ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

Variability in the Trophic Level Index in Lake Rotoehu from 1990-2021

ERI REPORT NUMBER 170

By
Whitney Woelmer and Deniz Özkundakci

A client report prepared for
Bay of Plenty Regional Council

2024

Email: wwoelmer@waikato.ac.nz

Te Tumu Whakaora Taiao - Environmental Research Institute
Division of Health, Engineering, Computing & Science,
University of Waikato, Private Bag 3105,
Hamilton 3240, New Zealand

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

Cite report as:

Woelmer W., Özkundakci D. 2024. Variability in the Trophic Level Index in Lake Rotoehu from 1990 to 2021. ERI Report 170, a client report prepared for Bay of Plenty Regional Council. Environmental Research Institute – Te Tumu Whakaora Taiao, Division of Health, Engineering, Computing & Science, University of Waikato, Hamilton, New Zealand. 44 pp. DOI: 10.15663/ERI.Report.170 ISSN 2463-6029 (Print) & ISSN 2350-3432 (Online)

Cover picture:

Lake Rotoehu (Photo: Grant Tempero)

Disclaimer:

The information and opinions provided in the Report have been prepared for the Client and its specified purposes. Accordingly, any person other than the Client uses the information and opinions in this report entirely at their own risk. The Report has been provided in good faith and on the basis that reasonable endeavours have been made to be accurate and not misleading and to exercise reasonable care, skill and judgment in providing such information and opinions.

Neither the University of Waikato nor any of its employees, officers, contractors, agents or other persons acting on its behalf or under its control accepts any responsibility or liability to third parties in respect of any information or opinions provided in this Report.

Reviewed by:

Name: Maggie Armstrong

Title: Postdoctoral Researcher

Affiliation: University of Waikato

Approved for release by

Name: Charles Lee

Academic Co-Director

Environmental Research Institute

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

Executive Summary

Setting an appropriate Trophic Level Index (TLI) target for lakes in the Rotorua Te Arawa region is critical for informing short- and long-term management decisions. Due to high variability in the observed TLI of Lake Rotoehu, coupled with contrasting understanding of historical water quality in the lake, the Bay of Plenty Regional Council requested a robust analysis of the drivers of the TLI over time to help inform the suitability of the current TLI target. To undertake this, we first quantified the uncertainty around the TLI in the 1990s, when the current TLI target was set based on the "best" observed water quality within the long-term monitoring program, to understand how a lower sampling frequency from 1990-2000 (~every 2-3 months) impacted TLI estimates. We found that while sampling fewer times in the year in the 1990s increased uncertainty around TLI estimates by 0.2-0.3 TLI units, a marked shift in water quality from a TLI of 3.6 in 1992 to 4.5 in 1993 remained evident.

To understand if the main drivers affecting the TLI were different during the 1990s as opposed to later decades, we calculated the Pearson correlation coefficient between the TLI and several driver variables separately for each decade. Drivers in this analysis included meteorological variables, lake water level, and in-lake water quality not in the TLI calculation (e.g., bottom water nutrients, water temperature, stratification metrics), and the amount of aluminium sulphate dosed to the lake. While several variables were consistently important across all decades (e.g., bottom water nutrients, bottom water temperature, and mean monthly air temperature), average monthly water level and minimum windspeeds were only important in the 1990s, indicating that these variables had been related to major shifts in water quality seen during that time.

Lastly, to better understand the relative importance of multiple drivers of the TLI and whether this importance changes over time, we conducted a moving window analysis. We fit autoregressive models, meaning the models included the previous month's TLI estimate, as well as a single driver variable over a moving window of ~8 years, where each window moved forward one month at a time. We found that air temperature, bottom water temperature, and bottom water dissolved reactive phosphorus (DRP) were most often the top predictors of the TLI. However, the relative importance of average monthly water level increased sharply to the most important driver during a time period which also corresponded to very high water levels. Additionally, examination of model parameters over the simulation period demonstrated that the strength and magnitude of the relationship between the TLI and individual drivers changed over time, indicating that that relationships are not fixed through time.

Overall, this work highlights the importance of re-evaluating the underlying relationships between the TLI and drivers over time, emphasising the dynamic nature of Lake Rotoehu. We demonstrate a clear shift in water quality between 1992 and 1993, likely driven by low water levels and windspeeds, which may have induced a pattern of increased external loading that still exists. Importantly, many of the variables which emerged as important for the TLI (water level, windspeed) will continue to vary with changing climate and are outside of the control of management. In light of this, while the current TLI target (3.9) may be feasible for Lake Rotoehu given historical observations and possible reference conditions, achieving this target consistently may be very difficult going forward due to catchment and climate pressures which have put Lake Rotoehu outside of an undisturbed reference state. This work can be built upon through future analyses which continue quantifying the changing relationships between the TLI and drivers over time, increasing the use of high-frequency buoy data to better inform water quality dynamics at shorter time scales, and testing our understanding of the drivers of water quality through predictions. The overall aim for future research would therefore be to establish water quality targets that not only align with community aspirations but are also technically feasible, especially in light of global change.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

Table of Contents

1. **Introduction** 7
2. **Methods** 8
   2.1 Water quality and driver data collection 9
   2.2 Calculation of Trophic Level Index 10
   2.3 Quantifying uncertainty in the TLI due to sampling frequency 10
   2.4 Decadal Driver Analysis 11
   2.5 Analysing drivers of TLI over time 14
3. **Results** 15
   3.1 Greater uncertainty in TLI during the 1990s due to low sampling frequency 15
   3.2 Observational trends in selected driver variables 16
   3.3 Differences in drivers by decade 18
   3.4 Ability of drivers to predict the TLI is variable over time 19
   3.5 Importance of TLI drivers changes over time 20
   3.6 Air temperature, bottom water temperature, and bottom water DRP are most often top predictors 21
   3.7 Magnitude and direction of relationships between drivers and TLI change over time 22
4. **Discussion** 23
   4.1 Potential causes of water quality declines in early 1990s 23
   4.2 Understanding key variables which influence the TLI 24
   4.3 Implications of changing driver importance 24
   4.4 Caveats of methodology 26
   4.5 Potential future work 27
   4.6 Conclusions 28
5. **Acknowledgements** 30
6. **References** 31
7. **Supplemental Material** 33

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

List of Figures

Figure 1. Time series of Trophic Level Index (TLI) from 1990 to 2021. The coloured points represent monthly TLI estimates to visualise intra-annual variability of TLI, while the black line represents the annual TLI calculated from the average of each monthly TLI component (e.g., mean across all monthly Secchi observations used to calculate TLI_{Secchi} then used to calculate the overall TLI, see Eqns. 1-5). Each hydrological year covers the period from July 1 to June 30. 8

Figure 2. Conceptual figure showing (1) how relationships between the TLI and driver variables were calculated over a fixed window (FW), and (2) how the moving window (MW) analysis quantifies changing relationships over time to (3) estimate variable importance for (4) all driver variables which show strong relationships with the TLI. 9

Figure 3. Times series of annual TLI values with estimated uncertainty (grey shading) due to limited sampling frequency from 1991 to 2000. 16

Figure 4. Time series of observations of selected driver variables by decade. Only driver variables with significant correlations (alpha < 0.05) and Pearson's r > 0.3 with monthly TLI estimates are shown. The amount of alum dosed to the lake is also shown here but was not included in the decadal analysis, only the moving window analysis. Units and descriptions of all variables can be found in

Figure 5. Pearson's correlation coefficient (r) between monthly TLI estimates and driver variables. Only correlation coefficients which were significant ($p < 0.05$) and with a value greater than 0.3 are shown. 19

Figure 6. Time series of model performance (as measured by R^2) of autoregressive models with a single driver added, with each driver denoted by a different colour. Note that each data point represents model performance over a time series of 100 monthly observations, where the x-axis represents the first observation in that 100-observation window. The model 'none' included no drivers (only the autoregressive terms). 20

Figure 7. Time series of the difference between the best performing model (shown at 0) and all other models. Colour represents the different drivers included in each model. 20

Figure 8. Time series of the relative rank of each model calculated using the absolute difference in R^2 across models. Colour represents the different driver included in each model. 21

Figure 9. Bar plots showing the percent of the simulation that each driver model sat at each rank (with 1 being the best performing model and 10 being the worst). "none" represents the model with no driver, and only autoregressive components. 22

Figure 10. Parameter coefficients for individual driver models tested, where the value of the parameter indicates the magnitude and direction of the relationship between the driver and the TLI response. 23

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

List of Tables

Table 1. Potential drivers of lake water quality and trophic state included in this study for the decadal analysis and hypothesized mechanism by which each driver could influence changes in the TLI. Because the TLI is calculated from monthly samples, variables collected on a higher frequency were either aggregated or the value on the day of sampling was used. 12

Table 2. Number of sampling events that occurred within each hydrological year at Rotoehu from 1991 to 2021 15

Table 3. Average percent of the TLI included in the 95% confidence interval 15

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

1. Introduction

The Rotorua Te Arawa lakes provide critical ecosystem services to the Bay of Plenty region, including provisioning of water for cultural practices, recreational activities, and drinking water supply. However, these valuable ecosystem services are threatened by human development in the surrounding catchment as well as exacerbating effects of climate change, such as warmer average air temperatures and increased likelihood of extreme weather events (Srinivasan et al. 2021).

The Bay of Plenty Regional Council has monitored many of the Rotorua Te Arawa lakes on an approximately monthly basis for nearly thirty years (increasing from every 2-3 months between 1990 and 2000, to monthly from 2000 to present), providing an invaluable foundation for understanding changes in water quality. One way of measuring holistic changes in water quality is through the use of a trophic state indicator, which incorporates multiple measures of water quality. One such indicator is the Trophic Level Index (TLI), which combines observations of total nitrogen (TN), total phosphorus (TP), chlorophyll *a* (chl *a*), and Secchi depth to estimate overall lake trophic state and water quality (Burns et al. 2009). TLI estimates serve as a measure of water quality over a specified period, usually one hydrological year (from July to June of the following year), and are a useful tool for management as they can be compared across lakes and across years within the same lake. Within the Rotorua Te Arawa lakes, TLI targets have been set for each lake to inform management activities aimed at improving water quality in the lakes (i.e., decreasing the observed TLI to meet the management target). However, determining the appropriate target TLI for each lake poses significant challenges. Many TLI targets in the Rotorua Te Arawa lakes were set based on historically observed water quality (typically from the 1990s) or desired water quality (Memo to Bay of Plenty Regional Council, February 2013), and may or may not be reflective of achievable water quality goals based on an estimated reference condition presuming undisturbed conditions (i.e., condition in the absence of human influence, Abell et al. 2020).

Determining an appropriate TLI for Lake Rotoehu has been particularly difficult for a number of reasons. Lake Rotoehu is a relatively shallow (max depth 13.5 m, mean depth 8.2 m), polymictic lake which regularly experiences algal blooms in recent history (Bay of Plenty Regional Council, unpublished data), and typically has exceeded the TLI target of 3.9 (Hamill 2022). Management efforts to decrease the TLI and meet the target include nutrient reduction through land management (Lake Rotoehu Action Plan, 2007), temporary aeration/destratification of the lake with the intention of preventing hypolimnetic anoxia and subsequent internal loading (McBride et al. 2015, Tempero 2015), and regular alum dosing to immobilise phosphorus from 2011-2021 (Tempero and Hamilton, 2016, Williams et al. 2023).

Paleolimnological studies of sediment cores in Lake Rotoehu indicate that increases in cyanobacterial abundance contributing to algal blooms are more recent and coincide with human development in the catchment around the time of European settlement (Picard et al. 2022). However, because the figurative translation of ‘Rotoehu’ from Te Reo Maori means ‘dirty lake’, there is speculation that murky or poor water quality and algal blooms are a naturally occurring phenomenon in the lake, pre-dating human influence. Based on an analysis of undisturbed reference conditions, Abell et al. (2020) estimated that Rotoehu’s TLI could be as low as 3.08 (± 0.45), indicating the potential for better water quality than the current TLI target of 3.9, which was set based on water quality conditions in the early 1990s (Hamill 2022).

The TLI in Lake Rotoehu has shown substantial variation from the 1990s to present, with annual TLI values regular exceeding the TLI target of 3.9 (Figure 1). The lowest observed TLI occurred in 1991 and 1992 (annual TLI of 3.8 and 3.6, respectively), after which the TLI increased to above the target of 3.9 for a period of ~20 years. The TLI then decreased to 3.7 in 2013, meeting the

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

management target, but has since fluctuated substantially, including a maximum observed TLI of 5.3 in 2019 (Figure 1). Throughout most of the 30-year monitoring period, the TLI was above the target of 3.9 with an overall mean of 4.5 across this time period (Figure 1).

Understanding the causes of variation in the TLI from year to year is critical for determining why the lake may not be meeting its target TLI value. Because of the apparent step change (i.e., increase) in TN, TP, Secchi depth and chl *a* concentrations after 1992 (Figure S1), and questions around the suitability of the current TLI of 3.9 for Rotoehu, the Bay of Plenty Regional council requested an analysis to inform a potentially more appropriate TLI target for the lake. To address this question, we conducted a series of statistical analyses to identify the key drivers of the TLI and their relative importance. Overall, these analyses will help inform the appropriateness of the current TLI target by better understanding variability in and the underlying drivers of the Trophic Level Index in Lake Rotoehu.

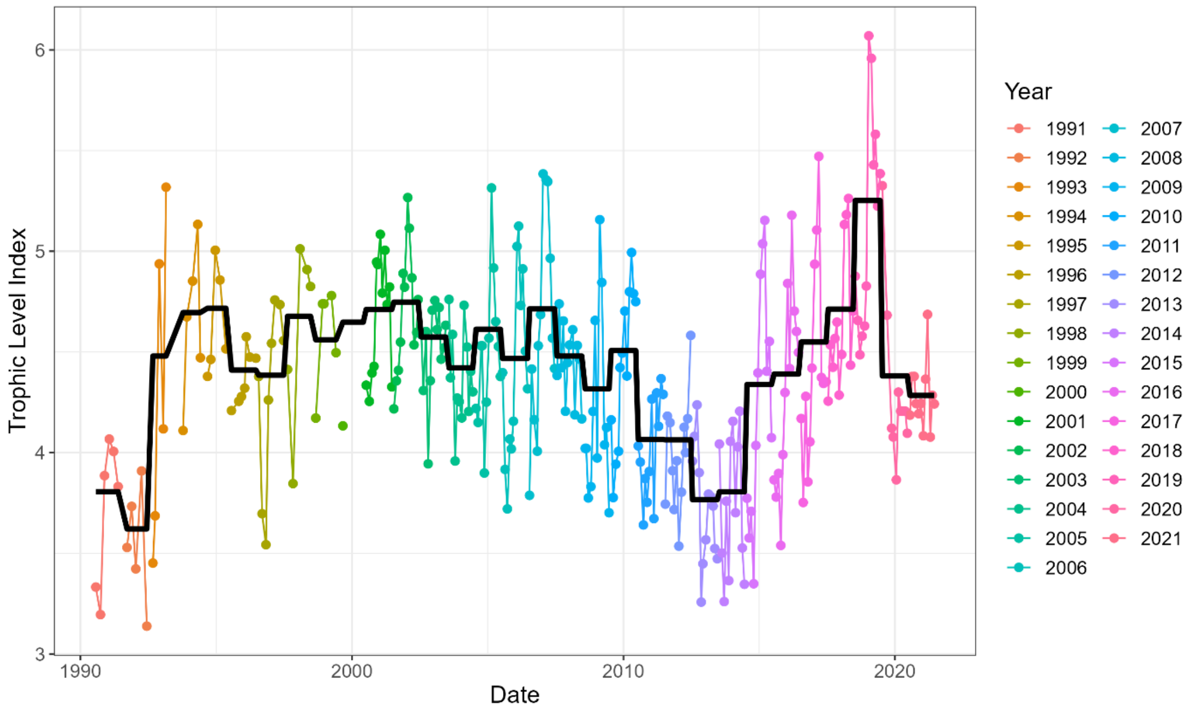

Figure 1. Time series of Trophic Level Index (TLI) from 1990 to 2021. The coloured points represent monthly TLI estimates to visualise intra-annual variability of TLI, while the black line represents the annual TLI calculated from the average of each monthly TLI component (e.g., mean across all monthly Secchi observations used to calculate TLI_{Secchi} then used to calculate the overall TLI, see Eqns. 1-5). Each hydrological year covers the period from July 1 to June 30.

2. Methods

Water quality and potential drivers of water quality were monitored by the Bay of Plenty Regional Council approximately every two to three months from 1990 to 2000, and monthly from 2001 to present. Using water quality measurements of TN, TP, chl *a*, and Secchi depth, the TLI can be estimated on a monthly or annual basis to represent changes in water quality. To better understand

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

changes in and drivers of water quality in Lake Rotoehu, we conducted a three-fold analysis which included:

1. Quantifying the level of uncertainty around TLI estimates in years with limited sampling
2. Identifying key drivers of the TLI in each decade (1990, 2000, 2010) using Pearson correlation coefficient within a fixed window (Figure 2, Step 1)
3. Analysing how the relative importance of key drivers of the TLI change over time using a moving window autoregressive linear modelling approach (Figure 2, Step 2-4)

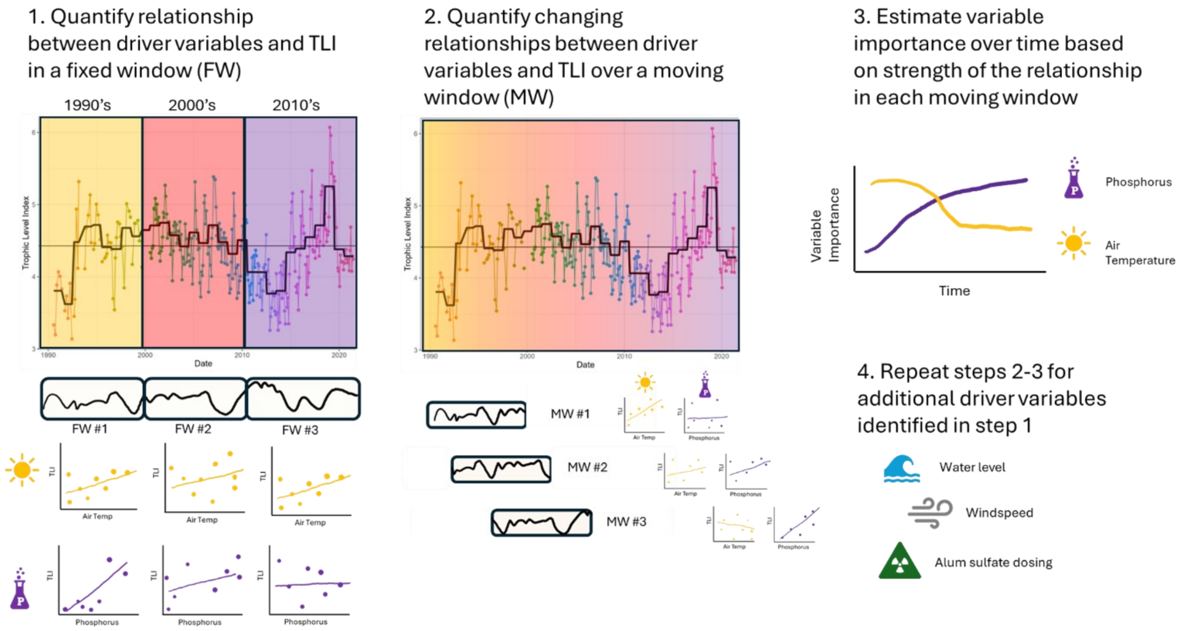

Figure 2. Conceptual figure showing (1) how relationships between the TLI and driver variables were calculated over a fixed window (FW), and (2) how the moving window (MW) analysis quantifies changing relationships over time to (3) estimate variable importance for (4) all driver variables which show strong relationships with the TLI.

2.1 Water quality and driver data collection

Water quality and physicochemical data were collected by the Bay of Plenty Regional Council (BoPRC) through a long-term routine monitoring program approximately every two to three months from 1990 to 2000, and monthly from 2001 to present. Data were collected at the deepest spot in the southern basin of the lake (identified in Land Air Water Aotearoa database as "Site 3"). Physicochemical variables, including water temperature and dissolved oxygen (DO), were measured from the surface to the bottom of the lake approximately every meter. Surface water samples were collected using an integrated tube sampler from the surface to a mid-range depth near to the thermocline, which typically ranged from 0 to 5.0 m, depending on the sampling day. Additionally, bottom water samples were collected using a discrete Van Dorn sampler, typically from 8.0 m. Water samples were analysed for both total and dissolved nutrient fractions, including dissolved inorganic nitrogen in the form of ammonium (NH_4) and nitrate-nitrite (NO_3), dissolved reactive phosphorus (DRP), and total nitrogen (TN) and total phosphorus (TP). Samples were also analysed for chl *a* as a metric of phytoplankton biomass. Secchi depth, as a measure of water clarity, was also measured.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

Further detail on field and laboratory sample collection and analyses can be found in Appendix A of Hamill (2022).

In addition to water quality data collected within the lake, we also collated data on several other potential drivers of water quality (see Table 1 for a list of drivers and potential mechanisms for influencing the TLI). Average water level was measured continuously using a pressure gauge at Kennedy Bay in Lake Rotoehu. To relate sub-daily water level estimates to monthly water quality data, we calculated the average water level over the previous month since sampling occurred. Weather data in the Lake Rotoehu region was collated from the ERA5 database (Hersbach et al. 2020). We calculated the mean, minimum, and maximum of all weather variables, except rainfall where we calculated the sum, over the previous month to relate to monthly water quality data. Using water temperature profiles from the day of sampling, we calculated thermocline depth and Schmidt stability, an indicator of the strength of stratification, using the 'RLakeAnalyzer' package (Winslow et al. 2019) implemented in the R statistical software. Aluminium sulfate (alum) was added to Lake Rotoehu by the BoPRC from 2011-2021 to decrease phosphorus loading to the lake (Tempero and Hamilton 2016). We calculated the sum of alum dosed to the lake over the previous month since each water quality sampling occurred.

While changes in land cover represent an important potential driver of changing water quality, data on land cover is only available every five years, and the available data from Land Air Water Aotearoa (LAWA) indicate minimal overall changes in land cover (Figure S2). We quantified the relationships between the TLI and the 5-yearly data from LAWA and found that all catchment variables showed a similar overall pattern, indicating that the relatively small changes across individual land use categories were not strongly related to changes in TLI (Figure S3), at least given the available data. As a result, we do not include land use categories in further analyses in this report.

2.2 Calculation of Trophic Level Index

The TLI for Lake Rotoehu was calculated for each hydrological year (e.g., the period from July 1 to June 30) from total nitrogen (TN), total phosphorus (TP), chlorophyll-*a* (chl), and Secchi depth measurements using Equations 1-5, below

$$TLI_{chl} = 2.22 + 2.54 \cdot \log(chl) \quad \text{Eq. 1}$$

$$TLI_{TN} = -3.61 + 3.01 \cdot \log(TN) \quad \text{Eq. 2}$$

$$TLI_{TP} = 0.218 + 2.92 \cdot \log(TP) \quad \text{Eq. 3}$$

$$TLI_{Secchi} = 5.56 + 2.6 \cdot \log\left(\frac{1}{Secchi} - \frac{1}{40}\right) \quad \text{Eq. 4}$$

$$TLI = \frac{TLI_{chl} + TLI_{TN} + TLI_{TP} + TLI_{Secchi}}{4} \quad \text{Eq. 5}$$

Equations 1-5 were determined following Hamill (2022). Annual TLI estimates were calculated by first taking the annual mean of each TLI component (e.g., chl, TN, TP, Secchi), and then calculating the TLI for the respective component (Eqns. 1-4), and finally calculating the overall TLI (Eq. 5). For the moving window analysis (Figure 2, Step 2), it was necessary to disaggregate the annual TLI into individual monthly TLI estimates, for which monthly TLI estimates were calculated for each available monthly sampling.

2.3 Quantifying uncertainty in the TLI due to sampling frequency

From 1990 to 2000, monthly monitoring occurred irregularly, ranging from 4 to 8 times per year, as compared to 12 times per year from 2001 onward (Table 2). During this period of irregular sampling, there was a noticeable shift in TLI values in the early 1990s, specifically between 1992 and

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

1993, when the TLI changed from 3.8 to 4.5 (Figure 1). It was uncertain whether this shift indicates significant degradation in water quality or if it is a result of inconsistencies in the sampling process. To investigate whether irregular sampling could account for this shift, we quantified the potential uncertainties in TLI values in years where fewer than 12 monthly samples were collected, using a bootstrapping approach. Using the higher frequency data set when $n = 12$ (i.e., monthly sampling between 2001-2021), we randomly sampled n monthly observations within each year, where n ranged from 4-8, which corresponded to the sampling frequencies between 1991 and 2001 (Table 2). From each of these subsamples, we re-calculated the TLI based on the subset of n samples. We bootstrapped this process 500 times for each sampling year (21 years from 2001-2021) and n sampling frequencies (ranging from 4-8). From this distribution of estimated TLI values, we calculated a 95% confidence interval for each subsampling frequency (n = 4-8) for the observed TLIs from 2001-2021. Using the estimated 95% confidence interval for each subsampling frequency, we calculated the percentage of the TLI values which were included in the confidence interval range due to subsampling frequency using the following equation

$$Percent\ Range = \frac{abs(CI_{upper} - CI_{lower})}{Observed\ Annual\ TLI} * 100 \qquad \text{Eq. 6}$$

We calculated the mean percent range across all years for each subsampling frequency to get a mean proportion of the observed TLI which occurred within the 95% confidence interval over the 2001-2021 period. To calculate the uncertainty around the TLI from 1991 to 2000, we added the corresponding percent range associated with the subsampling frequency which occurred in that year. For example, in 1991 the TLI was estimated as 3.7 but was only sampled 6 times. As a result, we added a range of 9.3% of the TLI, corresponding to a plausible TLI range in 1991 of 3.5-3.9.

2.4 Decadal Driver Analysis

To determine the strength of the relationship between the TLI and potential drivers across three decades, we conducted a correlation analysis broken down by decade, effectively quantifying the relationship between the monthly TLI and individual drivers of a fixed window of 8 years (Figure 2, Step 1). Specifically, we separated the time series into the 1990s, which included hydrological years 1991-1999; the 2000s, which included hydrological years 2000-2009; and the 2010s, which included hydrological years 2010-2019. Data from 2020 and onward were excluded from this analysis.

We calculated Pearson's correlation coefficient (r) between monthly TLI estimates and several candidate driver variables (Table 1) as a metric of the strength of the relationship between those variables and the TLI. We chose variables based on their hypothesised relationship with water quality in Lake Rotoehu (Table 1). For each weather variable, we tested multiple summary statistics (e.g., mean, minimum, maximum) over the previous sampling period. If multiple summary statistics were significant ($p < 0.05$) and $r > 0.3$, we only show results for the mean summary variable due to similar relationships between maximum and minimum weather variables and TLI. In addition to breaking down relationship by decade, we also calculated the Pearson's correlation coefficient across the whole time series (1990-2019) as a measure of the overall (rather than by decade) relationship between drivers and the TLI. Data on alum dosing was not used in the decadal driver analysis as dosing did not commence until 2011.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

Table 1. Potential drivers of lake water quality and trophic state included in this study for the decadal analysis and hypothesised mechanism by which each driver could influence changes in the TLI. Because the TLI is calculated from monthly samples, variables collected on a higher frequency were either aggregated or the value on the day of sampling was used.

Variable name (shorthand name)	Units	Hypothesised mechanism for influencing TLI	Temporal coverage and frequency	Measurement frequency	Aggregation
Air temperature (air temp)	°C	Warmer air temperatures can directly increase water temperature and phytoplankton growth, can increase stratification which can lead to internal loading	1990-present	Daily	Mean, minimum, maximum over previous month since sampling
Average water level over the previous month	Meters	Increased water levels can increase residence time and overall lake volume, leading to dilution of nutrient concentration (decreased TLI). However, higher water levels likely means increased nutrient loading from the catchment which can increase total nutrients and suspended sediment, and stimulate phytoplankton growth (increased TLI)	1990-present	Sub-daily	Mean
Bottom-water dissolved reactive phosphorus (bottom DRP)	µg/L	Higher bottom water DRP can lead to higher surface DRP if mixing occurs, which can promote phytoplankton growth	1990-present	~Monthly	NA
Bottom-water ammonium (bottom NH4)	µg/L	Higher bottom water NH4 can lead to higher surface NH4 if mixing occurs, which can promote phytoplankton growth	1990-present	~Monthly	NA
Rainfall (rain)	mm/day	Increased rain can decrease stratification and lead to internal mixing events. Through internal mixing or delivery of catchment-derived nutrients and suspended sediment this can increase within-lake nutrients, and stimulate phytoplankton growth (increased TLI). In contrast, increased rainfall could lead to	1990-present	~Monthly	Sum over previous month since sampling

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

		dilution of within-lake nutrients (decrease TLI)			
Aluminium sulfate dosed since the last sampling (alum dosed)	L/day	Increases in alum dosing should decrease available phosphorus and decrease phytoplankton growth (decrease TLI)			Sum over previous month since sampling
Schmidt stability	J/m^2	Higher Schmidt stability indicates increased strength in water column stratification, which can increase internal loading and available nutrients (increase TLI). Strong stratification events are also indicative of warm, calm periods, which could promote phytoplankton growth (especially cyanobacteria)	1990-present	Monthly	NA
Temperature at lake bottom, ~8.0 m (bottom water temp)	°C	Warmer bottom temperatures can stimulate phytoplankton growth (increase TLI). However, warmer bottom temperatures can indicate more frequent mixing and/or lower Schmidt stability, which can either reduce internal loading through oxygenation via regular mixing (decrease TLI) or allow mixing of bottom waters to introduce internal nutrients to the surface (increase TLI)	1990-present	Monthly	NA
Thermocline depth (thermo depth)	Meters	Deeper thermocline depth can indicate stronger stratification, potentially increasing internal nutrient loading. If mixing occurs following stratification, nutrients can be brought to the surface and stimulate phytoplankton growth (increase TLI)	1990-present	Monthly	NA
Windspeed	m^2/s	Increased windspeeds can increase mixing, leading to breakdown of stratification and prevention of internal nutrient loading due to anoxia (decrease TLI)	1990-present	Daily	Mean, minimum, maximum over previous month since sampling

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

2.5 Analysing drivers of TLI over time

To determine how the relationship between the TLI and key driver variables changed over time, we conducted a moving window analysis using autoregressive linear models (Figure 2, Step 2-4). Autoregressive models are used to represent dynamics which depend on the antecedent state (e.g., the TLI today depends on the TLI in the past). Using monthly TLI estimates, we fit autoregressive (AR) linear models, where models included all significant autoregressive lags and a single driver variable. Autoregressive lags are a way of representing autocorrelation in a time series (i.e., the degree of correlation between the TLI in one months and TLI in the previous month), where an autoregressive lag of one is the observation of TLI from the previous month, a lag of two is from two months prior, and so on. We used monthly TLI estimates, as opposed to annual TLI estimates, to better understand changes in water quality on a shorter time scale and disentangle seasonally varying effects. We fit models on a moving time window of n = 100 time steps to determine how the importance of drivers changed over time. We tested all potential drivers which had significant correlations greater than 0.3 from the decadal correlation analysis. We also included the amount of alum dosed to the lake over the previous month, which was not included in the decadal analysis. See Table 1 for a list of variables and potential mechanisms by which these variables could influence changes in the TLI. In summary, to estimate driver importance over a moving window, we subset the dataset down to 100 observations, fit all AR models for each driver variable, saved the model coefficients and R^2, and repeated this process moving one timestep (i.e., one month) forward in the time series. For example, the first iteration fit models on monthly observations from July 2000 to November 2008, while the second iteration on observations from August 2000 to December 2008, and so on.

For each simulation window, all significant lags were included. The significant autoregressive lags were selected using the ‘pacf()’ function in R. As a result, for each window, the significant lags may vary but all models at this timestep include the same lags. For example, for the window from July 2000 to November 2008, lags 1, 3, and 10 were all significant. As a result, all models for this window included all three autoregressive lags and a single driver variable. In contrast, from January 2001 to May 2009 only lags 1 and 10 were significant. As a result, all models fit for this window included lags 1 and 10, and a single driver variable. Because the identity of significant lags can vary with each simulation window, we also conducted our analysis only including the first lag and found no major differences in our results (Figure S4).

Following each model fit, we extracted the R^2 between the observed and predicted TLI from the model, as well as all model coefficients. Examining differences in R^2 across models provided a measure of driver variable importance, with higher R^2 of a given model indicating greater importance of that driver variable. Using differences in R^2 across drivers, we also calculated the rank across models. Because absolute differences between models varied, we used rank as a representation of relative model importance but note that some differences may not be representative of significant differences in model performance. As an alternative metric of model importance, we also examined corrected Akaike’s Information Criterion (AIC_c) and found similar relationships as using R^2 to represent model importance (Figure S5).

In addition, model coefficients which moderate the effect of each driver variable can indicate the magnitude and direction of the relationship between that driver and the TLI. Specifically, we looked at the model parameter which covered the individual driver in each model and plotted that over time (e.g., the parameter coefficient for the bottom water DRP model). We interpreted changes in parameter coefficients as a change in the magnitude and direction (e.g., negative or positive) of the relationship between drivers and TLI.

All analyses were conducted in the R statistical environment, version 4.2.2 (R Core Team 2022).

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

3. Results

3.1 Greater uncertainty in TLI during the 1990s due to low sampling frequency

From 1991 to 2000, sampling occurred infrequently, ranging from 4 to 8 months in each hydrological year (Table 2, Figure S6). Assuming that monthly monitoring data (i.e., 12 samples per year) yields reliable TLI values, decreased sampling frequency was found to lead to greater uncertainty around the calculated TLI (Table 3, Figure S7). A sampling frequency of 4 months per year led to a 13% uncertainty around the measured TLI, while sampling 8 months out of the year led to a 6.4% uncertainty around the measured TLI (Table 3). After adjusting for the sampling frequency (Table 2 and 3), TLI estimates in the 1990s uncertainty of TLI was likely in the range of 0.2-0.3 TLI units around the originally measured TLI (Figure 3). Despite this uncertainty of TLI estimates through the 1990s, a clear increase in the annual TLI was still evident from pre-1992, when the measured TLI ranged from 3.3 to 3.9, compared to the 1993-2000 period when the TLI increased sharply with an overall range over this period of 4.1 to 4.9 (Figure 3).

Table 2. Number of sampling events that occurred within each hydrological year at Rotoehu from 1991 to 2021

Hydrological Year	**# of times sampled**
1991	6
1992	5
1993	5
1994	5
1995	5
1996	7
1997	8
1998	5
1999	4
2000	5
2001-present	10-12

Table 3. Average percent of the TLI included in the 95% confidence interval

Number of sampling events	**Average percent range around observed TLI (%)**
4	13
5	11
6	9.3
7	7.8
8	6.4
10	3.8
11	2.2

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

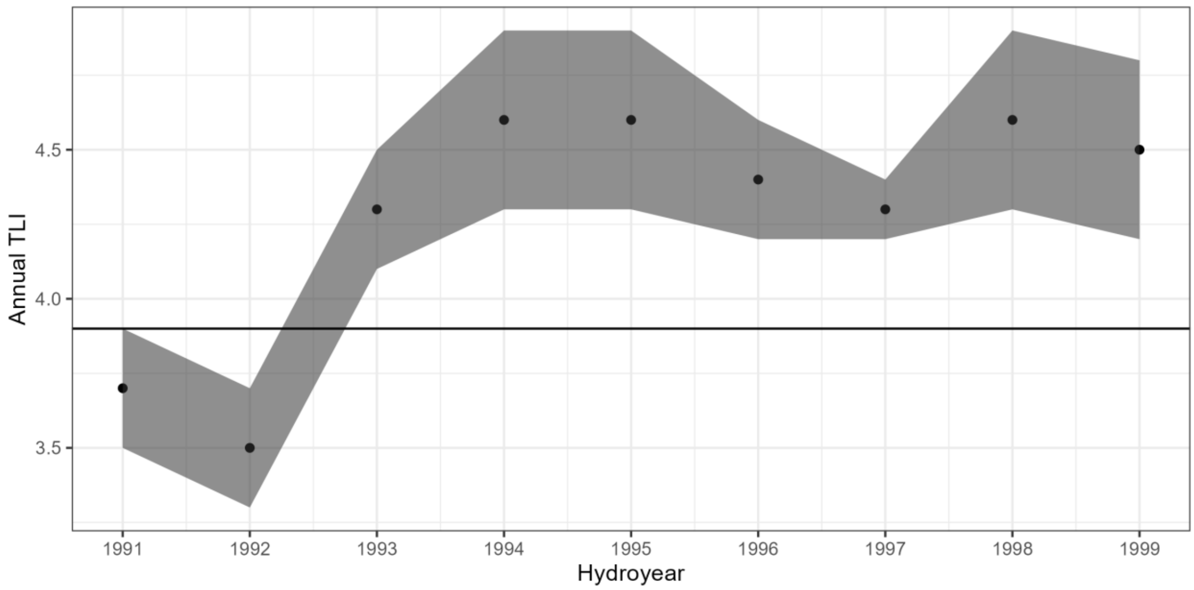

Figure 3. Times series of annual TLI values with estimated uncertainty (grey shading) due to limited sampling frequency from 1991 to 2000. The horizontal black line represent the current TLI management target (3.9)

3.2 Observational trends in selected driver variables

Across all three decades, most driver variables followed expected seasonal trends, with long-term trends present for some variables (Figure 4). Bottom water DRP and NH_4 concentrations hovered around 25 µg/L and 100-200 µg/L, respectively, for most of the time series, with occasional peaks in concentrations, which increased in frequency in later decades (Figure 4). Bottom water temperature and mean air temperature were variable throughout the year and exhibited expected seasonal trends (e.g., higher in the summer and decreasing through winter) and a slight increasing trend over the decades. Minimum windspeeds showed high variability throughout the year across all three decades. Monthly average water level showed the most dramatic differences across decades, with a major decrease in water level in the 1990s down to the lowest observation, relatively steady water levels in the 2000s, and many fluctuations in water level up to the local maximum in water level in the 2010s. Schmidt stability showed expected seasonal patterns with higher values in the summer across all three decades. Alum dosing did not begin until 2011 and remained high, with some fluctuations until 2021, when dosing ceased.

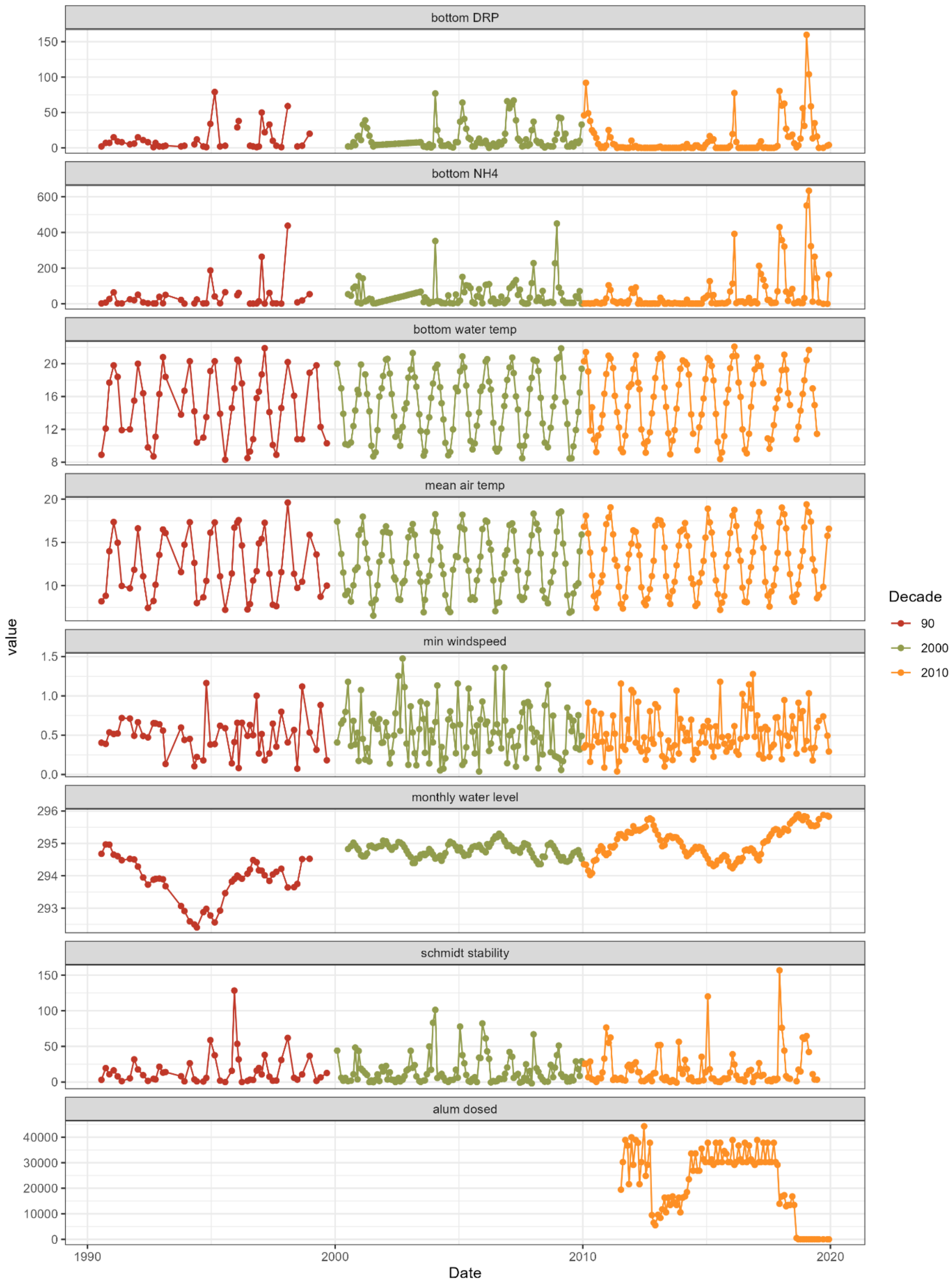

Figure 4. Time series of observations of selected driver variables by decade. Only driver variables with significant correlations (alpha < 0.05) and Pearson's r > 0.3 with monthly TLI estimates are shown. The amount of alum dosed to the lake is also shown here but was not included in the decadal analysis, only the moving window analysis. Units and descriptions of all variables can be found in Table 1.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

3.3 Differences in drivers by decade

The significant drivers of monthly TLI varied by decade (Figure 5), with some drivers which were only important in individual decades (Figure 5a, b, c). In the 1990s, the TLI was strongly correlated with three drivers which were not important in other decades: mean monthly water level, minimum windspeeds, and Schmidt stability (Figure 5a). The relationship with monthly water level and minimum windspeed were both negative, indicating that higher water levels or higher minimum windspeeds were correlated with lower TLI values. The 1990s exhibited the lowest and most variable water levels across all three decades with an overall range of 2.6m (mean = 292.4 m, 294.3 m, 294 m, respectively across decades; Figure S8), as opposed to 0.97 m and 1.9 m in the 2000s and 2010s, respectively (Figure S8). Differences in minimum windspeed across decades was less distinct than for average water level. However, minimum windspeeds were marginally lower in the 90s (average minimum windspeed = 0.5 m/s) compared to the 2000s (average minimum windspeed = 0.55 m/s) and 2010s (average minimum windspeed mean = 0.52 m/s) and were slightly less variable than other decades (i.e., more consistent in the 1990s; Figure S8). In contrast, Schmidt stability showed a positive relationship with monthly TLI in the 1990s, indicating that increases in Schmidt stability (and therefore, increasing strength of stratification) led to increases in the TLI (Figure 5a). Despite that observations of Schmidt stability were similar across decades (Figure S8), the relationship between the TLI and Schmidt stability was stronger in the 1990s than in other decades. Relationships between drivers and TLI in the 2000s were similar to the overall dataset, except that bottom water NH_4 was not strongly correlated with monthly TLI in this decade (Figure 5b). In contrast, in the 2010s, bottom water temperature was not strongly correlated with TLI but all other important variables across all decades were strongly correlated (Figure 5c). Overall, the magnitude of the correlation coefficients between the TLI and driver variables was similar across decades (Figure 5).

When analysing the full dataset across all three decades (e.g., Figure 5d), bottom water DRP, NH_4, and water temperature on the sampling day, as well as mean air temperature over the previous month were all significantly ($p < 0.05$) and highly ($r > 0.3$) correlated with monthly TLI. All four variables showed a positive relationship with TLI (Figure 5d), indicating that increases in these variables lead to higher TLI values. When looking across individual decades (e.g. Figure 5a, b, c compared to Figure 5d), only bottom water DRP and mean air temperature were consistently important in the 1990s, 2000s, and 2010s.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

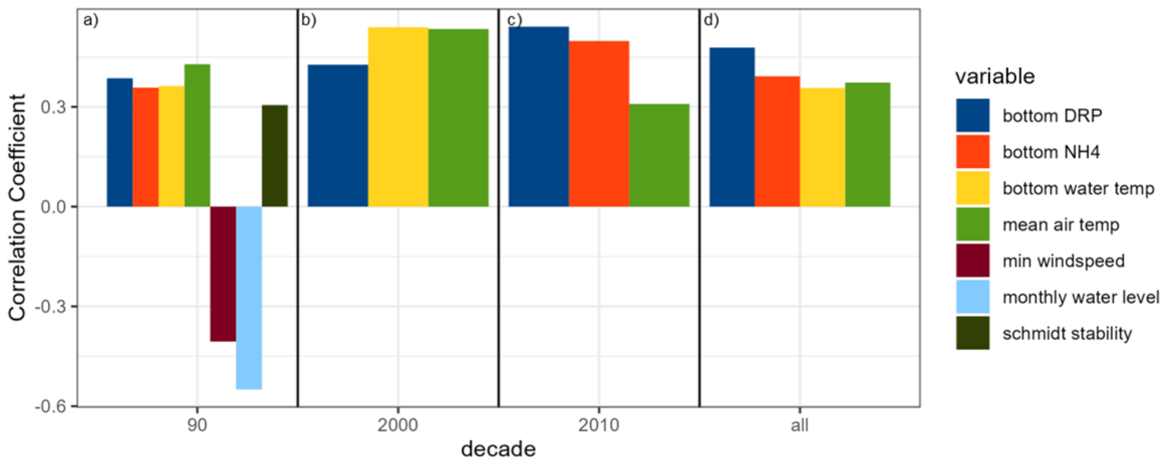

Figure 5. Pearson's correlation coefficient (r) between monthly TLI estimates and driver variables. Only correlation coefficients which were significant ($p < 0.05$) and with a value greater than 0.3 are shown.

3.4 Ability of drivers to predict the TLI is variable over time

Through our moving window analysis, we found that, across all models, our ability to predict the TLI (as measured by changes in R^2) was variable over time (Figure 6). Model performance was lowest at the beginning of our simulations (average R^2 across models of ~0.4), and increased throughout the study period, with maximum R^2 of 0.7 in models covering the time period 2011-2019. However, there were several time periods where the model performance declined sharply, such as model simulations beginning in 22 June 2006, when the maximum R^2 was 0.56, to simulations beginning in 31 October 2006, where the maximum R^2 declined to 0.45 over a period of several months, and then increased back to an R^2 of 0.57. We state the R^2 values for the highest performing model but note that all models followed a similar pattern (Figure 6). Throughout the simulations, comparison to the "none" model, which includes only the autoregressive terms (but not driver variables), indicates the relative gain in prediction provided by including a driver variable. Our results demonstrate that overall, models with a driver variable had generally had better model performance than the "none" model, as this model never performed the best of all the models (Figure 6).

The relative difference between models also varied over time, indicating that there were differences in the ability of different drivers to predict the TLI (Figure 7). In the earliest simulations, there was a relatively large range across models (minimum $R^2 = 0.36$, maximum $R^2 = 0.48$). For example, the maximum difference between the best performing model (mean air temperature) and the worse performing model (average water level) was 0.15 R^2 units and occurred during the simulation beginning in 2001. In contrast, in later simulations, model performance was relatively similar across models, with an overall range in performance of ~0.05 R^2 units. For example, in simulations beginning in 2009, model performance ranged from a minimum R^2 of 0.60 to a maximum R^2 of 0.63 (Figure 7).

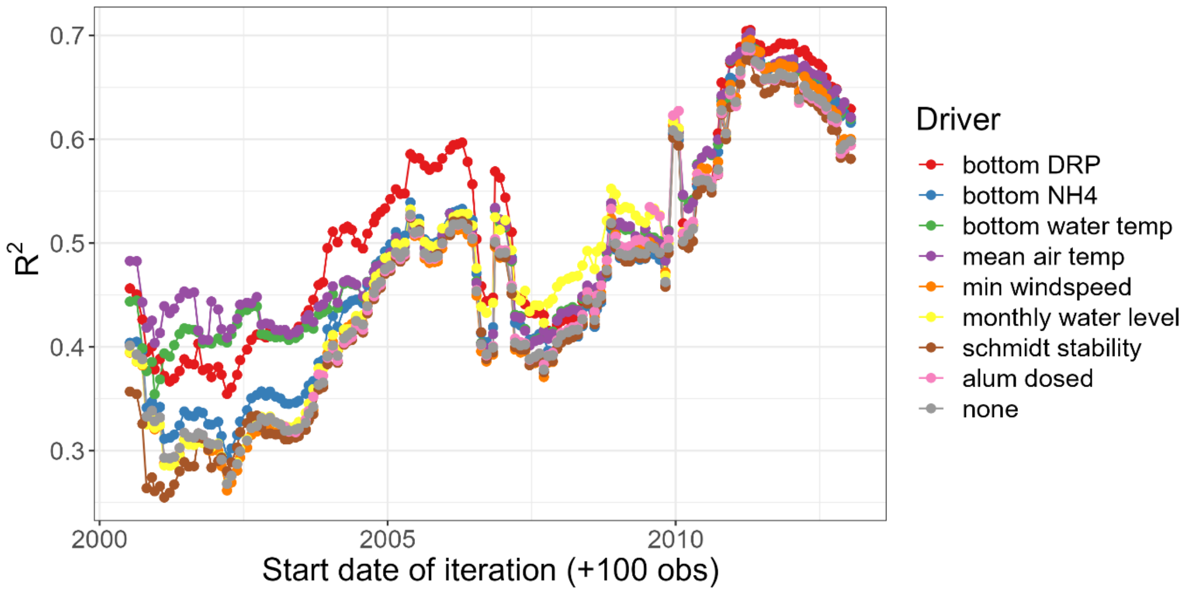

Figure 6. Time series of model performance (as measured by R^2) of autoregressive models with a single driver added, with each driver denoted by a different colour. Note that each data point represents model performance over a time series of 100 monthly observations, where the x-axis represents the first observation in that 100-observation window. The model 'none' included no drivers (only the autoregressive terms).

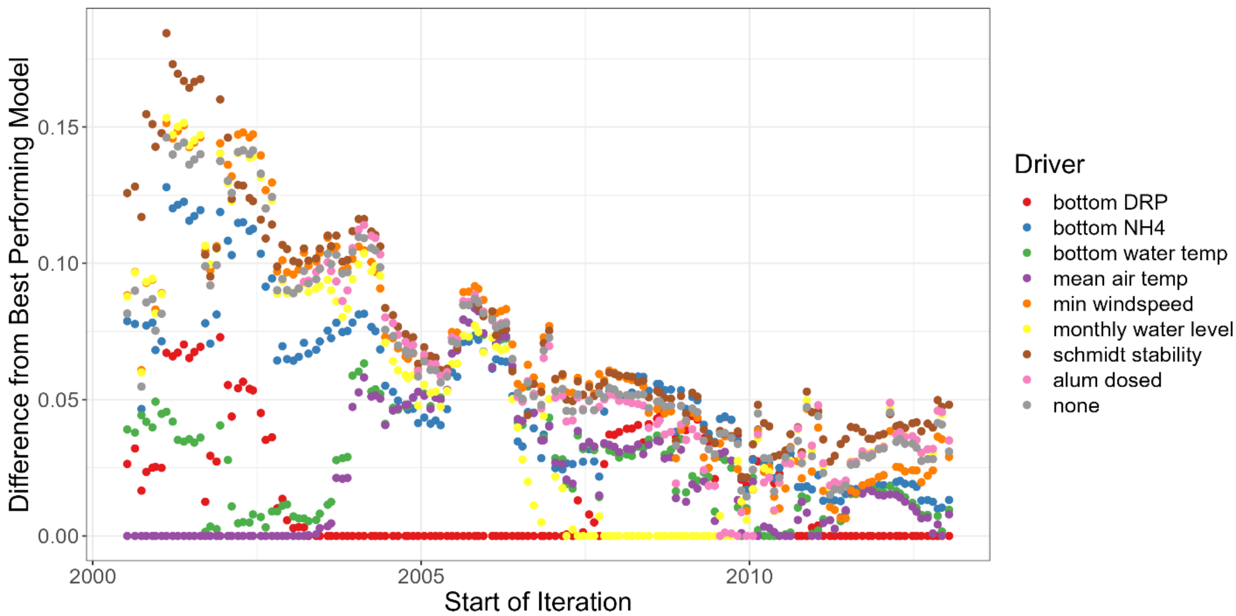

Figure 7. Time series of the difference between the best performing model (shown at 0) and all other models. Colour represents the different drivers included in each model. The y-axis represents differences in R^2 and is therefore unitless.

3.5 Importance of TLI drivers changes over time

By assigning a rank to each moving window autoregressive linear model based on the difference in their R^2 values, we found that the relative importance of each driver changed over time (Figure 8). Air temperature was most important in the beginning and end of the study period but

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

decreased to a rank of 6 (out of 9) in moving windows covering the period 2006 to 2015. Note that each point in the moving window covers a period of 8 years. For example, the moving window beginning January 2006 incorporates data from January 2006 to May 2014. In contrast, average water level was often a low-ranking model except for moving windows covering the time period from 2007 to 2018, where average water level became the best performing model. Bottom water phosphorus concentrations were almost always in the top three models, except for a brief period in the moving windows covering 2008 to 2018 where this variable became one of the worst ranking models (Figure 8). Bottom water NH_4 also showed a similar pattern in rank over time, but had lower overall rankings than bottom water DRP. Bottom water temperature was relatively consistent in its ranking over time, with a decrease to rank 5 in moving windows covering the period mid-2005 to 2015, but was rank 1 for a short time in moving windows covering the period from late 2009 until 2018. Alum dose rates to the lake was often an unimportant variable, but briefly increased in ranking in windows which began in late 2009 and early 2010 (e.g., these windows cover the full time period from late 2009 until 2018). The model 'none' which included no drivers, but only the autoregressive terms was always of rank 6 or worse, indicating that models with drivers included frequently outperformed the autoregressive terms alone.

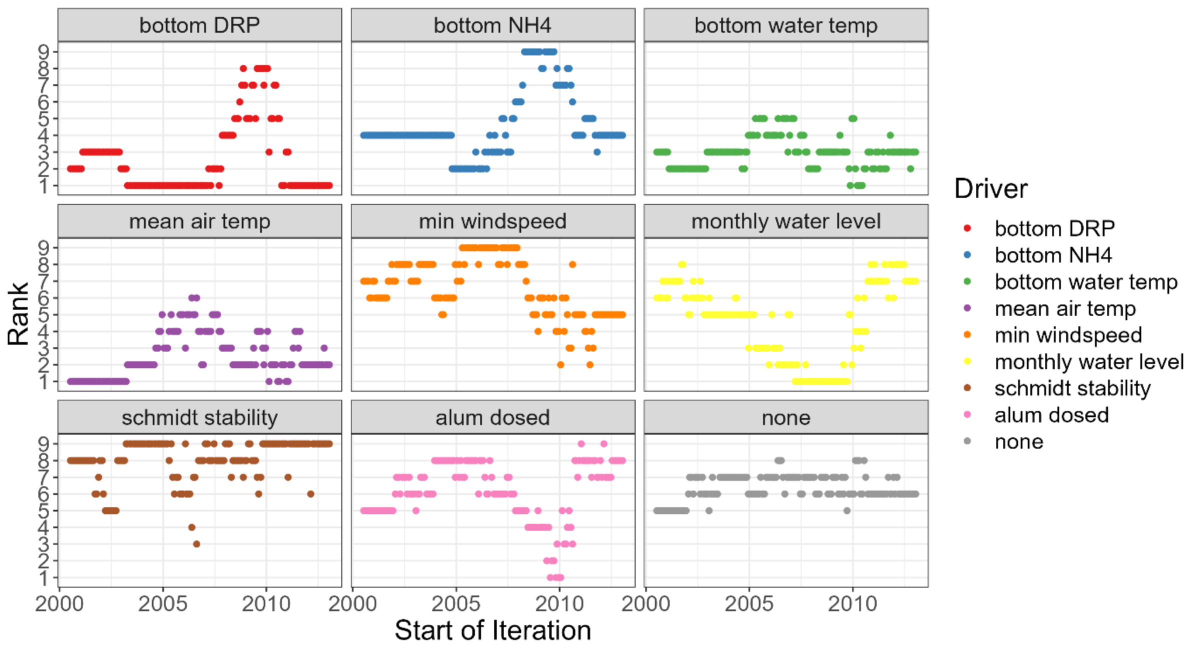

Figure 8. Time series of the relative rank of each model calculated using the absolute difference in R^2 across models. Colour represents the different driver included in each model.

3.6 Air temperature, bottom water temperature, and bottom water DRP are most often top predictors

Mean air temperature, bottom water phosphorus concentrations, and bottom water temperature were most often the best performing models (Figure 9). These three models held a rank of 3 or greater more often than any other models (Figure 9) at 75%, 78%, 78% of the time, respectively. Bottom water NH_4 concentrations was often a high-ranking model but was never of rank 1, and held medium ranks most often (50% of the simulation at rank 5). Monthly average water level was a low-ranking model most of the simulation but held a rank of 1 for a short period of time (19%).

Some models were consistently top-ranking models, while others held both top- and bottom-ranking positions at different points in the simulation period (Figure 9). For example, while bottom water DRP concentration was of rank 1 for 50% of the study period, it also ranked 7 or lower 10% of the time. In contrast, bottom water temperature was only the best ranking model 3% of the time but never decreased lower than a rank of 5. Air temperature performed similarly as rank 1 for 25% of the time but never worse than rank 6. In contrast, some drivers never performed at high ranks. Alum dosing and minimum windspeed were of a rank 5 or worse for 83% and 91% of the time, respectively. Schmidt stability was the worst performing model overall, being at rank 8 or worse 78% of the time.

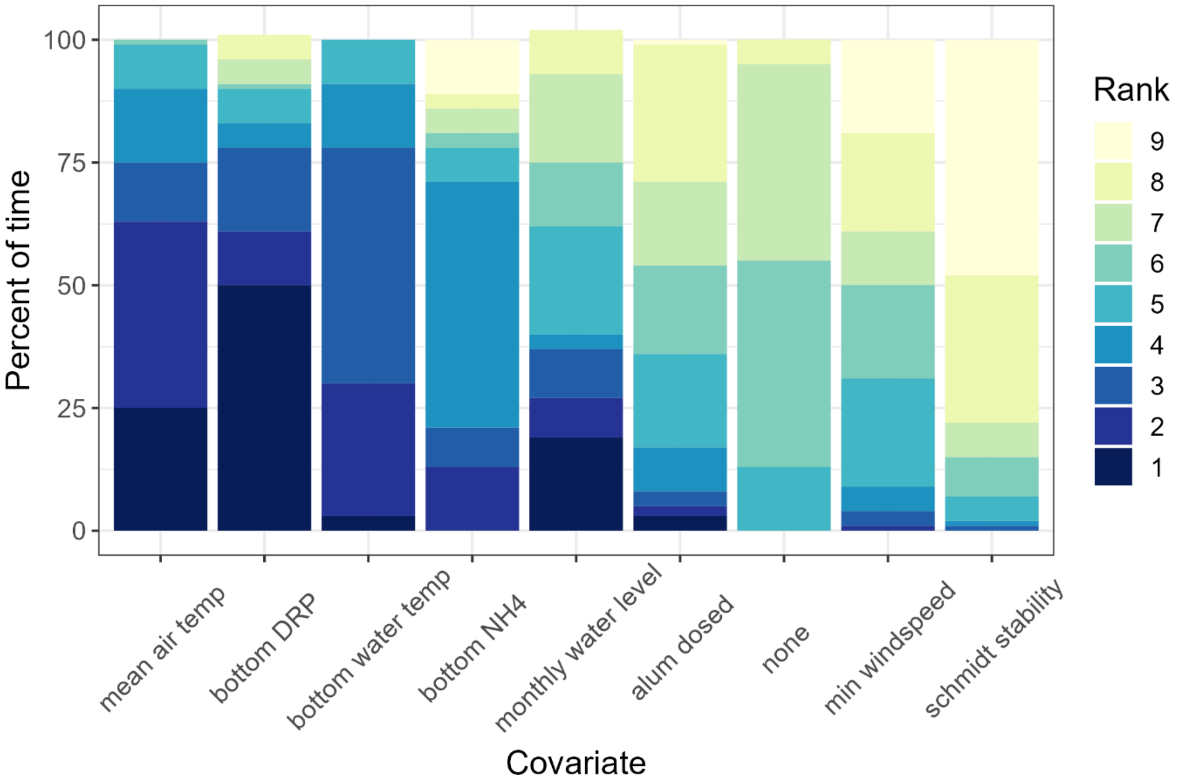

Figure 9. Bar plots showing the percent of the simulation that each driver model sat at each rank (with 1 being the best performing model and 10 being the worst). “none” represents the model with no driver, and only autoregressive components.

3.7 Magnitude and direction of relationships between drivers and TLI change over time

Parameter coefficients representing the relationship between the driver and the TLI changed over the simulation period (Figure 10). Correspondingly, changes of parameter values over time for each model (Figure 10) generally mirror the relative rank of each model over time (Figure 8). For example, parameter values for bottom water DRP concentrations were highest (max = 0.0097) in moving windows beginning in 2005, after which they decline to near zero and slightly negative (min = -0.001) during moving windows beginning around 2010 (Figure 10, bottom DRP). This corresponded to the same time period during which bottom water DRP concentrations decreased in rank (Figure 8, bottom DRP).

Some drivers (minimum windspeed, Schmidt stability, and amount of alum dosed) all showed changes from a positive to a negative relationship with the TLI over the study period. For both

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

minimum windspeed and Schmidt stability, the most negative parameter values (-0.17 and –0.001, respectively; Figure 10) also corresponded with relative increases in the rank of these variables (Figure 8). This indicates that when the relationship between these two variables is negative, the relative importance of that variable increases. For alum dosing, parameter values were primarily negative (min = -5.75 e^{-6}) until moving windows beginning in ~2007, after which there was a steep increase in parameter values until ~2010 (max = 8.4 e^{-6}) after which parameter values decreased to near zero. Similar to windspeed and Schmidt stability, the maximum parameter value corresponded to an increase in rank for the alum dosing model.

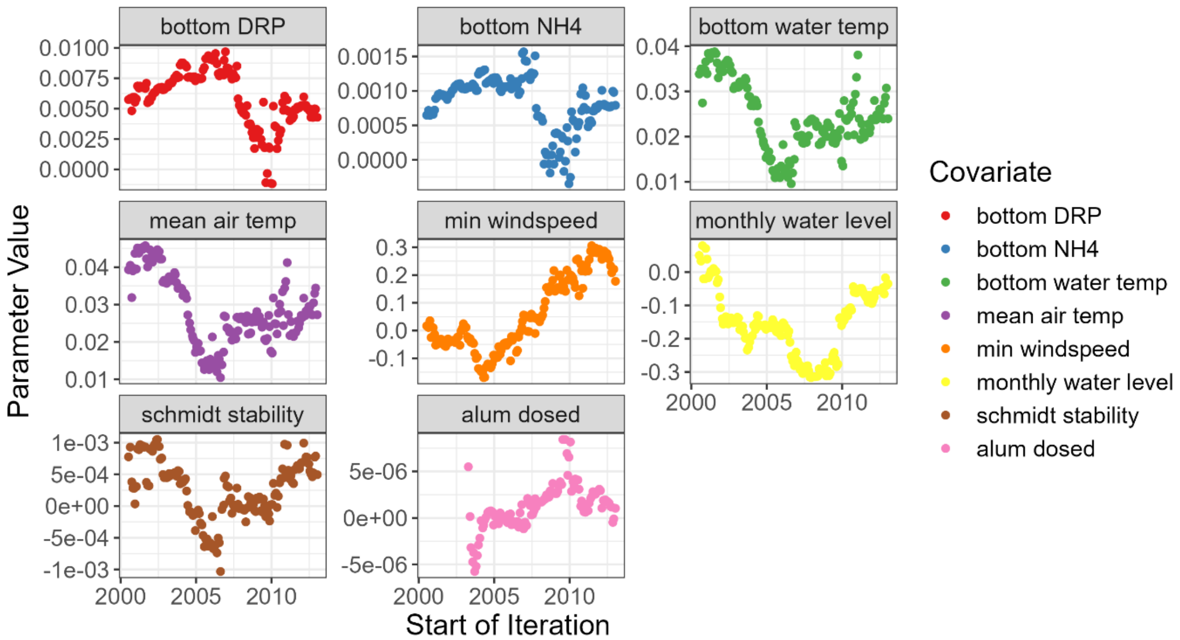

Figure 10. Parameter coefficients for individual driver models tested, where the value of the parameter indicates the magnitude and direction of the relationship between the driver and the TLI response.

4. Discussion

4.1 Potential causes of water quality declines in early 1990s

Regardless of the uncertainty around the 1990s TLI estimates, there remains a clear increase in TLI to above the TLI target (3.9) between 1992 and 1993, after which the TLI remains high for the rest of the decade. Previous examination of this trend in the Rotoehu Action Plan (2007) posits the shift in water quality from 1992 to 1993 was likely due to declines in water level and a stilling period with very low wind speeds. Our decadal correlation analysis supports this explanation, with water level and windspeed showing significant negative relationships with the TLI which were not present in other decades. Indeed, water levels and windspeeds were lower overall in the 1990s as compared to other decades (Figure 4, Figure S8). Importantly, water levels declined in the early 1990s, and exhibited the lowest recorded water level across these decades in 1994, which supports the hypothesis that the processes leading to increases in the TLI were coincident with water level declines. Mechanistically, lower windspeeds may have decreased mixing throughout the water column. In combination with lower water levels, which would have concentrated nutrients within a smaller amount of water in the hypolimnion this may have led to increased internal nutrient loading and phytoplankton production, ultimately leading to higher TLI estimates.

We also found that Schmidt stability was significantly positively related to TLI in the 1990s but not in other decades. Schmidt stability is an indicator of the strength of thermal stratification, and when it is strong enough for a long enough period, can lead to anoxic conditions and internal loading (Ladwig et al. 2021, Wagner et al. 2022). While we did not see significant differences in Schmidt stability across decades (Figure S8), the importance of this variable in the 1990s may indicate that together with lower water levels and less wind-driven mixing, the potential for internal nutrient loading was increased. We did not measure substantial internal loading during the 1990s as compared to other decades (Figure S9) but note that monitoring during this period only occurred 4-8 times per year. In polymictic lakes such as Lake Rotoehu, mixing events can occur frequently and may have occurred between sampling events, leading to an underestimation in the amount of internal loading which the lake experienced.

4.2 Understanding key variables which influence the TLI

In both the decadal correlation analysis and the moving window analysis, we found that air temperature, bottom water temperature, and bottom water nutrient concentrations (DRP and NH_4) were consistently important for determining the TLI. There are several mechanisms by which we suggest these variables have influenced the TLI. Warmer mean air temperatures can lead to increased phytoplankton production (Paerl and Huisman 2008), which can drive changes in all four components of the TLI (Chl *a*, Secchi depth, TN, TP), leading to decreased water quality overall. Bottom water temperatures may be related to the TLI through several possible mechanisms. Given that the relationship between TLI and bottom waters was positive, it is plausible that (i) warmer waters provide more thermal habitat for phytoplankton (i.e., a similar mechanism as air temperature), or (ii) warmer bottom waters indicate mixing of the water column and entrainment of hypolimnetic, nutrient-rich waters to the surface which can increase TN, TP, and subsequently phytoplankton production, and decrease Secchi depth. For mechanism (i) to occur, bottom waters would also need to have high enough levels of photosynthetically active radiation for phytoplankton to photosynthesise at these depths, which is unlikely given the average Secchi depth of 2.8 min Lake Rotoehu (although we note that a full calculation of light availability at depth is necessary to robustly rule out this mechanism). For mechanism (ii) to occur, we would expect to also see indicators of changing thermal stratification (e.g., Schmidt stability). While we did not identify a strong relationship between Schmidt stability and the TLI in this analysis, (Figure 5, Figure 9), we did observe time periods where bottom and surface water temperatures were very similar, indicating a high likelihood of water column mixing (Figure S10). In addition, it is possible that mixing events occurred which were not measured by sampling only 4-8 times per year in the 1990s. Without high-frequency data such as buoy profiling systems during this time period, it is not possible to robustly distinguish the mechanism by which bottom water temperatures influence the TLI.

Lastly, the positive relationship between bottom water nutrients and the TLI can also occur through several mechanisms. First, it may be that increased bottom water dissolved nutrients simply contribute to the total fractions of nutrients in the surface water measurements included in the TLI (if bottom water nutrients were mixed to the surface). However, we found that bottom and surface water dissolved nutrients were often distinct (Figure S9), indicating that bottom water dissolved nutrients acted as a nutrient source, fuelling phytoplankton growth. Both of these mechanisms require, first, anoxic conditions leading to the release of sediment-bound nutrients into the water column, and second, mixing of bottom waters with surface waters. Again, without high-frequency temperature and oxygen data, it is difficult to determine the timing and magnitude of anoxic and mixing events in the lake.

4.3 Implications of changing driver importance

Understanding when variables increase in importance can shed light on the underlying mechanisms for changes in the TLI. For example, monthly water level, which was most often of low

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

importance, increased to the most important model in moving windows ranging from 2010-2015 (Figure 8). This corresponded to a period with especially high water levels in Lake Rotoehu (Figure 4) and the lowest recent TLI value, observed in 2013 (Figure 1). Coupled with our finding that especially low water levels in the 1990s were correlated with major increases in the TLI, this provides support for the ability of unusually high or low water levels to impact water quality in Lake Rotoehu in a predictable way. Especially since Lake Rotoehu has no surface outlet and primarily drains through groundwater outlets within the lake bottom, lake levels will likely continue to fluctuate naturally and follow major precipitation and drought events in the future. This provides a natural opportunity to test our hypothesis on the relationship between water level and water quality in Rotoehu in future.

Around the same time period that monthly water level increased in importance, the magnitude of the relationship between bottom water DRP and NH_4 decreased (Figure 10). This weakening of the relationship between the TLI and bottom water nutrients indicates that bottom water nutrients were less important in determining overall TLI, through phytoplankton, Secchi depth, and/or total nutrients, during this time period. Over this moving window beginning in 2007 and covering the time period through 2016, we observed fewer peaks in bottom water nutrients, and a period of relatively low concentrations from 2010 to 2015 (Figure 4). While we cannot infer a direct mechanism for the decrease in bottom water nutrients, it is likely that the relationship between the TLI and bottom water nutrients weakened during this time period due to lower internal loading and subsequent mixing leading to lower phytoplankton production at the surface. In addition, this time period also corresponds to short-term lows in TLI and, therefore, improved water quality (Figure 1), supporting the idea that prolonged periods of low bottom water nutrients can contribute to improved water quality.

The most negative parameter values for Schmidt stability correspond to a short-lived increase in importance of Schmidt stability (simulations beginning in 2005-2006). A negative relationship between Schmidt stability and TLI indicates that as Schmidt stability increases, TLI decreases (i.e., better water quality). Also associated with increased Schmidt stability is the likelihood of anoxic conditions and subsequent internal loading. This may indicate that when Schmidt stability is high (as indicated by a negative relationship with TLI), its relative importance to the TLI is greater.

The relationship between alum and TLI was negative for a few years (indicating that alum was causing a decrease in P), but became positive from moving windows starting in 2006-2010 and then decreased again in moving windows starting from 2010-2012. This may indicate that alum dosing was no longer effective after the first few years of application, which can be possible in ecosystems where macrophytes are abundant (Eager 2017). When the rank for alum dosing was most positive (moving windows covering the period from 2009– 2018), the relationship between alum and the TLI was positive, indicating that increased alum dosing led to an increase in the TLI. However, we note that alum dynamics also closely followed the "none" (autoregressive only) model during this time period, which likely means that the direction of this relationship was driven primarily by autocorrelation in the ecosystem, rather than by dynamics in alum dosing alone. Overall, changes in the nature of the relationship between the TLI and alum are likely also reflective of the moving window nature of this analysis, as each point in the analysis covers a period of 8 years. The increase in importance of alum corresponds to the moving window covering a time period where all years present in that window experienced alum dosing to the lake, while the decrease in parameter values corresponds to the moving window covering a time period after 2020, when dosing ceased. This reflects the statistical nature of the analysis, where the ranking may be artificially inflated due to having more variability in alum dosing, as opposed to previous windows which experienced no dosing.

In addition to changes in the importance of specific drivers, the overall ability of all models to predict the TLI was variable over time, as indicated by changes in R^2 (Figure 6). While we do not perform a systematic analysis of the cause of this pattern, major ecosystem disturbances can sometimes cause decreased predictability by driving erratic or stochastic changes in ecosystem dynamics. For example, unprecedented decreases in water level during the 1990s, prior to the first moving window beginning in 2000, may have provided a disturbance to the ecosystem leading to low predictability in the beginning of the moving window analysis, which predictability increased as the ecosystem recovered. Overall, these results indicate that there are certain time periods where the information contained in our models (e.g., autoregressive models with a single covariate) is less able to capture the underlying ecosystem dynamics in Lake Rotoehu. This prompts more future work to examine why ecosystem dynamics may be more or less predictable over time.

4.4 Caveats of methodology

Some aspects of our study methodology may affect the interpretation of our results. First, we acknowledge that we compare multiple models using differences in R^2, and it is difficult to identify whether some of these small differences are ecologically meaningful. We also compared models using AICc, and found that relative differences in AICc were much larger (20-80) than differences in R^2 for the same model comparisons (Figure S5). Importantly, these differences were larger than the commonly used threshold of 2 AIC units difference used to distinguish among model performance in many studies, so we are reasonably confident that differences in R^2, albeit small ones, likely reflect actual differences in predictability.

We also note that while only a single driver was included within each model in the moving window analysis, some drivers may be correlated with each other, which could result in overinterpretation of the effect of a single model. For example, higher air temperatures may also occur at the same time as higher bottom water temperatures, meaning the effect of each of these models cannot be disentangled. We checked for correlations among all our tested driver variables and found that air temperature, water temperature, and Schmidt stability all had correlations > 0.5 (Table S1). Additionally, bottom water DRP and NH_4 concentrations were highly correlated ($r = 0.68$, Table S1). However, all other variables had correlations < 0.5, providing support for our interpretation of model results as independent.

In addition, there are several other factors which may also have a substantial influence on Lake Rotoehu's TLI, but which we were not able to include in this analysis due to data limitations of our time series approach. First, our analysis is limited to focusing on primarily climatological and internally-varying water quality variables (e.g., nutrients), rather than other anthropogenic factors which influence water quality. Specifically, we lacked high-resolution data on changes in catchment characteristics, especially dynamics in land cover such as the amount of nitrogen-fixing plants in the catchment, fertiliser application, pine plantation harvesting, and stock loading data, which undoubtedly influence the external nutrient budget of Lake Rotoehu. Second, Lake Rotoehu also has substantial populations of invasive macrophytes, which cycle large amounts of nitrogen and phosphorus within the lake. From 2006 to 2016, macrophytes such as Hornwort were harvested from the lake and estimated to have removed up to 4,000 kg of nitrogen and 800 kg of phosphorus, respectively, in years where harvesting took place (Scholes 2013; Horne, 2020). Harvesting has ceased since 2017 and it remains poorly understood how the remaining macrophyte populations continue to influence whole-lake nutrient cycling.

Lastly, it is important to acknowledge that Lake Rotoehu experiences substantial volcanic influence, similar to many of the lakes in the Te Arawa region. These influences include geothermal inputs at the Waitangi Soda Springs, the primary inflow to Lake Rotoehu, which carries high levels of phosphorus, silica, nitrate, ammonia, and iron, as well as water temperatures which are much higher than ambient temperatures in the lake (Eager 2017). Research conducted on the influence of the geothermal inputs at Waitangi shows that alum dosing at the stream has effectively reduced soluble phosphorus loads by 50% (Eager 2017), although in-lake phosphorus concentrations remain elevated.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

However, macrophytes may also influence the effectiveness of alum dosing in these regions due to changes in pH as a result of photosynthetic activity during the day (Eager 2017). Altogether, the complex biogeochemistry associated with geothermal inputs and photosynthetic activity of macrophytes may have important implications for soluble nutrients which drive overall water quality in Lake Rotoehu which were not robustly examined in this analysis.

4.5 Potential future work

This work provides many opportunities to build our understanding of past, current, and future water quality in Lake Rotoehu. We recommend several historical re-analysis approaches which could further improve our understanding of changing water quality in Lake Rotoehu. Specifically, this can be done on the scale of decades to centuries using additional paleolimnological analyses from sediment cores that have been collected by Geological and Nuclear Sciences (GNS) Limited and Cawthron Institute as part of the Lakes380 research programme. Note that there are currently no paleolimnological methods that measure the TLI or its components directly and existing statistical methods to infer water quality variables from paleolimnological methods can be uncertain. However, existing methods could shed additional light on when changes in ecosystem dynamics have occurred on paleoecological time scales.

In addition, on the scales of years to decades, we suggest developing studies using ecosystem models to hindcast water quality. These methods would leverage long-term historical monitoring datasets and high-frequency monitoring data to calibrate and initialise ecosystem models to explore when and how water quality in the lake has changed. This would require careful parametrisation of relevant processes that simulate the sediment-water interaction to ensure that triggers and changes in internal nutrient loading are adequately captured.

In addition, several opportunities exist for expanding our understanding of the current water quality dynamics in lakes. We suggest a more thorough examination of the existence of thresholds in the relationships between water quality and drivers, and identifying the likelihood of regime shifts as lake ecosystems are pushed outside of the range of historical conditions (Dodds et al. 2010). For example, methods such as piecewise regression (e.g., Toms and Lesperance 2003) could be useful in determining breakpoints which indicate when relationship between the TLI and a specific driver changes. This would necessitate better utilising high frequency data to improve understanding of lake dynamics and drivers both within Lake Rotoehu, as well as across other lakes in the Rotorua region and across the world to understand commonalities and differences in the drivers of water quality. Future analyses which can extend our current findings to determine what factors influence the relative importance of variables, and identifying thresholds in these relationships, would be incredibly valuable in developing pre-emptive models which can identify when drivers are important at what time.

Lastly, we suggest more emphasis on developing forecasts of future water quality to help test our understanding of the importance of drivers for changing water quality and develop decision support tools for future management decisions. For example, forecasts of lake level, water temperature, fish migration timing, and algal bloom risk which were co-developed with stakeholders in five case studies in Europe and Australia show potential for providing advance notice of appropriate management actions (Jackson-Blake et al. 2022). Forecasts can be developed by instantiating our understanding of the relationship between, for example, air temperature and phytoplankton, and testing our predictions using real-time buoy data. This can be done on the scale of days to years and could include predictions of how the relative importance of driver variables might change over time. In addition, projections of water quality further into the future, on the scale of decades, may also help inform management decisions by providing a scenario-based projection of how water quality will change as, for example, air temperatures increase.

4.6 Conclusions

This study provides several insights into the current TLI target for Lake Rotoehu of 3.9 based on water quality in the early 1990s. Despite the appreciable uncertainty around TLI estimates from the 1990s due to a low sampling frequency, we are confident there was a shift in water quality between 1992 and 1993. Decreases in water level and wind stilling (Figure 4) were likely drivers of this shift, which could have induced substantial internal loading that has increased over several decades (Figure S9). A lack of systematically collected water quality data prior to 1991 limits our ability to conclude whether 1991 and 1992 were typical hydrological years for Lake Rotoehu or if they represent unusually good water quality relative to historical conditions prior to 1991. External drivers of water quality (e.g., climate) during 1991-1992 were not unusual for Lake Rotoehu compared to later decades (Figure 4, Figure S8). Regardless, the decrease in water quality remained persistent through the 2000s and part of the 2010s, with a mean observed TLI over this period of 4.4, which is above the current target of 3.9.

In contrast, estimates of reference conditions in Lake Rotoehu (e.g., without human influence) range from 2.64-3.55, with the upper limit reaching close to the TLI target. We also recognise that aiming for a TLI target close to a reference condition could be challenging and perhaps not advisable, given the current pressures on the lake from the catchment and climate drivers. Significant changes in Lake Rotoehu's catchment over the past several decades might impede achieving management targets that represent pre-human disturbance conditions. Alterations in land management practices in agricultural areas often lead to substantial enrichment of benthic sediment nutrient stores, leading to sustained internal loading for prolonged periods. In such cases, setting a management target which corresponds to pre-human disturbance conditions may be highly ambitious. Moreover, climate change, acting as a continuous 'shifting driver,' can be expected to alter the achievable conditions that environmental managers can aim for in numerous lake ecosystems.

Additionally, paleolimnological analyses of sediment cores in Lake Rotoehu pre-dating Māori settlement indicate that although bloom-forming cyanobacteria have been present in the lake for centuries, increases in the relative abundance of cyanobacteria coincided with human-induced land use change (Picard et al. 2022). Together, these results suggest that improved water quality in Lake Rotoehu may have existed historically (supported by Abell et al. 2020) but reaching target TLI values may not currently be feasible in the short term due to the cycling of internal nutrient loading which can last for decades or centuries (Shatwell and Kohler 2019).

Overall, we interpret this evidence to indicate that achieving a TLI of 3.9 is within the realm of possibility in this lake, but consistently meeting this target will likely provide very difficult even with substantial catchment interventions which necessitate years of lake recovery. Specifically, we believe it may be possible that a disturbance caused by low lake levels and stilling in the 1990s induced a cycle of internal loading in the lake that can take decades to reverse as found in other lakes (Shatwell and Kohler 2019, Hanson et al. 2023). In addition, we emphasise that many of the drivers identified as important in this study (e.g., water level, wind speed, air and water temperature) are outside of the control of management activities and will continue to change as climate change exacerbates. Overall, the large variability in the TLI which has been observed over the past 30 years will likely continue given the multi-faceted relationships between the TLI and increasingly variable climatic and anthropogenic drivers.

Altogether, our work demonstrates the importance of examining the strength of relationships between water quality and water quality drivers over time. We demonstrate that while bottom water nutrients (DRP and NH_4) and air temperature predominantly drive water quality, these relationships can weaken under certain conditions. In addition, we identify the importance of some drivers (e.g.,

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

water level and windspeed) as increasing when these variables are at extreme values, highlighting the value of continual monitoring of many potential drivers of water quality and emphasizing that extreme conditions may alter the way that lakes respond to environmental drivers. Analyses like this which integrate the long-term historical monitoring efforts with increasingly available high-frequency buoy data both at Rotoehu, as well as other lakes across New Zealand, will continue to help improve our understanding of variability in and drivers of water quality.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

5. Acknowledgements

We thank Maggie Armstrong and Charles Lee for reviewing this report. We thank the Bay of Plenty Regional Council for funding and providing the data used in our analysis.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

6. References

Bruere, A. 2023. Water Level: Lake Rotoma and Rotoehu. Bay of Plenty Regional Council Presentation May 2023. https://atlas.boprc.govt.nz/api/v1/edms/document/A4383149/content

Dodds, Walter K., William H. Clements, Keith Gido, Robert H. Hilderbrand, and Ryan S. King. "Thresholds, Breakpoints, and Nonlinearity in Freshwaters as Related to Management." *Journal of the North American Benthological Society* 29, no. 3 (September 2010): 988–97. https://doi.org/10.1899/09-148.1.

Eager, C. A. 2017. Biogeochemical characterisation of an alum dosed stream: implications for phosphate cycling in Lake Rotoehu (Masters thesis, The University of Waikato).

Hamill, K. 2022. Trophic Level Index Review of targets and variability for Rotorua Lakes. RiverLake Report prepared for Bay of Plenty Regional Council.

Hanson, Paul C., Robert Ladwig, C. Buelo, E. A. Albright, A. D. Delany, and C. C. Carey. 2023. Legacy Phosphorus and Ecosystem Memory Control Future Water Quality in a Eutrophic Lake. JGR Biogeosciences. 128, e2023JG007620. https://doi.org/10.1029/2023JG007620

Hersbach, Hans, Bill Bell, Paul Berrisford, Shoji Hirahara, András Horányi, Joaquín Muñoz-Sabater, Julien Nicolas, et al. 2020. The ERA5 Global Reanalysis. *Quarterly Journal of the Royal Meteorological Society* 146, no. 730: 1999–2049. https://doi.org/10.1002/qj.3803.

Horne, Hannah. Weed harvesting in the Rotorua Te Arawa Lakes 2006-Present. Prepared for the Bay of Plenty Regional Council. May 2020.

Jackson-Blake, Leah A., François Clayer, Elvira de Eyto, Andrew S. French, María Dolores Frías, Daniel Mercado-Bettín, Tadhg Moore, et al. "Opportunities for Seasonal Forecasting to Support Water Management Outside the Tropics." *Hydrology and Earth System Sciences* 26, no. 5 (March 14, 2022): 1389–1406. https://doi.org/10.5194/hess-26-1389-2022.

Ladwig, R., Hanson, P. C., Dugan, H. A., Carey, C. C., Zhang, Y., Shu, L., Duffy, C. J., and Cobourn, K. M. 2021. Lake thermal structure drives interannual variability in summer anoxia dynamics in a eutrophic lake over 37 years, Hydrol. Earth Syst. Sci., 25, 1009–1032, https://doi.org/10.5194/hess-25-1009-2021.

Lake Rotoehu Action Plan. 2007. The Rotorua Lakes Protection and Restoration Action Programme. A Bay of Plenty Regional Council, Rotorua District Council, and Te Arawa Lakes Trust joint project. Environmental Publication 2007/19

McBride CG, Tempero GW, Hamilton DP, Cutting BT, Muraoka K, Duggan IC and Gibbs M. 2015. Ecological Effects of Artificial Mixing in Lake Rotoehu. Environmental Research Institute Report No. 59. The University of Waikato, Hamilton. 70 pp.

Memorandum to Bay of Plenty Regional Council. Trophic Level Indices for Rotorua Lakes. 14 February 2013.

Paerl, Hans W., and Jef Huisman. 2008. Blooms Like It Hot. *Science* 320, no. 5872: 57–58. https://doi.org/10.1126/science.1155398.

Picard, Maïlys, Xavier Pochon, Javier Atalah, John K. Pearman, Andrew Rees, Jamie D. Howarth, Christopher M. Moy, et al. "Using Metabarcoding and Droplet Digital PCR to Investigate

Drivers of Historical Shifts in Cyanobacteria from Six Contrasting Lakes.” *Scientific Reports* 12, no. 1 (July 27, 2022): 12810. https://doi.org/10.1038/s41598-022-14216-8.

Scholes, P. File Note: Hornwort harvest Lake Rotoehu. Prepared for Bay of Plenty Regional Council. 20 March. 2013.

Shatwell, T., and J. Köhler. 2019. Decreased Nitrogen Loading Controls Summer Cyanobacterial Blooms without Promoting Nitrogen-Fixing Taxa: Long-Term Response of a Shallow Lake. *Limnology and Oceanography* 64, no. S1: S166–78. https://doi.org/10.1002/lno.11002.

Srinivasan, R., T. Carey-Smith, and G. Macara. 2021. Trend Analysis on Frequency of New Zealand Climate Extremes. *Weather and Climate* 41, no. 1: 18–29. https://doi.org/10.2307/27127987.

Tempero, G.W. 2015. Ecological Monitoring of Artificial Destratification Effects in Lake Rotoehu: 20142015. ERI Report No. 57. Client report prepared for Bay of Plenty Regional Council. Environmental Research Institute, Faculty of Science and Engineering, University of Waikato, Hamilton, New Zealand. 32 pp.

Tempero, G.W. and Hamilton, D.P. 2016. Lake Rotorua and Lake Rotoehu: Total and Noncrystalline Aluminium Content in Bottom Sediments. ERI Report No. 89. Client report prepared for Bay of Plenty Regional Council. Environmental Research Institute, Faculty of Science and Engineering, University of Waikato, Hamilton, New Zealand. 44 pp.

Toms, Judith D., and Mary L. Lesperance. “Piecewise Regression: A Tool for Identifying Ecological Thresholds.” *Ecology* 84, no. 8 (2003): 2034–41. https://doi.org/10.1890/02-0472.

Wagner, N. D., Osburn, F. S., Robbins, C. J., Ernst, M. R., Owens, J., Powers, S. M., & Scott, J. T. 2023. Lake stability and anoxia dynamics revealed from high frequency vertical profiling in a eutrophic polymictic reservoir. *Inland Waters*, *13*(2), 167-181.

Williams, E., A. Woodward, and K. Ratana. 2023. Resource consent RC65966 – Alum dosing in Rotoehu. Independent Technical Peer Review Client report prepared for Ngati Tamateatutahi-Ngati Kawiti with funding from Bay of Plenty Regional Council. NIWA

Winslow L, Read J, Woolway R, Brentrup J, Leach T, Zwart J, Albers S, Collinge D. 2019. rLakeAnalyzer: Lake Physics Tools. R package version 1.11.4.1. https://CRAN.R-project.org/package=rLakeAnalyzer

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

7. Supplemental Material

Tables: 1
Figures: 10

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

Table S1. Correlation coefficients across all variables included in the moving window analysis. Correlations > |0.5| are highlighted in orange.

	Monthly TLI	Bottom water DRP (ug/L)	Bottom water NH_4 (ug/L)	Bottom water temp (C)	Mean air temperature (C)	Minimum windspeed	Average water level (m)	Schmidt stability (J/m^2)	Sum of Aluminium sulfate dosed
tli_monthly	1								
bottom_DRP_ugL	0.51	1							
bottom_NH4_ugL	0.43	0.68	1						
Bottom water temp	0.27	0.43	0.39	1					
air_temp_mean	0.29	0.48	0.43	0.97	1				
windspeed_min	-0.02	-0.08	-0.05	-0.16	-0.16	1			
monthly_avg_level_m	-0.04	0	0.1	-0.12	-0.11	0.09	1		
schmidt_stability	0.09	0.47	0.45	0.52	0.6	-0.06	0.08	1	
sum_alum	-0.17	-0.24	-0.06	0.01	-0.02	0.05	-0.03	-0.09	1

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

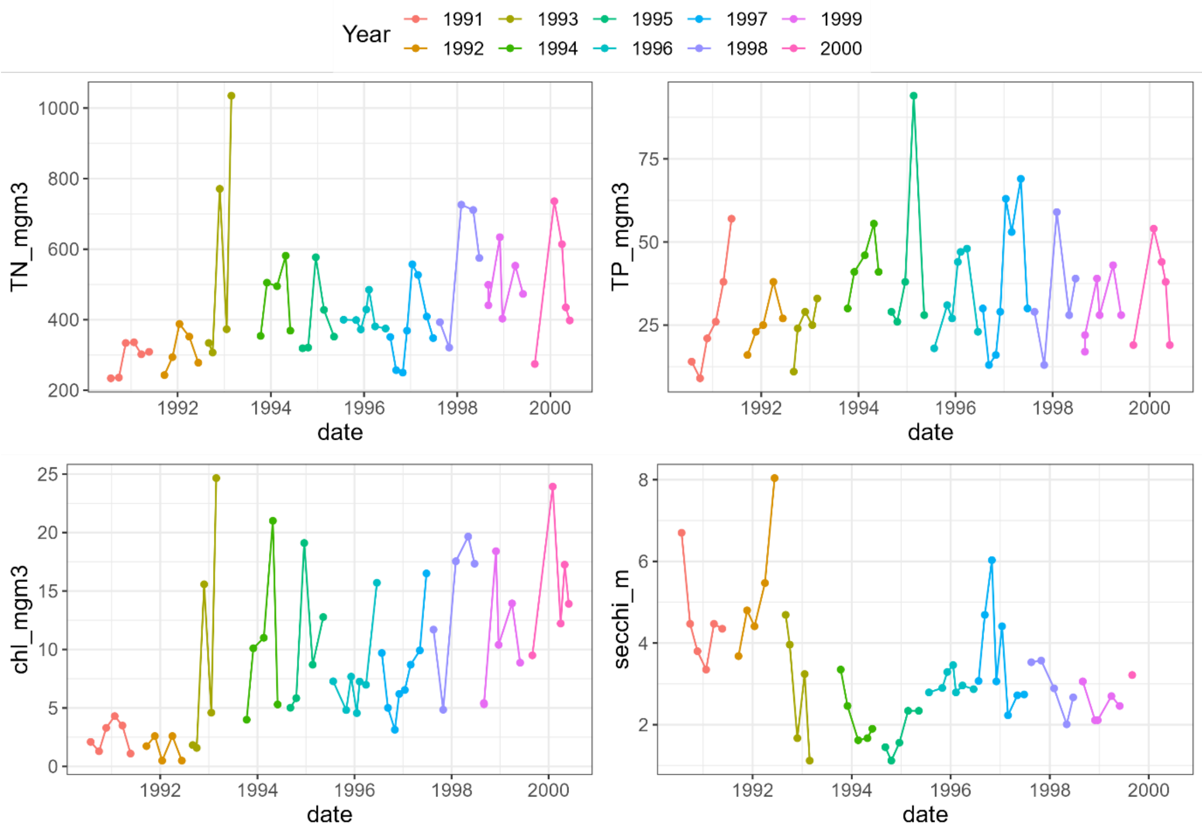

Figure S1. Time series of TLI-contributing variables (TN, TP, chl-a, Secchi depth) during the period of 1990-2000.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

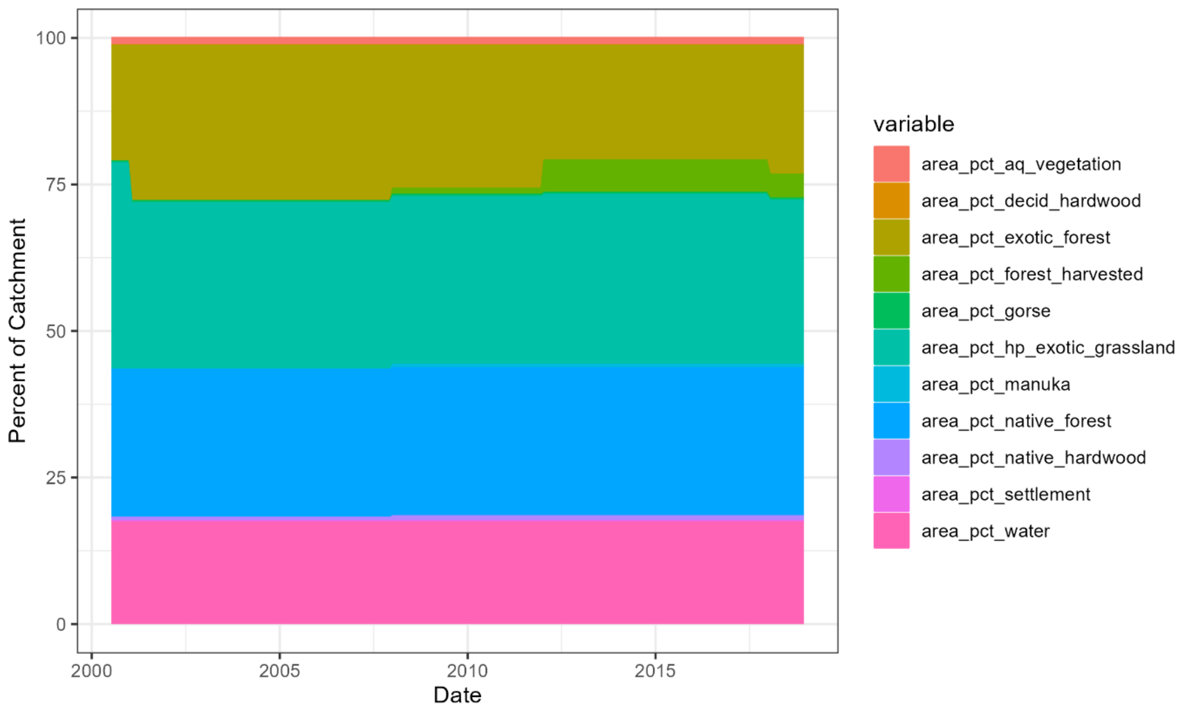

Figure S2. Time series of land cover data available for Lake Rotoehu's catchment from Land Water Air Aotearoa (LAWA)

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

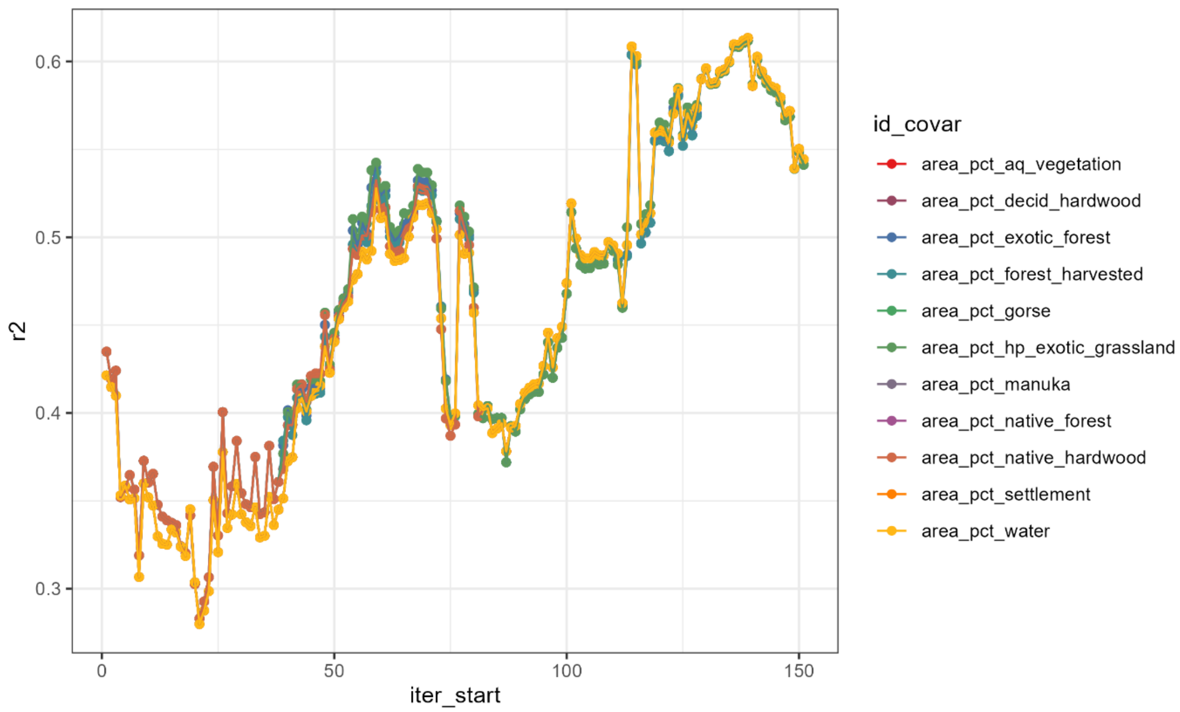

Figure S3. Time series of model performance (R^2) of autoregressive models which include a single driver (shown by different colours), where all drivers shown here are metrics of land cover

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

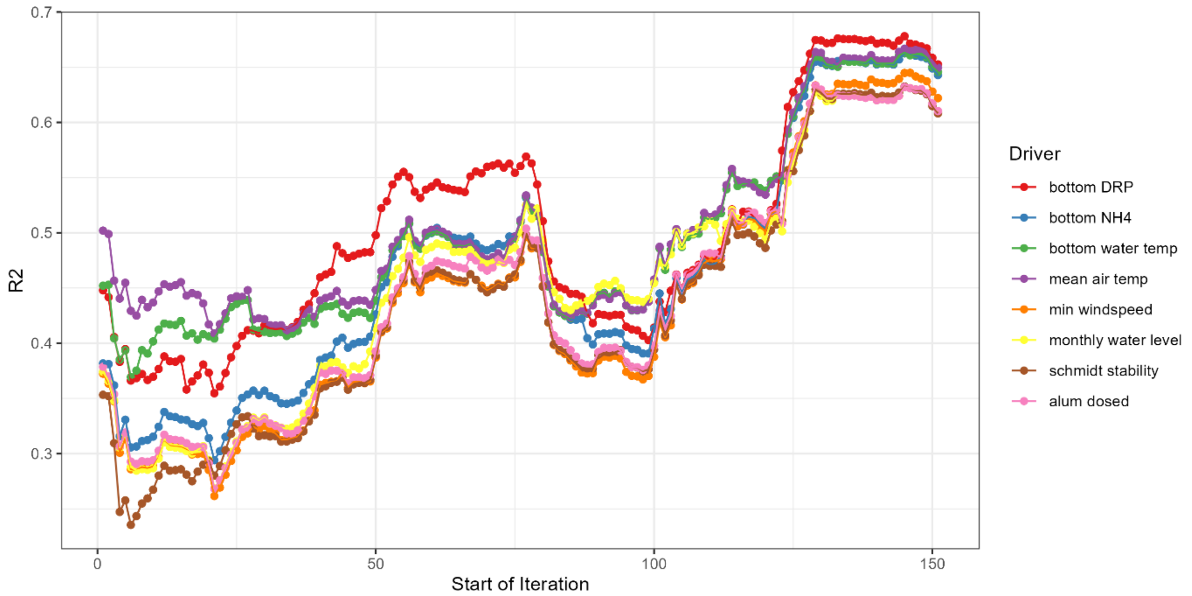

Figure S4. Time series of model performance using only the first autoregressive lag (lag = 1) for all driver variables

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

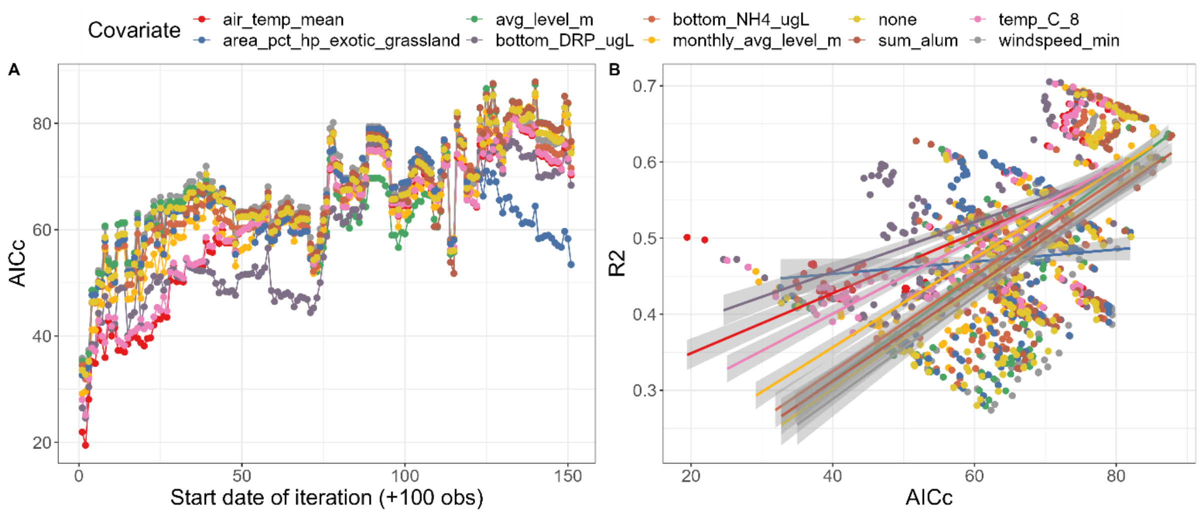

Figure S5. A) Time series of model performance using corrected Akaike's Information Criterion (AICc) of autoregressive models which include a single driver (shown by different colours), and B) a comparison of model performance between AICc and R^2. For both panels, colour represents the driver included in the model

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

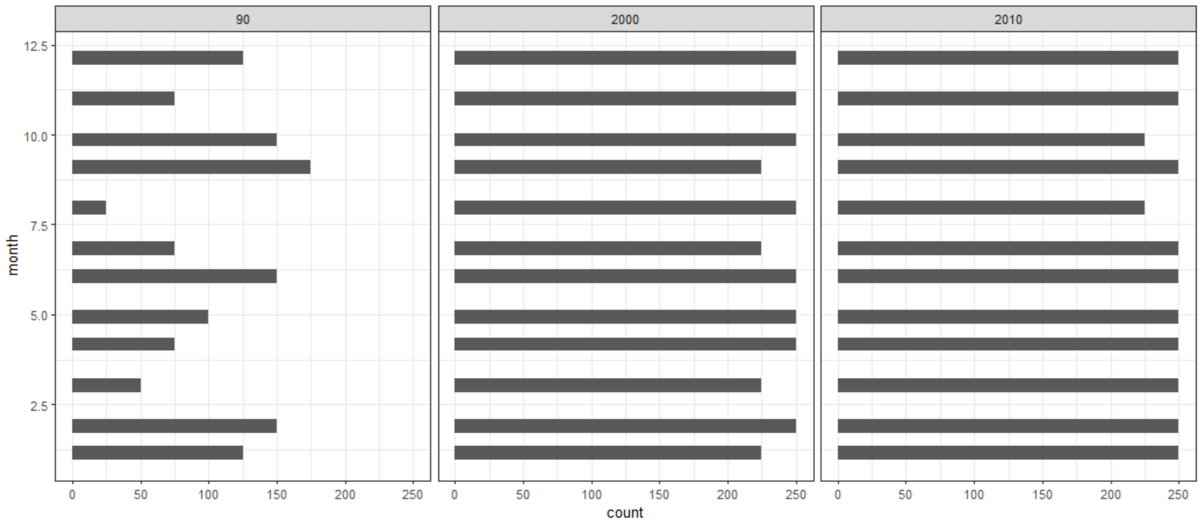

Figure S6. Histogram of sampling months in each decade from the Bay of Plenty Regional Council's long-term monitoring program in Lake Rotoehu.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

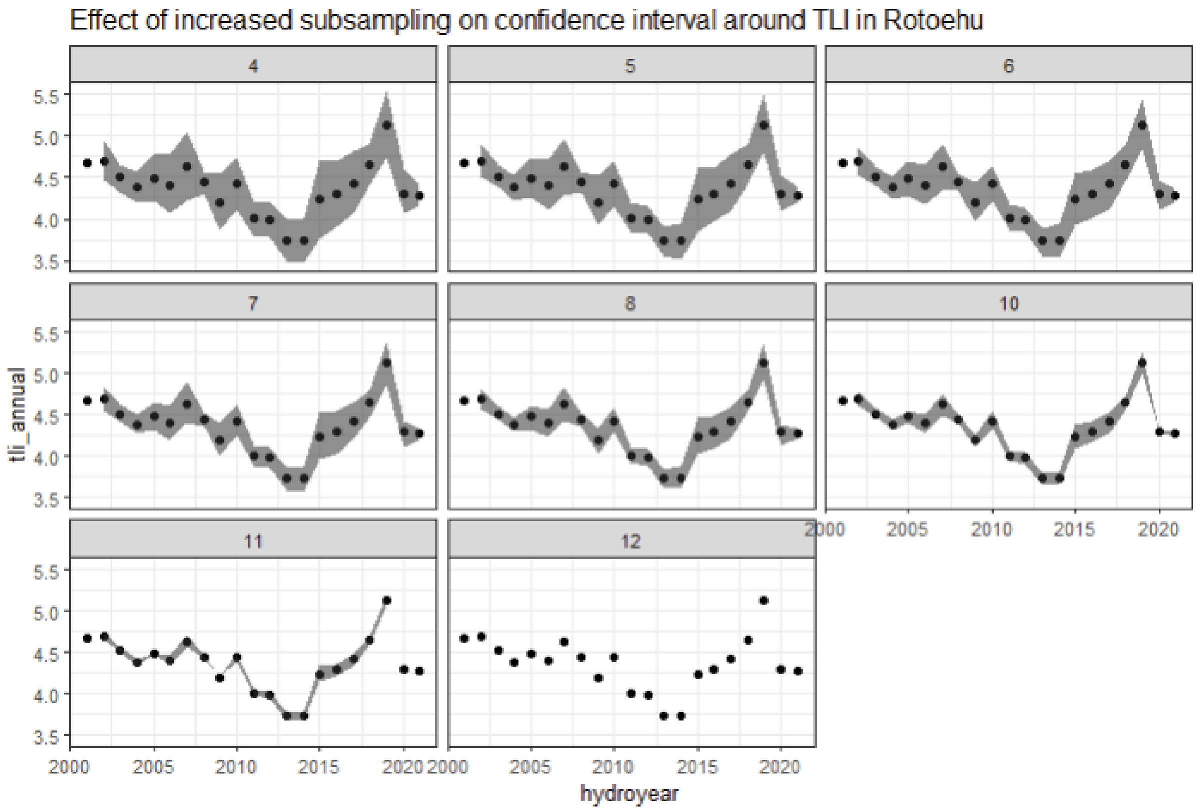

Figure S7. Effect of subsampling from the available observations in each hydrological year on the calculated 95% confidence interval, where each panel represents the number of subsamples taken to calculate the estimated TLI. The grey shaded area represents the 95% confidence interval from the bootstrapped distribution which sampled n times from all available observations each year (n = 12 total) and the black dots represent the actual TLI calculated for that year.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

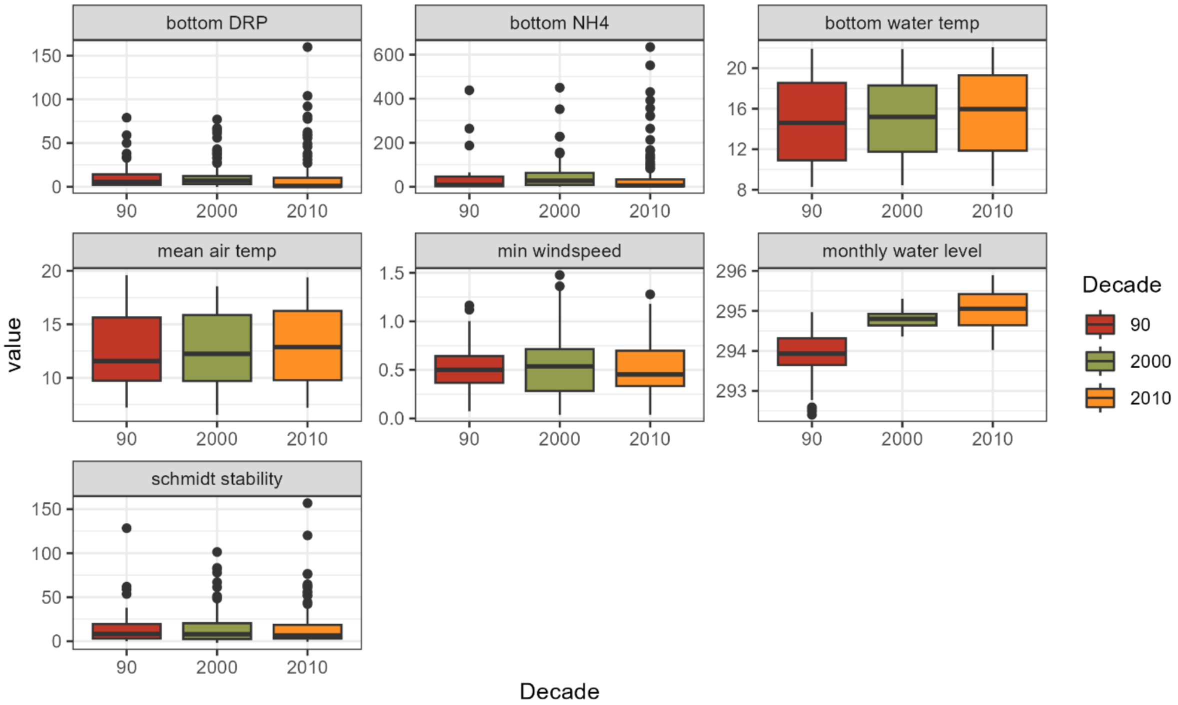

Figure S8. Boxplots of observations of selected driver variables broken down by decade. Only driver variables with significant correlations (alpha < 0.05) and Pearson's r > 0.3 are shown.

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

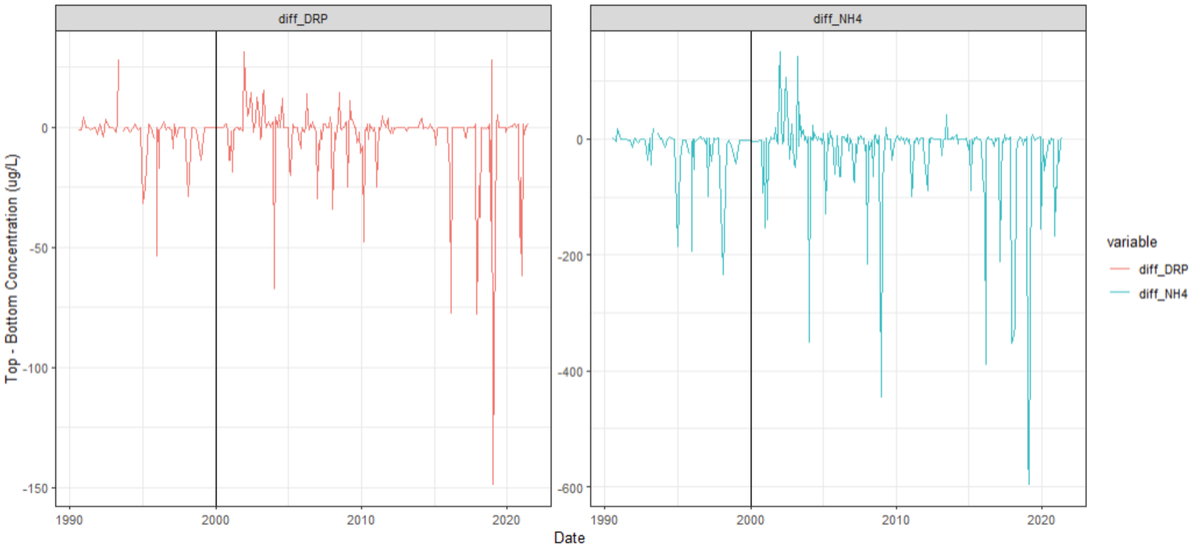

Figure S9. Difference between surface and bottom nutrient samples as an indication of the occurrence of internal nutrient loading. The black vertical line represents hydrological year 2000

ISSN 2463-6029 (Print)
ISSN 2350-3432 (Online)

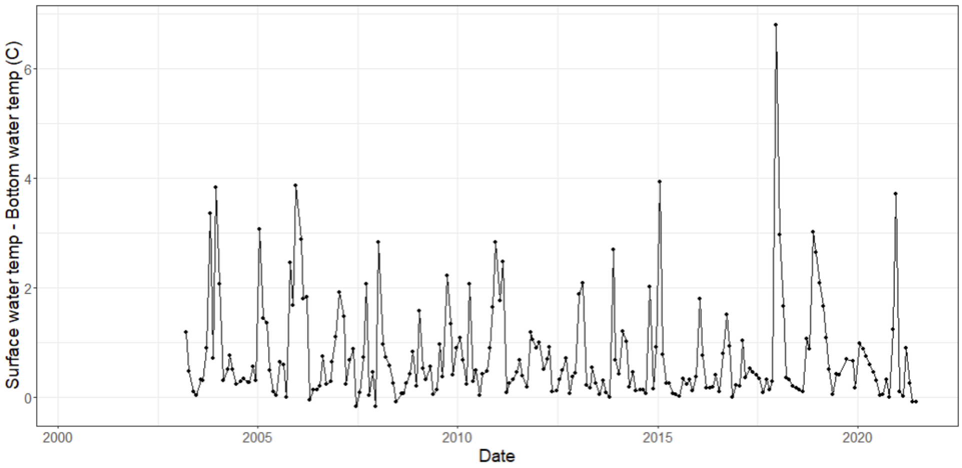

Figure S10. Difference in temperature between the surface and bottom waters of Lake Rotoehu