

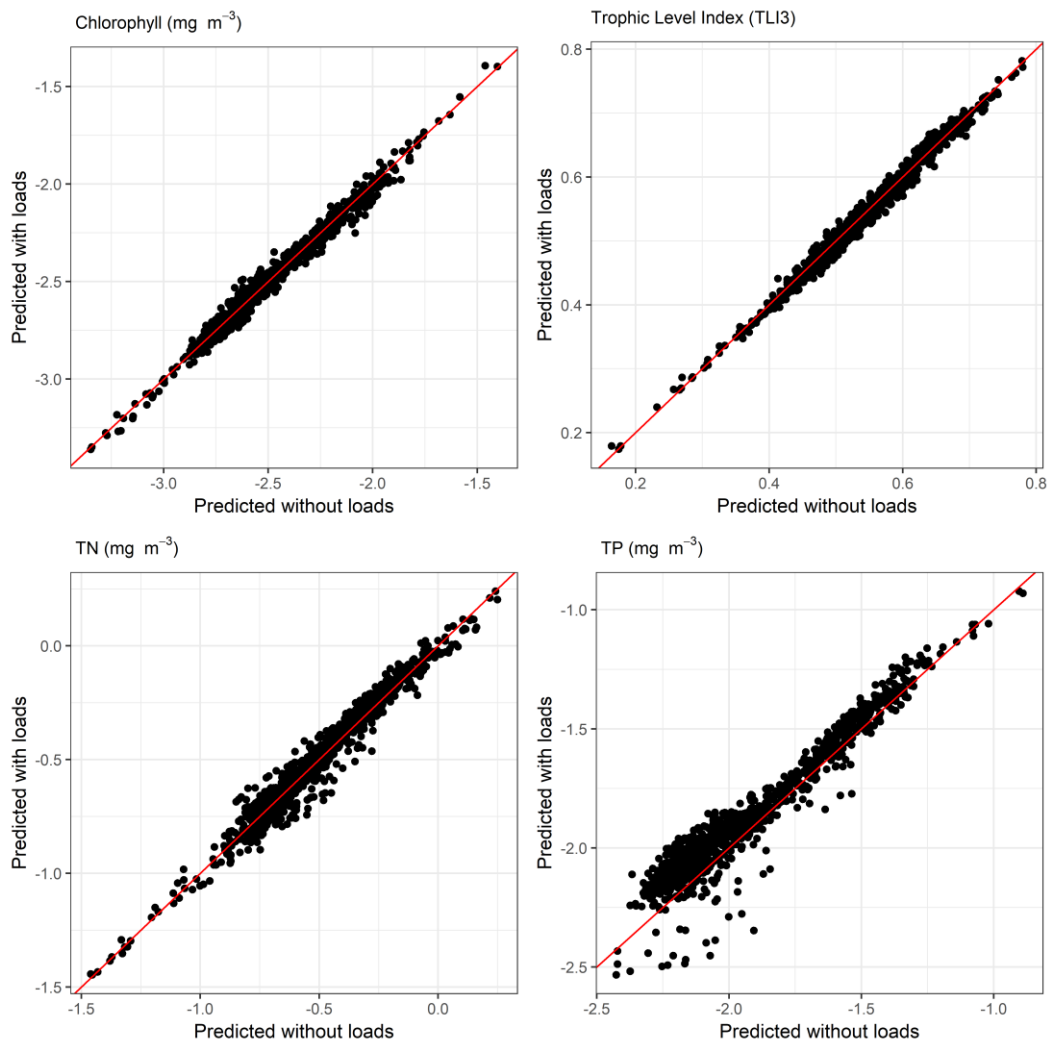


Figure 7. Comparison of predicted water quality for 978 lakes (those with load predictions) from RF models.. The solid red line is one-to-one. Units for the variables are the log₁₀ of the original units.

Appendix B Mapped predictions from random forest models

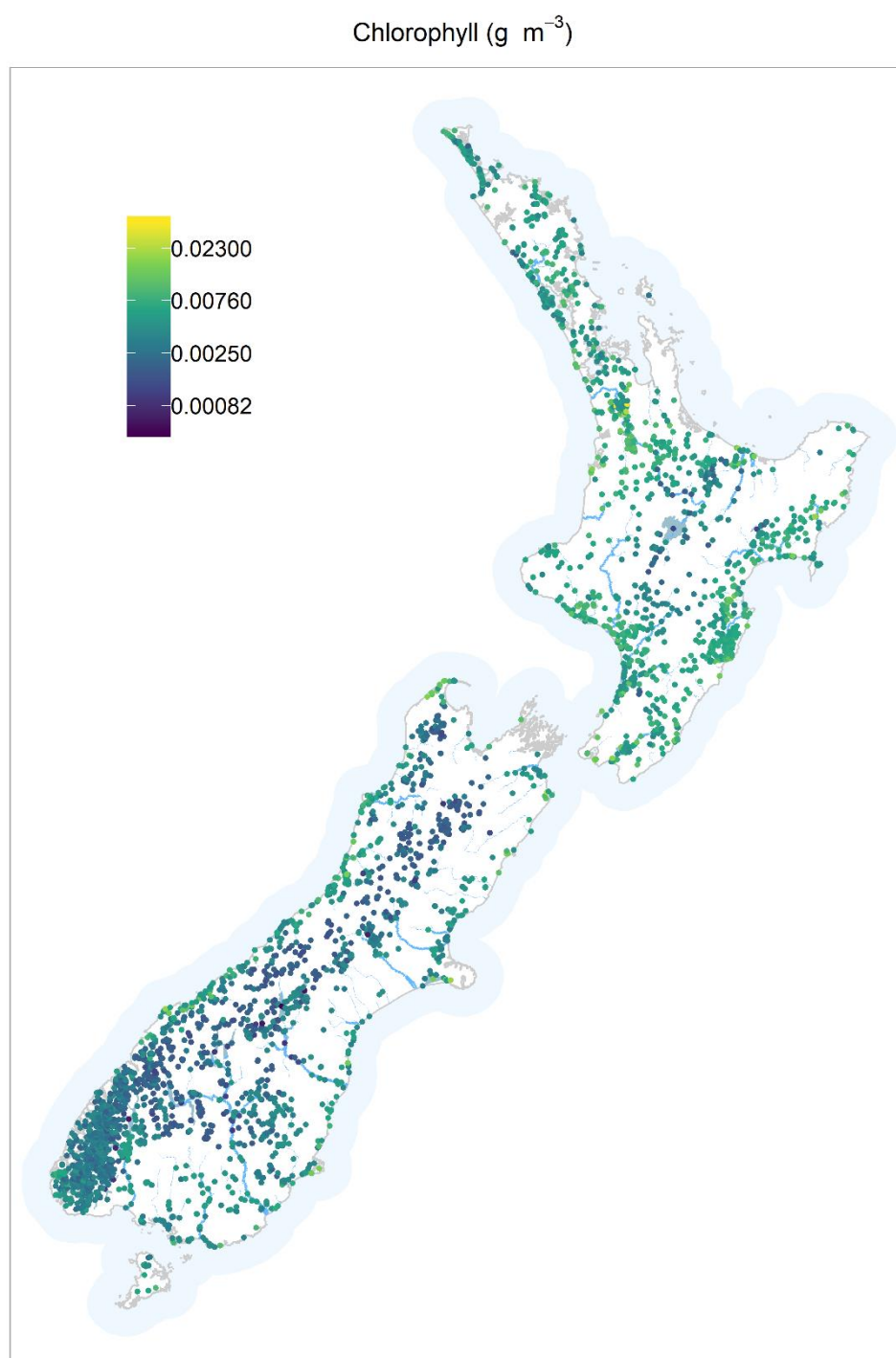


Figure 8. Predicted chlorophyll for New Zealand lakes. The lakes are indicated by points located at the lake centre

NH₄N (g m⁻³)

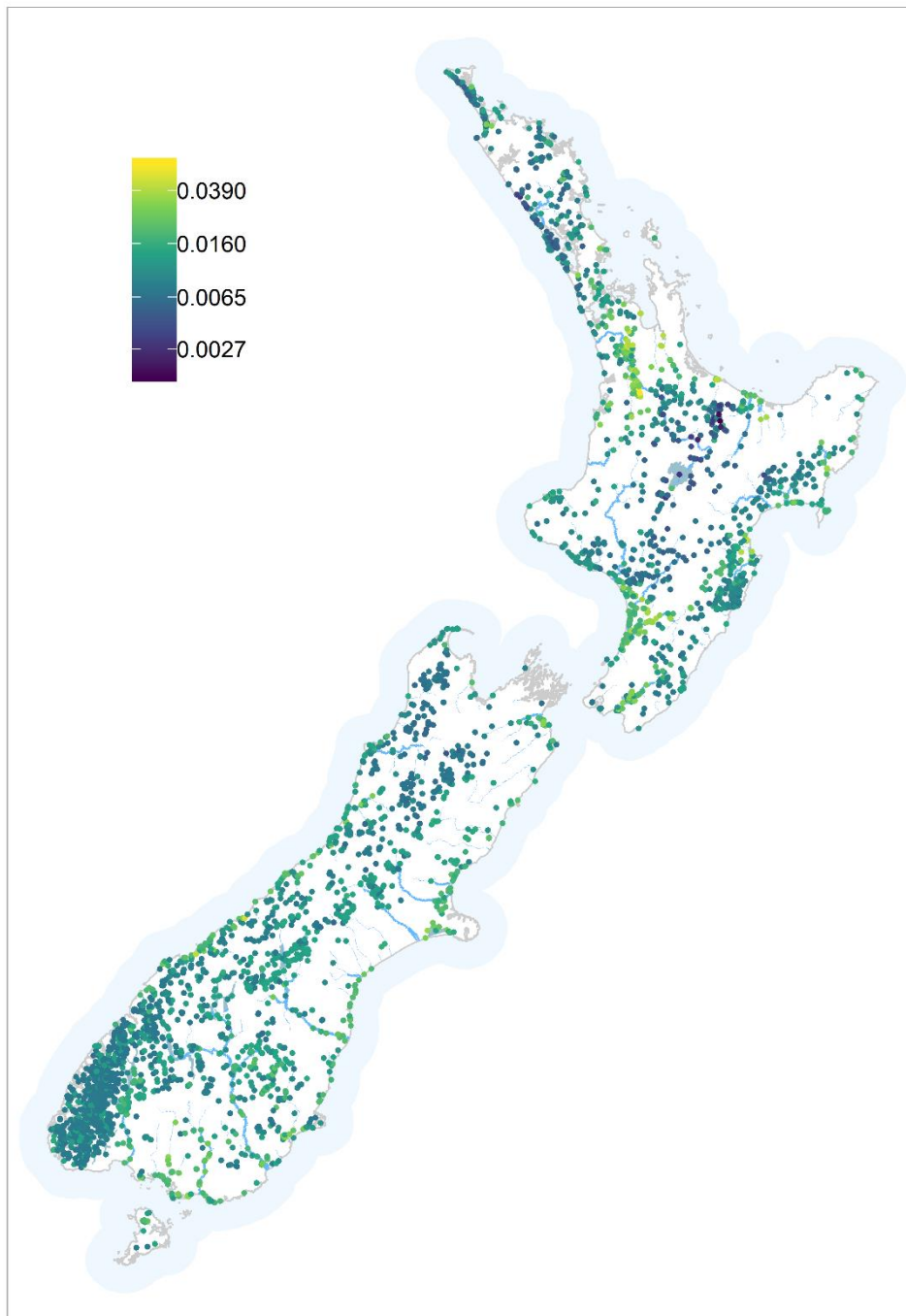


Figure 9. Predicted NH₄N for New Zealand lakes. The lakes are indicated by points located at the lake centre

Secchi Depth (m)

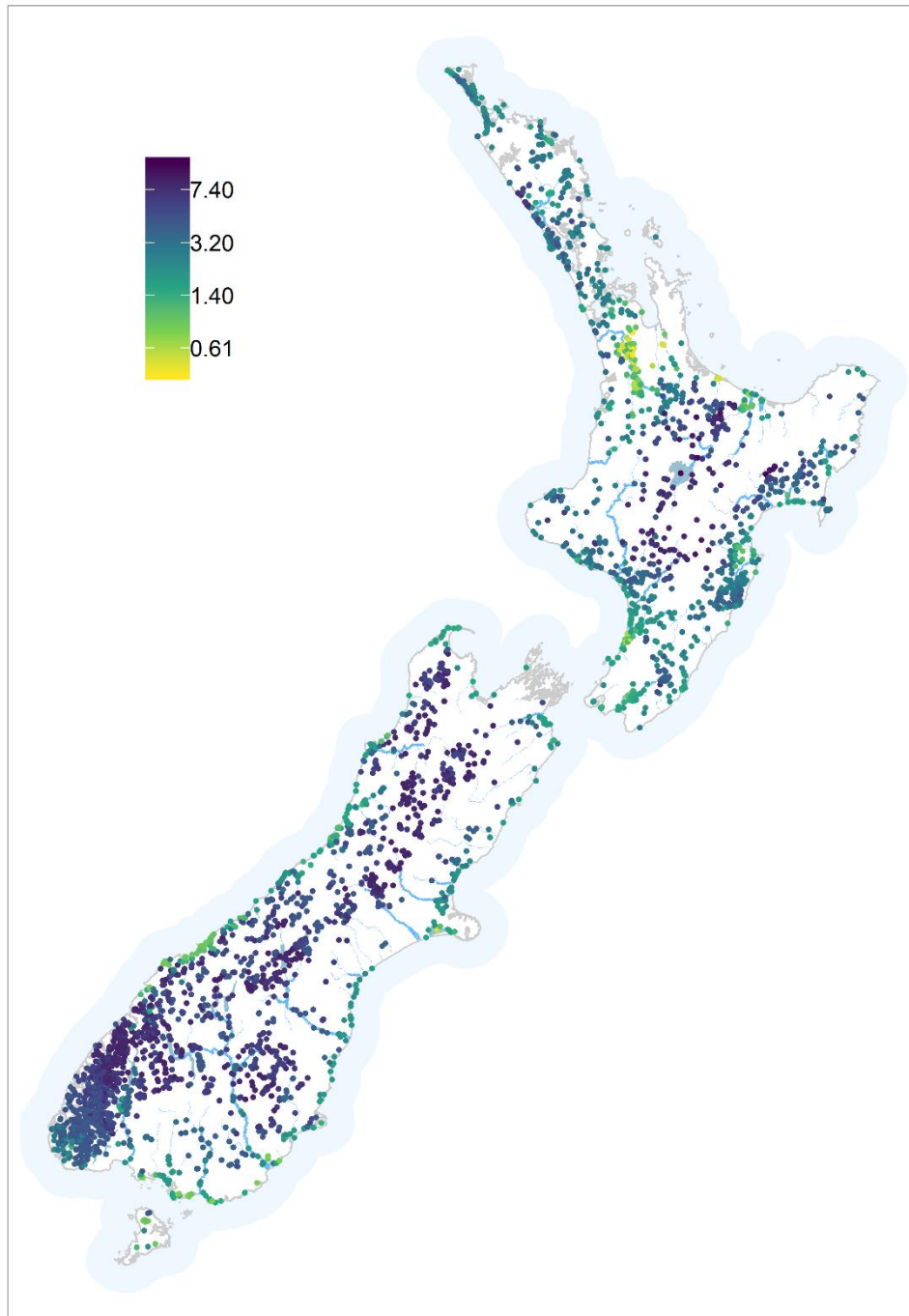


Figure 10. Predicted SECCHI for New Zealand lakes. The lakes are indicated by points located at the lake centre

Trophic Level Index (TLI3)

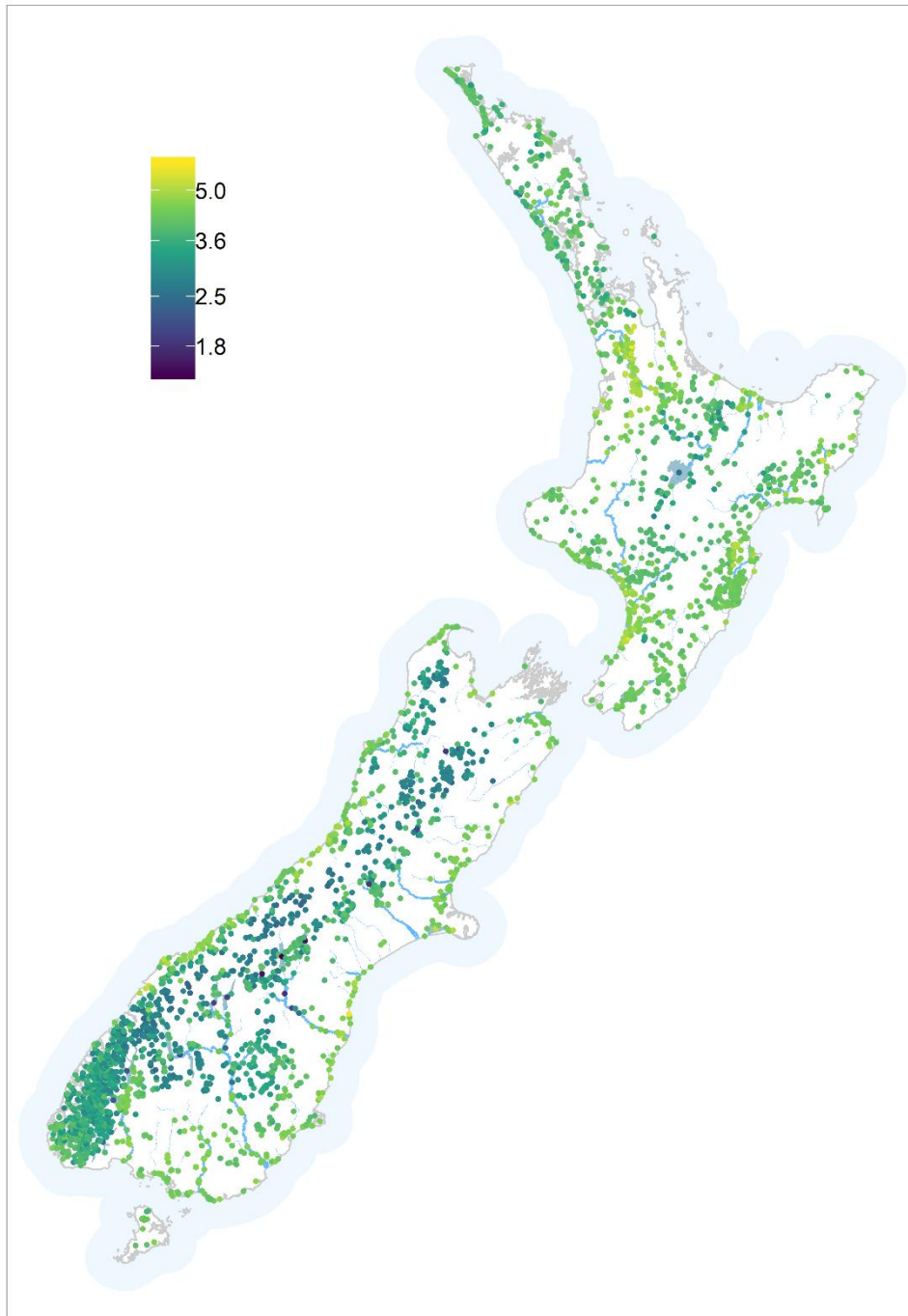


Figure 11. Predicted TLI3 for New Zealand lakes. The lakes are indicated by points located at the lake centre

TN (g m^{-3})

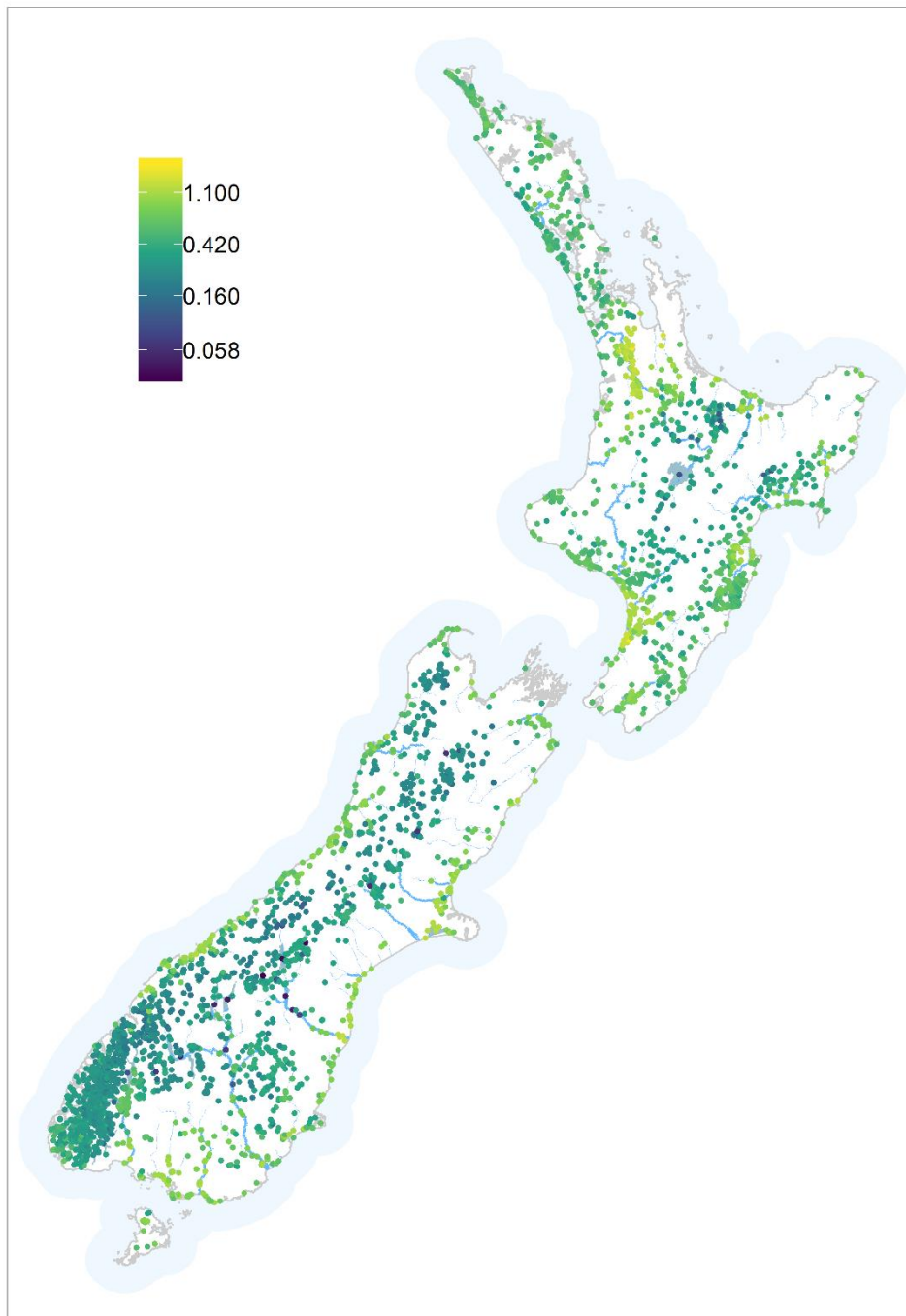


Figure 12. Predicted TNI for New Zealand lakes. The lakes are indicated by points located at the lake centre

TP (g m^{-3})

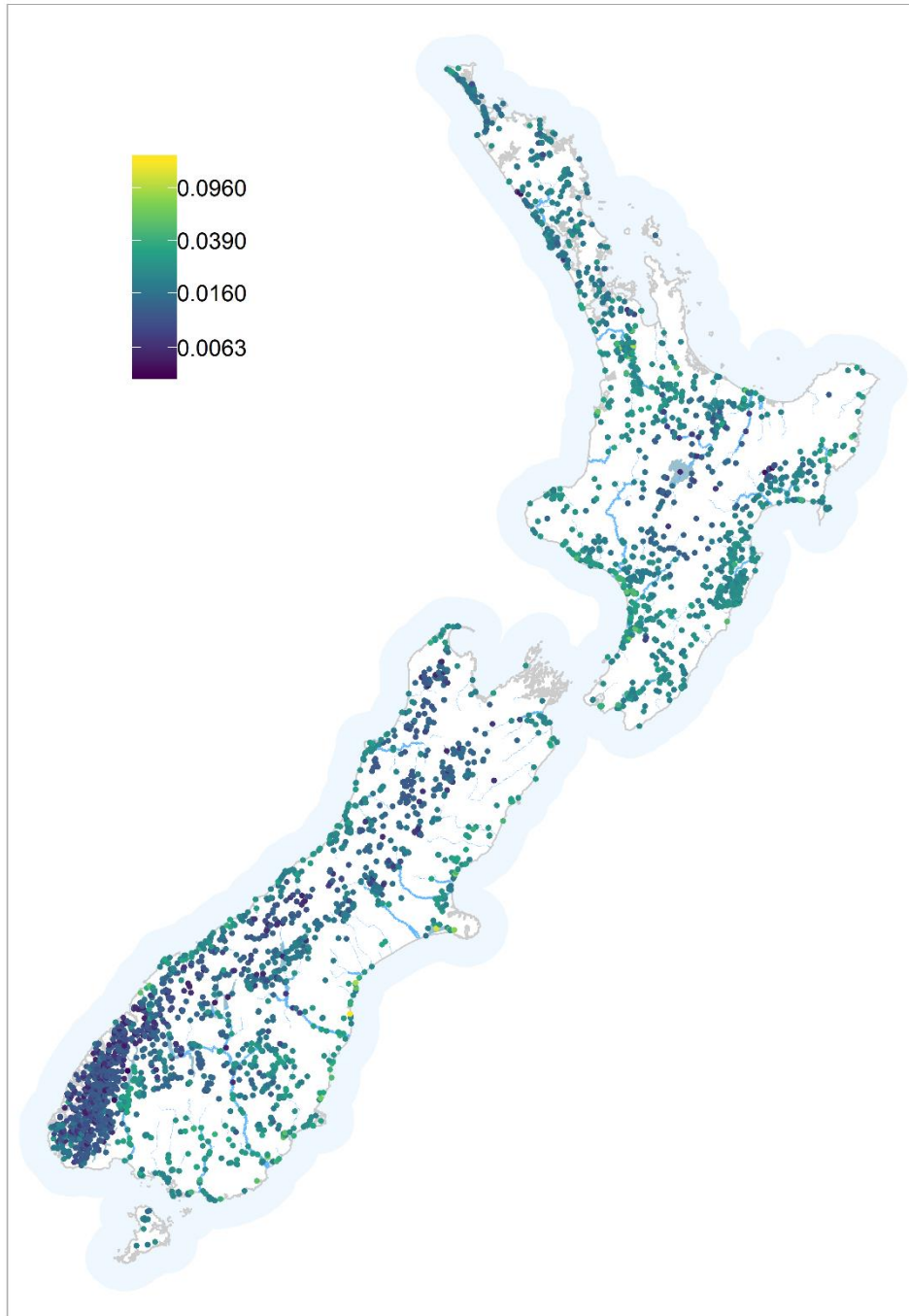


Figure 13. Predicted TP for New Zealand lakes. The lakes are indicated by points located at the lake centre