



Earth Sciences
New Zealand

Spatial modelling of river water-quality state

Incorporating monitoring data from 2020 to 2024

Prepared for The Ministry for the Environment

November 2025

Prepared by:
David Wood
Doug Booker
Rachel Smith

For any information regarding this report please contact:



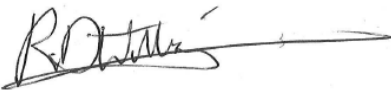
David Wood
Water Quality Scientist
Freshwater Modelling Group
+64 3 343 8050
david.wood@niwa.co.nz

New Zealand Institute for Earth Science Limited
PO Box 8602
Riccarton
Christchurch 8440

Phone +64 3 348 8987

Client Report No: 2025348CH
Report date: November 2025
Project No: MFE25504

Revision	Description	Date
Version 1.0	Final version sent to client	07 November 2025

Quality Assurance Statement		
	Reviewed by:	Aidin Jabbari
	Formatting checked by:	Terry Smith
	Approved for release by:	Richard Williams

© New Zealand Institute of Earth Science Limited (“Earth Sciences New Zealand”) 2025. All rights reserved. This publication may not be reproduced or copied in any form without the permission of the copyright owner(s). Such permission is only to be given in accordance with the terms of the client’s contract with Earth Sciences New Zealand. This copyright extends to all forms of copying and any storage of material in any kind of information retrieval system.

Whilst Earth Sciences New Zealand has used all reasonable endeavours to ensure that the information contained in this document is accurate, Earth Sciences New Zealand does not give any express or implied warranty as to the completeness of the information contained herein, or that it will be suitable for any purpose(s) other than those specifically contemplated during the project or agreed by Earth Sciences New Zealand and the client.

Contents

Executive summary	5
Purpose	5
Methods	5
Results	6
1 Introduction	7
2 Data	8
2.1 River attribute state data	8
2.2 Predictor data	12
3 Modelling methods	16
3.1 Random forest models	16
3.2 Model performance	17
3.3 Representativeness of monitoring sites used in random forest models	18
3.4 Model predictions	19
4 Results	20
4.1 Model performance	20
4.2 Modelled relationships	22
4.3 Monitoring site representativeness	27
4.4 Model predictions	30
5 Discussion	48
5.1 Comparison with previous studies	48
5.2 Model uncertainty	48
5.3 Alternative modelling methods	49
6 Acknowledgements	51
7 Glossary of abbreviations and terms	52
8 References	54

Tables

Table 2-1:	River attributes, measurement units and site numbers used to develop random forest models.	8
Table 2-2:	Predictors used in random forest models of river attribute states.	13
Table 3-1:	Performance ratings for statistics used in this study.	18
Table 4-1:	Performance of the 16 attribute state models.	20
Table 4-2:	Rank order of importance of predictors retained in the random forest models for at least one attribute state.	23
Table 4-3:	Comparisons of the minimum and maximum observed and predicted values of water quality attribute states.	30

Figures

Figure 2-1:	Locations of river monitoring sites used for modelling the current state of the 16 attributes.	11
Figure 4-1:	Comparison of observed attribute state versus values predicted by each of the random forest models.	21
Figure 4-2:	Partial plots for the 12 most important predictors in random forest models of current attribute state.	25
Figure 4-3:	Partial plots for the 12 most important predictors for the median MCI random forest model.	26
Figure 4-4:	The distributions of predictors across all segments in the digital river network and at water quality sites	28
Figure 4-5:	Distributions of predictors across all segments in the digital river network and at invertebrate sites	29
Figure 4-6:	Predicted median clarity in New Zealand rivers.	32
Figure 4-7:	Predicted median turbidity in New Zealand rivers.	33
Figure 4-8:	Predicted median ammoniacal nitrogen $\text{NH}_4\text{-N}$ in New Zealand rivers.	34
Figure 4-9:	Predicted median ammoniacal nitrogen $\text{NH}_4\text{N-adj}$ adjusted for pH in New Zealand rivers.	35
Figure 4-10:	Predicted 95 th percentile ammoniacal nitrogen $\text{NH}_4\text{N-adj}$ adjusted for pH in New Zealand rivers.	36
Figure 4-11:	Predicted median NNN concentration in New Zealand rivers.	37
Figure 4-12:	Predicted 95 th percentile NNN concentration in New Zealand rivers.	38
Figure 4-13:	Predicted median TN concentration in New Zealand rivers.	39
Figure 4-14:	Predicted median DRP concentration in New Zealand rivers.	40
Figure 4-15:	Predicted 95 th percentile DRP concentration in New Zealand rivers.	41
Figure 4-16:	Predicted median TP concentration in New Zealand rivers.	42
Figure 4-17:	Predicted median <i>E. coli</i> in New Zealand rivers.	43
Figure 4-18:	Predicted 95 th percentile <i>E. coli</i> in New Zealand rivers.	44
Figure 4-19:	Predicted proportion of samples exceeding 260 <i>E. coli</i> 100 ml ⁻¹ in New Zealand rivers.	45
Figure 4-20:	Predicted proportion of samples exceeding 540 <i>E. coli</i> 100 ml ⁻¹ in New Zealand rivers.	46
Figure 4-21:	Predicted median MCI scores in New Zealand rivers.	47

Executive summary

Purpose

The New Zealand Ministry for the Environment (MfE) and Stats NZ Tātauranga Aotearoa (Stats NZ) use river water quality data originating from state-of-the-environment (SoE) monitoring to inform policy development and meet their requirements for environmental reporting on the freshwater domain under the Environmental Reporting Act 2015. Analysis of monitored data can describe water quality at SoE monitoring sites but does not provide insight into conditions at unmonitored sites. Raw monitored data for each water quality variable are often summarised over a specified period using a specified summary statistic. The combination of variable and statistic is called an attribute. This report supplements analysis of monitored data by providing model-based predictions of current water quality attribute states for all rivers across New Zealand.

Methods

This report provides model-based predictions of current water quality attribute states for each of approximately 593,000 unique river segments that comprise New Zealand's national digital river network. These predictions are based on the observations of the state of water quality in rivers for the period 2020–2024. Four comparable reports were produced in 2010, 2016, 2019 and 2022, using data for the periods 2003–2007, 2009–2013, 2013–2017 and 2016–2020, respectively. This report is the second in a series of reports prepared for the Ministry for the Environment on the topic of national-scale state and trends in river freshwater quality. The first report described data collation and analysis that provided site-specific river water quality state and trends for approximately 1,000 river monitoring sites operated by Regional Councils and Earth Sciences New Zealand. The river water quality data acquired and processed for the first report were used as input to the predictions presented in this report.

The predicted water quality states presented in this report were generated using random forest models, an advanced form of regression-tree modelling. Unlike single regression trees, which can be unstable and sensitive to small changes in data, random forests build many trees and base their predictions on the average outcome, improving accuracy and reliability. This approach is particularly well-suited to water quality analysis because it can handle complex and variable datasets, works well when predictors are intercorrelated, and does not require strict statistical assumptions about the data. Random forests are also less prone to overfitting than many other modelling methods, making them a robust tool for predicting water quality states.

Random forest models were developed for ten water quality variables: visual clarity (Clarity), turbidity (Turbidity), ammoniacal nitrogen ($\text{NH}_4\text{-N}$), pH-adjusted ammoniacal nitrogen ($\text{NH}_4\text{-N-adj}$), nitrate + nitrite-nitrogen (NNN), total nitrogen (TN), dissolved reactive phosphorus (DRP), total phosphorus (TP), *Escherichia coli* (*E. coli*), and the macroinvertebrate community index (MCI). These variables were combined with a selection of statistics (e.g., median and 95th percentile) describing aspects of the distribution of the observed values. With the exception of turbidity and ammoniacal nitrogen, the attributes are defined by the National Policy Statement for Freshwater Management (NPS-FM). Model predictors comprised 40 variables, mapped to the digital river network (DN2.4). These predictors were selected to represent climatic,

geological, topographic, land cover, and hydrological conditions across New Zealand rivers and their catchments.

The observational data used in the random forest models consisted of site attribute states from monthly and quarterly measurements (and annual invertebrates for MCI scores) for the period 2020–2024. These data came from 695–992 monitoring sites (depending on the attribute state). The sites are reasonably well distributed across the North Island and South Island, although there are some gaps in inaccessible areas. To assess the extent to which the monitoring sites used for observational data capture the range of environmental conditions in New Zealand, we compared density plots of the distributions of predictor values for the monitoring sites with those for all river segments in New Zealand. The monitoring sites were reasonably representative, with moderate over-representation of low-elevation, low-slope locations with large proportions of intensive agricultural land cover. Locations with higher elevation, higher slope, and more natural landcover were therefore less well represented.

Results

The random forest models generally performed well in predicting attribute states, as indicated by high variance explained, close observed versus predicted agreement, low bias (no tendency to systematically over- or under-estimate), and low uncertainty. Model performance, importance/selection of predictors and predicted values representing the period 2020–2024 produced for this report were commensurate with those produced by previous work that predicted attribute states for the period 2016–2020.

The six most important predictors, when averaged across models of nutrients (NNN, NH₄-N, NH₄-N-adj, TN, DRP, TP), clarity, turbidity and *E. coli* were: proportion of catchment occupied by pasture (and other high production land), proportion of catchment occupied by alluvial geology, catchment elevation, slope and rainfall, as well as the density of stock. Collectively, the models suggest that attribute state is worse in low-elevation, low-slope land under intensive land use compared to elsewhere.

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

1 Introduction

State-of-the-environment (SoE) monitoring for river water quality involves collecting, analysing, and reporting on physical, chemical, and biological data to assess health and trends over time. SoE monitoring involves regularly measuring a suite of standard variables at representative sites to provide a comprehensive and unbiased understanding of water quality condition (state) and its changes (trends).

Only a minority of New Zealand rivers are monitored as part of the SoE process. In the companion report, (Booker et al. 2025) noted that SoE monitoring was conducted at approximately 1,000 locations on rivers, and that there are over 540,000 river segments in the digital river network. The New Zealand Ministry for the Environment (MfE) and Stats NZ Tauranga Aotearoa (Stats NZ) asked Earth Sciences New Zealand to estimate the state of water quality in all river segments in the digital network. These estimates can then be used to inform policy and for other purposes.

Specifically, Earth Sciences NZ was asked to update the river water quality spatial modelling report prepared by Whitehead et al. (2022a). This update would take into account the summary statistic representing conditions up to the end of 2024 from (Booker et al. 2025). Other model inputs, referred to as predictors, come from the Land Cover Database (LCDB), the Fundamental Soil Layers (FSL), the Digital Network (DN), and the Agriculture Production Statistics (APC) from Stats NZ. MfE requested that, for each water quality attribute, the state will be predicted for all segments in the digital river network. MfE also requested that predictive performance be assessed, including estimates of modelling uncertainty.

We describe the data preparation, predictor selection, site representativeness assessment, modelling process, and model performance evaluation. The results section presents national predictions, identifies key predictors, and quantifies model performance. The discussion compares these models with previous ones, addresses uncertainties, and explores alternative methods. We also provided MfE with the model outputs for every river segment of a national digital river network in the supplementary file “RiverRF_WQModel_Predictions_2025-11-04csv”.

2 Data

Predictive modelling aims to establish a relationship between response variables (such as river attribute state data) and predictor variables (which include input data). Once this relationship is defined, it can be used to estimate attribute states for segments on rivers where direct measurements are not available, provided that predictor data are accessible.

To achieve this, response and predictor data must be combined into a training dataset. This integration is facilitated by using unique river segment identifiers from the river digital network, ensuring that data from each river segment are correctly matched.

The remainder of this section describes the river attribute state data and the predictor variables used in the modelling process.

2.1 River attribute state data

River attribute state data, for a five-year period (2020 to 2024), except for MCI (July 2019 to June 2024), were derived from monitoring results collated by (Booker et al. 2025). The attribute state data consisted of physical, chemical, microbiological, and invertebrate variables summarised over the five-year periods using a specified statistic (median etc) from river monitoring sites in council SoE networks and the NRWQN sites (Table 2-1). Detailed methods for processing the water quality observations are given in (Booker et al. 2025).

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

Attribute type	Variable	Abbreviation	Statistic	Units	Sites
Physical	Visual clarity	Clarity	Median	m	695
	Turbidity	Turbidity	Median	NTU or FNU	940
Chemical	Ammoniacal nitrogen	NH ₄ -N	Median	mg L ⁻¹	992
	Ammoniacal nitrogen adjusted for pH	NH ₄ -N-adj	Median, 95 th percentile	mg L ⁻¹	974
	Nitrate + nitrite-nitrogen	NNN	Median, 95 th percentile	mg L ⁻¹	991
	Total nitrogen (unfiltered)	TN	Median	mg L ⁻¹	990
	Dissolved reactive phosphorus	DRP	Median, 95 th percentile	mg L ⁻¹	992
	Total phosphorus (unfiltered)	TP	Median	mg L ⁻¹	990
Microbiological	<i>Escherichia coli</i>	<i>E. coli</i>	Median, 95 th percentile	cfu or MPN 100 ml ⁻¹	965
			>260	% exceedances	899
			>540	% exceedances	854
Invertebrate	Macroinvertebrate Community Index	MCI	Median	unitless	874

Visual water clarity is a measure of light attenuation due to absorption and scattering by dissolved and particulate material in the water column. Clarity is monitored because it affects primary production, plant distributions, animal behaviour, aesthetic quality and recreational values, and because it is correlated with suspended solids, which can impede fish feeding and cause riverbed sedimentation. Visual clarity in rivers is generally measured *in situ* as the horizontal sighting range of a black disc (Ministry for the Environment 1994). At a few sites, clarity is measured adjacent to the river with water samples in clarity tubes.

Turbidity refers to light scattering by suspended particles. Turbidity is generally measured *in situ* with hand-held nephelometers or with a bench-top nephelometer in a laboratory, using grab samples of water from the monitoring site. Both types of nephelometers are calibrated with standard light-scattering solutions (e.g., formazin). The sensor reading is not absolute light scattering, but light-scattering relative to the standard solution, in “nephelometric turbidity units” (NTU) or “formazin nephelometric units” (FNU) depending upon the light source wavelength and calibration standard being used. However, different instruments using the same methods can result in numerically different results with the same water quality (Davies-Colley et al. 2021). Nephelometric turbidity is generally inversely correlated with visual water clarity (Davies-Colley and Smith 2001), but unlike visual clarity, turbidity measurements do not account for the optical effects (i.e., absorption) of dissolved materials.

The five nutrient species (NNN, $\text{NH}_4\text{-N}$, DRP, TN and TP) were included in our analysis because they influence the growth of benthic river algae (periphyton) and vascular plants (macrophytes), and because nitrate and ammonia can be toxic to aquatic organisms at elevated concentrations. Nutrient enrichment from point and non-point source discharges is strongly associated with intensive land use in New Zealand (Larned et al. 2016b; Snelder et al. 2018). Nutrient enrichment can promote excessive growth of “nuisance” periphyton and macrophytes that can, in turn, degrade river habitat, increase daily fluctuations in dissolved oxygen and pH, impede flows, block water intakes, and cause water colour and odour problems. At elevated concentrations, nitrate and ammonia can be toxic to animals, including river fish and invertebrates (Hickey 2013; Hickey 2014). Mechanisms of nitrate and ammonia toxicity include reduced oxygen transport by haemoglobin, carcinogenic nitrosamine formation, and disruption of ion transport across cell membranes (Camargo et al. 2005). Water quality in rivers also of interest because it is associated with delivery of nutrients to receiving environment such as lakes and estuaries (Dudley et al. 2020; Plew et al. 2020). When in solution, ammonia occurs in two forms: the ammonium cation (NH_4^+) and unionised ammonia (NH_3); the relative proportions of the forms are strongly dependent on pH (and temperature). Unionised ammonia is more toxic to fish than ammonium, hence the total ammonia toxicity increases with increasing pH (and/or temperature) (ANZECC and ARMCANZ 2000). The NPS-FM 2020 attributes related to ammoniacal-N concentrations in freshwater require a correction to account for pH and temperature. Despite this requirement, the results in the current report are not temperature-corrected following the methods of (Whitehead et al. 2022b) which did not apply temperature correction due to insufficient temperature data.

The concentration of the bacterium *Escherichia coli* (*E. coli*) is used as an indicator of human or animal faecal contamination, which is associated with the risk to humans arising from infection or illness from waterborne pathogens during contact-recreation.

In addition to the physical, chemical, and microbiological variables described above, we used the New Zealand Macroinvertebrate Community index (MCI) as a biotic indicator of general river

health. MCI scores are calculated using tolerance values for the macroinvertebrate taxa present in benthic samples, using presence/absence data which are widely available. Tolerance values are weighting factors that correspond to the relative abundance of taxa along stressor gradients. We used the non-quantitative MCI rather than the quantitative (qMCI) or semi-quantitative (sqMCI) forms of MCI because some council datasets do not include invertebrate abundance data (Stark and Maxted 2007). Most MCI data were supplied by the collecting agency as calculated scores rather than raw invertebrate data.

The frequency of water quality monitoring can vary depending on the variable and location. Water quality monitoring often occurs through monthly or quarterly measurements, but can also occur annually (e.g., for invertebrates used to calculate MCI scores).

The number of water quality observations per site for which measurements were available ranged from 18 (quarterly observations for five years with two missing observations) to 60 (monthly observations for five years with no missing observations), with a median of 57, except for clarity, which had a median of 55. For the annually sampled MCI the minimum, median, and maximum number of observations were 4, 5, and 10 samples per site, respectively. In addition, we shifted the 2020–2024 time period by six months (1 July 2019 – 30 June 2024) in order to prevent splitting summer samples into two calendar years, and only included samples if they occurred from December to March to align with the NPS-FM MCI attribute (New Zealand Government 2024). The final dataset used for random forest modelling included 1382 distinct sites, all of which were mapped to 1342 digital river segments (nzsegment).

Predictive performance of random forests can be improved by transforming the response variable before modelling. Transformation can be particularly beneficial in cases where the response variable is strongly skewed, which was the case for many of the water quality variables. Transformation of skewed variables is beneficial because predictions from random forests are derived from a weighted mean of observations. Attribute data were therefore transformed before inclusion in the modelled dataset. Attribute states were log-transformed (\log_{10}), except for *E. coli* exceedances (proportion of samples exceeding 260 or 540 counts/100 mL), which were logit-transformed, and MCI, which was untransformed.

Data transformation led to the exclusion of some river attribute combinations from the training dataset where data transformation created numeric values that cannot be modelled. After the logit-transformation was applied, sites that did not exceed the *E. coli* concentration thresholds within the analysis period were excluded from the modelling process because numeric values of minus infinity cannot be modelled with random forests. Specifically, 7% and 12% of sites were excluded from models of the proportion of *E. coli* concentration data that exceeded 260 or 540 cfu or MPN 100 mL⁻¹ within the analysis period respectively.

The geographic distribution of river monitoring sites used for modelling is shown in Figure 2-1. The sites are reasonably well-distributed, although there are gaps in the central North Island and the west coast of the South Island, often related to mountainous terrain. There is a high degree of overlap among the sites used for physical, chemical, and microbiological water quality monitoring, as some or all the corresponding variables are measured at each site in council SoE programmes. However, measurements of clarity were not available from Otago, Wellington, and Auckland regions. There is less overlap among sites used for invertebrate monitoring; several councils operate separate programmes for monitoring physical-chemical water quality and invertebrates, with variable levels of site overlap between programmes.

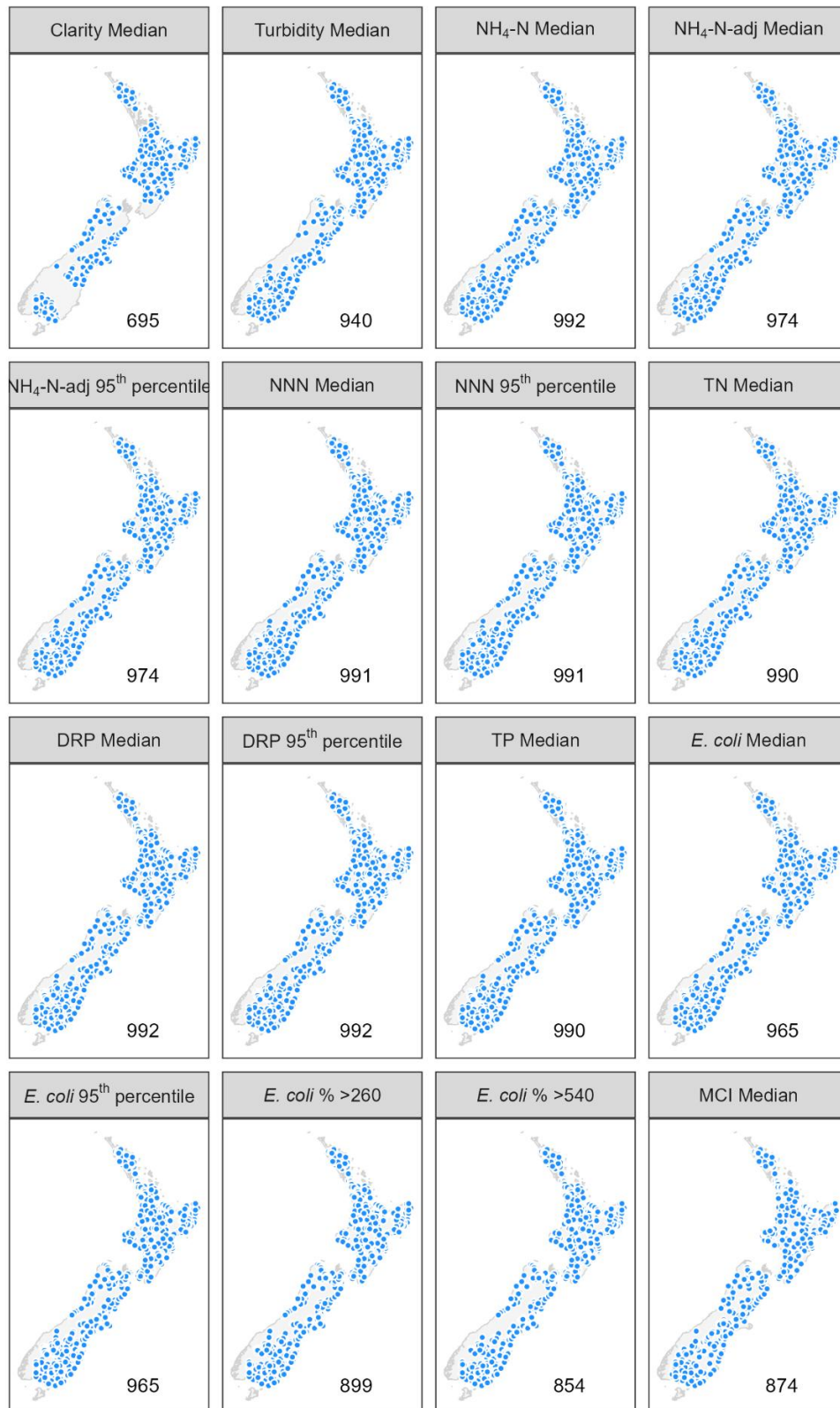


Figure 2-1: Locations of river monitoring sites used for modelling the current state of the 16 attributes. The number in the lower right of each panel corresponds to the number of sites included in each attribute state model after data transformation.

2.2 Predictor data

We used the digital river network version 2.4 (DN2.4) (Whitehead and Booker 2020) to provide the spatial framework for the random forest models of river attribute state. The river network and catchment boundaries were derived from a Digital Elevation Model (DEM) with a spatial resolution of 50 m. The digital network represents New Zealand's rivers as ~593,000 segments (bounded by upstream and downstream confluences) and their corresponding catchments. Each segment in the digital network ("nzsegment") has a unique nzsegment identifier. The links between each nzsegment and its catchment, between adjacent nzsegments and between adjacent catchments facilitate analyses of upstream-downstream connectivity and the accumulation of catchment characteristics in the downstream direction.

Spatial data layers describing the climate, topography, geology, vegetation, infrastructure, and hydrology of New Zealand were used to calculate predictor variables mapped onto DN2.4. This mapping process links spatial data layers to DN2.4, thus generating variables that describe the environmental characteristics of the segment and its catchment. The predictor variables and data sources used to calculate each predictor variable are described in Table 2-2. More information about the variables mapped on to DN2.4 can be found in Booker (2023) and Booker et al. (2024).

We selected 40 spatial variables (Table 2-2) for predictors in spatial models of the sixteen attribute states listed in Table 2-1. The predictors were selected based on their expected statistical relationships with water quality and experience from previous data-driven modelling studies, as well as data availability. Julian et al. (2017) noted the negative impact of intensive land use on water quality, though the relationship is complex. Snelder et al. (2022b) noted the impact of climate variability in determining water quality trends. Larned et al. (2016b) noted elevation serves as a proxy for land cover: low elevations are mostly agricultural, while high elevations are forested or alpine. Elevation and climate are also related, for example higher altitudes are associated with lower temperature. Timperley (1983) highlighted the influence of geology, noting the volcanic zone in the North Island having elevated phosphorus, whereas sedimentary geology is associated with lower phosphorus (McDowell et al. 2013). Consequently, land cover, land use intensity, climate, topography, and geology are all expected to be associated with river water quality. Previous experience with national-scale modelling of water quality using data-driven approaches (Unwin et al. 2010; Larned et al. 2016a; Whitehead 2018; Snelder et al. 2022a; Whitehead et al. 2022a), identified similar factors and informed our selection of predictors. Predictors were generated from the most recent available data, including the recently published Land Cover Database version 6.0 (LCDB6¹).

¹ [LCDB v6.0 - Land Cover Database version 6.0, Mainland, New Zealand | LRIS Portal](#)

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

Predictor class	Data sources used for calculation of predictor	Predictor Description	Abbreviation	Unit
River	DN2.4	Segment area	area	km ²
	DN2.4, Land Information New Zealand coastline layer (LINZ 2025)	Distance to coast	distCoast	km
	DN2.4	Segment mean elevation	elev	m ASL
Catchment topography	DN2.4, Land Environments of New Zealand Slope layer (Leathwick et al. 2003)	Mean Catchment slope	catSlope	Degrees
	DN2.4	Catchment area	catArea	km ²
	DN2.4	Catchment elevation	catElev	m ASL
Climate and flow	DN2.4, Virtual climate station network (NIWA ND) for the years 2000-01-01 to 2025-01-01	Mean segment summer (December) solar radiation	solarRadSum	W/m ²
		Mean segment winter (June) solar radiation	solarRadWin	W/m ²
		Mean segment summer (December) air temperature	airTempSum	degC
		Mean segment winter (June) air temperature	airTempWin	degC
		Mean segment summer (December) wind speed	windSpeedSum	m/s
		Mean segment winter (June) wind speed	windSpeedWin	m/s
		Mean catchment summer (December) air temperature	catAirTempSum	degC
		Mean catchment winter (June) air temperature	catAirTempWin	degC
		Mean annual catchment potential evapotranspiration	catPET	mm
		Mean catchment rain days >10 mm	catRainDays	Days/year
		Mean Catchment rainfall	catRain	mm/year
		Mean catchment coefficient of variation of annual rainfall	catRainVar	Ratio
	DN2.4 and predictor variables catArea, catPET, catRainDays,	Mean flow at catchment outlet	Qmean	m ³ /s

	catRainVar, catElev, catPsize following methods of Booker & Woods (2014)	Median flow at catchment outlet	Qmed	m ³ /s
Geology	DN2.4, Land Environments of New Zealand geology layers (Leathwick et al. 2002)	Mean catchment phosphorous content of regolith	catPhosphorus	Ordinal
		Mean catchment calcium content of regolith	catCalcium	Ordinal
		Mean catchment induration (hardness) of regolith	catHardness	Ordinal
		Mean catchment particle size of regolith	catPsize	Ordinal
	DN2.4, New Zealand Fundamental Soil Layer (LRIS 2024)	Proportion of catchment occupied by peat	catPeat	Proportion
		Proportion of catchment occupied by alluvial FSL types K, S, L, Z, Ts and Tl	catAlluvial	Proportion
Land cover	DN2.4, Landcover database version 6, as provided by MfE, which is based on analysis of satellite imagery from the 2020–2023 summer ²	Proportion of catchment that is bare (LCDB6 classes 12, 14, 15, 16)	usBare	Proportion
		Proportion of catchment in exotic forest (LCDB6 class 64 and 71)	usExoticForest	Proportion
		Proportion of catchment in indigenous forest (LCDB6 class 54 and 69)	usIndigenousForest	Proportion
		Proportion of catchment occupied LCDB6 classes 10, 68, 70	usMisc	Proportion
		Proportion of catchment occupied by combination of high producing exotic grassland short-rotation cropland, orchard, vineyard and other perennial crops (LCDB6 classes 40, 30, 33)	usPastoral	Proportion
		Proportion of catchment occupied in scrub (LCDB6 classes 50, 51, 52, 55, 56, 58)	usScrub	Proportion
		Proportion of catchment in low producing grassland (LCDB6 class 41, 43, 44)	usTussockGrassland	Proportion
		Proportion of catchment in built-up areas urban	usUrban	Proportion

² <https://iris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/>

		parkland, surface mines, dumps and transport infrastructure (LCDB6 classes 1, 2, 6, 5)		
		Proportion of catchment occupied by wetlands (LCDB6 classes 45, 46, 47)	usWetlands	Proportion
Land use intensity	DN2.4, Agricultural Production Census (APC) obtained from StatsNZ ³ intersected with Land use layers (Harris et al. 2023)	Catchment density of total stock units (SU)	TotalSUDensity	SU/m ²
		Proportion of total stock units attributable to beef cattle in catchment	usBeef	Proportion
		Proportion of total stock units attributable to dairy cows in catchment	usDairy	Proportion
		Proportion of total stock units attributable to deer in catchment	usDeer	Proportion
		Proportion of total stock units attributable to sheep in catchment	usSheep	Proportion

³ <https://www.stats.govt.nz/information-releases/agricultural-production-statistics-year-to-june-2024-final/>

3 Modelling methods

3.1 Random forest models

We modelled each attribute state as a function of predictors using random forest models (Breiman 2001). Most attribute states were log-transformed (\log_{10}), except for *E. coli* exceedances (proportion of samples exceeding 260 or 540 counts/100 mL), which were logit-transformed, and MCI, which used the untransformed data.

A random forest model is an ensemble of individual classification and regression trees (CART). In regression, CART partitions observations into groups that minimise response variance, based on binary splits derived from predictor variables. CART models require no distributional assumptions and can automatically fit non-linear relationships and high-order interactions. However, single regression trees are prone to instability and may not find globally optimal splits (Hastie et al. 2009). Random forests overcome these issues by averaging predictions from many trees grown on bootstrap samples of the data, with a random subset of predictors considered at each split. This randomisation and averaging improve prediction accuracy while retaining CART's flexibility.

Each random forest produces an asymptotic generalisation error, the prediction error for unseen data, which stabilises as the number of trees increases, reducing the likelihood of overfitting (Breiman 2001). We used default settings: 500 trees per forest, with one-third of the total predictors available at each split. This aligned with previous water quality studies (Snelder et al. 2022a; Whitehead et al. 2022a). Performance typically improves with tree number, most gain occurring in the first 100 trees, with computational limits being the main constraint (Probst and Boulesteix 2018).

Unlike linear models, random forest models cannot be expressed as a simple equation, so it can be challenging to see the relationship, if any, between the predictor and response variable. To assess predictor importance, random forest models permute response values for out-of-bag (OOB) observations and measure the resulting loss in prediction accuracy. Importance is quantified as the increase in mean squared error (MSE) between predictions based on permuted versus original OOB observations, averaged across all trees and normalised by the standard deviation of these differences (Cutler et al. 2007). OOB observations are observations not used to fit the model.

Partial dependence plots (PDPs) depict the marginal effect of a predictor on the response when other predictors are held constant (typically at their mean values). Although correlations or interactions among predictors can distort the interpretation, PDPs provide a useful approximation of modelled relationships (Cutler et al. 2007).

Random forest models can include any of the available predictors selected during fitting. Including marginally important or correlated predictors does not reduce predictive performance but may complicate interpretation. Random forest models produce an estimate of the importance of each predictor based on the OOB approach. Then 10-fold cross validation was used to for model selection. In brief, we applied a backward elimination process to remove the least important predictors from the initial saturated models. The mean squared error (MSE) and its standard error were estimated using 10-fold cross-validation. The least important predictors were removed iteratively, with MSE recalculated at each step. The final “reduced” model was

defined as the simplest model whose error was within one standard error of the minimum error (“one standard error rule”; Breiman et al. 2017). This approach yields a parsimonious model with equivalent predictive performance. Importance values were not recalculated at each step to avoid overfitting (Svetnik et al. 2004), though Speiser et al. (2019) noted that Svetnik’s approach is computationally intensive; this approach was retained to be consistent with previous reports.

All calculations were conducted in R (R Core Team 2022) using the randomForest package (Liaw and Wiener 2002), with supporting packages including tidyverse for data manipulation and graphing (Wickham et al. 2019), sf for spatial data handling (Pebesma and Bivand 2023), and pdp for partial dependence plots (Greenwell 2017).

3.2 Model performance

A key question about any model is how well it performs in predicting unmeasured observed values. A key advantage of random forests is that model error can be estimated directly from the data using out-of-bag (OOB) observations, eliminating the need for a separate test dataset as required by many other modelling approaches. For each tree, predictions are made for samples excluded from its bootstrap selection (the OOB data). Prediction accuracy for these OOB samples provides an unbiased estimate of model error and variable importance.

Model performance was assessed using predictions for out-of-bag (OOB) samples (i.e. data not used to train each tree), which provide an internal, approximately independent validation. We summarised the models using four statistics: regression R^2 , Nash-Sutcliffe Efficiencies (NSE), per cent bias (PBIAS) and root mean square deviation (RMSD).

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

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

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

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

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

The root mean square deviation (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). RMSD can be used to evaluate the prediction intervals of the expected value of the observation.

The 95% prediction intervals for values predicted by our models for individual segments can be obtained using the following approximations. Equation 1 can be used for calculating the intervals for the TLI predictions. Equation 2 can be used for calculating the intervals for the water quality variables for which the variables were \log_{10} transformed prior to model fitting, and the prediction uncertainty (RMSD) values have been reported in the \log_{10} transformed space. Equation 3 can be used for calculating the intervals for attribute states for which the variables were logit-transformed prior to model fitting (*E. coli* % >260, *E. coli* % >540) and the prediction uncertainties are reported in the logit-transformed space.

$$95\% PI = x \pm 1.96 \times RMSD \quad (1)$$

$$95\%_0 PI = 10^{[\log_{10}(x) \pm 1.96 \times RMSD]} \quad (2)$$

$$95\% PI = \frac{e^{[\text{logit}(x) \pm 1.96 \times RMSD]}}{(1 + e^{[\text{logit}(x) \pm 1.96 \times RMSD]})}, \text{ where } \text{logit}(x) = \log\left(\frac{x}{1-x}\right) \quad (3)$$

In all equations, x is the estimated value in the original units, RMSD is the model error in the transformed space, and 1.96 is the standard normal deviate or Z-score for probability ($0.025 \leq Z \leq 0.975$). The prediction intervals for the \log_{10} -transformed variables, when expressed in the original units, are asymmetric, and their values vary in proportion to the predicted water quality value. For example, if we let x be a predicted value for water clarity of 1 m, the lower and upper 95% confidence intervals are 0.37 and 2.7 m, respectively, assuming RMSD = 0.22.

3.3 Representativeness of monitoring sites used in random forest models

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

Density plots illustrating the distribution of the 12 most important predictors in the random forest models (i.e., the predictors with the greatest explanatory power) for sites where monitoring took place and sites (segments that comprise DN2.4) that were not monitored. Two sets of comparable density plots were derived. The first represented data from all sites that monitored at least one water quality attribute state, excluding MCI (997 unique sites). The second set of comparable histograms represented the 874 invertebrate monitoring sites that were used for modelling MCI scores. Separate density plots were constructed due to the limited overlap in physical-chemical water quality and invertebrate monitoring sites, as noted in Section 2.1. Note that the representativeness of monitoring sites is different from model bias, which is defined in Section 3.2. Model bias is a measure of systematic error in model predictions (i.e., over- or under-estimation).

3.4 Model predictions

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

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

where $\hat{\varepsilon}$ are the residuals of a random forest model. The predictions were back-transformed by raising them to the power of 10, then corrected for retransformation bias by multiplying by S . Predictions of *E. coli* exceedance values (%>260, %>540) were back-transformed using the inverse-logit function. The back-transformed and corrected predictions for all river segments in New Zealand were projected on a national map for each attribute state.

4 Results

4.1 Model performance

Assessment of OOB predictions from random forest models indicated that for eleven of the 16 attribute states were considered to be very good (NNN 95th percentile, TN Median, TP Median, MCI Median) or good (Clarity Median, NH₄-N-adj 95th percentile, NNN Median, DRP Median, DRP 95th percentile, *E. coli* Median, *E. coli* 95th percentile) as indicated by R^2 and NSE and the criteria of Moriasi *et al.* (2015) (Table 4-1, Figure 4-1). The remaining seven models (Turbidity Median, NH₄-N Median, NH₄-N-adj Median, *E. coli* >260, *E. coli* % >540) had satisfactory performance. All 16 models had very low bias (PBIAS; Table 4-1) as indicated by the close match between the line representing the regression of the observed versus OOB predicted values (blue line in Figure 4-1) and the one-to-one line (red dashed line in Figure 4-1). This indicates that predicted values generally give a good estimate of the average observed values.

Table 4-1: Performance of the 16 attribute state models. Performance was determined using independent predictions (i.e., sites that were not used in fitting the models) generated from the out-of-bag (OOB) observations. R^2 = coefficient of determination, NSE = Nash-Sutcliffe efficiency, PBIAS = percent bias, RMSD = root mean square deviation. Units for RMSD are the log₁₀ or logit transformed units of the attribute state except for MCI, which were not transformed. The colours indicate the performance ratings indicated in Table 3-1.

Attribute	N	R^2	NSE	PBIAS	RMSD	Rating
Clarity Median	695	0.66	0.65	2.35	0.23	Good
Turbidity Median	940	0.60	0.59	-0.11	0.32	Satisfactory
NH ₄ -N Median	992	0.56	0.55	0.31	0.29	Satisfactory
NH ₄ -N-adj Median	974	0.59	0.58	0.16	0.28	Satisfactory
NH ₄ -N-adj 95 th percentile	974	0.64	0.63	0.48	0.32	Good
NNN Median	991	0.63	0.62	-0.47	0.47	Good
NNN 95 th percentile	991	0.76	0.76	1.08	0.29	Very good
TN Median	990	0.77	0.76	1.46	0.25	Very good
DRP Median	992	0.62	0.62	0.12	0.30	Good
DRP 95 th percentile	992	0.64	0.64	0.49	0.30	Good
TP Median	990	0.71	0.70	0.16	0.25	Very good
<i>E. coli</i> Median	965	0.68	0.67	-0.01	0.35	Good
<i>E. coli</i> 95 th percentile	965	0.64	0.64	0.05	0.40	Good
<i>E. coli</i> % >260	899	0.59	0.59	1.33	1.05	Satisfactory
<i>E. coli</i> % >540	854	0.54	0.54	1.05	0.92	Satisfactory
MCI Median	874	0.72	0.72	0.10	9.57	Very good

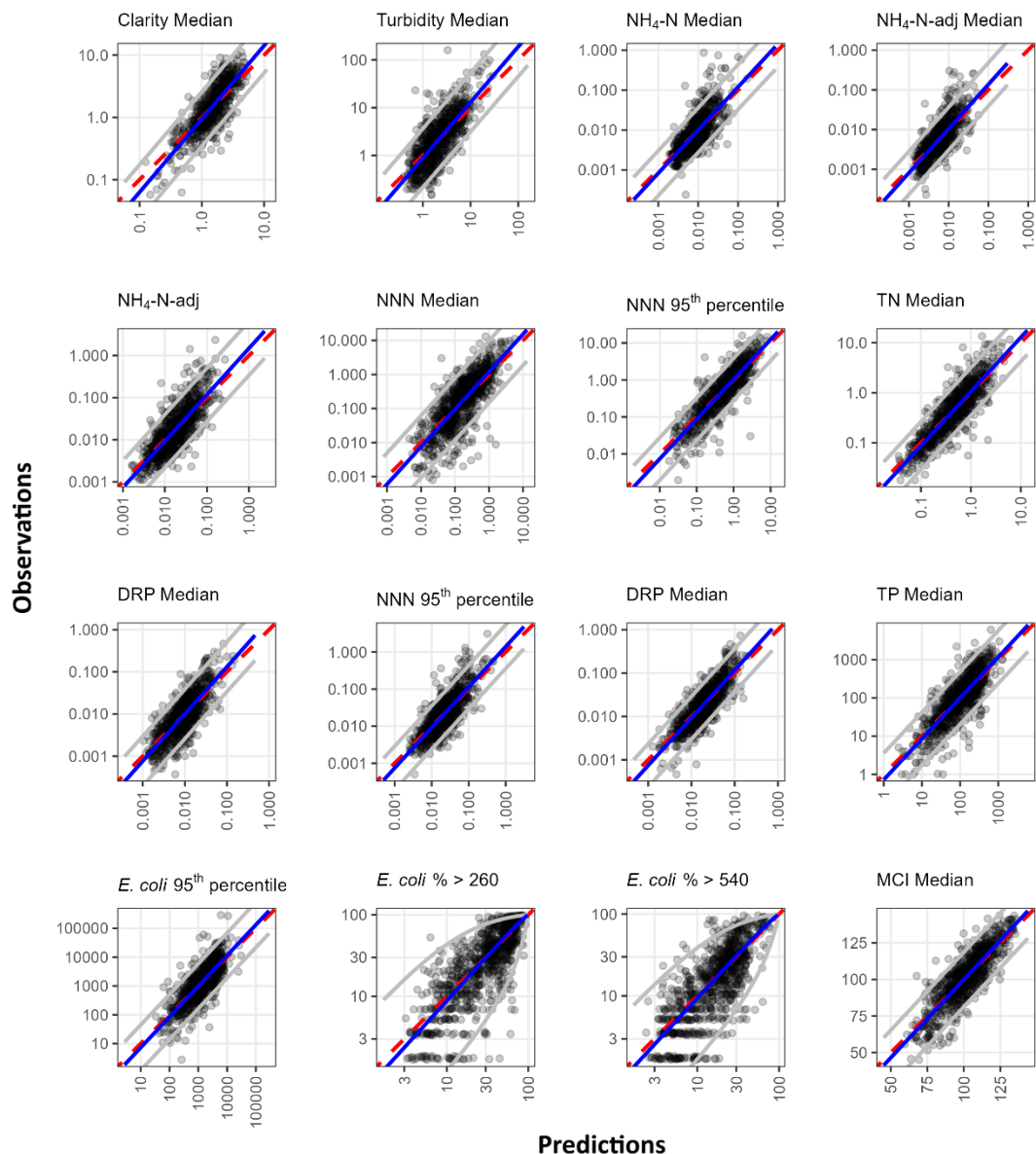


Figure 4-1: Comparison of observed attribute state versus values predicted by each of the random forest models. Note that the observed values are plotted on the Y-axis and predicted values on the X-axis, following Piñeiro et al. (2008). Blue solid line: best fit linear regression of the observed and predicted values. Red dashed line: one-to-one line. The grey lines represent the upper and lower prediction intervals, the regions where the actual value is expected to lie for a given prediction value, note all the axis are plotted on a log scale, the numbers are in the appropriate units for each variable.

Figure 4-1 displays the prediction intervals, the range where a future observation is expected to lie for any predicted value, as illustrated in the upper and lower grey lines (prediction interval). In some cases, the prediction intervals cover a small range relative to the observed values, such as the case of TN Median. In some cases, the prediction interval is very wide, such as for *E. coli* % > 260, covering the range of observed values and explaining the predictive performance rating of satisfactory (Table 4-1). The prediction intervals show that the models cannot explain some of the observed variation.

4.2 Modelled relationships

In all the random forest models, the predictors with high importance reflected strong associations between water quality and land cover, geology, catchment topography, climate and flow as well as land use intensity. These associations are broadly similar to the finding in previous studies of modelling water quality state (Snelder et al. 2022b; Whitehead et al. 2022a).

The proportion of catchment occupied by pastoral agriculture (usPastoral), which is in the land cover prediction class, was the most important predictor in terms of average rank (Table 4-2). It was ranked top in nine out of 16 attributes and in the top eight of all others, with the exception of clarity (ranked 26th). Two other land cover predictors appeared in the top ten. One was the proportion of catchment covered in indigenous forest (usIndigenousForest), which was ranked 7th, and the other the proportion of catchment in built-up areas (usUrban), which was ranked 9th. The partial plots (Figure 4-2 and Figure 4-3) indicate that higher proportions of catchment occupied by pastoral agriculture, built-up areas, and higher land use intensity (as indicated by stock unit density – TotalSUDensity, ranked 6th) are associated with worse water quality, that is higher levels of all water quality attributes, except clarity and MCI where worse quality is indicated by lower levels. Built-up areas were not an important factor for predicting water clarity, DRP median, TP median or turbidity. Higher proportions of catchments occupied by indigenous forest tended to be associated with better water quality.

The second most important predictor (Table 4-2) was the proportion of the catchment occupied by alluvial FSL (catAlluvial), though only seven of 16 random forests included this geology class variable as a predictor. This predictor was particularly important for clarity, turbidity, and NH₄-N-adj. Based on the result from the partial plots (Figure 4-2 and Figure 4-3), catchments occupied with a low proportion of alluvial material tended to have low clarity and high turbidity. Catchments with a high proportion of alluvial material tended to have low levels of NH₄-N-adj.

Catchment predictors describing elevation (catElev) and mean catchment slope (catSlope) ranked 3rd and 4th overall (Table 4-2). The importance values showed that catElev and catSlope were important predictors for all the random forest models, with the exception of catElev for clarity median and DRP median. The partial plots show that MCI increases with slope and elevation with maximum values of 20° slope and 250 m elevation. There is a general tendency for water quality variables (except for clarity and DRP median, which are not included) to decrease, showing improved quality with increasing catchment elevation up to 1000 m. Similar tendencies are noted for slope, though TP shows maximum values around 2–5°.

Several climate variables are important predictors including mean annual catchment rainfall (catRain), mean catchment winter (June) air temperature (catAirTempWin), mean segment winter (June) air temperature (airTempWin), mean average number of days with over 10 mm rain per year (catRainDays), and mean segment winter (June) solar radiation (solarRadWin), which were ranked 5th, 8th, 10th, 11th and 12th in terms of importance. Mean annual rainfall between 500 and 2000 mm/year showed the greatest variation in water quality attributes (Figure 4-2 and Figure 4-3). Higher rainfall tended to be associated with better water quality, lower values of nutrient attributes and higher values of clarity and MCI, though there was not a simple monotonic relationship for many of the attributes. For some attributes, such as DRP median and turbidity median, the water quality was positively related to rainfall at very low rainfall, and then exhibited a negative relationship with rainfall (better water quality with higher rainfall) above 1000 mm/year for mean annual rainfall. Higher winter air temperatures tend to be

associated with poorer water quality. Though winter solar radiation was identified as an important predictor, water quality and MCI were only very weakly associated with it.

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

Predictor	Clarity Median	Turbidity Median	NH ₄ -N Median	NH ₄ -N-adj Median	NH ₄ -N-adj 95 th percentile	NNN Median	NNN 95 th percentile	TN Median	DRP Median	DRP 95 th percentile	TP Median	E. coli Median	E. coli 95 th percentile	E. coli % >260	E. coli % >540	MCI Median
usPastoral	26	8	3	2	1	1	1	1	8	1	2	1	1	1	1	4
catAlluvial	1	3	6	1	2	28	-	-	-	-	10	-	-	-	-	-
catElev	-	5	2	5	5	8	8	3	-	18	19	2	2	2	2	3
catSlope	7	20	1	4	6	3	5	2	2	2	1	5	4	6	9	8
catRain	4	4	21	23	20	19	4	13	3	7	4	-	8	-	14	7
TotalSUDensity	8	6	10	9	7	11	6	11	-	17	15	4	5	4	3	20
usIndigenousForest	17	11	15	3	3	4	7	6	-	8	14	10	6	9	8	1
catAirTempWin	14	15	5	6	4	10	10	-	4	3	12	13	3	8	7	22
usUrban	-	-	11	11	8	6	11	7	-	19	-	6	9	5	5	16
airTempWin	3	2	9	8	10	7	13	10	9	6	16	-	-	-	18	15
catRainDays	13	9	26	25	-	12	9	15	6	4	6	-	16	-	-	5
solarRadWin	10	10	8	7	9	15	18	12	-	14	9	12	13	17	-	6
distCoast	28	18	-	-	-	-	-	-	-	-	-	11	7	10	6	12
usSheep	20	-	-	-	-	9	12	8	-	-	-	18	-	12	-	-
usBare	-	13	-	-	-	23	28	9	12	20	3	3	-	3	4	28
elev	-	-	20	13	26	22	14	-	-	-	-	7	12	11	10	2
catArea	11	14	-	-	-	-	-	-	-	-	-	-	-	-	-	13
usDairy	27	-	18	-	16	2	2	4	-	9	-	20	-	16	12	-
usExoticForest	19	-	24	16	13	5	3	5	-	-	-	14	-	-	-	24
catPET	16	26	19	14	12	20	19	14	-	12	13	-	18	15	11	11
solarRadSum	5	7	13	20	18	21	-	-	13	16	11	-	-	-	-	18
catCalcium	-	-	14	10	14	-	17	-	14	15	-	-	19	20	-	-
catRainVar	2	1	7	26	19	26	24	16	5	5	8	17	10	14	19	21
usDeer	9	12	-	-	24	-	-	-	-	-	-	15	15	7	16	-
catAirTempSum	-	25	12	17	22	17	22	-	7	10	5	8	11	19	15	17
catPhosphorus	18	16	16	22	21	14	15	20	-	13	17	-	-	13	-	9
windSpeedWin	6	17	-	12	23	-	25	-	-	-	-	-	-	-	17	27
airTempSum	21	19	22	19	17	18	16	17	11	-	18	-	14	-	-	14
catPsize	-	-	25	-	27	16	20	-	1	11	7	-	-	-	-	26
usScrub	-	-	28	-	-	13	23	18	-	-	-	9	-	-	-	19
catHardness	15	21	4	18	11	24	21	19	-	-	-	-	-	-	-	23
usMisc	-	-	-	-	-	25	26	-	-	-	-	19	20	18	13	10

Predictor	Clarity Median	Turbidity Median	NH ₄ -N Median	NH ₄ -N-adj Median	NH ₄ -N-adj 95 th percentile	NNN Median	NNN 95 th percentile	TN Median	DRP Median	DRP 95 th percentile	TP Median	E. coli Median	E. coli 95 th percentile	E. coli % >260	E. coli % >540	MCI Median
usWetlands	24	24	23	15	15	-	-	-	-	-	20	-	-	-	-	-
Qmed	22	27	-	-	-	-	-	-	-	-	-	-	17	-	-	-
usTussockGrassland	-	22	17	24	25	27	27	-	10	-	-	16	-	-	20	25
Qmean	23	23	-	-	-	-	-	-	-	-	-	-	-	-	-	-
windSpeedSum	25	-	-	21	-	-	-	-	-	-	-	-	-	-	-	-
catPeat	12	28	27	27	-	-	-	-	-	-	-	-	-	-	-	-
usBeef	-	-	-	28	28	-	-	-	-	-	-	-	-	-	-	-

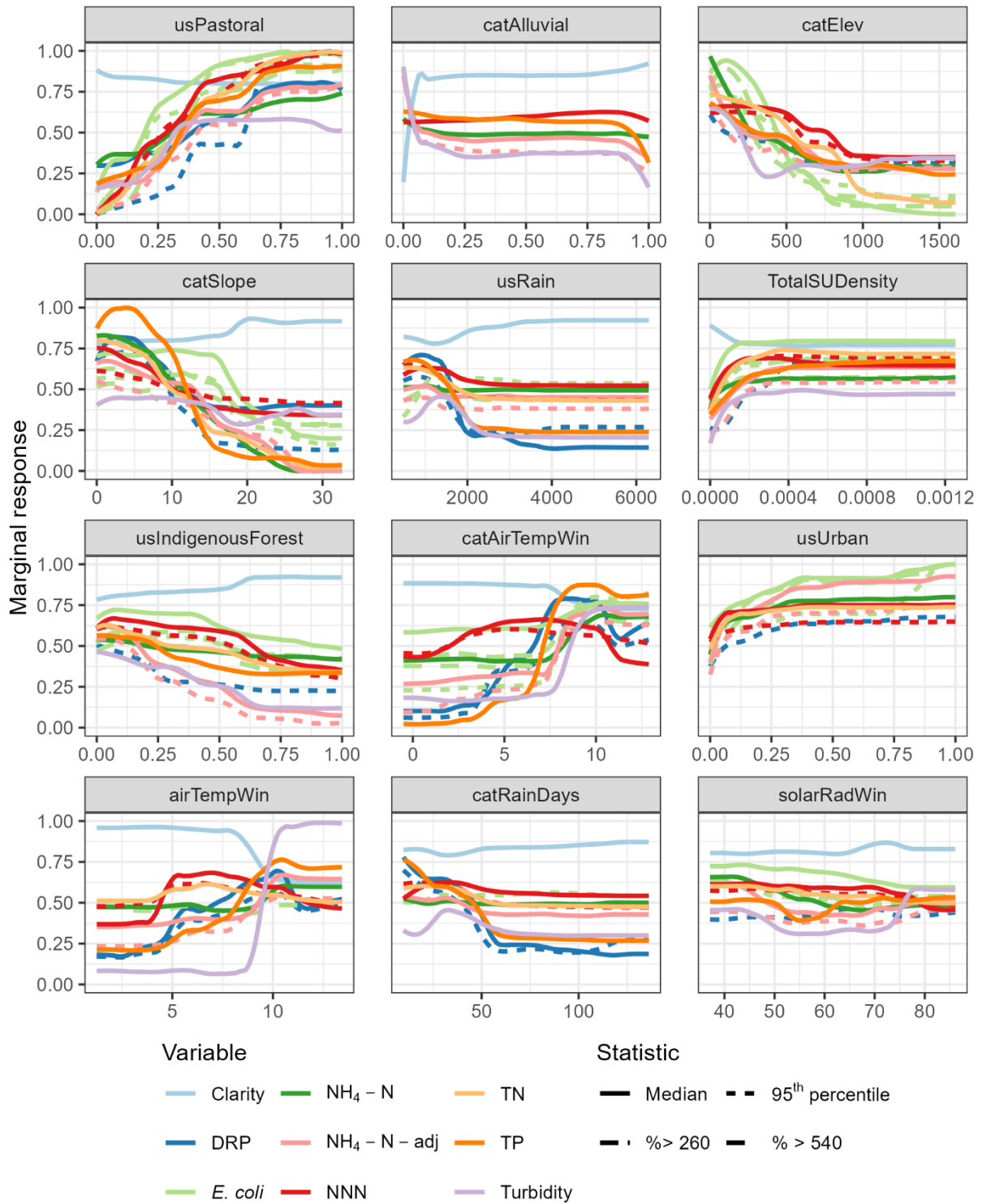


Figure 4-2: Partial plots for the 12 most important predictors in random forest models of current attribute state. Colours represent water quality variables, with the statistic indicated by line type (i.e., the combination of colour and line type represents an attribute). Each panel corresponds to one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

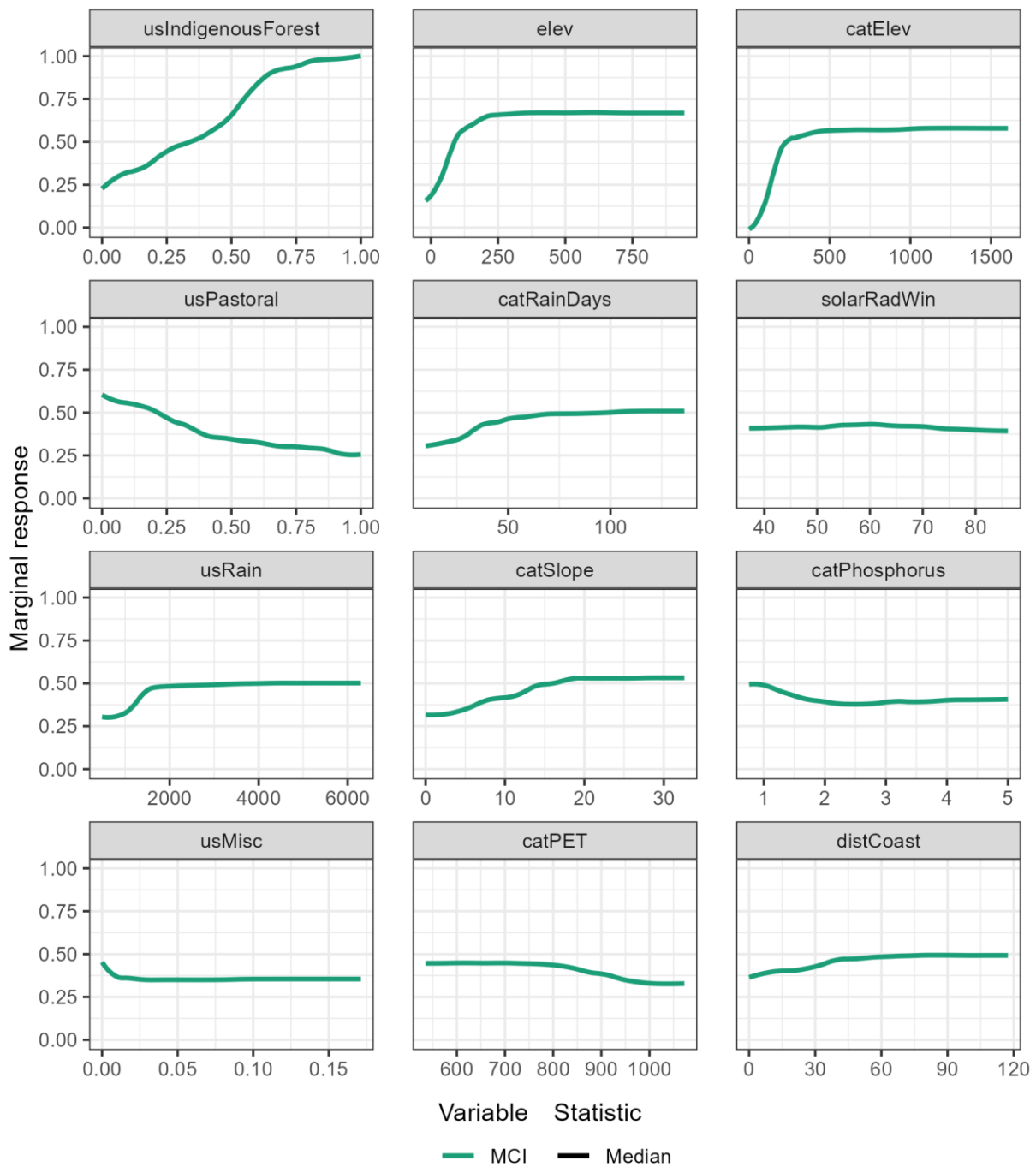


Figure 4-3: Partial plots for the 12 most important predictors for the median MCI random forest model. Each panel corresponds to one predictor, with predictors ordered from most (top left) to least (bottom right) important. The Y-axis scale represents the absolute value of the marginal response (i.e., the scale represents the marginal change in MCI values). The amplitude of each plot (i.e., the range of the marginal response shown on the y-axis) is directly related to a predictor's importance. Units on X-axes are in Table 2-2. The attribute for MCI only includes the median statistic.

4.3 Monitoring site representativeness

The distributions of the top 12 ranked predictors were similar for river segments with water quality and MCI monitoring sites compared to non-monitored segments (Figure 4-4 and Figure 4-5). The predictors shown in the density plots were those subsequently found to have the highest average importance for the models that included these predictors (Table 4-2).

There were several cases of moderate over- and under-representation of monitoring sites compared to the river network (regardless of river size or position of segments within catchments). Water-quality sites were over-represented in environments characterised by low catchment elevations (catElev) and low catchment slopes (catSlope), warmer winter temperatures (airTempWin, catAirTempWin), and low rainfall (usRain, catRainDays), suggesting that the monitoring network does not capture information from mountainous areas. Water quality sites were under-represented in areas with very low and high levels of pastoral land use (usPastoral), indigenous forest (usIndigenousForest) and alluvial deposits (catAlluvial).

Invertebrate monitoring sites were over-represented in low elevation and low slope rivers (elev, catElev, catSlope) close to the coast (distCoast), fewer very sites with many rainy days (catRainDays), had higher levels of winter solar radiation (solarRadWin), potential evapotranspiration (catPET) and were under-represented in areas with very low and high levels of pastoral land use (usPastoral) and indigenous forest (usIndigenousForest) .

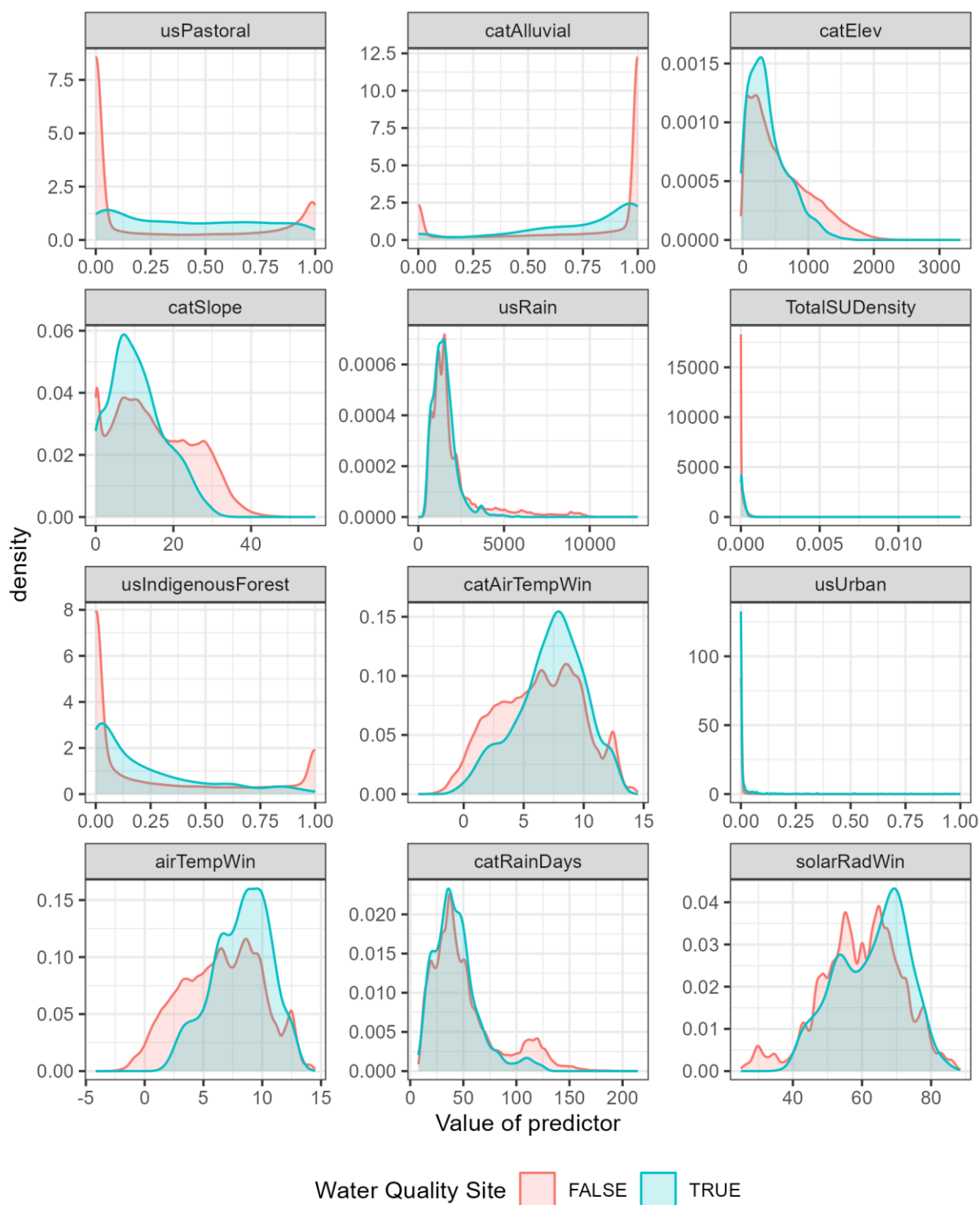


Figure 4-4: The distributions of predictors across all segments in the digital river network and at water quality sites (grey and blue histograms, respectively). Similarities in the distributions shown in the two histograms in each panel provide an indication of the degree to which environmental variation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

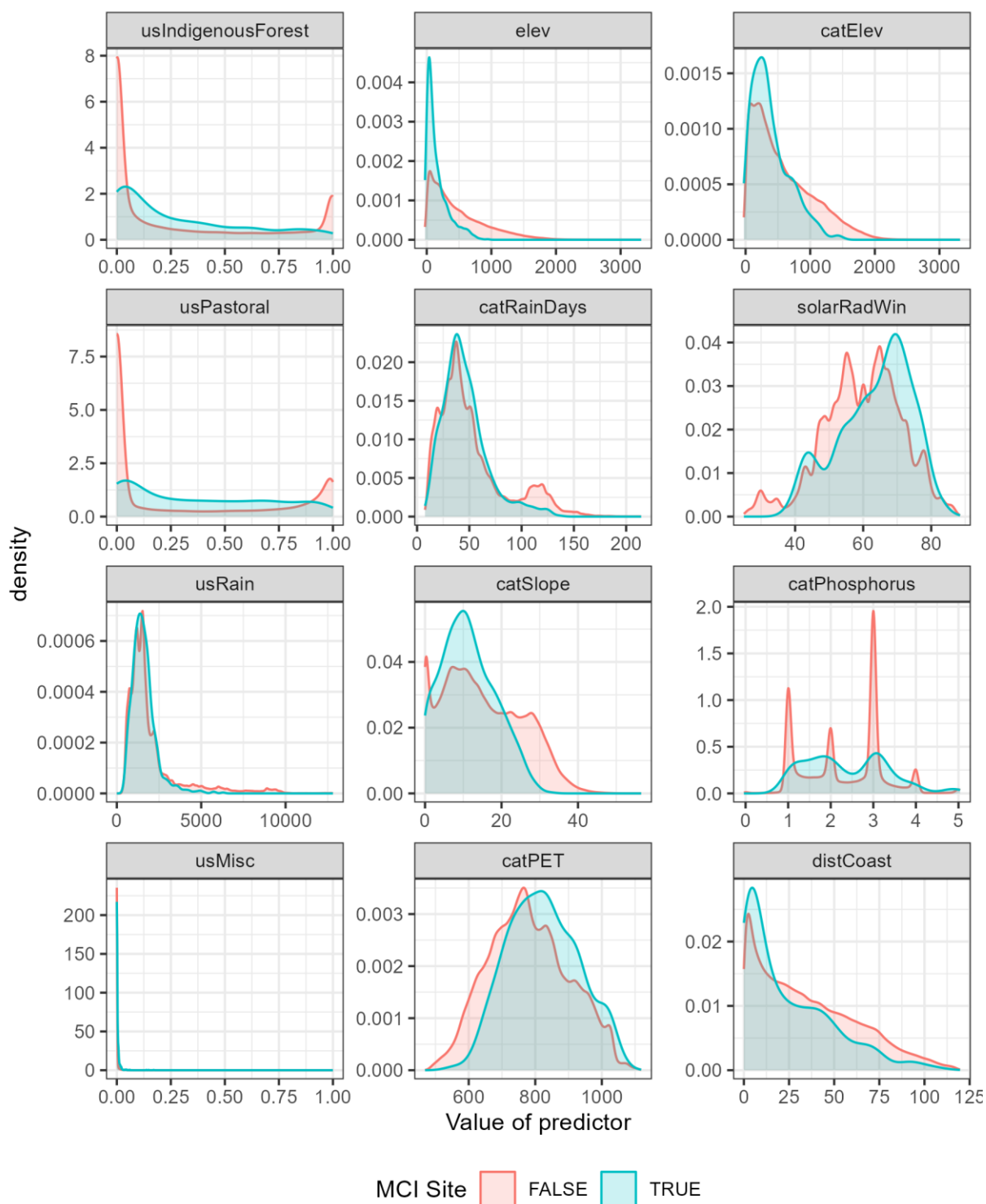


Figure 4-5: Distributions of predictors across all segments in the digital river network and at invertebrate sites (grey and blue histograms, respectively). Similarities in the distributions shown in the two histograms in each panel provide an indication of the degree to which environmental variation across the monitoring sites represents environmental variation across the New Zealand river network; complete representativeness would be indicated by exact matches between the histograms. These 12 predictors were the most important in the random forest model of MCI and are ordered from most (top left) to least (bottom right) important (Table 4-2).

4.4 Model predictions

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

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

Attribute state (units)	units	Observed Values		Predicted Values	
		Minimum	Maximum	Minimum	Maximum
Clarity Median	m	0.06	11.1	0.09	7.41
Turbidity Median	mg L ⁻¹	0.15	160	0.29	55.41
NH ₄ -N Median	mg L ⁻¹	0	0.86	0	0.3
NH ₄ -N-adj Median	mg L ⁻¹	0	0.3	0	0.12
NH ₄ -N-adj 95 th percentile	mg L ⁻¹	0	2.35	0	0.62
NNN Median	mg L ⁻¹	0	13.1	0	9.35
NNN 95 th percentile	mg L ⁻¹	0	15.9	0.01	11.69
TN Median	mg L ⁻¹	0.02	13.5	0.03	8.78
DRP Median	mg L ⁻¹	0	0.47	0	0.21
DRP 95 th percentile	mg L ⁻¹	0	3.12	0	0.99
TP Median	mg L ⁻¹	0	0.74	0	0.37
<i>E. coli</i> Median	cfu or MPN 100 ml ⁻¹	1	6050	1.64	2093
<i>E. coli</i> 95 th percentile	cfu or MPN 100 ml ⁻¹	2.76	287000	19.33	73743.23
<i>E. coli</i> % >260	Unitless	1.67	98.31	2.27	95.95
<i>E. coli</i> % >540	Unitless	1.67	96.43	2.06	87.98
MCI Median	Unitless	45.13	143.8	54.13	139.89

The mapped predictions for attribute states describing nutrients (NH₄-N, NH₄-N-adj, NNN, TN, DRP, TP), *E. coli*, and turbidity have similar coarse-scale spatial patterns, with relatively higher values in lower elevation areas (Figure 4-6 to Figure 4-21). This is the inverse of clarity and MCI, which tend to have high values at higher elevations.

The areas on the east coasts of the North and South Islands, as well as inland regions like Waikato, Wairarapa Valley, Rangitikei-Manawatu coastal plain, Taranaki Ring Plain, Auckland Region of the North Island, and the Southland Plain in the South Island, tend to have higher levels of nutrients, *E. coli*, and turbidity. These areas also generally exhibit lower water clarity and MCI. In contrast, predicted nutrient and *E. coli* attribute states are generally low in major

mountain ranges (e.g., Southern Alps, Kahurangi, Kaimanawa, and Tararua Ranges), in large areas of the Department of Conservation estate (e.g., Fiordland, Westland, Te Urewera, Egmont, Whanganui and Tongariro National Parks), and in smaller, native forest-dominated areas of Northland and the Coromandel Peninsula.

The low elevation areas characterised by high nutrient and *E. coli* attributes coincide with land used for intensive agriculture and with most of New Zealand's urban centres. High-intensity agricultural and urban land currently accounts for 60% of the land area below 350 m elevation (Larned et al. 2016b). Within these areas, there are some finer-scaled differences in predicted attribute state. The Canterbury Plains are characterised by high TN and NNN concentrations, and intermediate TP and DRP concentrations (though some areas in south Canterbury have higher TP and DRP), and the Waikato-Hauraki Plains area is characterised by high concentrations of both nitrogen and phosphorus.

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

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

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

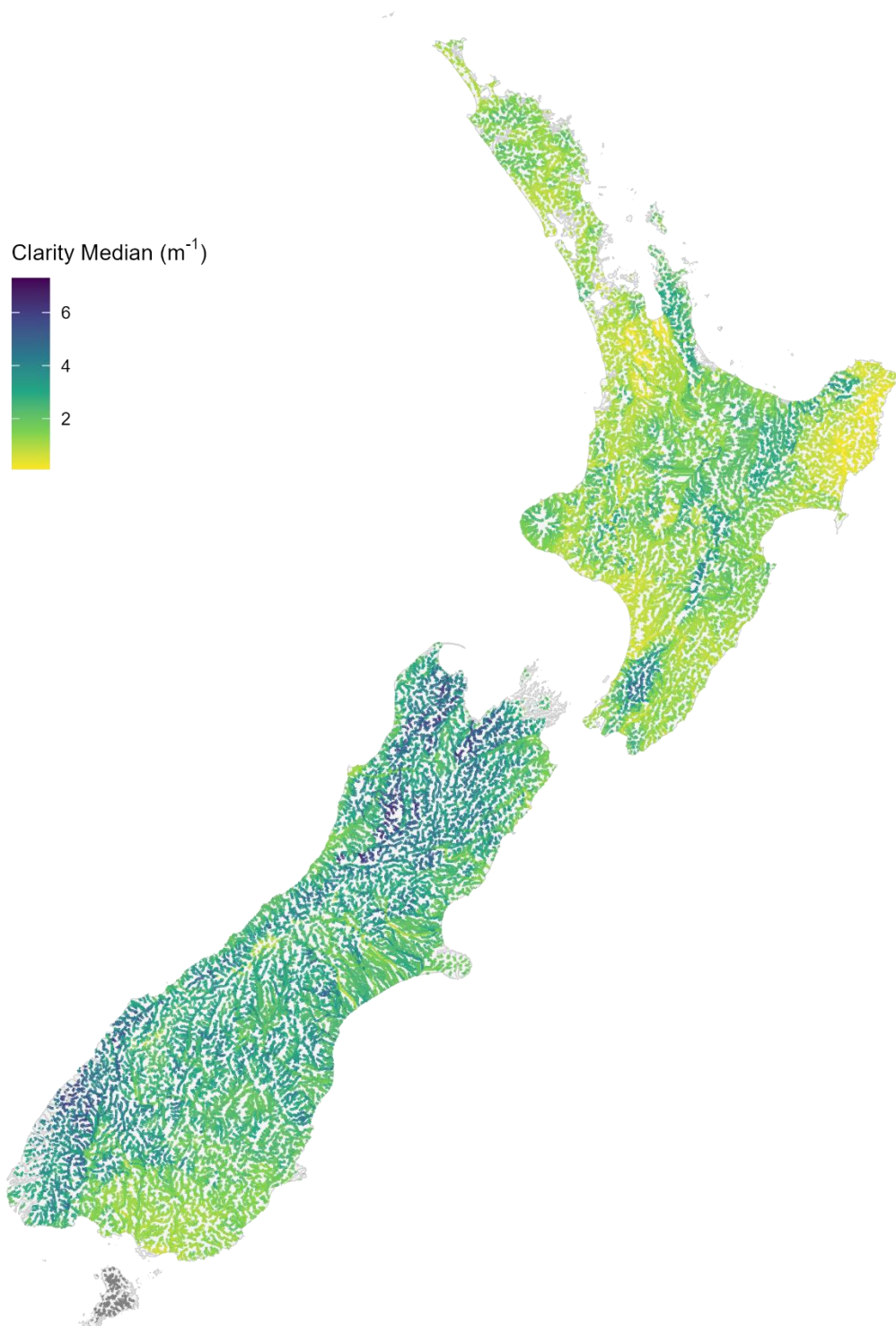


Figure 4-6: Predicted median clarity in New Zealand rivers. Map shows all nzsegments of Order 3 and higher.

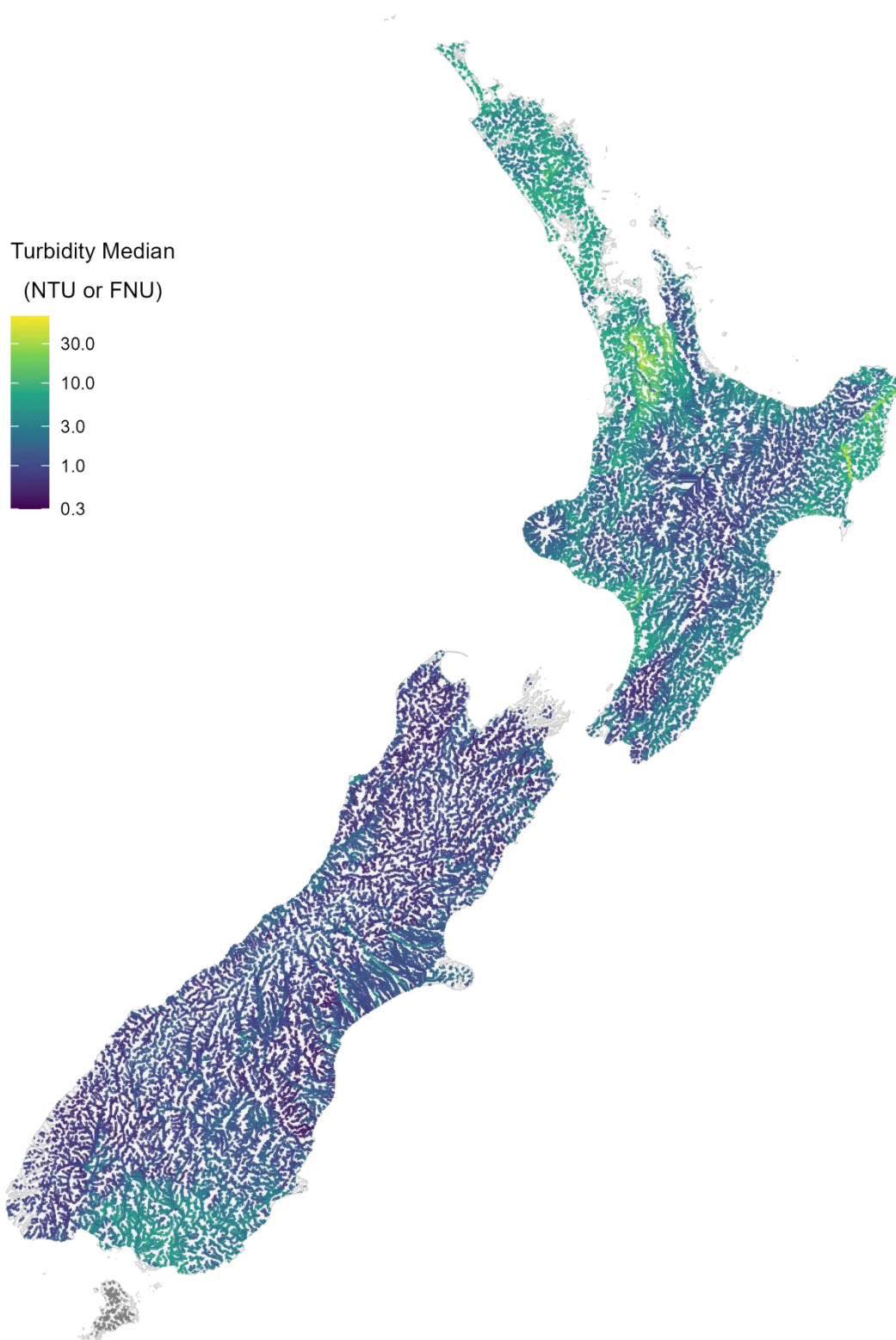


Figure 4-7: Predicted median turbidity in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a \log_{10} colour scale.

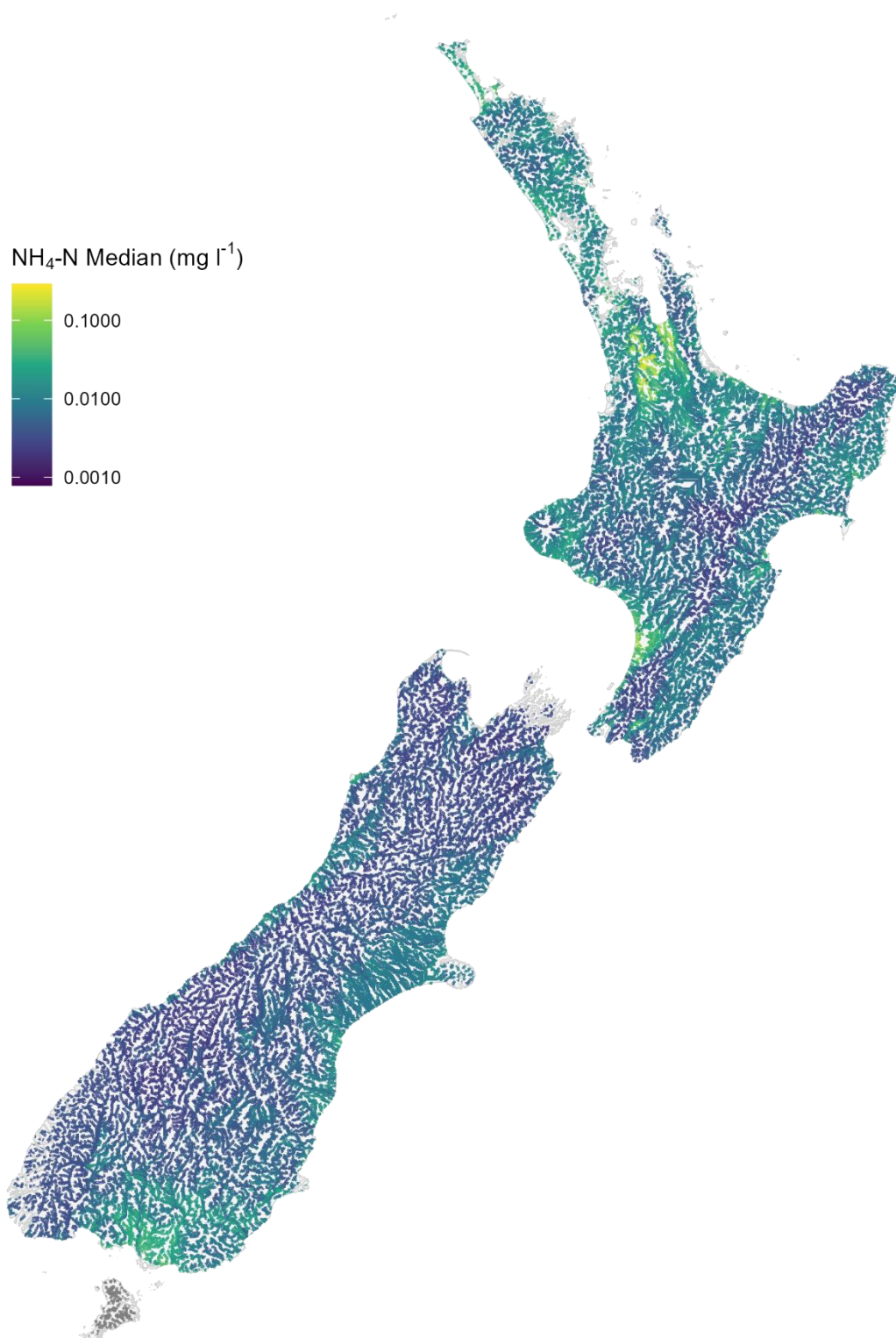


Figure 4-8: Predicted median ammoniacal nitrogen $\text{NH}_4\text{-N}$ in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a \log_{10} colour scale.

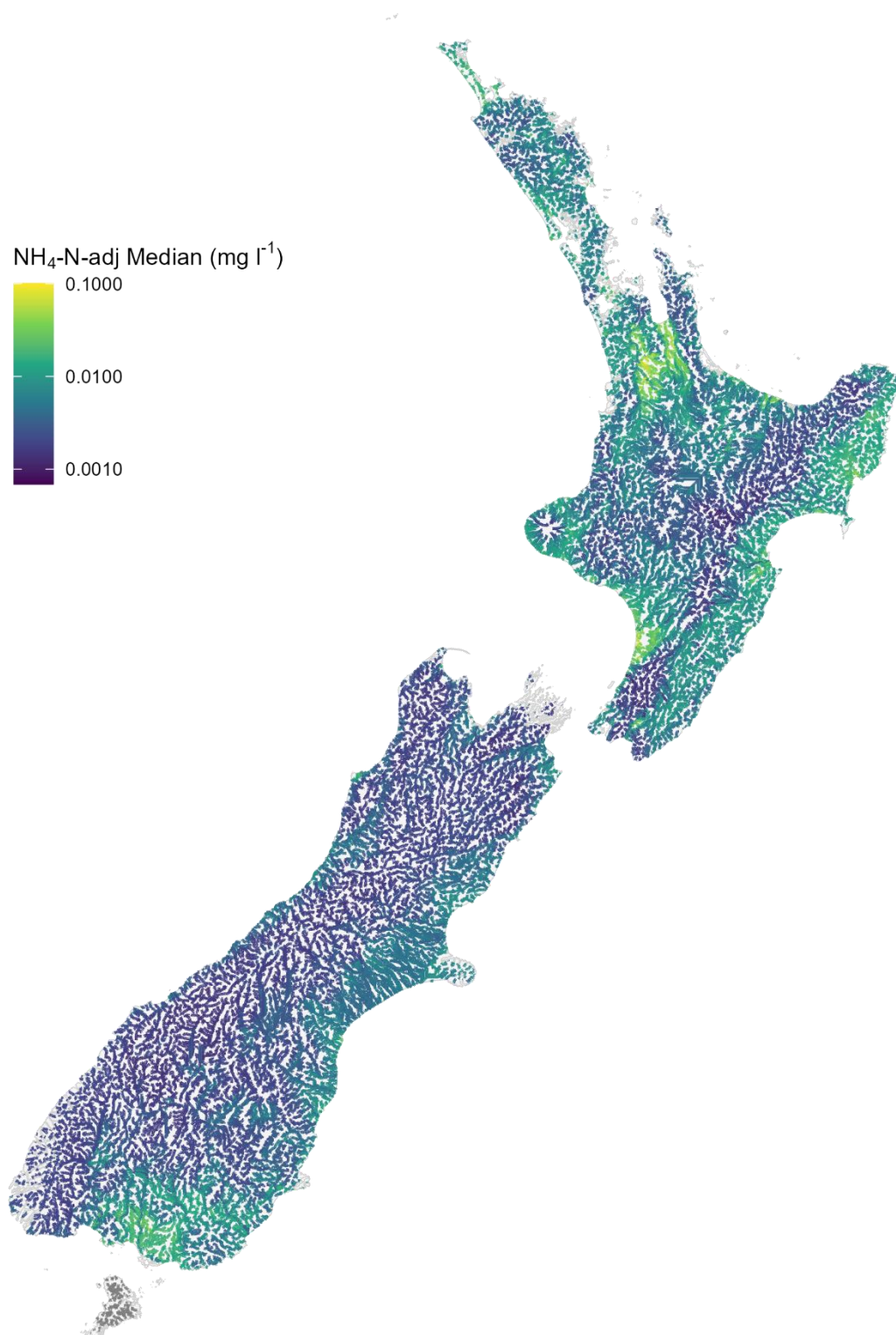


Figure 4-9: Predicted median ammoniacal nitrogen $\text{NH}_4\text{-N-adj}$ adjusted for pH in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a \log_{10} colour scale.

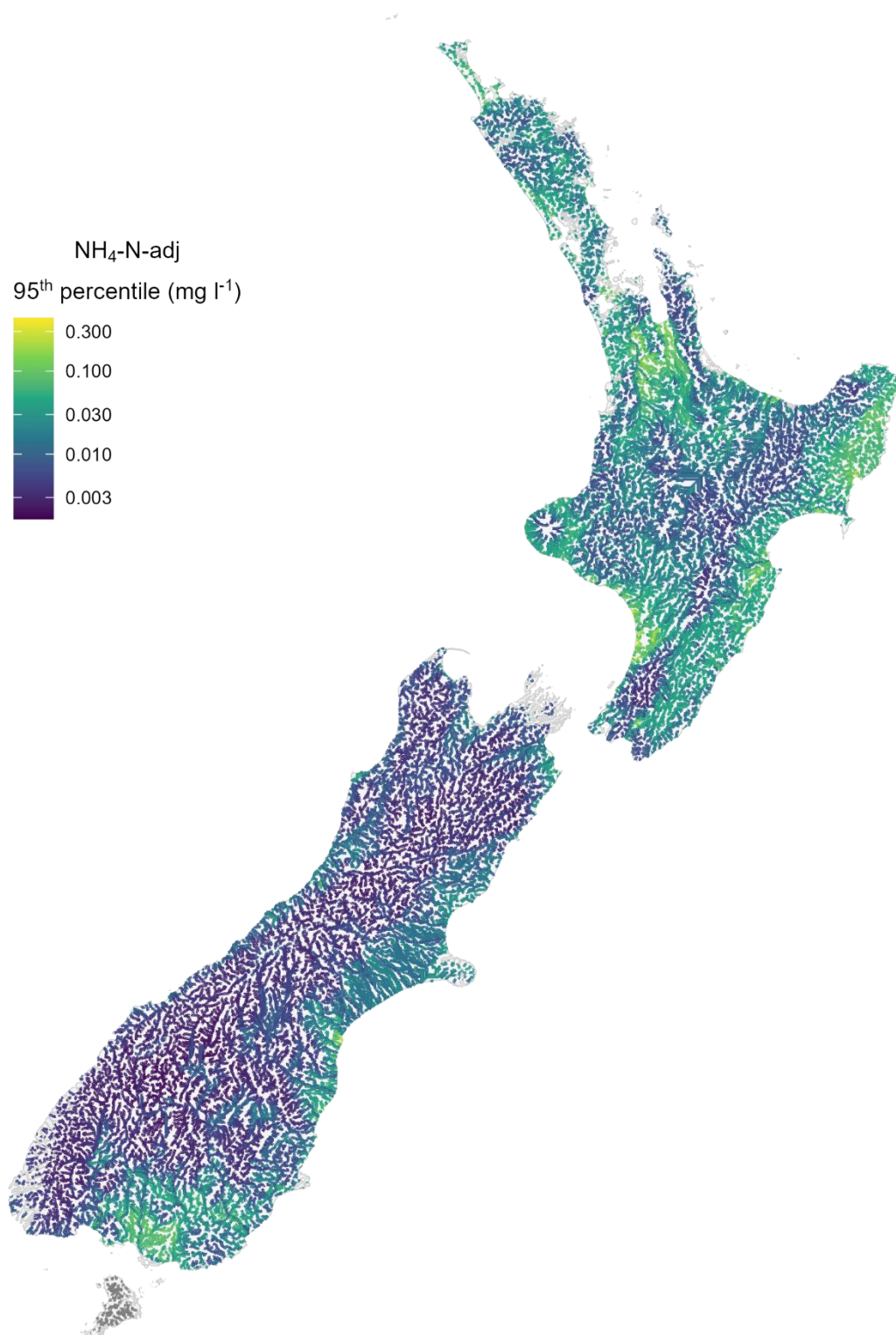


Figure 4-10: Predicted 95th percentile ammoniacal nitrogen $\text{NH}_4\text{-N-adj}$ adjusted for pH in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a \log_{10} colour scale.

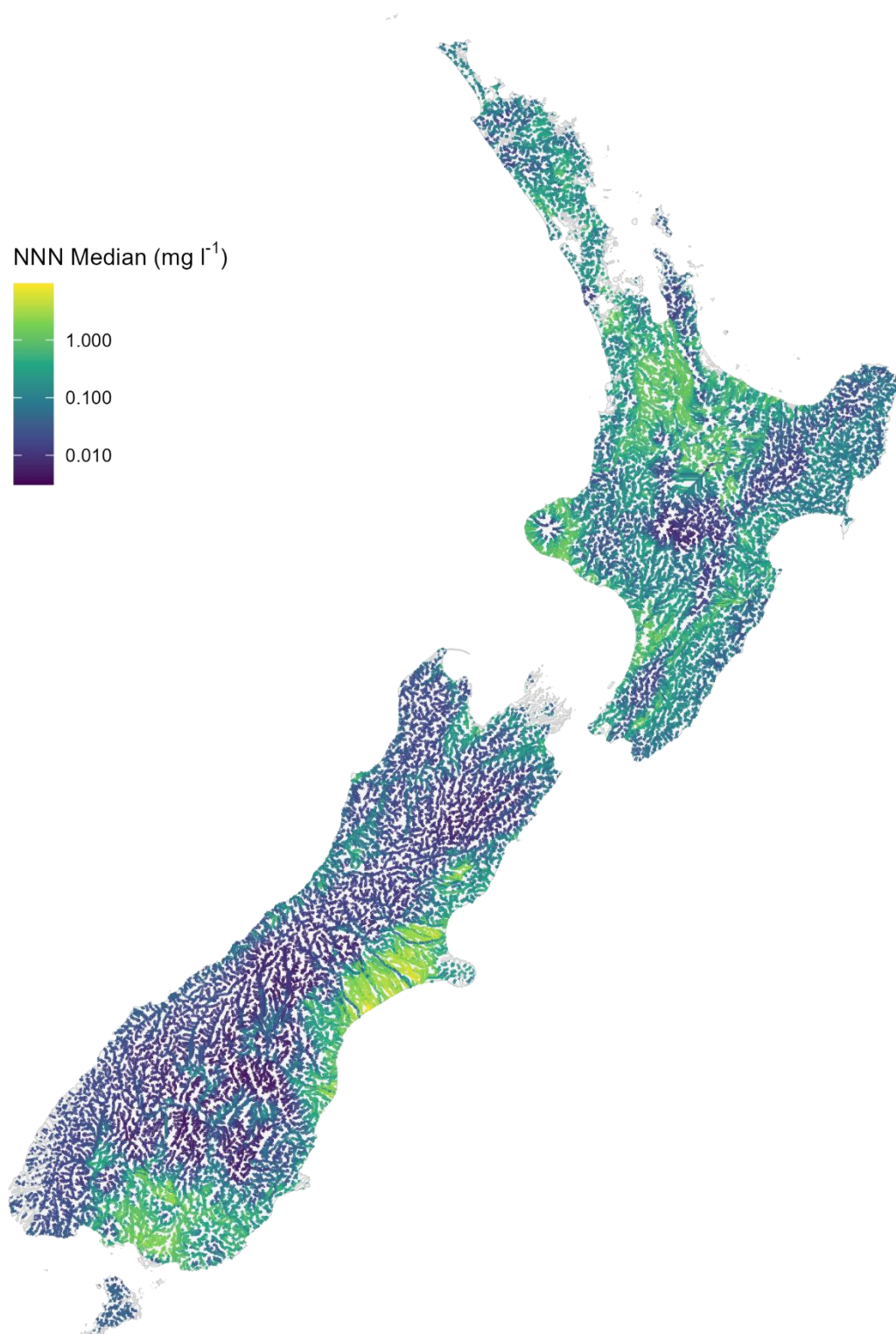


Figure 4-11: Predicted median NNN concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a \log_{10} colour scale.

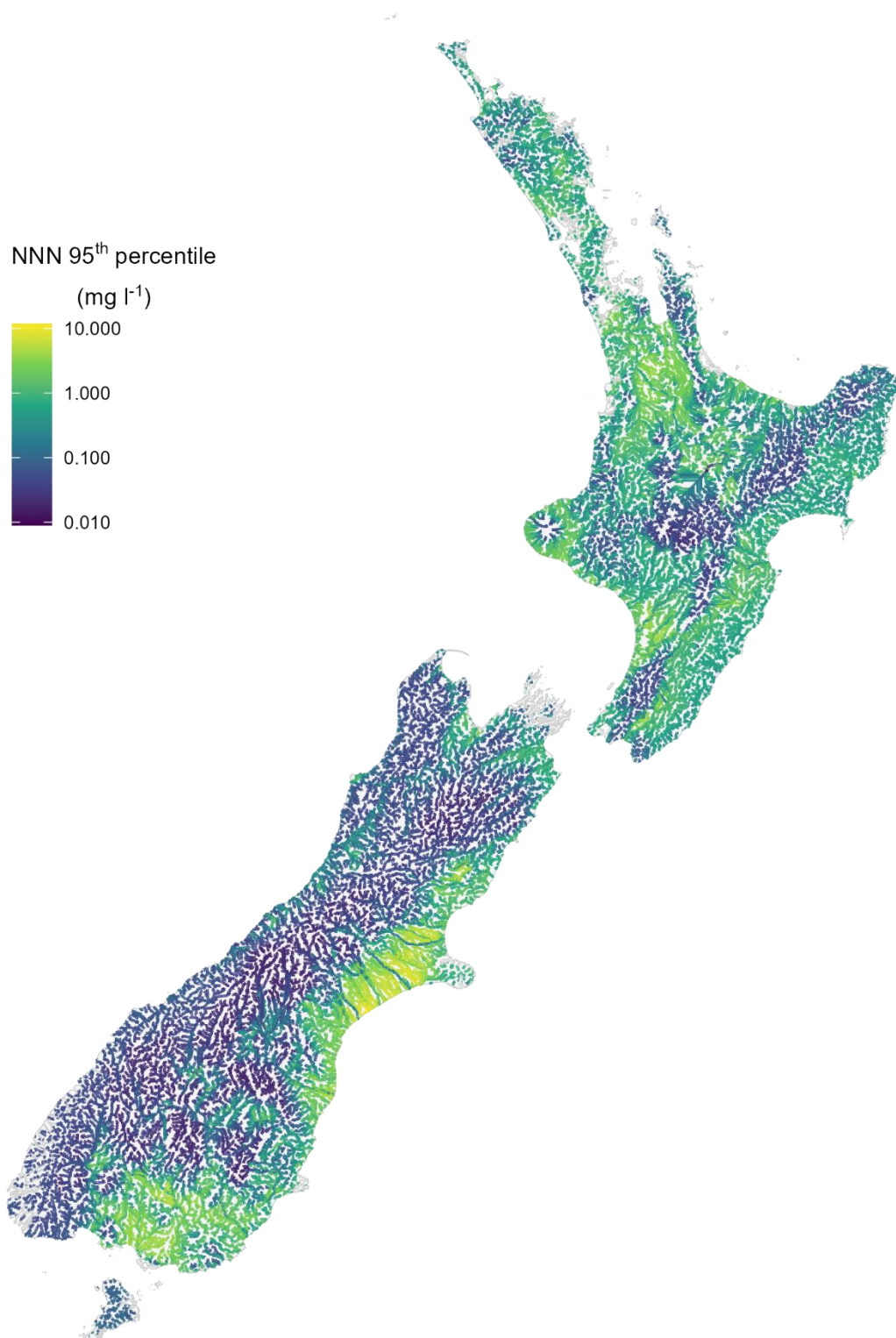


Figure 4-12: Predicted 95th percentile NNN concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a log₁₀ colour scale.

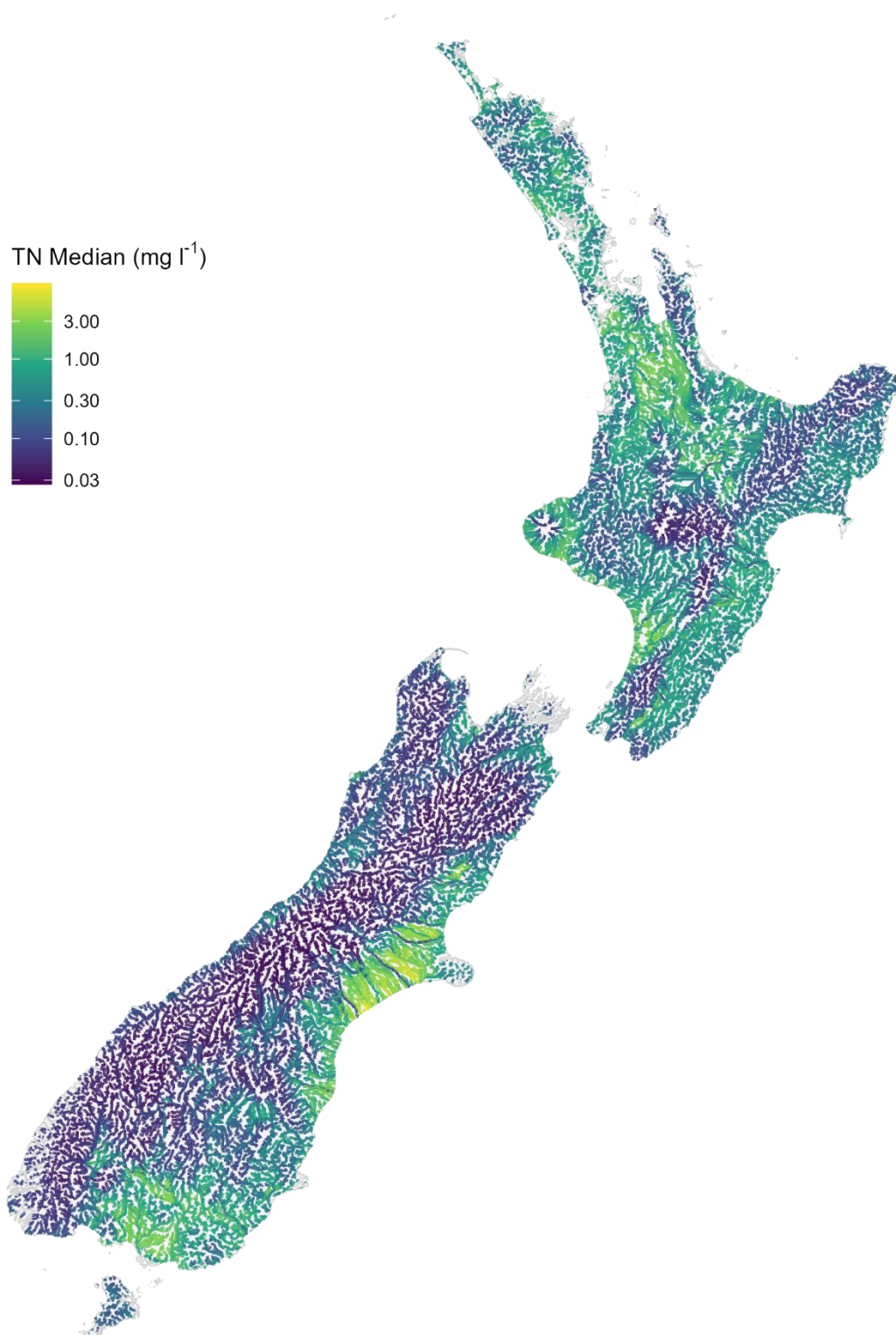


Figure 4-13: Predicted median TN concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a \log_{10} colour scale.

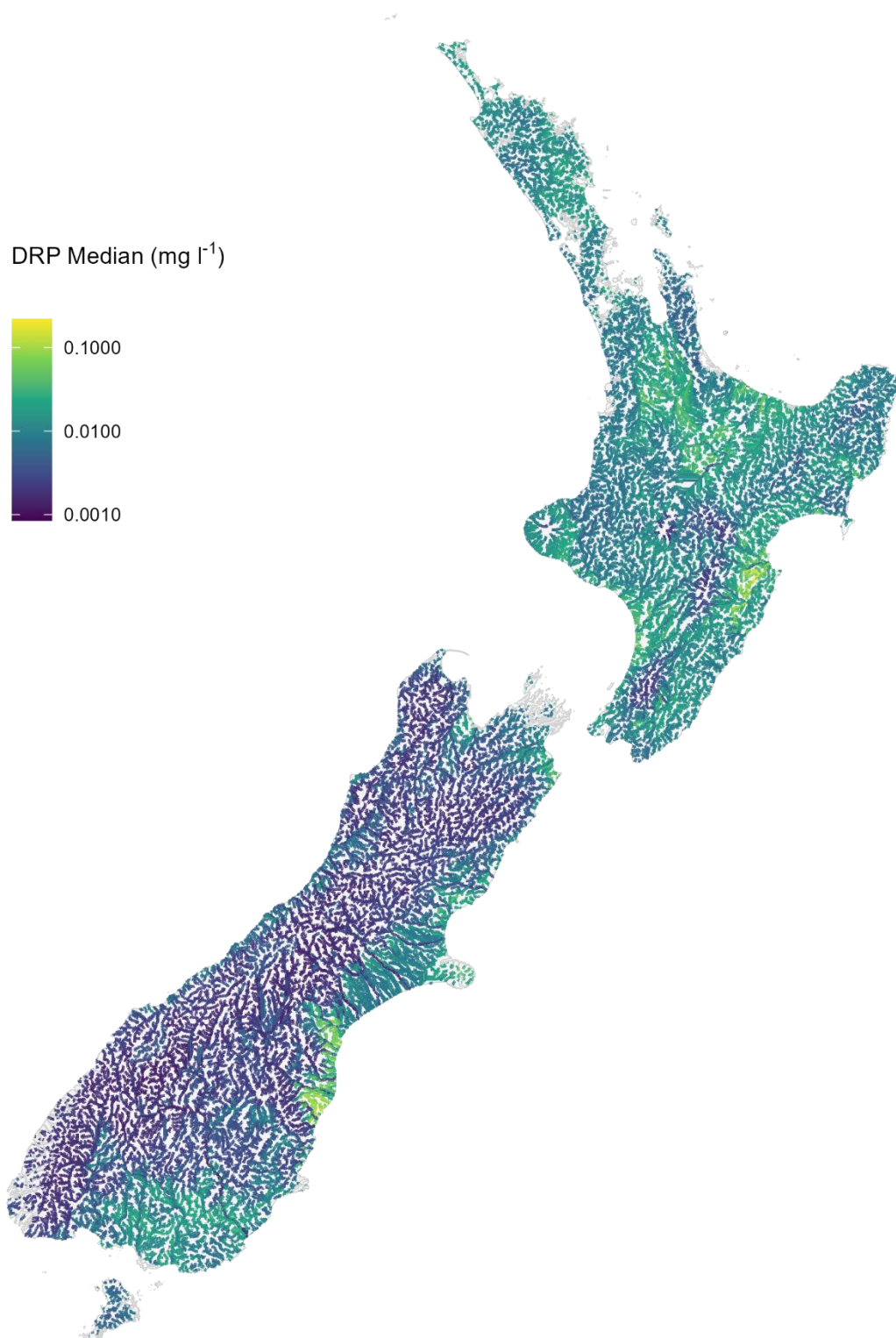


Figure 4-14: Predicted median DRP concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a \log_{10} colour scale.

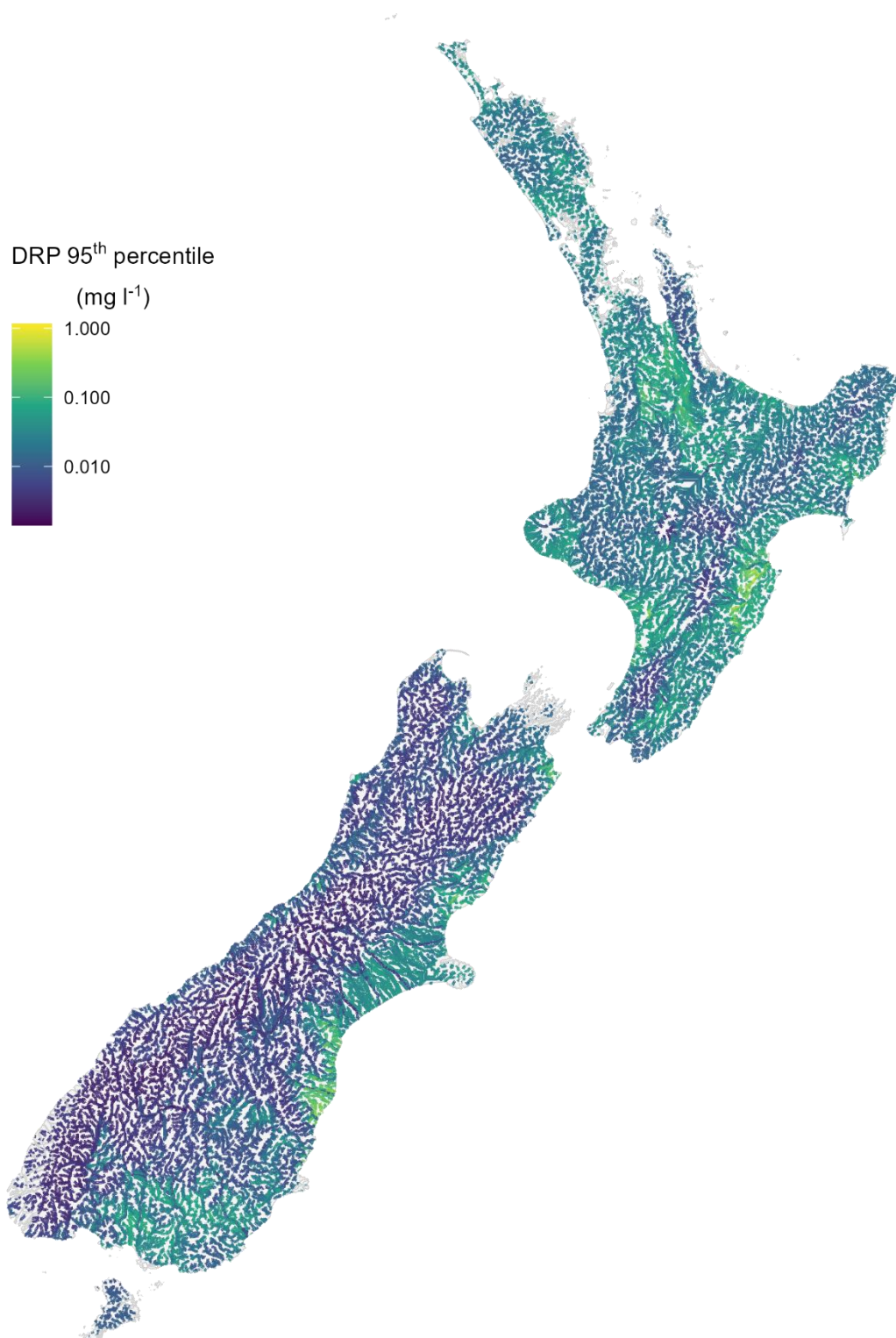


Figure 4-15: Predicted 95th percentile DRP concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a log₁₀ colour scale.

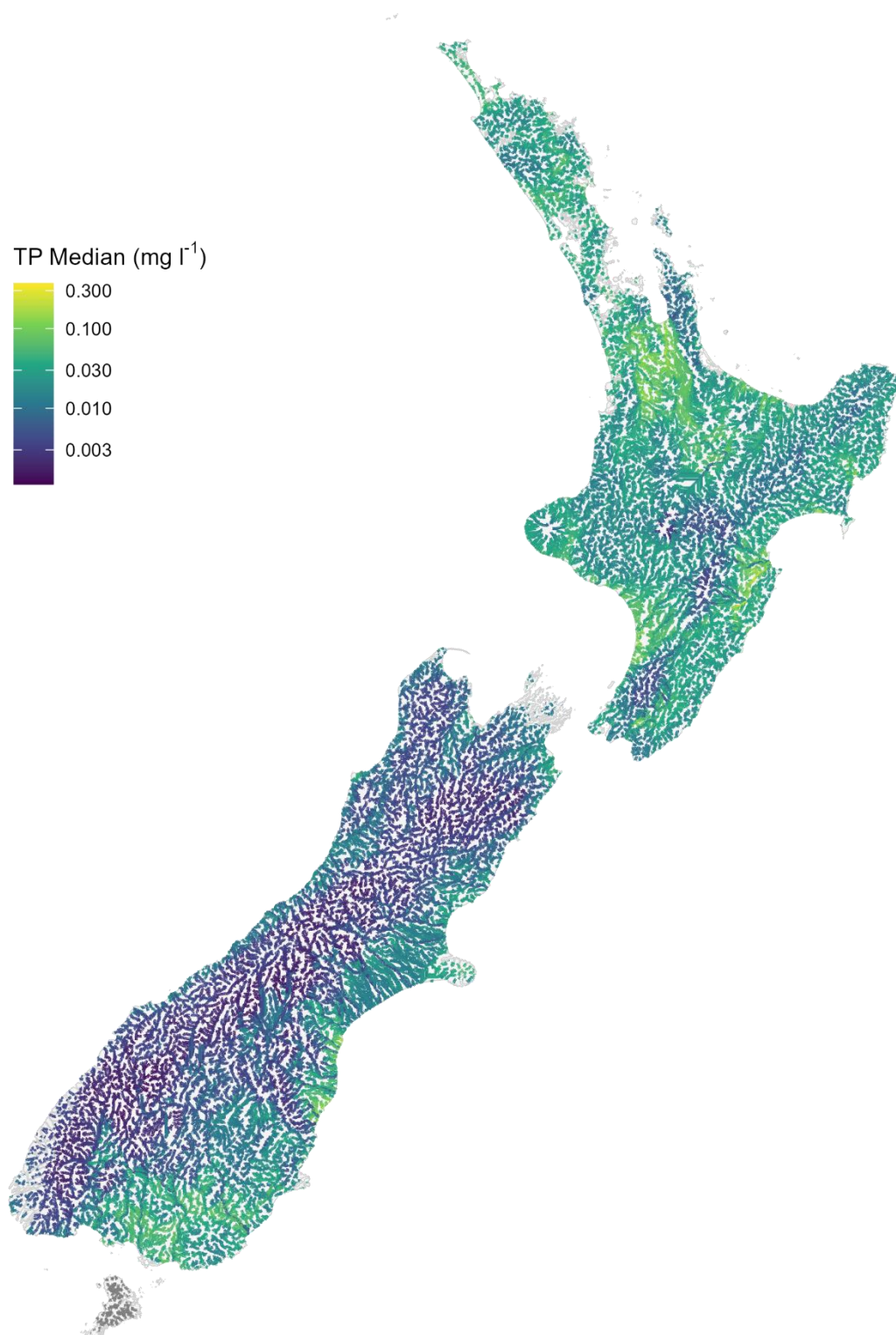


Figure 4-16: Predicted median TP concentration in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a \log_{10} colour scale.

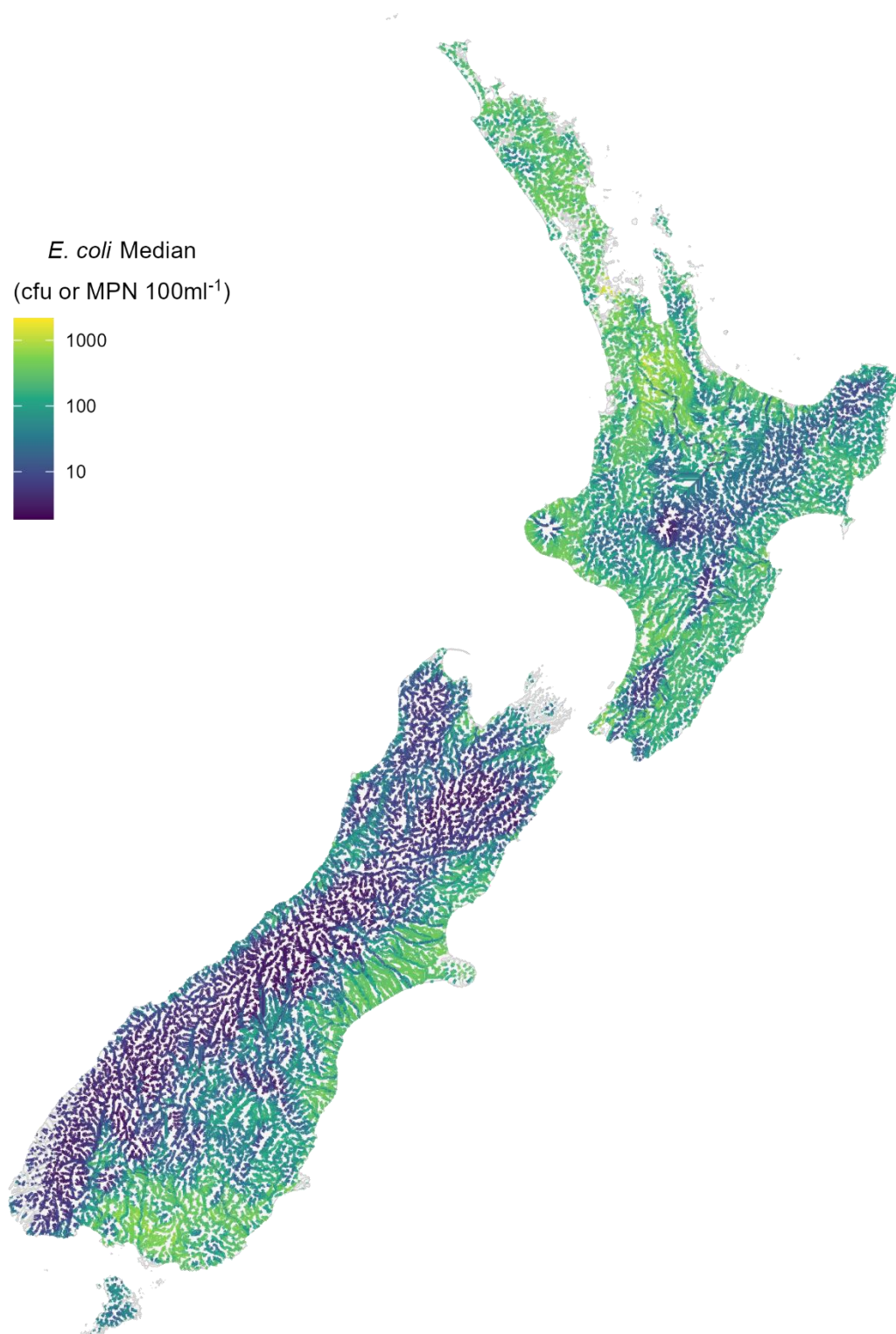


Figure 4-17: Predicted median *E. coli* in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a log₁₀ colour scale.

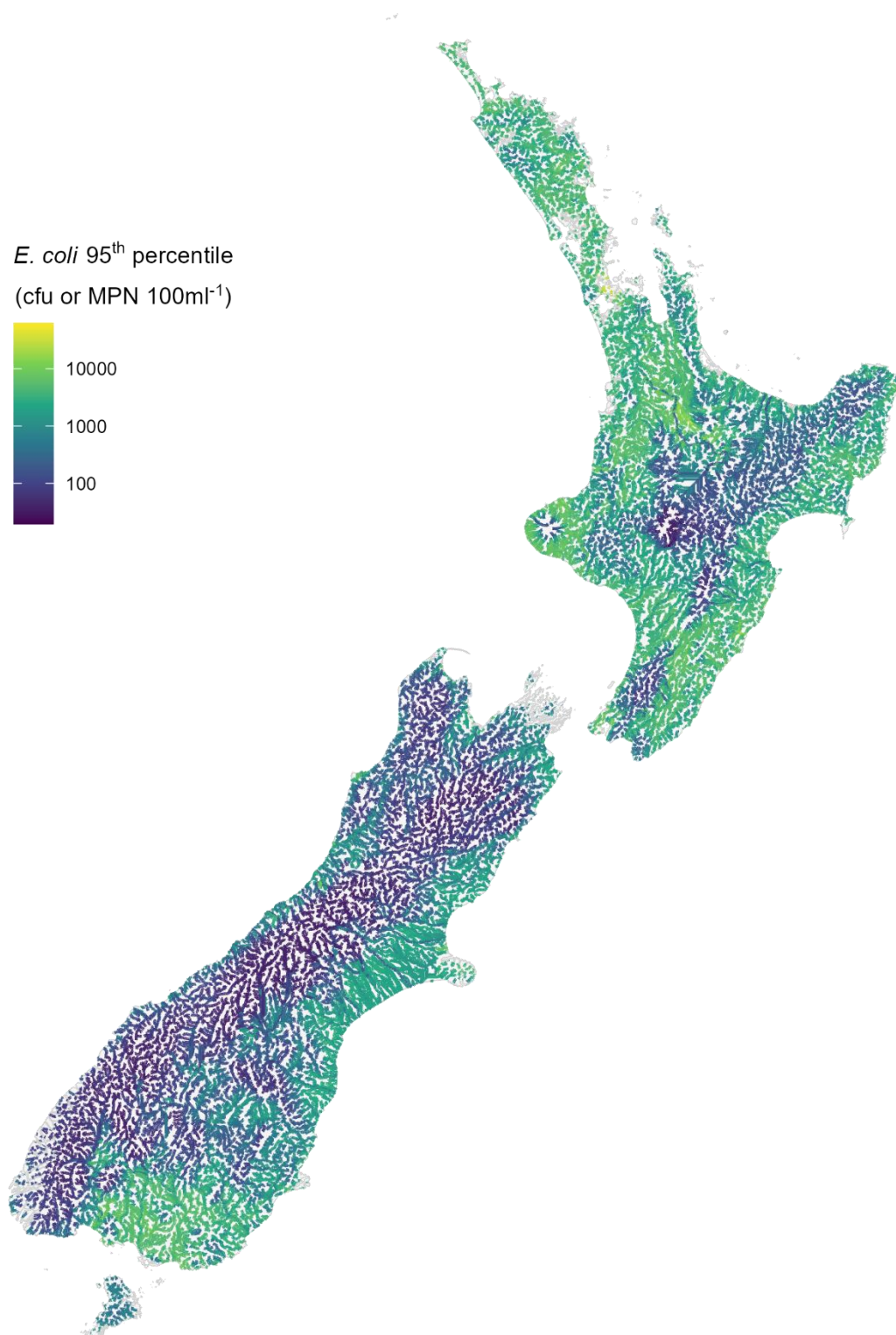


Figure 4-18: Predicted 95th percentile *E. coli* in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a log₁₀ colour scale.

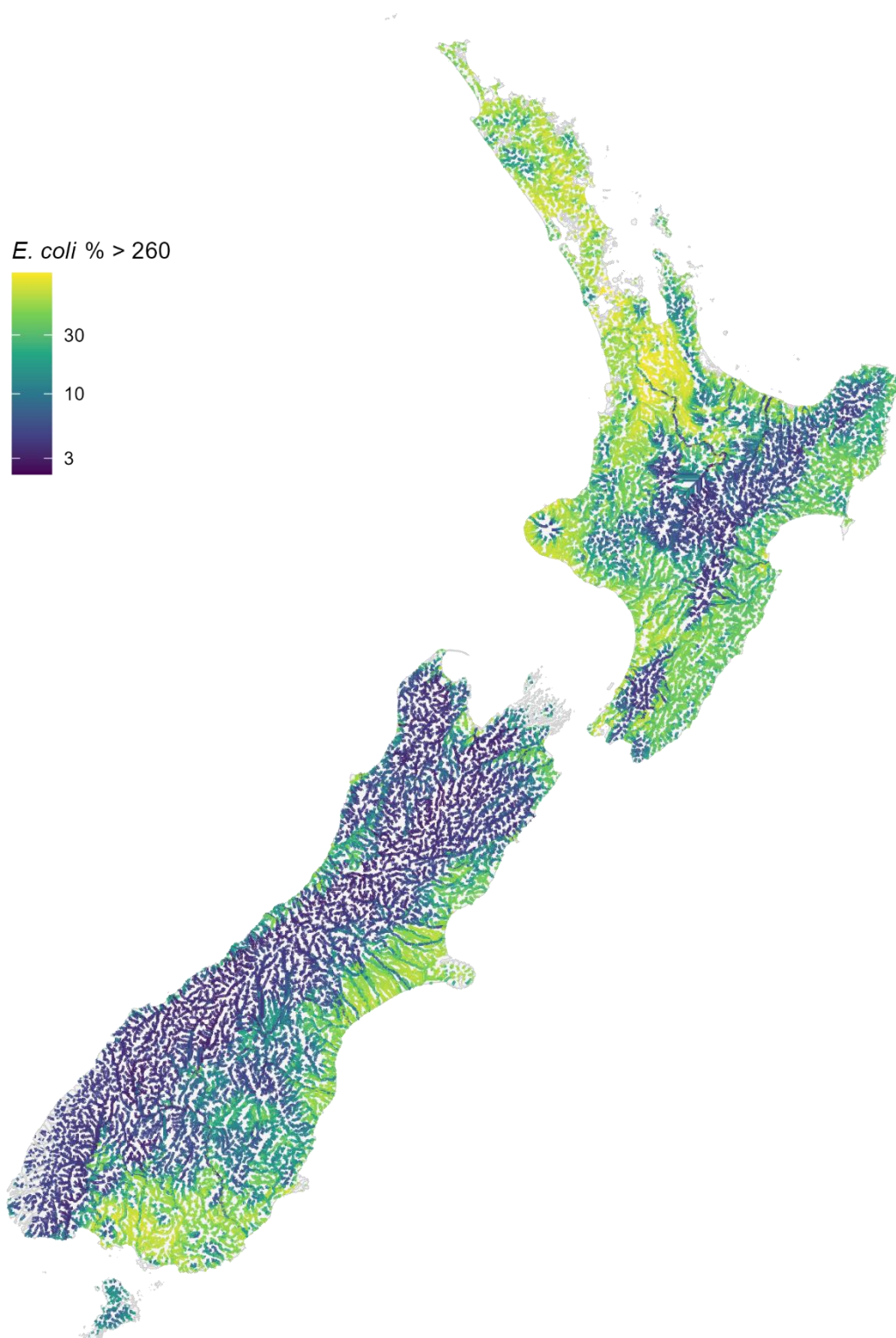


Figure 4-19: Predicted proportion of samples exceeding 260 *E. coli* 100 ml⁻¹ in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a log₁₀ colour scale.

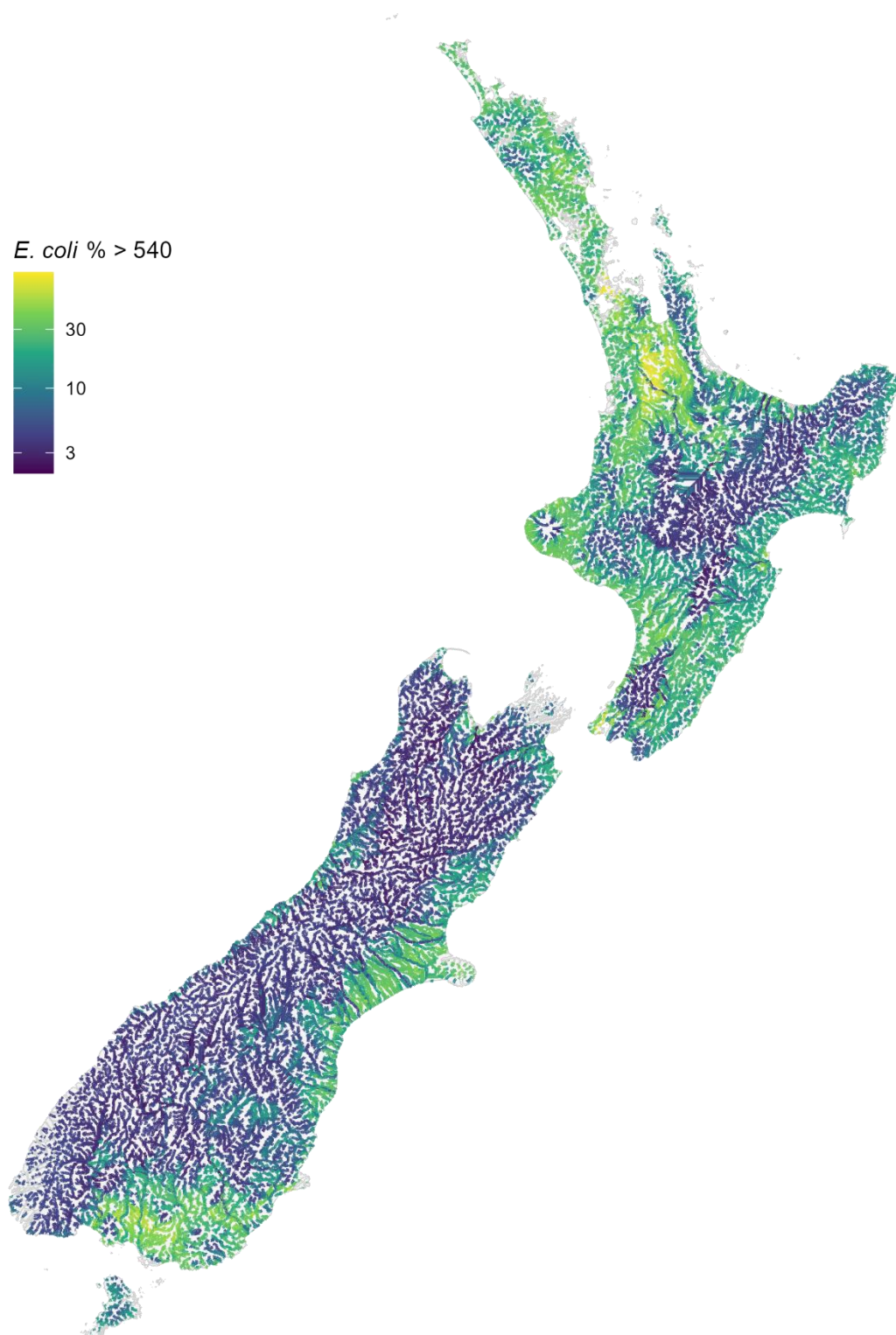


Figure 4-20: Predicted proportion of samples exceeding 540 *E. coli* 100 ml⁻¹ in New Zealand rivers. Map shows all nzsegments of Order 3 and higher. Values are represented on a log₁₀ colour scale.

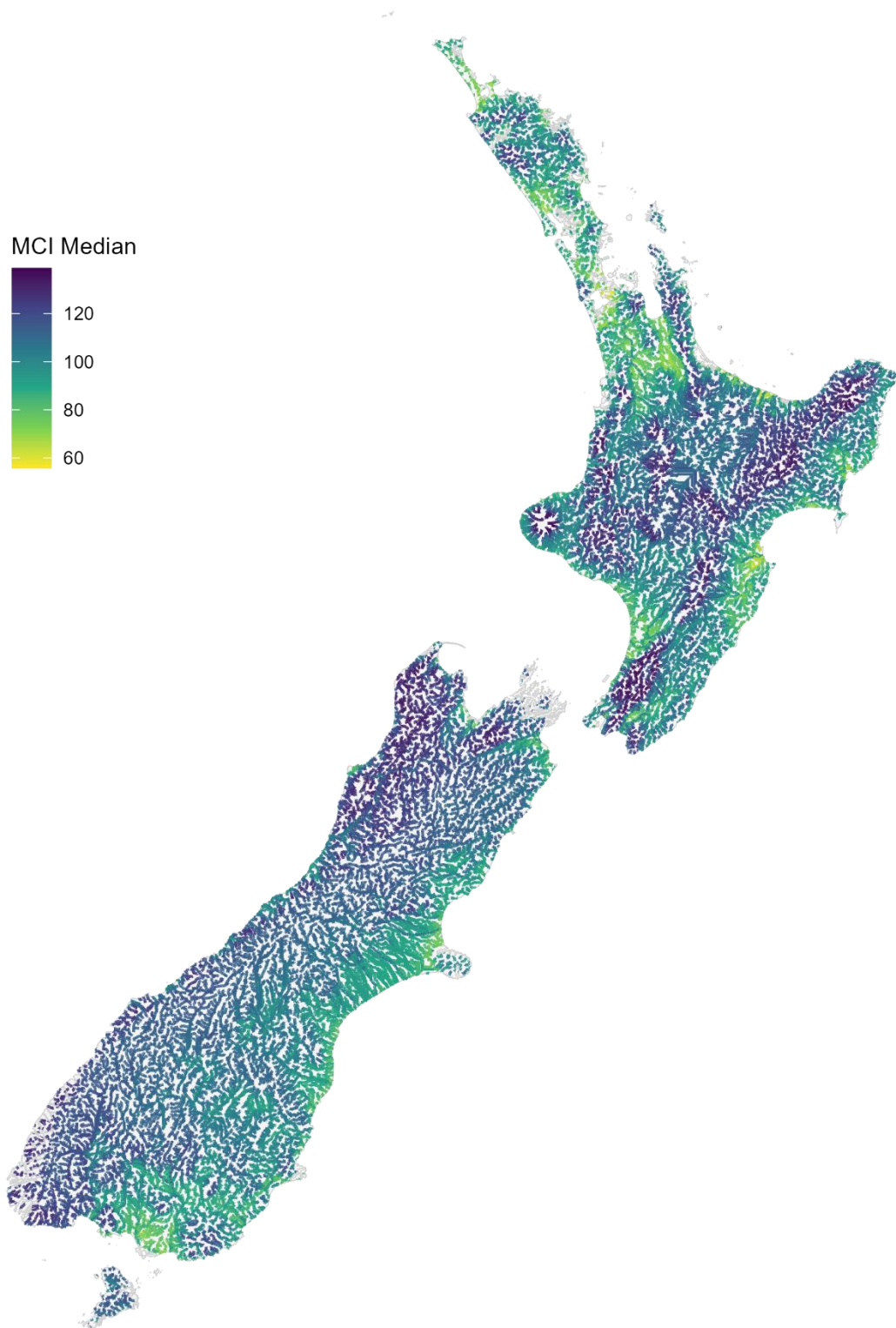


Figure 4-21: Predicted median MCI scores in New Zealand rivers. Map shows all nzsegments of Order 3 and higher.

5 Discussion

5.1 Comparison with previous studies

The models of current river attribute state presented in this study update previous modelling work carried out by Unwin et al. (2010), Larned et al. (2016a), Whitehead (2018), and Whitehead et al (2022). The models in the previous three reports were based on water quality data from 1996–2007, 2009–2013, 2013–2017 and 2016–2020, respectively, while the current models are based on water quality data from 2020–2024. The results of the current study are generally consistent with those of earlier studies, with similar structures of the models (as indicated by the relative importance of predictor variables and directions of partial plots), although some of the predictors have changed (e.g., we used the latest available information on landcover and elevation). In addition, the performance of the models in the present study (as indicated by percent variance explained) was generally comparable to model performance from earlier studies (Unwin et al. 2010; Larned et al. 2016a; Whitehead 2018; Whitehead et al. 2022a).

Improvements in the modelling methodology and predictor variables between the 2010 and 2016 studies (see Larned et al. 2016) increased the performance of the random forest models. In the current study, we used the same modelling procedures and most of the predictor variables as (Whitehead 2018). However, we generated new landcover predictors using 2023 landcover data (LCDB6), instead of the 2008 (LCDB3), 2012 (LCDB4), 2018 (LCDB5) landcover data used in the previous reports. This spatial layer represents the most current landcover data available at a national scale. We also included predictors of land use intensity associated with four different stock types (beef, dairy, deer, sheep) following Snelder et al (2022).

In a previous study, Clapcott et al. (2013) fitted a random forest model to site median MCI scores and reported a cross-validated R^2 of 0.63. Clapcott et al. (2013) also reported a cross-validated R^2 of 0.64 with an alternative technique, boosted regression trees. The equivalent R^2 statistic for the MCI model in the current study was 0.72 (Table 4-1). The improvement in performance in the current study compared to previous studies may reflect an increase in number of sites at which water quality data are available and the inclusion of the land use intensity predictors.

5.2 Model uncertainty

In this study, we modelled broad-scale patterns in attribute state using catchment characteristics and segment-scale descriptors as predictor variables. Because the processes determining water quality at any location are complex, some unexplained variation in our models is to be expected. Predictions made for individual locations are associated with uncertainties characterised by model RMSD (Table 4-1). However, the level of model bias as estimated by PBIAS, alongside the smearing estimate (S), which ranged from 1.02 to a maximum of 1.13 for NNN median, both indicated that the level of bias for each attribute state was relatively low.

Random forest model predictive performance differed among attribute states (Table 4-1). This variation may be attributed to differences in number of monitored sites, uncertainty in predictor data, and the biophysical processes that control the water quality variables, as well as measurement errors. Some biophysical processes may be poorly represented by our catchment-averaged spatial predictors. For example, concentrations of dissolved and total

nitrogen and phosphorus in rivers are influenced to differing degrees by adsorption-desorption processes, deposition and suspension, and biological assimilation, transformation and removal; these mechanisms are not explicitly represented in the random forest models. Measurement errors, associated with differences in laboratory and data collection processes, as well as instrumentation (for example weak numerical comparability between turbidity measurement instruments (Davies-Colley et al. 2021)) cannot be accounted for in the random forest models. The absence of predictors and measurement errors that account for these and other processes means that some level of unexplained variation is inevitable.

We used the criteria of Moriassi *et al.* (2015) to categorise the performance of the 16 models as “very good”, “good”, “satisfactory” and “unsatisfactory” (Table 3-1). 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.

5.3 Alternative modelling methods

The random forest method that we used to model attribute state is well-suited to generation of spatial predictions using data from monitoring sites that represent a wide range of environmental conditions. However, it is not the only method available. Alternative statistical models include generalised additive models (GAMs; Hastie et al. 2001), artificial neural networks (Joy and Death 2004), and boosted regression trees (e.g., Elith et al 2008). We did not employ these alternatives, but it is possible that some water quality applications would be better served by models developed by one of the alternative methods. In particular, if it is important to identify areas with potentially extreme water quality values, models such as GAMs that can extrapolate beyond the range of the fitting data would be useful, although such predictions may lead to spurious results.

Alternative methods for modelling *E. coli* exceedances, which would allow inclusion of all available observed data, include: a) develop a model whose response is classes indicating whether sites had any exceedances within the assessment period; b) application of Generalised Linear Models (GLM) with an appropriately link function to represent proportion data; or c) application of hurdle models that first models whether there is an exceedance and then model the proportion of exceedances.

In addition, models that incorporate biophysical processes (e.g., Elliot et al 2016) are available. In some circumstances, process models are better suited than machine learning statistical models (e.g., random forests) to inform a particularly aspect of environmental policy. We considered random forest models to be a reliable and robust tool for predicting attribute state for national scale reporting for three general reasons:

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

time-consuming and complex to calibrate or parameterise than purely empirical models.

2. Random forest model predictions can be mapped across spatial scales, from individual river segments to the entire county. These maps provide a useful description of spatial patterns in water quality attribute state for environmental reporting purposes.
3. Among empirical modelling methods that generate associations between water quality and environmental characteristics, random forest models have several advantages: they are minimally affected by multi-collinearity among predictor variables, they are robust against over-fitting, they are unaffected by variation in data distributions, and it is possible to use techniques such as a variable importance plot to assist with model interpretability. Random forest models cannot predict beyond the range of the observations, which may limit their utility in some applications. In the present study, limiting model predictions to the range of observations was a positive attribute as it ensured that those predictions were conservative. See Booker and Whitehead (2018) for further discussion of this topic.

6 Acknowledgements

Many thanks are due to many staff across the various organisations who provided water quality data. We also thank Matt Wilkins for support in providing predictor data, and to Caroline Frazer, who prepared the attribute data.

7 Glossary of abbreviations and terms

Ammoniacal nitrogen adjusted for pH (NH ₄ -N-adj)	The NPS-FM values for ammoniacal nitrogen are set at a pH of 8 and a temperature of 20 °C. the raw values have to be adjusted to allow for comparison.
Ammoniacal nitrogen (NH ₄ -N)	Ammoniacal nitrogen is a form of nitrogen in water that exists as either ammonia (NH ₃) or ammonium (NH ₄) and is usually derived from human or animal waste and at high concentrations can be toxic. NH ₄ -N is the concentration of ammoniacal nitrogen measured as nitrogen (N)
Attribute states	The combination of water quality variables and statistics (e.g., 95 th percentile of dissolved reactive phosphorus).
Digital network	Digital representation of New Zealand rivers
Dissolved reactive phosphorus (DRP)	Phosphate that is dissolved and available for aquatic plants and algae growth. High levels of DRP can cause rapid weed growth or algal blooms, which can choke aquatic life.
<i>Escherichia coli</i>	A bacterium found in warm-blooded animals' faeces. Indicates the level of faecal contamination and risk to human health due to drinking or accidentally ingesting water contaminated with faecal matter.
Macroinvertebrate Community Index (MCI)	Macroinvertebrate Community Index (MCI) is a method for assessing the health of streams by sampling the macroinvertebrates (small invertebrates like insects, worms, and snails) that live in them. A higher MCI score indicates a healthier ecosystem.
monitoring sites	Location on the river where samples are taken for the purpose of monitoring water quality.
Nitrate + nitrite-nitrogen (NNN)	Sum of nitrate-nitrogen (NO ₃ -N), nitrite-nitrogen (NO ₂ -N). At high concentrations, it can become toxic to fish and macroinvertebrates. It can also cause excessive algal growth.
NPS-FM	National Policy Statement for Freshwater Management. Contains attribute tables to define and measure the health of freshwater bodies.
NRWQN	National River Water Quality Network - monitoring sites operated by Earth Sciences New Zealand for State of the Environment Reporting
nzsegment identifier	Unique identifier for each section of the river represented on the digital network.
nzsegment	Section of the river represented on the digital network version 2.4.
random forest	An advanced form of regression-tree modelling that builds many decision trees and combines their predictions to improve accuracy and reduce overfitting.
River segments	A section of the river is represented on the digital network.
SoE	State of the Environment.

Total nitrogen (TN)	The sum of all nitrogen species in water, including nitrate-nitrogen ($\text{NO}_3\text{-N}$), nitrite-nitrogen ($\text{NO}_2\text{-N}$), ammoniacal-nitrogen ($\text{NH}_4\text{-N}$) and organic-nitrogen, measured as nitrogen (N). At high concentrations, it can become toxic to fish and macroinvertebrates. It can also cause excessive growth of algae.
Total phosphorus (TP)	The combined amount of all phosphorus forms in water, including both dissolved reactive phosphorus (DRP), which is easily taken up by plants, and phosphorus bound to sediments or particles.
Turbidity	The cloudiness or haziness in a fluid caused by individual small particles (suspended solids) and water colour.
Visual clarity	A measure of how far underwater visibility extends. Low clarity may reduce the ability of fish to see prey and detect predators, and reduce light penetration. It may also reduce or inhibit the growth of aquatic plants and their ability to produce food and oxygen for species that depend on them

8 References

- ANZECC, ARMCANZ (2000) Australian and New Zealand Guidelines for Fresh and Marine Water Quality. In: ACT (Ed), Canberra.
- Booker, D. (2023) National river digital networks and the River Environment Classification: future pathways for stewardship, maintenance, and upgrading of products and services for national benefit. *NIWA Client Report*, 2023272CH: 39.
- Booker, D., Wilkinson, C., Wilkins, M. (2024) Digital networks: challenges, solutions, and case studies to inform nationwide integrated freshwater-land mapping. *NIWA Client Report*, 2024138CH: 90.
- Booker, D.J., Smith, R.G., Wood, D., Fraser, C.E., Snelder, T.H. (2025) Water quality state and trends in New Zealand rivers; analysis of national data ending in 2024. *ESNZ Client Report*, 2025331CH: 95.
- Booker, D.J., Whitehead, A.L. (2018) Inside or Outside: Quantifying Extrapolation Across River Networks. *Water Resources Research*, 54(9): 6983–7003.
<https://doi.org/10.1029/2018WR023378>
- Breiman, L. (2001) Random forests. *Machine Learning*, 45(1): 5–32.
- Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J. (2017) *Classification and regression trees*. Chapman and Hall/CRC.
- Camargo, J.A., Alonso, A., Salamanca, A. (2005) Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates. *Chemosphere*, 58(9): 1255–1267.
<https://doi.org/10.1016/j.chemosphere.2004.10.044>
- Close, M.E., Davies-Colley, R.J. (1990) Baseflow water chemistry in New Zealand rivers 2. Influence of environmental factors. *New Zealand Journal of Marine and Freshwater Research*, 24(3): 343–356. 10.1080/00288330.1990.9516429
- Cutler, D.R., Edwards Jr, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J. (2007) Random forests for classification in ecology. *Ecology*, 88(11): 2783–2792.
- Davies-Colley, R., Hughes, A.O., Vincent, A.G., Heubeck, S. (2021) Weak numerical comparability of ISO-7027-compliant nephelometers. Ramifications for turbidity measurement applications. *Hydrological Processes*, 35(12): 12. 10.1002/hyp.14399
- Davies-Colley, R., Smith, D. (2001) Turbidity suspended sediment and water clarity. *J Am Water Resour Assoc*, 37: 1085–1101.
- Duan, N. (1983) Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association*, 78(383): 605–610. 10.1080/01621459.1983.10478017
- Dudley, B.D., R. Burge, O., Plew, D., Zeldis, J. (2020) Effects of agricultural and urban land cover on New Zealand’s estuarine water quality. *New Zealand Journal of Marine and Freshwater Research*, 54(3): 372–392.

- Eden, D., Parfitt, R. (1992) Amounts and sources of phosphate in hill country rocks, south-eastern North Island, New Zealand. *Soil Research*, 30(3): 357–370.
- Elith, J., Leathwick, J.R., Hastie, T. (2008) A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4): 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>
- Elliott, A.H., Semadeni-Davies, A.F., Shankar, U., Zeldis, J.R., Wheeler, D.M., Plew, D.R., Rys, G.J., Harris, S.R. (2016) A national-scale GIS-based system for modelling impacts of land use on water quality. *Environmental Modelling & Software*, 86: 131–144. 10.1016/j.envsoft.2016.09.011
- Greenwell, B.M. (2017) pdp: An R package for constructing partial dependence plots. *The R Journal*, 9: 421–436. <https://doi.org/10.32614/RJ-2017-016>
- Hastie, T., Tibshirani, R., Friedman, J.H. (2009) *The elements of statistical learning: data mining, inference, and prediction*. Springer, New York. <https://go.exlibris.link/C2wM0hC2>
- Hickey, C. (2013) Derivation of indicative ammoniacal nitrogen guidelines for the National Objectives Framework. *NIWA Memorandum*, MFE13504.
- Hickey, C.W. (2014) Updating nitrate toxicity effects on freshwater aquatic species. *NIWA Client Report*, HAM2013-009: 39. Q:\LIBRARY\ClientRept\E-copies Client reports
- Joy, M.K., Death, R.G. (2004) Predictive modelling and spatial mapping of freshwater fish and decapod assemblages using GIS and neural networks. *Freshwater Biology*, 49(8): 1036–1052. <https://doi.org/10.1111/j.1365-2427.2004.01248.x>
- Julian, J.P., de Beurs, K.M., Owsley, B., Davies-Colley, R.J., Ausseil, A.G.E. (2017) River water quality changes in New Zealand over 26 years: response to land use intensity. *Hydrol. Earth Syst. Sci.*, 21(2): 1149–1171. 10.5194/hess-21-1149-2017
- Larned, S., Snelder, T., Unwin, M. (2016a) Water quality in New Zealand rivers: Modelled water quality state. *NIWA Client Report*, CHC2016-070: 40. Q:\LIBRARY\ClientRept\E-copies Client reports\CHRISTCHURCH
- Larned, S.T., Snelder, T., Unwin, M.J., McBride, G.B. (2016b) Water quality in New Zealand rivers: current state and trends. *New Zealand Journal of Marine and Freshwater Research*, 50(3): 389–417. 10.1080/00288330.2016.1150309
- Leathwick, J.R., Overton, J.M., McLeod, M. (2003) An Environmental Domain Classification of New Zealand and Its Use as a Tool for Biodiversity Management. *Conservation Biology*, 17(6): 1612–1623. <https://doi.org/10.1111/j.1523-1739.2003.00469.x>
- Liaw, A., Wiener, M. (2002) Classification and Regression by randomForest. *R news*, 2(3): 18–22. <https://CRAN.R-project.org/doc/Rnews/>
- LINZ (2025) NZ Coastlines and Islands Polygons (Topo 1:50k). <https://data.linz.govt.nz/layer/51153-nz-coastlines-and-islands-polygons-topo-150k/>
- LRIS (2024) FSL Particle Size Classification. Manaaki Whenua Landcare Research. <https://lris.scinfo.org.nz/layer/48112-fsl-particle-size-classification/>

McDowell, R., Snelder, T., Cox, N., Booker, D., Wilcock, R. (2013) Establishment of reference or baseline conditions of chemical indicators in New Zealand streams and rivers relative to present conditions. *Marine and Freshwater Research*, 64(5): 387–400.

McGroddy, M.E., Baisden, W.T., Hedin, L.O. (2008) Stoichiometry of hydrological C, N, and P losses across climate and geology: An environmental matrix approach across New Zealand primary forests. *Global Biogeochemical Cycles*, 22(1). <https://doi.org/10.1029/2007GB003005>

Ministry for the Environment (1994) Guidelines for the management of the colour and clarity of water. Water Quality Guidelines 2.,

Moriasi, D.N. (2007) Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, 50(3): 885–900. 10.13031/2013.23153

Moriasi, D.N., Gitau, M.W., Pai, N., Daggupati, P. (2015) Hydrologic And Water Quality Models: Performance Measures And Evaluation Criteria. *Transactions of the ASABE*, 58(6): 1763–1785. <Go to ISI>://WOS:000368341900024

Nash, J.E., Sutcliffe, J.V. (1970) River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3): 282–290. [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6)

New Zealand Government (2024) National Policy Statement for Freshwater Management 2020 (October 2024). In: Ministry for the Environment (Ed), Wellington: 75.

NIWA (ND) Virtual Climate Station Network (VCSN) data technical description. <https://niwa.co.nz/climate-and-weather/virtual-climate-station-network-vcsn-data-technical-description>

Pebesma, E., Bivand, R. (2023) *Spatial data science: With applications in R*. Chapman and Hall/CRC.

Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M. (2008) How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling*, 216(3): 316–322. <https://doi.org/10.1016/j.ecolmodel.2008.05.006>

Plew, D.R., Zeldis, J.R., Dudley, B.D., Whitehead, A.L., Stevens, L.M., Robertson, B.M., Robertson, B.P. (2020) Assessing the Eutrophic Susceptibility of New Zealand Estuaries. *Estuaries and Coasts*, 43(8): 2015–2033. <https://doi.org/10.1007/s12237-020-00729-w>

Probst, P., Boulesteix, A.-L. (2018) To tune or not to tune the number of trees in random forest. *Journal of Machine Learning Research*, 18(181): 1–18.

Snelder, T., Fraser, C., Whitehead, A. (2022a) Spatial modelling of lake water quality state: Incorporating monitoring data for the period 2016 to 2020, LWP Client Report Number: 2021-15.

Snelder, T.H., Fraser, C., Larned, S.T., Monaghan, R., De Malmanche, S., Whitehead, A.L. (2022b) Attribution of river water-quality trends to agricultural land use and climate variability in New Zealand. *Marine and Freshwater Research*, 73(1): 1–19. <https://doi.org/10.1071/MF21086>

Snelder, T.H., Larned, S.T., McDowell, R.W. (2018) Anthropogenic increases of catchment nitrogen and phosphorus loads in New Zealand. *New Zealand Journal of Marine and Freshwater Research*, 52(3): 336–361. 10.1080/00288330.2017.1393758

Speiser, J.L., Miller, M.E., Tooze, J., Ip, E. (2019) A comparison of random forest variable selection methods for classification prediction modeling. *Expert Systems with Applications*, 134: 93–101. <https://doi.org/10.1016/j.eswa.2019.05.028>

Stark, J.D., Maxted, J.R. (2007) A user guide for the macroinvertebrate community index. *Cawthron Report*, 1166: 58.

Svetnik, V., Liaw, A., Tong, C., Wang, T. (2004) Application of Breiman's Random Forest to Modeling Structure-Activity Relationships of Pharmaceutical Molecules. *Multiple Classifier Systems*, Berlin, Heidelberg, 2004//.

Timperley, M.H. (1983) *Chemical Geology*, 38(3): 287–306. [https://doi.org/10.1016/0009-2541\(83\)90060-8](https://doi.org/10.1016/0009-2541(83)90060-8)

Unwin, M., Snelder, T., Booker, D., Ballantine, D., Lessard, J. (2010) Predicting water quality in New Zealand rivers from catchment-scale physical, hydrological and land use descriptors using random forest models. *NIWA Client Report*, CHC2010-037: 50.

Whitehead, A. (2018) Spatial modelling of river water-quality state Incorporating monitoring data from 2013 to 2017, 2018360CH: 41.

Whitehead, A., Booker, D. (2020) NZ River Maps: An interactive online tool for mapping predicted freshwater variables across New Zealand. <https://shiny.niwa.co.nz/nzrivermaps/>

Whitehead, A., Fraser, C., Snelder, T. (2022a) Spatial modelling of river water quality state: Incorporating monitoring data from 2016 to 2020. *NIWA Client Report*, 2021303CH: 47.

Whitehead, A., Fraser, C., Snelder, T., White, R., Walter, K., Woodward, S., Zammit, C. (2022b) Water quality state and trends in New Zealand rivers analyses of national rivers data ending in 2020, 2021296CH: 73.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D.A., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. (2019) Welcome to the tidyverse. *Journal of Open Source Software*, 4(43): 1686.