Stories in the Sediment

AN ANALYSIS OF PHYTOPLANKTON PIGMENTS WITHIN LAKE SEDIMENT TO PREDICT AND RETRODICT WATER QUALITY IN NEW ZEALAND LAKES.

A thesis presented in partial fulfilment of the requirements for the degree of

Master of Science
in
Ecology

At Massey University, Manawatū,

New Zealand.

CARRIE PAGE

2022

Abstract

When lakes experience an increase in nutrient availability, the phytoplankton and the primary productivity of the lake will also increase. This increase provides a robust means of signifying lake trophic fluctuations. The phytoplankton from the surface of the lake (photic zone) will sediment out, leading to the accumulation of both planktonic and benthic phytoplankton remains at the bottom of the water column. It is these past fluctuations in phytoplankton biomass, which accumulate in the lake sediments, provide indications of past environmental conditions and lake health. This research aimed to assess the potential of phytoplankton pigments preserved/captured within sediments as indicators of lake water quality in both neoliminological and paleolimnological lake sediments from a dataset of 223 New Zealand Lakes (≈ 6% of the lakes in New Zealand) was used for the analysis of surface sediments. These lakes ranged from low elevation lakes (<10 m) to high elevation lakes (up to 1,839 m) and included a range of geomorphic classifications. The catchments ranged from 35,288 m^2 to 704,470,618 m^2 and included shallow lakes (<10 m) to deep lakes (up to 445 m).

In addition to accessing the use of hyperspectral imaging (HSI) techniques as a method for detecting phytoplankton pigments within sediments. The assessment of calibrating chlorophyll-a (chl-a) detected by HSI in lake core sediment samples to chl-a quantified by analytical chemistry methods (High performance liquid chromatography and spectrophotometry analyses), found that the use of spectrophotometry without acidification provided more consistent results (with an error rate of less than 7.5%) when compared to spectrophotometry analysis with acidification. Additionally, the use of spectrophotometry without acidification for chl-a calibration revealed the potential for a universal equation to be researched.

Within lakes, high trophic levels are positively correlated with cyanobacterial dominance. One of the complications of high trophic levels is cyanobacterial blooms, which can be toxic. A reliable pigment indicator for the presence of cyanobacteria is phycocyanin. Therefore, the use of HSI was assessed as an analytical technique utilised for detecting and quantifying the concentration of phycocyanin within lake core sediment samples. This study revealed that phycocyanin could not be detected within the lake sediments within this research. Suggesting that phycocyanin was not incorporated into the lake sediment within the lakes assessed. Additionally, the HSI signal thought to be detecting phycocyanin is potentially measuring chlorophyll-a within the lake core rather than phycocyanin.

To predict lake water quality through the lake trophic level index (TLI) several machine learning models were created (regression trees, random forest models, and boosted regression trees) The random forest model was created using the quantification of key phytoplankton pigments within surface sediments plus five static lake physical characteristics this model was the most accurate (within 10% of the TLI). This model provides a predictive tool to access lake TLI using a single sample of surface sediment. This model was then applied to lake core sediment samples to retrodict lake water quality. The assumption of many degraded lakes throughout New Zealand, is that this is of anthropogenic origin. The retrodicted TLI's suggests, that while anthropogenic influence is exacerbating the degradation of the lakes, prior to this the trophic levels of these lakes did not fluctuate beyond one trophic level (i.e., moving from oligotrophic to mesotrophic). Additionally, apparent in the retrodiction of the lakes is the integration of cyanobacteria indicator pigments into the sediments relatively recently. The integration of these pigments coincides with the arrival of Europeans to the respective areas.

Acknowledgments

I would like to thank Jonathon Puddick (JP), who is not only my mentor but also my friend. If it were not for your support this would never have been written, let alone started. It was fun delving into the 'rabbit holes' with you, even though we are both as bad as each other. I look forward to our next collaboration. Susie Wood, you have been an inspiration, and thank you for entertaining some of my more eccentric ideas. From Massey University I would like to thank Ian Henderson and Katherine Holt, your guidance and supervision has been invaluable for this research.

I would like to thank Claire Shepherd and Marcus Vandergoes for HSI analysis. For technical assistance in the laboratory, I would like to thank from Cawthron Institute: Natsumi Nishikawa, Steven Liu, Ronan Griffin, Sophie Young, and Lucy Thompson, as well as, from GNS Science Lizette Reyes. In addition to all the Lakes 380 team, especially those who work in the background planning, organising fieldwork and ensure we are safe in the process, sample and data management, subsample collection, laboratory management and sample analysis, etc.

I would like to thank and acknowledge the support of: Ngāti Kahungunu ki Wairarapa, Greater Wellington Regional Council and the landowners (Lake Nganoke and Pounui); Ngāti Hauiti, Horizons Regional Council and the landowners (Lake Oporoa); Te Arawa Lakes Trust, Bay of Plenty Regional Council, and the Department of Conservation (Lake Okataina); and Ngāti Whātua o Kaipara, Auckland Council, and the Department of Conservation (Lake Rototoa).

It has been an amazing experience to have been a part of the Lakes 380 team and contributing to some of the amazing research being conducted. For a full description of the goals, research, and teams within Lakes 380 please see https://lakes380.com/.

This work was supported by the New Zealand Ministry for Business, Innovation and Employment Endeavour Fund research programme (C05X1707; Our lakes' health: past, present, and future)

Contents

List of Figures

List of Tables

List of Equations

Abbreviations Used

Abbreviation	Explanation
AIC	Akaike's information criteria
Bagging / bagged / bag	Bootstrap aggregation
Chl-a	Chlorophyll-a
CI	95% confidence interval
cp	Complexity Parameter, which controls the size of a regression tree
cv	5-fold Cross Validation
DAD	Diode-array detection
DEM	New Zealand national digital elevation model
GIS	Geographical information system
HPLC	High performance liquid chromatography
HPLC-DAD	High-performance liquid chromatography with diode-array detection
HSI	Hyperspectral imaging
Lakes380	'Our Lakes' Health: Past, Present, Future' Ministry of Business, Innovation and Economy Endeavour Fund programme
LCDB	Land cover database
LC-UVD	Liquid chromatography method with ultraviolet detection
MAE	Mean absolute error
MSE	Mean squared error
Pigments-only	Model created with the variables of thirteen phytoplankton pigments: twelve measured by HPLC-DAD (Alloxanthin, Chlorophyll-a via HPLC, Chlorophyll-b, Canthaxanthin, Diadinoxanthin, Diatoxanthin, Echinenone, Fucoxanthin, Lutein, Peridinin, Violaxanthin, and Zeaxanthin); and one by spectrophotometry (Chlorophyll-a)
Pigments +	Model created with the variables of thirteen phytoplankton pigments: twelve measured by HPLC-DAD (Alloxanthin, Chlorophyll-a via HPLC, Chlorophyll-b, Canthaxanthin, Diadinoxanthin, Diatoxanthin, Echinenone, Fucoxanthin, Lutein, Peridinin, Violaxanthin, and Zeaxanthin); and one by spectrophotometry (Chlorophyll-a) plus five physical lake characteristics that would not change through time, or would only change very slowly (Elevation, Lake Area, Lake Volume, Maximum Depth, and Catchment Area).
RABD	Relative absorption band depths
$RABD_{615}$	Suspected phycocyanin relative absorption band depth
$RABD_{660\text{-}670}$	Chlorophyll-a relative absorption band depth
RMSE	Root mean squared error
Spectro	Spectrophotometry

SSE	Sums of squares error
TLI	Trophic level index
UV-VIS	Ultraviolet–visible spectrophotometry
ʎ-em	Fluorescence emission wavelength
ʎ-ex	Fluorescence excitation wavelength

Chapter 1 Introduction

When lakes experience an increase in nutrient availability, the phytoplankton and the primary productivity of the lake will also increase. This increase provides a robust means of signifying lake trophic trends (Willén, 2000). The phytoplankton from the surface of the lake (photic zone) will sediment out, leading to the accumulation of both benthic and planktonic autotrophs remains at the bottom of the water column (Bellinger & Sigee, 2015). It is these past fluctuations in phytoplankton biomass, which accumulate in the lake sediments and provide indications of past environmental conditions and lake health. These sedimented remains indicate that phytoplankton pigments captured and preserved within the sediments can be used as indicators of lake water quality in both neoliminological and paleolimnological lake sediments.

Paleolimnology analysis of sediments generally relies on several methods of identification of the phytoplankton. Traditional microscopy and identification of phytoplankton is commonly complemented with pigment analysis using high-performance liquid chromatography with diode-array detection (HPLC-DAD). HPLC-DAD has proven to be a robust and reliable method of pigment analysis, suited for analysing large numbers of samples (Sanz et al., 2015). As chlorophylls and carotenoids are often the only fossil remains for non-siliceous phytoplankton they are of value to paleolimnology studies (Hodgson et al., 1997).

In the past decade increasingly more research is being conducted on the identification and quantification of phytoplankton fossil pigments through Hyperspectral Imaging (HSI) (Butz et al., 2015; Butz et al., 2017; Makri et al., 2020; Schneider et al., 2018; Wolfe et al., 2006; Zander et al., 2021). HSI looks to become a valuable paleolimnological technique, as when compared to High performance liquid chromatography (HPLC) techniques, it is more economical on laboratory resources and is non-destructive on samples. However, current research is based on the calibration of individual lake sediments using high-performance liquid chromatography with diode-array detection (HPLC-DAD).

Commonly associated with eutrophic lakes and poor water quality, cyanobacterial blooms are of increasing concern (Paerl & Paul, 2012). Phycocyanin is an important phytoplankton pigment water quality indicator, as it is indicative of cyanobacteria (Bellinger & Sigee, 2015; Randolph et al., 2008; Vincent et al., 2004). Thus, the detection of phycocyanin is becoming an important step in the identification of Lake health.

The analysis of phytoplankton pigments within lake sediment core samples provides a step toward the reconstruction of both neoliminological and paleolimnological water quality. The neoliminological analysis could potentially provide a predictive tool to access lake trophic level indexes (TLI) using a single surface sediment sample, rather than requiring monitoring data trends. The use of a single sample will allow for identification of lake TLI incorporating seasonal fluctuations with a shorter time frame than what is currently required when utilising monitoring data trends. This could be a valuable tool for regional councils to aid in the identification of degraded lakes, allowing for prioritisation of rehabilitation projects. These predictive models, if refined, may also provide an idea of how to balance the various land uses within a catchment to not exert too much pressure on the lake.

1.1 Objectives and aims

The overarching aim of this research was to assess the potential of phytoplankton pigments preserved/captured in sediments as indicators of lake water quality in both neoliminological (modern) and paleolimnological (ancient) settings in New Zealand lakes. Another important aim was to test HSI as a method for detecting phytoplankton pigments within sediments. These aims were addressed through the following objectives: 1) comparing sediment pigment measurements made by HSI against measurements made by the 'traditional' methods of HPLC and spectrophotometry; and 2) using pigment analyses from neoliminological sediments to build a predictive/retrodictive model for lake trophic status, and then applying that to the paleolimnological sediments.

The first research objective evaluates the limitations of utilising HSI to analyse chl-a and phycocyanin in lake sediment cores. Three questions drove this research:

A) Does the chlorophyll-a (chl-a) HSI signal from lake sediment cores provide accurate and reliable measurements of chl-a compared to those achieved through traditional analytical methods?
B) Are between-lake differences small enough that a universal calibration factor can be applied to HSI lake sediment core measurements?
C) Does the phycocyanin HSI signal specifically measure phycocyanin within lake sediment cores?

To answer these questions, lake sediment cores were analysed by HSI as well as by HPLC-DAD, spectrophotometry (for chl-a) and fluorimetry (for phycocyanin). Comparison of the HSI and analytical measurements were undertaken to determine how accurate the chl-a HSI measurements were, whether the same relationship was

observed across five New Zealand lakes and if phycocyanin could be specifically measured by HSI.

The second research aim analysed lake surface sediments for key phytoplankton pigments to infer the water quality. Two questions drove this research:

A) whether the quantification of key phytoplankton pigments within surface sediments can be used to create a predictive model of lake TLIs?
B) Can this predictive model be used to retrodict TLIs using a lake sediment core?

To answer this question combinations of key phytoplankton pigment concentrations within the surface sediments, and land use proportions of the lake's catchment were used to create machine learning models. Selected models were then used to predict and retrodict TLI's of selected lakes.

Chapter 2 Literature Review

Paleolimnology is a multidisciplinary science which aims to reconstruct environmental histories and their catchments with the aim of further understanding environmental changes. Sediments at the bottom of a lake will accumulate chronologically, thus providing a record of the lake and the surrounding catchment within the layers. Paleolimnology will analyse a lake's history through, among other things, the remains of aquatic and terrestrial organisms preserved within the sediments. to reconstruct historical environmental conditions (Smol et al., 2002).

Some of the biological proxies utilised consist of, but are not limited to: pollen and spores (Bennett & Willis, 2002), as well as plant macrofossils (Birks, 2002) to infer past flora in the lake area; diatom remains (Battarbee et al., 2002) and biogenic silica (Conley & Schelske, 2002) to indicate diatom abundance as well as diatom productivity; chrysophyta remains as environmental indicators which can infer conditions such as pH changes, historical climate, lake-level changes, salinity, etc (Zeeb & Smol, 2002); charcoal as a fire proxy, which alongside pollen data can suggest relationships regarding the climate, past flora, fire as an ecosystem process, and historical anthropogenic activities (Whitlock & Larsen, 2002); and sedimentary pigments which can be utilised in research assessing historical phytoplankton abundance, community composition, and primary production of the lake (Leavitt & Hodgson, 2002).

2.1 Phytoplankton and Pigments

The name phytoplankton refers to motile (either actively or passively), free-floating, photosynthetic, aquatic microorganisms (Pal & Choudhury, 2014). Phytoplankton include both Prokaryota (cyanobacteria) and Eukaryota (microalgae) and can be classified further into divisions or phyla according to various characteristics. However, there is no universally accepted classification system for all organisms within these phyla (Barsanti & Gualtieri, 2014; Willén, 2000). This thesis will focus on the phyla described in Table 2.1 which allows for the classification of phytoplankton phyla without distracting from the research of this paper with classification specifics.

Table 2.1. *Algae Phyla classifications.*

Algae Phyla classifications derived from varieties utilised by Pal and Choudhury (2014), Bellinger and Sigee (2015), and Barsanti and Gualtieri (2014). This classification is based upon the phytoplankton microscopal appearance, biochemical, and cytological characteristics.

Phytoplankton Phyla	Common Name
Bacillariophyta	Diatoms
Chlorophyta	Green algae
Chrysophyta	Golden algae
Cryptophyta	Cryptomonads
Cyanophyta	Blue-green algae/ Cyanobacteria
Dinophyta	Dinoflagellates
Euglenophyta	Euglenoids
Rhodophyta	Red algae
Xanthophyta	Yellow-green algae

Analysis of phytoplankton pigments differs from traditional assessments performed through studying Bacillariophyta communities, as these contain siliceous structures which remain within the sediment and can be examined through various microscopy techniques. Often the fossil pigments (the pigments retained in lake sediments) are all that remains identifiable from non-siliceous phytoplankton communities.

All phytoplankton are characterised by their ability to perform photosynthesis; however, the pigments utilised for this process differ between different phyla, and even genera, of phytoplankton. Numerous phytoplankton pigments have been identified to-date, with the pigments falling into three main categories: chlorophylls, carotenoids, and phycobilins. These fossil pigments can indicate the phytoplankton community at the time the sediment was laid down, which can then be used as indicators of a wide variety of past lake conditions, from changes in the physical structure of the lake (Cohen, 2003) to anthropogenic impacts, such as eutrophication, and land-use practices (Leavitt & Hodgson, 2002).

2.1.1 Chlorophylls

Of the phytoplankton pigments, chlorophylls are found in all major phyla of phytoplankton (Bellinger & Sigee, 2015; Sanger, 1988; Yackulic, 2017) as it is required for photosynthesis. Several variations of chlorophyll can be found in phytoplankton (Table 2.2), most notably chlorophyll-a (chl-a), which is known as the characteristic green pigment within plants and can be found in almost all photosynthetic phytoplankton. It is particularly useful as a proxy for total phytoplankton biomass (Bellinger & Sigee, 2015; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011) and for primary productivity

analyses. Chlorophyll-b is present in phytoplankton phyla of the 'green' lineage, including euglenophyta and chlorophyta. Chlorophyll-c_1, -c_2, and c_3 are present in phytoplankton of the 'red' lineage, including cryptophyta, bacillariophyta, dinophyta, chrysophyta, and xanthophyta. These 'red' and 'green' lineages are denoted by the phytoplankton's plastid (Falkowski et al., 2004).

While chlorophylls-a, -b and -c are the three main chlorophyll pigments, recent research has identified two other chlorophyll pigments, chlorophyll-d and -f. Chlorophyll-d was originally isolated from the cyanobacteria *Acaryochloris marina* and is a pigment which absorbs light at the far-red end of the spectrum (Li et al., 2012) and is present in rhodophyta. Chlorophyll-f also absorbs at the far-red end of the spectrum and has been isolated from cyanobacteria when grown under 750 nm far-red light (Nürnberg et al., 2018). The chemical structures of chl-a, -b, -c, and -d are shown in Figure 2.1. Chlorophyll-f discovered by Chen et al. (2010) was originally found in stromatolites (sedimentary formations formed by cyanobacteria). However, as chlorophyll-f is not known to be found in phytoplankton outside of stromatolites or laboratory cultures and chlorophyll-d has not been isolated outside of marine phytoplankton (Larkum & Kühl, 2005; Li et al., 2012), they will not be considered further in this thesis.

After the algal cell has degraded, following post-mortem deposition on the lakebed, chl-a and -b, or their distinctive degradation products, remain detectable (Leavitt & Hodgson, 2002; Wolfe et al., 2006), providing useful pigments for paleolimnological analysis (Cohen, 2003). After the chlorophyll pigment is deposited within the sediments it is likely to degrade into either pheophytin or pheophorbide. Both chlorophyll and its degradation products can be used as a proxy for lake productivity (Butz et al., 2017).

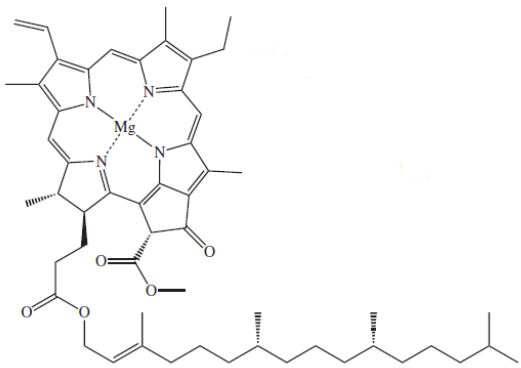
Table 2.2. *Chlorophylls reported from different phytoplankton phyla in literature.*

A summary of chlorophyll isomers reported from different phytoplankton phyla in literature.

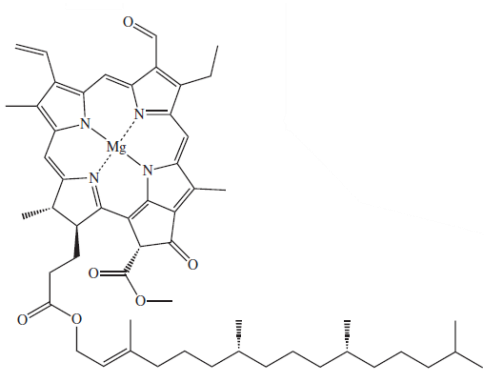
	Euglenophyta	Cryptophyta	Bacillariophyta	Dinophyta	Chrysophyta	Cyanobacteria	Chlorophyta	Rhodophyta	Xanthophyta	Reference
Chlorophyll-a	X	X	X	X	X	X	X	X	X	(Bellinger & Sigee, 2015; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011)
Chlorophyll-b	X						X			(Bellinger & Sigee, 2015; Cohen, 2003; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011)
Chlorophyll-c_1			X		X				X	(Bellinger & Sigee, 2015; Leavitt & Hodgson, 2002; Roy et al., 2011)
Chlorophyll-c_2		X	X	X	X				X	(Bellinger & Sigee, 2015; Cohen, 2003; Leavitt & Hodgson, 2002; Roy et al., 2011);
Chlorophyll-c_3			X		X					(Bellinger & Sigee, 2015; Leavitt & Hodgson, 2002; Roy et al., 2011).
Chlorophyll-d						X				(Li et al., 2012)

Figure 2.1. *Structures of Main Chlorophyll Pigments Highlighted in this Section. Chemical Structures of various chlorophyll compounds as depicted in Roy et al. (2011).* ***A)*** *chlorophyll-a;* ***B)*** *chlorophyll-b;* ***C)*** *chlorophyll-c$_1$;* ***D)*** *chlorophyll-c$_2$;* ***E)*** *chlorophyll-c$_3$*

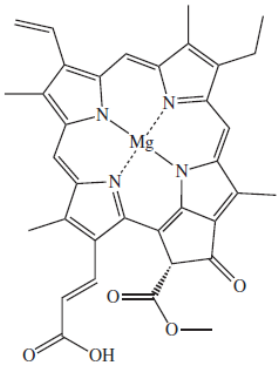
A)



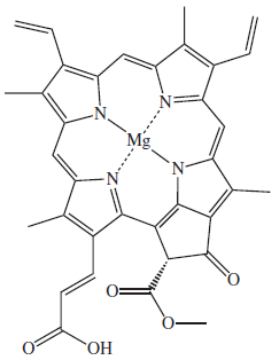
B)



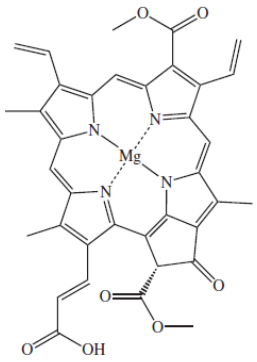
C)



D)



E)



2.1.2 Carotenoids

The carotenoids consist of carotenes (which do not contain oxygen) and xanthophylls (which contain oxygen). Carotenoids have various functions within the organism, from aiding photosynthetic organisms to harvest light (fucoxanthin and β-carotene), and facilitating energy transfer within the cell (peridinin) to protecting the organism from high levels of light (echinenone), for example, dissipating excess energy when chlorophylls become too excited (Zeaxanthin) (Takaichi, 2011). Additionally, carotenoids are generally less labile than chlorophylls (Buchaca & Catalan, 2007; McGowan, 2007).

The most common carotene is β-carotene (β,β-carotene) which is a dominant pigment in chlorophyta, rhodophyta, and one group of dinophyta, and can be found as a minor pigment in almost all other phytoplankton groups (Roy et al., 2011), indicating a common cyanobacterial ancestor for the inheritance of all extant plastids (Roy et al., 2011). Additionally, β-carotene is one of the more relatively stable photosynthetic pigments (McGowan, 2007). Other carotenes include α-carotene (β,ε-carotene) and ε-carotene (ε,ε-carotene). α-carotene is a minor, or trace pigment present in the phytoplankton phyla; chlorophyta, xanthophyta, cryptophyta, chrysophyta, rhodophyta (Bellinger & Sigee, 2015; Leavitt, 1993; Leavitt & Hodgson, 2002) and cyanobacteria (Roy et al., 2011). ε-Carotene is a minor pigment in chlorophyta, dinophyta (Roy et al., 2011), chrysophyta, and bacillariophyta (Bellinger & Sigee, 2015). Nonetheless, as with chl-a, many carotenes are ubiquitous pigments in phytoplankton phyla. As such they provide limited information on phytoplankton community composition.

The xanthophylls are more structurally diverse than the carotenes (Sanger, 1988) and their stability is commonly dependent on their molecular structure, which can be negatively correlated with the number of functional groups present (Britton et al., 1995; Buchaca & Catalan, 2007; Damsté & Koopmans, 1997). This research focused on alloxanthin, canthaxanthin, diadinoxanthin, diatoxanthin, echinenone, fucoxanthin, lutein, myxoxanthophyll, peridinin, vaucheriaxanthin, violaxanthin, and zeaxanthin as biomarkers to gather paleolimnological information on phytoplankton phyla (Table 2.3) and water quality. The chemical structures of these selected xanthophylls are shown in Figure 2.2.

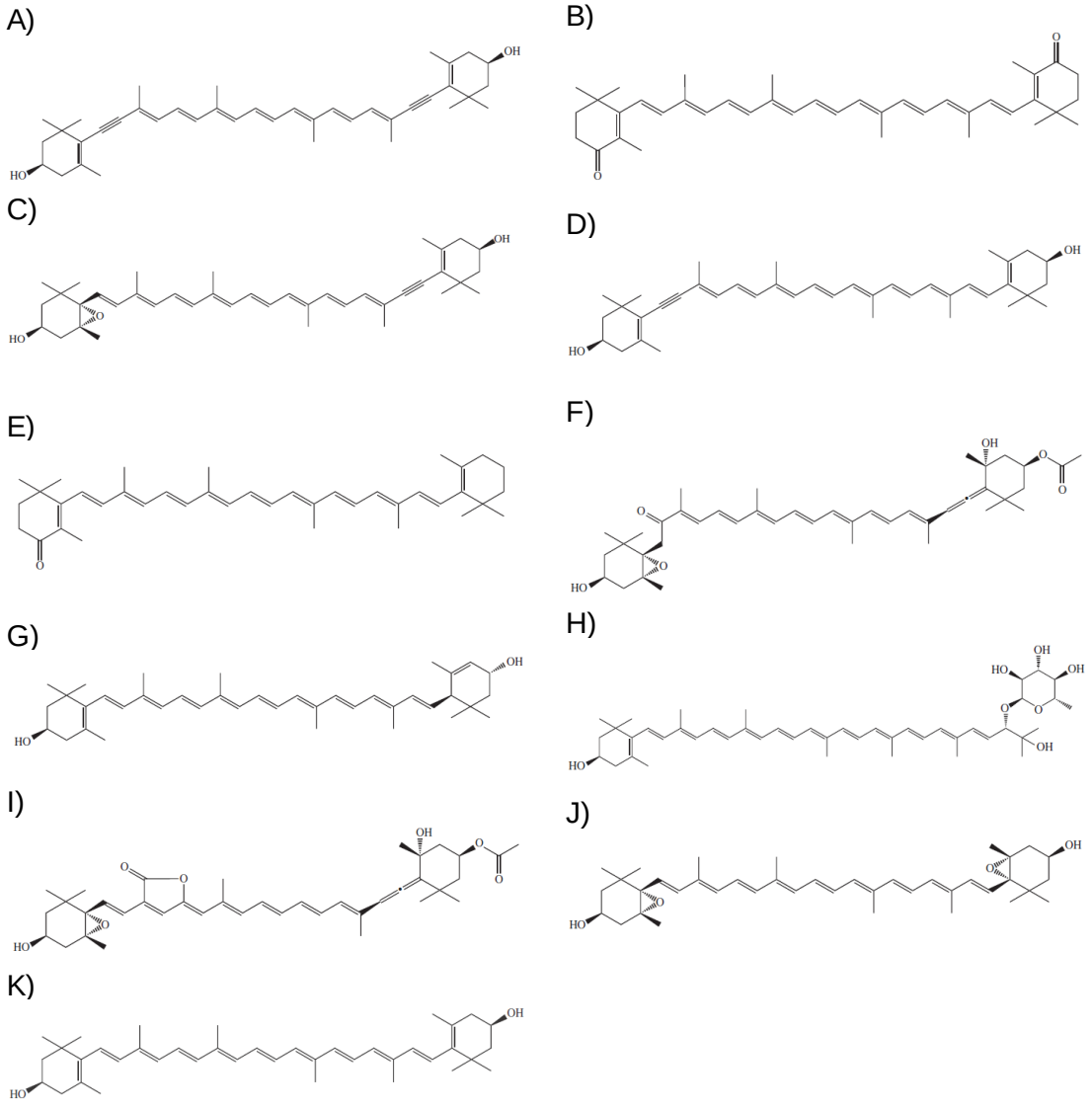
Table 2.3. *Xanthophylls reported from different phytoplankton phyla in literature. Phytoplankton groups where the respective pigment is generally encountered are reflected in the table.*

	Euglenophyta	Cryptophyta	Bacillariophyta	Dinophyta	Chrysophyta	Cyanobacteria	Chlorophyta	Rhodophytes	Xanthophyta	Reference
Alloxanthin		X					X			(Bellinger & Sigee, 2015; Buchaca & Catalan, 2007; Cohen, 2003; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011)
Canthaxanthin						X				(Buchaca & Catalan, 2007; Leavitt, 1993; Leavitt & Hodgson, 2002; Richardson, 1996; Roy et al., 2011)
Diadinoxanthin	X		X	X	X					(Buchaca & Catalan, 2007; Leavitt & Hodgson, 2002)
Diatoxanthin			X	X	X					(Buchaca & Catalan, 2007; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011).
Echinenone						X				(Buchaca & Catalan, 2007; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011)
Fucoxanthin			X		X					(Bellinger & Sigee, 2015; Buchaca & Catalan, 2007; Cohen, 2003; Leavitt, 1993; Leavitt & Hodgson, 2002).
Lutein	X						X			(Buchaca & Catalan, 2007; Cohen, 2003; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011)
Myxoxanthophyll						X				(Roy et al., 2011)
Peridinin				X						(Bellinger & Sigee, 2015; Cohen, 2003; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011)
Violaxanthin				X	X		X			(Bellinger & Sigee, 2015; Roy et al., 2011),
Zeaxanthin				X		X		X		(Bellinger & Sigee, 2015; Buchaca & Catalan, 2007; Cohen, 2003; Leavitt, 1993; Leavitt & Hodgson, 2002; Roy et al., 2011)

Figure 2.2. *Structures of Xanthophylls Highlighted in this Section.*

Chemical Structures of various xanthophyll compounds as depicted in Roy et al. (2011).

A) *Alloxanthin;* ***B)*** *Canthaxanthin;* ***C)*** *Diadinoxanthin;* ***D)*** *Diatoxanthin;* ***E)*** *Echinenone;* ***F)*** *Fucoxanthin;* ***G)*** *Lutein;* ***H)*** *Myxoxanthophyll;* ***I)*** *Peridinin;* ***J)*** *Violaxanthin;* ***K)*** *Zeaxanthin.*



Phytoplankton found within lakes with good water quality (oligotrophic) are bacillariophyta, chlorophyta, chrysophyta, and cryptophyta (Bellinger & Sigee, 2015; Katsiapi et al., 2016; Paul et al., 2012; Rawson, 1956). As such the pigments alloxanthin, diadinoxanthin, diatoxanthin, fucoxanthin, lutein, and violaxanthin can be expected to be present. Mesotrophic lakes bacillariophyta, chlorophyta, and cryptophyta (Bellinger & Sigee, 2015; Katsiapi et al., 2016; Paul et al., 2012; Rawson, 1956) will be present, not chrysophyta. However, this is not expected to alter the pigment composition expected to be present. In lakes with poorer water quality (eutrophic), cryptophyta, cyanobacteria, and euglenophyta (Bellinger & Sigee, 2015; Katsiapi et al., 2016; Paul et al., 2012; Rawson, 1956) are present. Therefore, the pigments canthaxanthin, diadinoxanthin, echinenone, lutein, myxoxanthophyll, and zeaxanthin can be expected to be present. In degraded lakes (hypereutrophic), cyanobacteria and euglenophyta are present (Bellinger & Sigee, 2015; Katsiapi et al., 2016; Paul et al., 2012; Rawson, 1956). Within these lakes canthaxanthin, diadinoxanthin, echinenone, lutein, myxoxanthophyll, and zeaxanthin can be expected to be present. Consequently, alloxanthin, diatoxanthin, fucoxanthin, and violaxanthin can be expected to be found in lakes with good water quality. While canthaxanthin, echinenone, myxoxanthophyll, and zeaxanthin can be expected in lakes with poor water quality.

Similar to chlorophylls, carotenoids can be found in lake sediments (Sanger, 1988; Vallentyne, 1957a), and there is currently no evidence to suggest that carotenes degrade significantly once the phytoplankton is buried in the sediments (Sanger, 1988). Vallentyne (1957a) even found evidence of α- and β-carotenes preserved within 20,000-year-old sediment. However, xanthophylls seem to degrade slightly more rapidly than carotenes (Sanger, 1988) but are less liable than chlorophylls. If degradation does occur, carotenoids will break down into colourless compounds (Leavitt & Hodgson, 2002) and will not be detectable through spectroscopic methods.

2.1.3 Anthocyanins and Phycobilins

Other forms of pigments found in phytoplankton include anthocyanins and phycobilins. There are difficulties analysing for anthocyanin and phycobilin pigments in sediments due to their high solubility in water (Leavitt & Hodgson, 2002; Sanger, 1988). This means that these pigments, including mycosporine-like amino acids and flavonoids (Leavitt & Hodgson, 2002), are generally destroyed before they can be integrated into the sediment layer. As such, they have not been widely utilised as paleolimnology indicators (Sanger, 1988). However, it has been noted in Leavitt and Hodgson (2002) that some of these pigments, or their degradation products, may be preserved in sediments within the bulk material of the detritus. This was noted in Yackulic (2017) where putative signals for

phycocyanin (a phycobilin) were detected in sediment core samples from Crater Lake, Colorado. Additionally, Favot et al. (2020) noted phycocyanin within the 588-642 wavelength when analysing sediment cores from Ontario, Canada, using visible near-infrared reflectance spectroscopy. Sorrel et al. (2021) also reported phycocyanin within a core sample from Lake Son Kol, Kyrgyzstan by analysing hyperspectral signals at $RABD_{615}$, following the analysis for phycocyanin as described by Yacobi et al. (2015). Phycocyanin is a blue-coloured pigment-protein accessory to chlorophyll which is indicative of cyanobacteria (Bellinger & Sigee, 2015; Randolph et al., 2008; Vincent et al., 2004).

2.1.4 Phytoplankton Pigments as Environmental Indicators

Due to short generation times, phytoplankton can react rapidly (i.e., within days) to geophysical and chemical changes in aquatic environments. When lakes experience an increase in nutrient availability, the biomass of phytoplankton and the primary productivity of the lake will also increase. This increase provides a robust means of signifying lake trophic fluctuations (Willén, 2000). The phytoplankton from the surface of the lake (photic zone) will sediment out, leading to the accumulation of both benthic and planktonic autotrophs remains at the bottom of the water column (Bellinger & Sigee, 2015). It is these past fluctuations in phytoplankton biomass, which accumulate in the lake sediments, which provide indications of past environmental conditions and lake health.

2.2 Phytoplankton Indicators of Water Quality

Traditionally, studies assessing environmental aspects of lakes have focused on the phytoplankton communities rather than the benthic communities. Predominantly this is because the phytoplankton provide the main phototrophic biomass within the lake and can be readily sampled (Bellinger & Sigee, 2015). This also means that the main body of research available is focused on phytoplankton ecological preferences. Algal indicators for the trophic state of lakes have typically been formulated by analysing the phytoplankton at either the species level (Flint, 1977; Reynolds, 1990) or the phyla level (Katsiapi et al., 2016; Nygaard, 1949; Ptacnik et al., 2008; Stockner, 1972), using various indicators and quotient formulae. Other trophic state indicators are provided by species diversity and richness metrics (Dodson et al., 2000; Nygaard, 1949; Watson et al., 1997). However, the most prevalent approach is the trophic level index (TLI). The TLI utilises measurements taken from chl-a concentrations (algal biomass), Secchi disk transparency (water clarity), and total phosphorous concentration (algal growth nutrient). These three parameters were utilised by Carlson (1977) and Chapra and Dobson (1981) who also added a primary production parameter. Carlson (1977) developed their index

using European and North American lake data, while Chapra and Dobson (1981) developed theirs using data from the Great Lakes of North America

Burns et al. (2000), updated the TLI framework for New Zealand, as the existing schemes were not deemed appropriate. The TLI from Carlson (1977) lacked definition in its higher trophic levels for New Zealand and the index developed by Chapra and Dobson (1981) was too detailed in scale for New Zealand lakes (Burns et al., 2000). Additionally, nitrogen concentration was added to the monitoring data for New Zealand lakes as many lakes show aspects of nitrogen limitation to growth (White et al, 1985), and total nitrogen is an essential variable in a TLI scheme for New Zealand lakes (Burns et al., 2000).

Lakes are often classified by their trophic status. Generally, low nutrient lakes possess low levels of phytoplankton, and the lake water is clear. These lakes are classified as microtrophic (lowest level) or oligotrophic (slightly higher level). Microtrophic and oligotrophic lakes are often relatively young lakes (geologically) and can be large lakes within small catchments, thus the inputs of nutrients are restricted (Schlesinger & Bernhardt, 2020). Kumar and Singh (1979) noted that some oligotrophic lakes can display a wide diversity of phytoplankton even though the overall concentration of each species is low.

As the nutrient levels rise, the lake becomes mesotrophic where moderate levels of phytoplankton and nutrients can be found. If the nutrient levels continue to rise, the lake then becomes eutrophic, here the water clarity will degrade, phytoplankton communities will change, and cyanobacteria can often dominate. The highest trophic level of a lake is hypertrophic/super-eutrophic, within these waters there is poor water clarity and often extreme algal growth (Larned et al., 2019). Additionally, as lakes increases in trophic status the complexity of the phytoplankton communities decreases (Rott, 1984) and eutrophic lakes rarely exhibit more than a few species of phytoplankton but rather high concentrations of these few species (Kumar & Singh, 1979) . Within New Zealand lakes, higher nutrient loads are positively correlated with higher trophic states and cyanobacterial dominance (Paul et al., 2012).

The process of lake eutrophication refers to the succession of the lake productivity. This can be either a natural process or of anthropogenic basis. Natural eutrophication of lakes is a process that occurs where the lake progresses from a state of low productivity (microtrophic) to one of high productivity (eutrophic). This process is a natural succession for lakes but occurs over centuries and will eventually conclude with the lake being filled in (Carpenter, 1981). Anthropogenic eutrophication occurs where human influence on the catchment (i.e., agricultural fertilizers, detergents, effluent, etc.) accelerates the

process of eutrophication (Schallenberg & van der Zon, 2019). Some of the complications of eutrophication include, but are not limited to: cyanobacterial blooms, which can be toxic; a decline in species diversity within the lake; water quality problems, such as diminishing water clarity, and potability issues; and hypoxia (Smith & Schindler, 2009).

Galbraith and Burns (2007) found that concentrations of nutrients and lake water quality are positively related to the extent of modification of the catchment within New Zealand lakes. Their study showed that catchment land use directly affected the water body trophic status. Paul et al. (2012), Peterson et al. (2007), and Jeppesen et al. (2005), were all able to demonstrate relationships between the phytoplankton within the lakes and the corresponding land use of the lake catchment. Particularly, Paul et al. (2012) correlated the land use of the respective catchment and the subsequent phytoplankton relationships to the trophic status of eleven New Zealand lakes located in the Rotorua lake district. Noting that lakes within catchments with higher nutrient loads from pastoral land use possessed a higher TLI and cyanobacterial dominance. Conversely, lakes with lower nutrient loads and a higher proportion of native land use within the catchment possessed a lower TLI and a chlorophyta dominance.

2.3 Phytoplankton Indicators in Paleolimnology

Fossil phytoplankton pigments such as chlorophyll are indicators for past algal biomass in lakes (Leavitt, 1993). Butz et al. (2017) were able to detect 'green pigments' (chl-a together with its diagenetic products) to infer lake productivity in a lake core dated 1071–1255 AD. Additionally, fossil carotenoids have also been used as indicators of past lake trophic conditions (Gorham et al., 1974; Guilizzoni et al., 1992; Lami et al., 2010; Sanger, 1988). Early research suggested that fossil pigments were detected in more abundance in sediments of eutrophic lakes (Cohen, 2003; Gorham, 1960; Vallentyne, 1957b) this suggests that ratios of specific pigments within sediments may indicate various trophic status of lakes (Gorham et al., 1974; Guilizzoni et al., 1992; Sanger, 1988). However, while fossil pigment concentrations can provide an indicator of algal production, Leavitt (1993) determined that >90% of pigments are degraded into colourless compounds prior to fossilisation. Primarily, this is because most pigments are exposed to rapid modifications through various microbial actions and diagenetic alteration due to oxidation. However, the majority of losses occur prior to burial within the sediments and at the surface of the sediments (Cohen, 2003), leaving the remains to be fossilised.

Sanger (1988) noted unavailability of evidence to suggest decomposition of carotenoids once buried and fossilised, and if gradual decomposition was to occur it would be similar to all the organic matter within the sediment. As noted earlier, Vallentyne (1957a) found

evidence of α- and β-carotenes preserved within 20,000-year-old sediment. Additionally, Züllig (1986) was able to isolate xanthophylls (including echinenone, lutein, and alloxanthin) from 14,000-year-old sediment. Even Watts and Maxwell (1977) were able to detect echinenone, zeaxanthin, and canthaxanthin in samples up to 5,000 years old before the echinenone and zeaxanthin showed degradation. Within the same core canthaxanthin was detected in sediments 340,000 years old. However, it should be noted that when comparing pigments in cores as old as the ones described above the stability of the pigment needs to be considered. Because if a less stable pigment is not detected it cannot be taken as indication in a change of phytoplankton community.

2.4 Sediment analysis methods

Paleolimnological analysis of sediments generally relies on several methods of identification of the phytoplankton. Traditional microscopy and identification of phytoplankton is commonly complemented with pigment analysis using high-performance liquid chromatography with diode-array detection (HPLC-DAD). HPLC-DAD has proven to be a robust and reliable method of pigment analysis, suited for analysing large numbers of samples (Sanz et al., 2015). As chlorophylls and carotenoids are often the only fossil remains for non-siliceous phytoplankton they are of value to paleolimnology studies (Hodgson et al., 1997).

2.4.1 Spectrophotometry

Spectrophotometric analyses can be used for determining concentrations of chl-a and phycocyanin. Chl-a analyses are a popular method for determining phytoplankton biomass in lakes and waterways. However, many of these methods are critiqued, as the chlorophyll concentrations can often differ depending on the phytoplankton species and the physiological state of the cells (Dos Santos et al., 2003). The more commonly performed method for analysing chl-a concentrations involves the spectrophotometric analysis of samples at the light wavelength 665 nm. The chl-a can then be quantified via a specific absorption coefficient, or the solution can be acidified prior to quantification. To differentiate the chl-a from its degradation products, the sample is often degraded with acid following the initial measurement. This acidification removes the magnesium ion creating pheophytin a. A second measurement following acidification is then subtracted from the original reading and adjusted using an extinction coefficient (absorption coefficient) to determine the level of chl-a present. A study conducted by Stich and Brinker (2004) compared the acidified and non-acidified analyses of chl-a and found that acidified analyses are underestimated when calibrated by high-performance liquid chromatography (HPLC). Additionally, the acidification impairs the accuracy of the

chl-a measurement by introducing chlorophyll-b, -c, and carotenoid artifact interference (Stich & Brinker, 2004).

Spectrophotometry can also detect phycocyanin, as noted in Favot et al. (2020) where phycocyanin was detected within spiked lake sediment samples using visible near-infrared reflectance spectroscopy within the 588-642 nm wavelengths. Phycocyanin has a characteristic fluorescence spectrum which does not interfere with chlorophyll, it makes analysis uncomplicated (Zamyadi et al., 2012). Fluorometrically phycocyanin will excite at 580-590 nm with an emission at 645-650 nm (Bellinger & Sigee, 2015; Roy et al., 2011). Additionally, Vincent et al. (2004) noted that phycocyanin excitation occurred at 630 nm and emission was at 660nm, which extends the range stipulated by Bellinger and Sigee (2015) and Roy et al. (2011).

2.4.2 HPLC

Where spectrophotometry measures the light absorbed by compounds at specific wavelengths; in this case, the light absorbed by phytoplankton pigments, HPLC-DAD separates out the target compounds for quantification. The HPLC will separate the compounds while the Diode-array detection (DAD) is used to quantify the eluted peak of the target compound. Like spectrophotometry, DAD's use the light spectrum absorbed by target compounds for detection. However, rather than just a single wavelength, HPLC-DAD can perform simultaneous acquisition of data across a range of wavelengths. Unfortunately, HPLC-DAD analyses can be expensive, time consuming, and demand specialised technical skill (Dos Santos et al., 2003). However, HPLC-DAD has the added advantage that the pigments are separated through chromatography prior to being analysed by UV-VIS, which has the advantage of increased specificity and the ability to measure pigments which possess similar light absorbance spectra. The confirmation of pigments through various retention times, absorbance spectra and co-chromatography is conducted with certified reference standards (Hodgson et al., 1997; Wright et al., 1991). Additionally, HPLC-DAD is also subject to low interference when analysing pigments and is more precise than some other traditional limnology methods, as it can distinguish and separate diverse phytoplankton pigments and their degradation products (Dos Santos et al., 2003).

2.4.3 Hyperspectral Imaging

Because of its reagent-free non-destructive nature (Butz et al., 2015; Das et al., 2005), HSI looks to become a valuable paleolimnological technique. Advances in technology and computer processing, in particular the increasing precision of spectrometers which are capable of quantifying adjacent, fine bands within wide spectral ranges (Goetz et al., 1985) will further improve this tool in the future.

The principle behind hyperspectral imaging (HSI) is the measurement of light wavelengths, measured as a series of narrow bands of wavelengths, detected by hyperspectral sensors (Ekpenyong, 2013). HSI measures light wavelengths reflected by various targets, as opposed to absorbed light measured by spectrophotometry, and can be used to measure inorganic and organic substances within sediment. For example, the measurement of inorganic clays within sediments can infer past climate and environmental conditions (Dianto et al., 2019), and the measurement of organic chlorophyll a can infer past primary productivity. The narrow sections of absorbed light from HSI are termed 'absorbance bands' and can be used as a diagnostic tool for specific substances. However, it must be noted that while concentrations for specific pigments can be determined in clean homogenised samples, there is a possibility for interferences-from and interactions-with substances also found in sediments when the reflectance peaks of specific substances overlap (Yackulic, 2017).

In the past decade increasingly more research is being conducted on the identification and quantification of phytoplankton fossil pigments through HSI (Butz et al., 2015; Butz et al., 2017; Wolfe et al., 2006). Rein and Sirocko (2002) originally developed a HSI technique to measure chl-a within marine sediments. Achieving an r^2 of 0.94. The subsequent work conducted by and Butz et al. (2015) and Butz et al. (2017) revealed that HSI could reliably identify and map bacteriopheophytin-a and sedimentary green pigments (combined chl-a with its diagenetic products) within lake sediment core samples. When this hyperspectral data was calibrated against HPLC-DAD pigment concentrations it produced an error of 10-15%. Other work to have calibrated HSI to chl-a concentration includes Makri et al. (2020), who achieved a calibration within 9%, and Zander et al. (2021) who calibrated to within 5.4%. Additionally, using HSI Yackulic (2017) and Favot et al. (2020) detected signals from the phycobiliprotein phycocyanin in sediment core samples from Crater Lake, Colorado and Blue Chalk Lake, Ontario, respectively.

A major challenge for hyperspectral analyses is the attribution of spectral properties to substances found in the sediment (Butz et al. (2015). This is primarily because lake sediments contain numerous other substances which could cause variations in spectral signals. Additionally, HSI lacks the ability to differentiate chl-a from its degradation products and to resolve trace pigments (Wolfe et al., 2006). HSI analyses are not ideal for sedimentary carotenoids, as they absorb at overlapping wavelengths making it unviable to distinguish between the carotenoids. Furthermore, HSI possesses limitations in the shorter wavelengths (400 – 470 nm) (Butz et al., 2015). As such the calibration of

spectral properties of the sample needs to be established quantitively. For this reason, the results obtained from the HSI should be calibrated using HPLC-DAD.

The $RABD_{660-670}$ was originally identified by Rein and Sirocko (2002) For HSI analysis to measure chl-a within marine sediments achieving a r^2 of 0.94. This analysis was then applied by Butz et al. (2015) on freshwater sediments achieving an error of 10-15% for the calibration of 'green pigments' (chl-a and pheophytin a) to HPLC-DAD analysed concentrations. This algorithm used a ratio between a weighted average for two absorption bands at each end of the specific absorption feature and the absorption band minimum (Butz et al., 2015). To select the most appropriate absorption bands for calculating the spectral indices Butz et al. (2015) utilised a continuum removal on the spectral endmembers. A continuum removal is a normalization procedure used for the quantifying absorption features where the overall concave shape of the spectra is removed. This resulted in a means to calculate an index of relative absorption band depths (RABD). Schneider et al. (2018) also calibrated HSI to 'green pigments' (chl-a together with its diagenetic products) with an error of 12%. Other work to have calibrated HSI to chl-a concentration includes Makri et al. (2020) achieved a calibration within 9%, and Zander et al. (2021) who calibrated to within 5.4%.

To identify phycocyanin within lake core sediment samples the $RABD_{615}$ signal can be used. $RABD_{615}$ was originally applied to HSI by Yackulic (2017) as this corresponds to phycocyanin peak absorption. Favot et al. (2020) noted phycocyanin within the 588-642 wavelength and Sorrel et al. (2021) also reported phycocyanin $RABD_{615}$, following the analysis for phycocyanin as described by Yacobi et al. (2015).

Chapter 3 Methods

3.1 Site selection

3.1.1 Sediment core samples

Lake sediment cores from five New Zealand lakes were provided for this study by the "Our Lakes' Health: Past, Present, Future" funded by the Ministry of Business, Innovation and Economy Endeavour Fund programme (referred to as Lakes380 hereafter). The five lakes are Lake Nganoke, Lake Okataina, Lake Oporoa, Lake Pounui, and Lake Rototoa (see Figure 3.1 for locations and Appendix C for specific characteristics of the lakes). These lakes were selected as they were study lakes for other Lakes380 projects, covered a range of elevations (from 14 to 358 m a.s.l.), and a range of depths, from shallow lakes at less than 10 m depth (Lake Nganoke) to deep lakes over 10 m depth, with the deepest at 78.5 m deep (Lake Okataina). These five lakes include two riverine lakes (Lake Oporoa and Lake Nganoke), one tectonic/landslide formed lake (Lake Pounui), one volcanic lake (Lake Okataina), and one wind/aeolian lake (Lake Rototoa).

Lake core samples were collected using an Uwitec gravity corer with hammer, with polyvinyl chloride barrels at 90 mm diameter and 2 m length. Prior to coring, the barrels were cleaned with 2% sodium hypochlorite. Once the core was retrieved, they were sealed and stored at 4°C in darkness until required (up to 4 weeks). The cores were prepared and analysed by HSI at GNS Science, Avalon, Lower Hutt, where they were halved manually and photographed, then sediment type and colour were described and recorded. One half of the core was analysed via HSI, the other half of selected cores were analysed by other analytical techniques including HPLC-DAD and spectrophotometry, with 40 – 42 sub-samples per core of approx. 2 g taken from the centre of other half at 1cm intervals and various depths, then stored at -20°C until required for analysis.

3.1.2 Surface sediment samples

A sample set of 223 New Zealand Lakes (≈ 6% of the lakes in New Zealand) was used for the analysis of surface sediments (see Figure. 3.2 for locations). These lakes ranged from low elevation lakes (<10 m) to high elevation lakes (up to 1,839 m) and included coastal shoreline/lagoons, glacial, riverine, swamp/wetland formed, tectonic/landslide formed, volcanic, and wind/aeolian lakes. The catchments ranged from 35,288 m^2 to 704,470,618 m^2 and included shallow lakes (<10 m) to deep lakes (up to 445 m) (see Appendix C for specific characteristics of the lakes). These have been divided into two subsets for the statistical analysis: A) Lakes with TLI values calculated by the Lakes 380 team (95 lakes), and B) Lakes with unknown TLI values (127 lakes).

Figure 3.1. *Locations of the five lakes where core samples used for this study were collected from.*

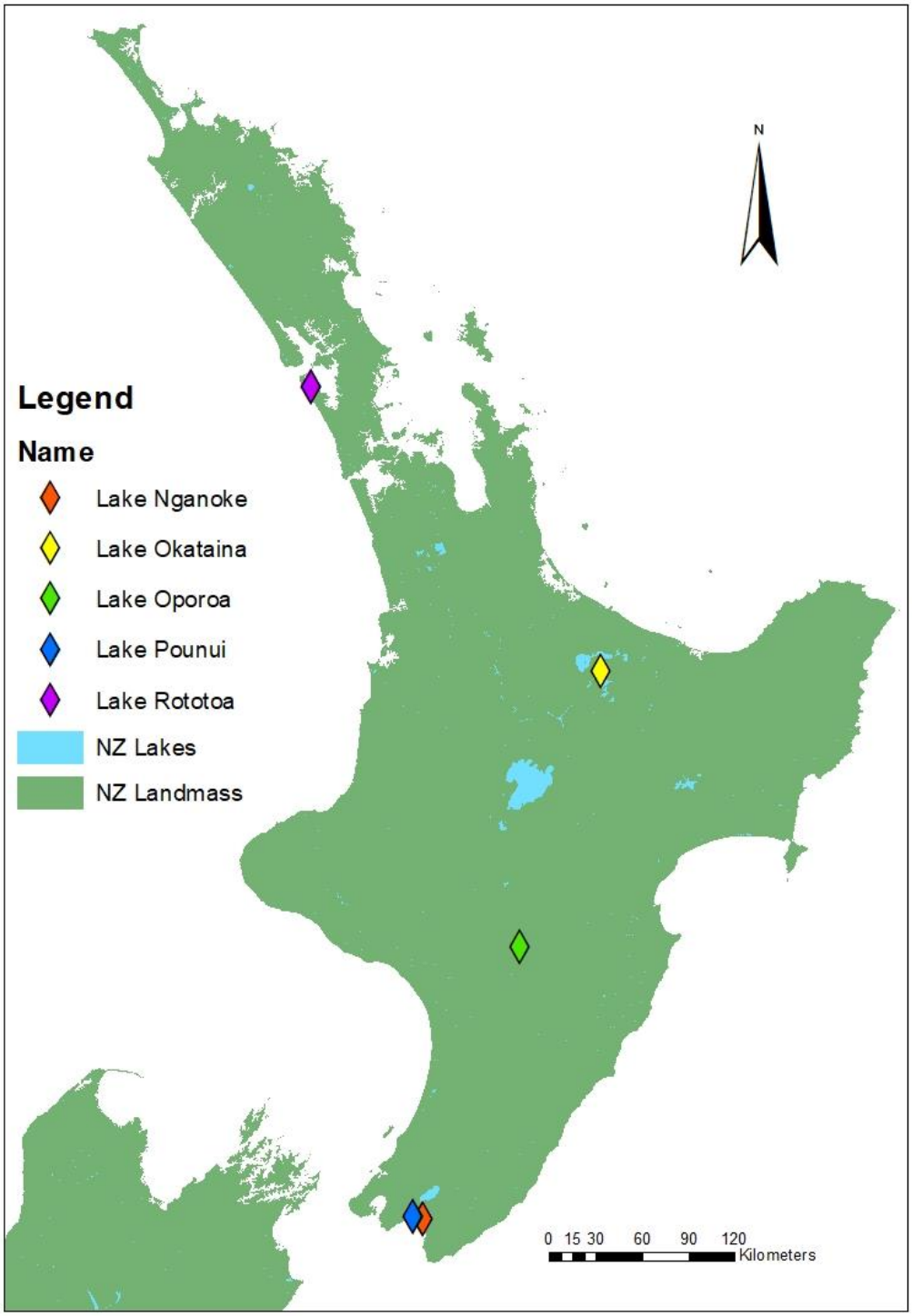
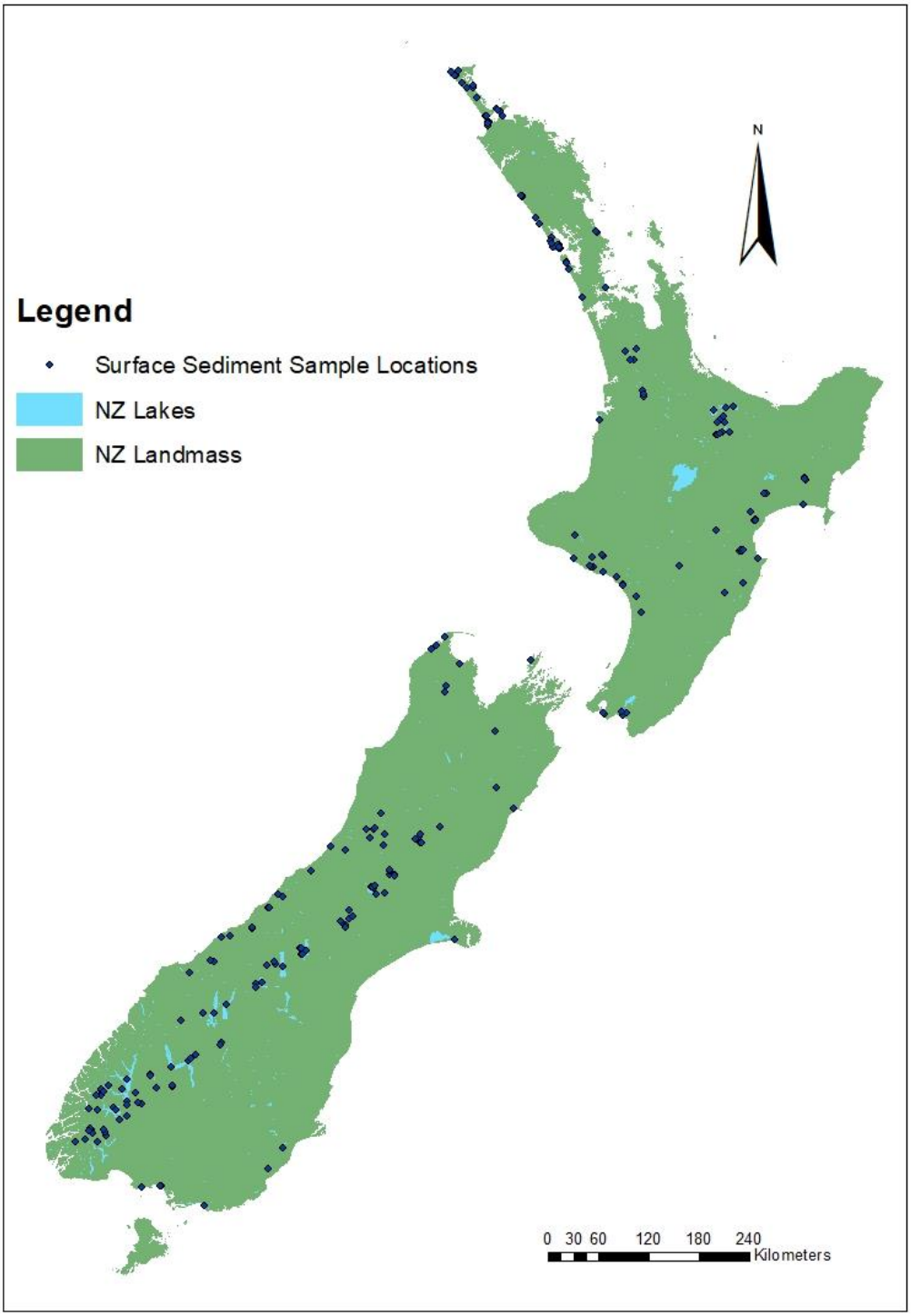


Figure 3.2. *Locations of the 223 lakes where surface sediment samples used for this study were collected from.*



3.2 Hyperspectral Imaging

Claire Shepherd, at GNS Science, Avalon, Lower Hutt conducted the hyperspectral analyses, with $RABD_{660\text{-}670}$ and $RABD_{615}$ values supplied for this analysis. After the lake sediment cores were halved, the half intended for HSI was scraped to ensure a flat, smooth surface. This core half was then scanned using a Specimen SISU core scanner with a sCMOS-50-V10E spectral camera.

Following the calculation from Butz et al. (2015) (Equation 3.1) HSI image pre-processing and normalisation was performed, with spatial resolution set to 1mm/pixel at a spectral resolution of 660-670 nm for chl-a and 615 nm for phycocyanin sampled at 48 µm intervals. These results were then converted to spectral indices, calculated in R (version 3.4.0), possessing a range of values between 1 - 2.5. The spectral indices were then binned (averaged to 1 cm resolution) to correspond with the analytical chemistry resolution from the second half of the core.

Equation 3.1. *RABD Spectral Index Equation as Described in (Butz et al., 2015)*

This calculation determines a weighted average using the two bands situated at each end of the absorption feature. In addition to ratio calculated between this weighted average and the absorption band minimum.

$$RABD_{Feature\,MIN} = \left(\frac{X \times R_{Left} + Y \times R_{Right}}{X + Y}\right) \div R_{Feature\,MIN}$$

"[W]here $RABD_{Feature\ MIN}$ *= relative absorption band depth at absorption feature minimum,* R_{Left} *= reflectance at the start of the absorption feature,* R_{Right} *= reflectance at the end of the absorption feature,* $R_{Feature\ MIN}$ *= reflectance at the minimum of the absorption feature,* X *= number of spectral bands between* $R_{Feature\ MIN}$ *and* R_{Right}*, and* Y *= number of spectral bands between* $R_{Feature\ MIN}$ *and* R_{Left}*" Butz et al. (2015, p. 7).*

3.3 Pigment Analysis

Sediment samples from both the lake cores and surface sediments were thawed at 4°C then subsampled for each respective analysis. 40 – 42 samples per core were selected and analysed via spectrophotometry, HPLC-DAD, and fluorescence. This amount of core samples analysed was limited due to the batch size and run-time required by the HPLC (55 min per sample). The water content of each sample was determined by weighing a subsample of sediment (≈0.5 g; to 4 d.p.), freeze-drying and weighing again.

3.3.1 Pigment extraction for HPLC-DAD and spectrophotometric analysis

The carotenoids and chlorophylls in the sediment samples were extracted by adding 5 mL of solvent to ≈1 g subsample of wet sediment. Initially 100% methanol was used for the extraction of the surface sediment samples. However, this was later changed to 5 mL 100% acetone for the extraction of the lake sediment core samples. The samples were vortexed, then sonicated for 30 min with ice to avoid methylation of the pigments.

Samples were clarified by centrifugation at 3,000 × g and 10°C for 10 min. The supernatant was transferred to a 20 mL glass vial and the process of adding 5 mL of solvent, clarification, and transfer was repeated two more times for a final extraction volume of 15 mL. The extract was dried under nitrogen gas at 40 °C and stored away from light at −20°C prior to resuspension and analysis.

Prior to analysis the dried pigment extracts were resuspended in 0.5 mL of 100% acetone for the lake sediment core samples and 100% methanol for the surface sediment samples. The resuspended extract was centrifuged at 12,000 × g for 5 min to remove particulates. 200 µL of the resuspension was diluted into 800 µL 90% Ethanol for chl-a analysis by spectrophotometry while the remainder of the undiluted resuspension was analysed by HPLC-DAD.

3.3.2 Spectrophotometric Analysis of Chlorophyll-a

3.3.2.1 Spectrophotometry with Chlorophyll-a acidification

The chl-a spectrophotometric analysis of the extracts was based on the standard operating procedure at Cawthron Institute (Puddick et al., 2018; Puddick et al., 2019). For analysis, 200 µL of the sample diluted in 90% Ethanol was plated in quadruplicate onto a 96-well plate (Corning, 96-well cell culture cluster 35960 flat bottom with lid) along with quadruplicates of a 20 mg/mL $CuSO_4$ solution for path-length calculation, a 1% H_2SO_4 path-length calculation blank, and a blank 90% Ethanol solution. Path-length calculation solutions were calibrated spectrophotometrically (Eppendorf BioSpectrometer 6137) prior to analysis with the calibration blanked using the 1% H_2SO_4 solution (Appendix A for solution preparation). Path-length was calculated following Equation 3.2.

Equation 3.2. *Calculation for Path-length of Extract in the Sample Well (Puddick et al., 2018; Puddick et al., 2019).*

$$l\ (cm) = \frac{\left[\left(\frac{[A_{665} - Blank_{665}]}{CuSO_{4\,665}}\right) + \left(\frac{[B_{750} - Blank_{750}]}{CuSO_{4\,750}}\right)\right]}{2}$$

Where, l = the path length (cm); A_{665} = Absorbance of $CuSO_4$ on the EnSpire platereader at 665 nm; B = Absorbance of $CuSO_4$ on the EnSpire platereader at 750 nm; Blank = 1% H_2SO_4 at 665 and 750 nm; $CuSO_4$ = as measured in the Eppendorf spectrometer at 665 and 750 nm.

Chl-a analysis was performed via spectrophotometry (Enspire Multimode Platereader, Thermo Fisher Scientific; Waltham, MA, USA) by shaking for 30 sec then obtaining a wavelength absorbance reading at 665 nm then at 750 nm. 5µL of 1M HCl was added to each sample on the plate and reloaded into the spectrophotometer. The samples were shaken for 30 seconds, and a series of readings taken every 2 minutes at 665 nm and

750 nm. Once the readings show no further degradation, the final reading is then used for the calculation. Chl-a in the extract was calculated following Equation 3.3a for concentration in the extract, then Equation 3.3b for the wet weight chl-a concentration in the lake sediment sample.

Equation 3.3. *Calculations for Wet Weight Chlorophyll-a Concentration via spectrophotometry with acidification (Puddick et al., 2018; Puddick et al., 2019).*

a) $$Chl-a\ (in\ extract;\ mg/L) = \frac{[(B_{665} - B_{750}) - (A_{665} - A_{750})] \times 28.66 \times dil}{l}$$

Where, l = the path length calculated in Equation 3.2 (cm); B_{665} = absorbance at 665 nm before acidification; B_{750} = absorbance at 750 nm before acidification; A_{665} = absorbance at 665 nm after acidification; A_{750} = absorbance at 750 nm after acidification; 28.66 = as defined by Sartory and Grobbelaar (1984) this absorbtion coefficient is established from the specific absorption coefficient of 83.4 L/g/cm and an acid ratio of 1.72, both for chl-a in ethanol;and dil = dilution factor of extracts (generally = 1 (no further dilution of extract), if detector saturation occurred further dilution ensued).

b) $$Chl-a\ (in\ sediment;\ \mu g/g) = \frac{Chl-a\ (in\ extract) \times ResVol \times 1{,}000}{Weight}$$

Where, Chl-a (in extract) = concentration of chl-a in the extract calculated in equation 3.3a (mg/L); ResVol = Resuspension volume (in L; generally, 0.0005 L or 0.5 mL); 1,000 = Conversion of mg into µg; and Weight = Weight of sediment sample (in g)

3.3.2.2 Spectrophotometry of Chlorophyll a without acidification

The pigment extracts were analysed via UV-VIS spectrophotometry for UV/visible light absorption throughout the range from 220 to 1,000 nm. The extracts were diluted in 100% methanol for the surface sediment samples and 100% acetone for the lake sediment core samples (1:10 initially for all samples. However, if detector saturation occurred further dilution ensued) and measured using a ThermoFisher Evolution 201 benchtop spectrophotometer (Waltham, MA, USA). Absorbance values at 662 and 665 nm were used to calculate Chl-a concentrations following the formula described by Lichtenthaler and Buschmann (2001). Chl-a in extract was calculated following Equation 3.4a for concentration in the extract, then Equation 3.4b for the wet weight chl-a concentration in the lake sediment sample. As the solution being analysed without acidification is a mixture of all pigment the chl-b absorbance peak can add to the chl-a absorbance peak at the measured wavelength. To counter act this addition, the influence of chl-b to the chl-a absorbance peak is subtracted.

Equation 3.4 *Calculations for Wet Weight Chlorophyll-a Concentration via spectrophotometry without acidification (Lichtenthaler & Buschmann, 2001).*

A) *100% Acetone:*

$$Chl-a(in\ extract;\ \mu g/mL) = \left(11.24\ A_{(661.6-750)} - 2.04\ A_{(644.8-750)}\right) \times dil$$

100% Methanol:

$$Chl-a(in\ extract;\ \mu g/mL) = 16.72\ A_{(665.2-750)} - 9.16\ A_{(652.4-750)}$$

Where, A(661.6-750) = absorbance for chl-a in acetone minus the measurement at 750 nm for turbidity normilisation; A(644.8-750) = absorbance for chl-b in acetone minus the measurement at 750 nm for turbidity normilisation; A(665.2-750) = absorbance for chl-a in methanol minus the measurement at 750 nm for turbidity normilisation; A(652.4-750) = absorbance for chl-b in methanol minus the measurement at 750 nm for turbidity normilisation, and dil = dilution factor of extracts (genreally 1:10 or 10). The path length of the cuvette (cm) was 1 cm and not specified as the equation was reduced.

B) $$Chl-a\ (in\ sediment;\ \mu g/g) = \frac{Chl-a\ (in\ extract) \times ResVol}{Weight}$$

Where, Chl-a (in extract) = concentration of chl-a in the extract calculated in equation 3.4a (mg/L); ResVol = Resuspension volume (in L; generally, 0.0005 L or 0.5 mL); Weight = Weight of sediment sample (in g).

3.3.3 HPLC-DAD Analysis of Key Phytoplankton Pigments

HPLC-DAD analysis of the samples was conducted using an Agilent 1260 system (Santa Clara, CA, USA). The HPLC-DAD program used follows the methodology described in Puddick et al. (2021). Pigment separation on the HPLC-DAD was achieved using a C_{30} column (Develosil RP-Aqueous C30, 5-µm, 250 × 4.6 mm; Phenomenex, Torrance, CA, USA) at 30°C. The solvents used were (Appendix A for solution preparation):

- Solvent A: methanol + 0.1% triethylamine
- Solvent B: 40:60 methanol/isopropyl alcohol + 0.1% triethylamine

Prior to each injection the C_{30} column was washed with 90% solvent B for 5 min before being re-equilibrated with 100% solvent A for 5 min. Then 10 µL of sample was injected at a flow rate of 1 mL/min following the solvent system in Table 3.1.

Table 3.1. *Solvent System Used for Analysis*

Time (min)	Flow rate (mL/min)	Solvent A (%)	Solvent B (%)	Conditions
0	1	100	0	Injection
5	1	100	0	Stable state
40	1	35	65	Linear gradient
42	1	10	90	Steep gradient
47	1	10	90	Column wash
50	1	100	0	Steep gradient
55	1	100	0	Equilibration

Data was collected over 320 to 800 nm for spectrophotometric absorption, however the specific wavelengths for the pigment quantitation are described in Table 3.2.

Table 3.2. *Specific Wavelength Parameters for Data Collection of the Target Pigments*

	Pigment	Wavelength (nm)	Retention time (min)	Equivalence factor
Chlorophylls	Chl-a	655-675	23.1	NA
	Chlorophyll-b	460-480	15.2	NA
Xanthophylls	Alloxanthin	435-455	11.8	1.2
	Canthaxanthin	460-480	16.1	1.06
	Diadinoxanthin	435-455	8.0	1.21
	Diatoxanthin	435-455	10.9	1.26
	Echinenone	445-465	32.9	0.63
	Fucoxanthin	435-455	4.8	0.55
	Lutein	435-455	9.6	NA
	Myxoxanthophyll	460-480	12.7	1.26
	Peridinin	460-480	4.1	0.72
	Violaxanthin	435-455	5.9	0.88
	Zeaxanthin	435-455	10.3	1.56

A mixed standard with five known concentrations (20, 10, 5, 2, and 0.5 µg/mL) was created containing Chl-a (Sigma-Aldrich, St. Louis, MO, USA), chlorophyll-b (Sigma-Aldrich) and lutein (Carotenature, Münsingen, Switzerland) and analysed with each HPLC-DAD run. Additionally, each HPLC-DAD run contained four qualitative standards to verify the retention times for alloxanthin, canthaxanthin, diadinoxanthin, diatoxanthin, echinenone, fucoxanthin, myxoxanthophyll, peridinin, violaxanthin and zeaxanthin.

Prior to the standard mixtures being created, the target pigments were calibrated using extinction coefficients described in Roy et al. (2011), at 450 nm (Alloxanthin), 478 nm (Canthaxanthin), 445 nm (Diadinoxanthin), 452 nm (Diatoxanthin), 458 nm (Echinenone), 443 nm (Fucoxanthin), 445 nm (Lutein), 472 nm (Myxoxanthophyll), 475 nm (Peridinin), 443 nm (Violaxanthin), and 453 nm (Zeaxanthin) using a ThermoFisher Evolution 201 benchtop spectrophotometer (Waltham, MA, USA).

Rather than preparing a standard curve for all thirteen target pigments for routine quantification, equivalence factors were determined for Alloxanthin, Canthaxanthin, Diadinoxanthin, Diatoxanthin, Echinenone, Fucoxanthin, Myxoxanthophyll, Peridinin, Violaxanthin, and Zeaxanthin. The equivalence factor was determined in relation to lutein by analysing the known concentrations of these standards in conjunction with the known concentration of a lutein standard (Table 3.2).

Standard curves, or calibration curves are prepared to predict the unknown concentration of the target compound in a solution. These quantitative curves are regressions of relationships between prepared solutions with varying known concentrations (normally five different concentrations) and response (area of the feature's peak) using a least squares method. However, equivalence factors are used where a compound's feature peak is known to be a certain proportion of another. A standard mixture is used as quality control in analytical chemistry. Where a mixture of compounds is created with known concentrations and is analysed alongside each HPLC-DAD run. This ensures unknown variables are not interfering with the quantification of the samples.

3.3.4 Fluorescence Analysis of Phycocyanin

Phycocyanin has a characteristic fluorescence emission which does not interfere with chlorophyll, making analysis uncomplicated (Zamyadi et al., 2012). Fluorometrically phycocyanin will excite at 580-590 nm with an emission at 645-650 nm (Bellinger & Sigee, 2015; Roy et al., 2011). A qualitative presence/absence fluorometric analysis was developed to detect phycocyanin. To confirm that this fluorescence analysis was able to detect the presence of phycocyanin, aqueous extracts from 16 cyanobacteria cultures known to produce phycocyanin from the Cawthron Institute Culture Collection of Micro-

algae (Rhodes et al., 2016) were analysed with the florescence method. The recovery of the phycocyanin by fluorescence from the cyanobacteria cultures (Table 3.3), showed that the fluorescence method was able to detect phycocyanin concentrations within these cultures.

Table 3.3. *Cyanobacteria pigments recovered via fluorescence from culture extracts.*

Taxonomy	Zeaxanthin (ug/L)	Echinenone (ug/L)	Chl-a via HPLC (ug/L)	Phycocyanin (ug/L)
Microcystis aeruginosa	6.8	1.8	37.5	145.7
Microcystis aeruginosa	20.9	7.3	83.0	415.6
Microcystis sp.	1.7	3.5	40.9	64.0
Microcystis wesenbergii	9.4	1.6	31.3	198.4
Nodularia spumigena	1.3	18.7	124.6	358.9
Leptolyngbya sp.	19.0	0.9	74.5	357.3
Leptolyngbya sp.	22.5	8.5	99.2	297.1
Microcoleus autumnalis	21.7	3.7	77.1	469.5
Microcoleus autumnalis	10.0	10.8	133.0	217.1
Microcoleus glaciei	2.5	5.4	55.8	13.4
Nostoc sp.	0.9	1.0	45.0	352.8
Nostoc sp.	1.5	3.8	46.2	314.3
Planktotrix sp.	50.9	9.5	131.9	444.0
Planktotrix sp.	55.5	9.0	103.5	368.1
Scytonema cf crispum	2.6	3.5	21.5	261.8
Scytonema sp.	0.3	3.1	20.8	24.2

To determine the optimum number of extraction repetitions, sixteen samples of surface sediment from Lake Ponui were weighed (≈1 g) and batched into five groups of triplicate samples. These sediment samples were extracted in a phosphate buffer (50 mM, pH 7.2; 2 mL) (Appendix A for solution preparation) of which 2 mL was added to each sample, and sonicated for 30 min with ice, then frozen at −20°C. The frozen sample was sonicated for 30 min with ice and clarified by centrifugation at 12,000 x g for 5 min. Samples in batch one underwent this extraction once, batch two underwent this extraction twice, batch three underwent this extraction three times, and batch four was

four times. These samples were analysed using the fluorescence platereader method below. After the initial extraction, phycocyanin levels decreased notably in the subsequent extractions (some with zero pigment extracted), indicating that only one extraction cycle was required to extract the majority of the phycocyanin from the sediment.

The phycocyanin in the sediment samples were extracted by adding 2 mL of phosphate buffer to ≈1 g subsample of wet sediment. Samples were sonicated for 30 min with ice, then frozen at −20°C. The frozen sample was sonicated for 30 min with ice and clarified by centrifugation at 12,000 x g for 5 min to ensure all particulates were removed.

Sample extracts (200 µL) were analysed in duplicate on a black 96-well plate (Costar 96-well black flat bottom non-treated polystyrene) using a fluorescence platereader (Enspire Multimode Platereader, Thermo Fisher Scientific; Waltham, MA, USA), along with standards and blanks (also in duplicate). The phycocyanin was measured at an excitation wavelength (ʎ-ex) of 609 nm and an emission wavelength (ʎ-em) of 643 nm. The phycocyanin standards (Sigma-Aldrich; St. Louis, MO, USA) were resuspended in phosphate buffer and calibrated spectrophotometrically at 615 nm using extinction coefficients 5.92 L/g/cm (Bennett & Bogorad, 1973). A standard curve was prepared using phycocyanin at 20, 10, 5, 2, and 1 µg/mL.

To verify the detection of phycocyanin by HSI ($RABD_{615}$) in sediment core samples, a phycoliboprotein analysis via fluorescence was performed on lake core sediment samples from Lakes Nganoke, Okataina, Oporoa, and Rototoa.

3.4 Using ArcGIS to Extract Land-Use Proportions Within the Lake Catchment

Land-use data was collected using ArcGIS desktop 10.6.1 software by Esri, using the New Zealand National Digital Elevation Model (DEM) at 25-meter resolution for both the North Island (Landcare Research, 2018a) and the South Island (Landcare Research, 2018b). The ArcGIS Hydrology toolset was used to fill the DEM for each Island, calculate the flow direction, then the flow accumulation. Pour points for each lake were manually created based on the flow accumulation data and the Watershed tool was run for each lake to outline the individual watersheds (see Appendix B for the ArcGIS model).

These individual watersheds were then used as the ‘clip feature’ with the ArcGIS clip tool, with the LCDB v5.0 - Land Cover Database version 5.0, Mainland New Zealand (LCDB) (Landcare Research, 2019) as the Feature input overlaid with the respective watershed then clipped. This data was then exported as an Excel file and the shape area was converted into percentages of the various classifications (Appendix B for ArcGIS

model). The LCDB provides 33 land classifications (excluding the two classes specific to the Chatham Islands).

3.5 Statistical Analyses

All statistical analyses were conducted using RStudio Team (2021) version 4.1.1.

3.5.1 Hyperspectral Calibration

The HSI data the RABD value was calibrated to the concentration of chl-a obtained through either HPLC-DAD or spectrophotometry, through statistical regression following the calibration performed by Butz et al. (2015). The results for the chl-a within the selected lake cores were analysed as a wet weight then statistically compared to Hyperspectral $RABD_{660\text{-}670}$ results. Regressions and cross-validations were performed using the R packages caret (Kuhn, 2021) (an aggregator package for statistical analysis), dplyr (Wickham et al., 2021) (to read the "%>%" function) and ggplot2 (Wickham, 2016) (for model visualisation).

To determine whether HSI chl-a calibration required HPLC-DAD analysis, or it could be performed via spectrophotometry, a linear regression was performed on the chl-a concentrations obtained via HPLC-DAD and by spectrophotometry for the lake core sediment samples for each of the five lakes. Chl-a concentrations obtained via HPLC-DAD and via spectrophotometry were then applied to a linear regression against the $RABD_{660\text{-}670}$ values. The regressions were assessed to determine the optimum analyses for a calibration.

This optimum analysis was then applied to a calibration of the $RABD_{660\text{-}670}$ values where the combined concentrations of samples for all five lakes were assessed for linearity and outliers via a linear regression plot, and the regressions were assessed for linearity or quadratic relationships. A cross-validation was performed for the individual regression for each of the five lakes involving a leave one out cross validation (LOOCV), a k-fold cross validation (10-fold), a repeated k-fold cross validation with 10-folds repeated ten times (k-fold), and a bootstrap with 100 resamples for each lake.

To assess the feasibility of a universal $RABD_{660\text{-}670}$ calibration a linear regression was performed on a combined dataset of the chl-a concentrations for the core sediment samples for the five lakes. Lakes were then selectively removed from the dataset to ascertain individual influences on the combined dataset.

3.5.2 Phycocyanin Detection

Four lakes cores were analysed for phycocyanin using fluorescence in parallel with HSI, chl-a analyses (HPLC-DAD and spectrophotometry), and key xanthophyll pigments

using HPLC-DAD. The results for the phycocyanin within the selected lake cores were analysed as a wet weight then statistically compared using a linear regression to $RABD_{615}$ results. Statistical regressions were performed using the R package caret (Kuhn, 2021) and ggplot2 (Wickham, 2016).

$RABD_{615}$ values were linearly regressed against phycocyanin concentrations to determine relationships. Additionally, $RABD_{615}$ values were linearly regressed against chl-a via spectrophotometry concentrations, as well as Echinenone and Zeaxanthin concentrations determined via HPLC-DAD, to assess the likelihood of the $RABD_{615}$ signal picking up on these pigments rather than phycocyanin.

3.5.3 Creating a Predictive Model for Lake Trophic Status

Data normalisation was performed using base R functions (RStudio Team, 2021) with the package ggplot2 (Wickham, 2016). Distribution of pre- and post- transformation of model variables can be found in Appendix D. Pigment concentrations were normalised by transforming them using a natural logarithm of the given value plus one. The TLI values possessed a normal distribution, so were not transformed. Models were created using the surface sediment subset A (Lakes with known TLI's, see section 3.1.2). R codes for the predictive models can be found in Appendix E.

3.5.3.1 Regression Tree

Regression trees were created following Boehmke (2018c). The creation of the tree utilised the packages rsample (Silge et al., 2021) (for data splitting), dplyr (Wickham et al., 2021) (for data wrangling), rpart (Therneau & Atkinson, 2019) (for building the regression trees), rpart.plot (Milborrow, 2021) (for plotting the regression trees), and both ipred (Peters & Hothorn, 2021) and caret (Kuhn, 2021) (both for 'bagging' or validating the accuracy of the regression trees). Three regression trees were modelled to predict lake trophic level index (TLI), using 1) land use proportions from the lake's catchment (calculated by ArcGIS), 2) thirteen phytoplankton pigments: twelve measured by HPLC-DAD (Alloxanthin, chl-a via HPLC, Chlorophyll-b, Canthaxanthin, Diadinoxanthin, Diatoxanthin, Echinenone, Fucoxanthin, Lutein, Peridinin, Violaxanthin, and Zeaxanthin); and one by spectrophotometry (chl-a), termed the 'pigments-only data', and 3) thirteen phytoplankton pigments plus five physical lake characteristics that would not change through time, or would only change very slowly (Elevation, Lake Area, Lake Volume, Maximum Depth, and Catchment Area), termed the 'pigments +' data.

Data was first split into 'test' and 'training' (30:70) partitions. The initial tree was created using the 'rpart' package. In the background this package computes the values of which to prune the tree using a 10-fold cross validation. This will then compute the error of the

initial tree. Once created, the tree is then tuned using a grid search of hyperparameters which are manually supplied. The parameters tuned were the 'minsplit' and the 'maxdepth.' The minsplit is the minimum number of data points required to create a terminal node after it is forced to attempt a split, and the maxdepth is the maximum number of internal nodes between the terminal nodes and the root node. The optimum parameters were selected from the resulting table output, displaying the top results resulting in the lowest errors. This table will also display the optimum complexity parameter (cp) which controls the size of the regression tree. The minsplit, maxdepth, and cp values are used to create the optimum regression tree.

This optimised tree was bagged (bootstrap aggregation) to improve predictive performance. The process of bagging combines and averages multiple regression trees reducing the variability of a single tree and reducing overfitting. The 'ipred' package was used to bag the tree, where 25 bootstrap replicas were used to compute the root mean squared error (RMSE) on a portion of the training dataset. The optimised tree was also bagged using the 'caret' package. Caret preforms a 10-fold cross-validation on the tree and calculates the most influential variables (variable importance). The variable importance is computed by package assessing the total sums of squares error (SSE) decrease as each split in the tree occurs, over a given predictor, averaged over all the bagged trees created. The variable importance value is a relative mean decrease in SSE when compared to the most important variable on a 0-100 scale.

3.5.3.2 Random Forest Model

Random forest models are a modification of bagging regression trees. When bagging regression trees, they can be susceptible to correlation as they are not completely independent from one-another and possess similar structures. Random forest models overcome this correlation by creating a large collection of de-correlated trees. This results from this collection of trees is then averaged to produce the result (Boehmke, 2018b).

Random forest models were created following Boehmke (2018b). The creation of the models utilised the R packages rsample (Silge et al., 2021), RandomForest (Liaw & Wiener, 2002) (for basic implementation of the random forest model), ranger (Wright & Ziegler, 2017) (which provides faster implementation of the RandomForest package), caret (Kuhn, 2021) (an aggregator package for executing various machine learning models), dplyr (Wickham et al., 2021) (to read the "%>%" function) and igraph (Csardi & Nepusz, 2006) (for visualisation of the random forest model). Two random forest models

were developed to predict lake TLI using 1) using the pigments-only data, and 2) using the 'pigments +' data.

Data was first split into 'test' and 'training' (30:70) partitions. The initial tree was created using the 'RandomForest' package, throughout the model creation the RandomForest package will perform a built-in bootstrap validation on the model in the background. The model is then tuned using the 'tuneRF' function of the RandomForest package. To refine the tuning and find the optimum mtry, minimum node size (which controls the complexity of the trees), and sample size used for the model training a grid search of various hyperparameters was performed. Each hyperparameter combination was looped through and applied to 500 trees. The lowest ten RMSE results were displayed with the corresponding parameters.

Once the optimal hyperparameters are identified they can be input into the model parameters for an accurate error rate to be calculated. This also allows the calculation of the variable importance of each of the variables in the same manner as the rpart package. This optimal tree can them be used to predict using the test sub-sectioned data and thus calculate the final RMSE of the model.

3.5.3.3 Boosted Regression Tree

Where a random forest builds a collection of deep independent trees, a gradient boosted model (boosted regression tree) builds a sequential group of shallow weak trees, with each tree learning and being improved from the previous tree. This collection of trees is averaged sequentially to produce the result (Boehmke, 2018a). Boosted regression trees utilise stochastic gradient descents to optimise the loss function. This is achieved by overcoming potential non-uniform distributions of loss functions by sampling a fraction of the training observations, without replacement, and using this subsample to calculate the next time (Boehmke, 2018a).

The boosted regression tree model was created following Boehmke (2018a). The creation of the model utilised the packages rsample (Silge et al., 2021), gbm 2.1.8 (Greenwell et al., 2020) (for basic implementation of the gradient boosted regression tree), caret (Kuhn, 2021), and both pdp (Greenwell, 2017) and ggplot2 (Wickham, 2016) (both for model visualisation). One boosted regression tree was modelled to predict lake TLI using the using the 'pigments +' data.

Data was first split into 'test' and 'training' (30:70) partitions. The initial model of 10,000 trees was created using the 'gbm.fit' function of the 'gbm' package. This initial number of trees can be altered to achieve the lowest 5-fold cross validation (cv) error. This cv error is automatically calculated by the program in the background. The model was then tuned

using a hyperparameter grid search to identify the optimal parameters for the model. The parameters optimised are the 'shrinkage' (learning rate of the tree), 'interaction.depth' (complexity of the tree), 'n.minobsinnode' (minimum number of observations allowable for the trees terminal nodes) and the 'bag.fraction' (when <1 this allows for stochastic gradient descent). Using the results from the first grid search, a second grid search was conducted using the result from the first search to optimise the hyperparameters which will provide the minimum error.

The 'relative.influence' function of the gbm package is then used to calculate the relative influence. Like variable importance, relative influence calculates which variables possess the largest average decrease in mean squared error (MSE), and therefore are deemed more important. This model with its optimum parameters can then be used to predict on the test fraction of the data to calculate the RMSE using the 'caret' package.

To develop a deeper understanding of the data variables, partial dependence plots can be created to understand how the different response variables are affected by the different predictor variables.

3.5.4 Lake Health Predictions

The three machine learning models (regression tree, random forest model, and boosted regression tree) created to predict lake TLI using 'pigments +' data were compared using linear regression to determine which was more accurate at predicting lake TLI. After the model with the lowest error is identified, it is applied to the surface sediment subset B (see section 3.1.2), to predict the TLI values.

The model with the lowest error was then applied to the downcore sediment samples for Lake Nganoke, Lake Okataina, Lake Oporoa, Lake Pounui, and Lake Rototoa (see section 2.1.1). As the retrodicted lake TLI's do not possess reliable validation data, the pigments-only model with the lowest error was applied to the downcore sediment samples, to contrast the model with the lowest error and determine whether it presented a realistic scenario of historical lake health. Additionally, the retrodicted lake TLI's were contrasted with log transformed chl-a concentrations within the core sediments, and cyanobacteria indicator pigment concentrations (Table 2.3) (sum of Canthaxanthin, Echinenone, Myxoxanthophyll, and Zeaxanthin concentrations).

Chapter 4 Results

4.1 Hyperspectral Calibration

4.1.1 Chlorophyll-a Calibration

4.1.1.1 High-Performance Liquid Chromatography with Diode-Array Detection or Spectrophotometry Analysis

Linear regressions were performed to assess the relationship between chl-a concentrations determined by HPLC-DAD, and by spectrophotometry (both with acidification and without acidification). This revealed strong and statistically significant relationships in all the five lakes analysed ($R \geq 0.42$, $p < 0.05$; Table 4.1). Additionally, the 95% confidence interval (CI) reveals groupings of the lakes as indicated by overlap. The differences between the regressions can most likely be attributed to different chl-a diagenesis products (i.e., pheophorbide-a and pheophytin-a) within each lake. As spectrophotometry without acidification measures all chl-a diagenesis products these relationships will be different to HPLC-DAD and spectrophotometry with acidification as they specifically measure the chl-a compound. Additionally, the differing concentrations between HPLC-DAD and spectrophotometry with acidification, while it can be assumed that they are both measuring the chl-a compound, are different analytical techniques. HPLC-DAD is a highly specific analysis method targeting particular chemical compounds, while spectrophotometry with acidification is not as specific as HPLC-DAD and spectral interferences from other compounds within the lake sediment could be accounting for higher chl-a concentrations.

Additionally, this also shows that the chl-a has been preserved in a better state in three of the lakes (Lake Nganoke, Okataina, and Oporoa; see figure 4.1c). This preservation could be due to several factors unexplored in this research, such as, anoxic conditions, temperature, lake water depth, etc. A comparison of the linear regressions for all five lakes can be seen in figure 4.1.

Table 4.1. *Regression Relationships of Chlorophyll-a Concentrations Analysed by High-Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD) and Spectrophotometry both With and Without Acidification.*

A) *HPLC-DAD (µg/g) to spectrophotometry with acidification (µg/g).* ***B)*** *HPLC-DAD (µg/g) to those analysed by spectrophotometry without acidification (µg/g).* ***C)*** *Spectrophotometry without acidification (µg/g) to those analysed by spectrophotometry with acidification (µg/g).*

A)

Lake	R	r^2	p-value	Intercept	Gradient	CI
Nganoke	0.91	0.82	$< 2.2 \times 10^{-16}$	1.86	5.74	4.90, 6.60
Okataina	0.99	0.97	$< 2.2 \times 10^{-16}$	1.73	2.31	2.19, 2.43
Oporoa	0.84	0.69	1.6×10^{-11}	-0.16	3.04	2.39, 3.70
Ponui	0.55	0.28	1.8×10^{-3}	0.62	0.21	0.09, 0.34
Rototoa	0.86	0.73	1.4×10^{-12}	1.29	3.15	2.53, 3.77

B)

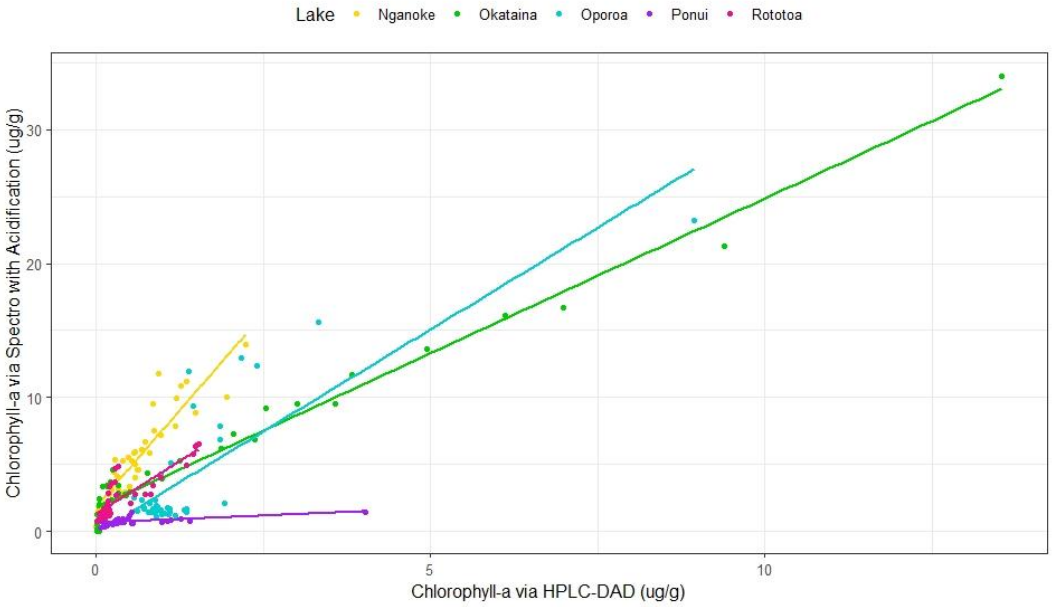
Lake	R	r^2	p-value	Intercept	Gradient	CI
Nganoke	0.77	0.58	2.8×10^{-9}	17.92	38.30	28.1, 48.5
Okataina	0.93	0.86	$< 2.2 \times 10^{-16}$	18.03	10.22	8.9, 11.5
Oporoa	0.72	0.50	2.0×10^{-7}	16.01	65.90	44.8, 87.0
Ponui	0.42	0.15	2.3×10^{-2}	23.53	4.70	0.7, 8.7
Rototoa	0.87	0.76	1.7×10^{-13}	10.91	19.33	15.8, 22.8

C)

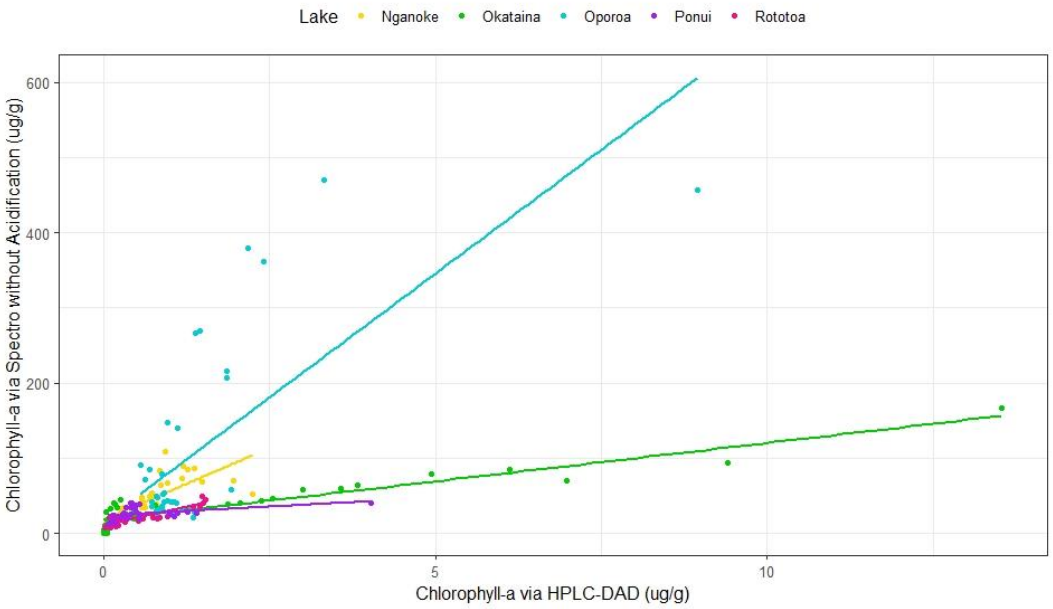
Lake	R	r^2	p-value	Intercept	Gradient	CI
Nganoke	0.92	0.84	$< 2.2 \times 10^{-16}$	2.50	7.20	6.2, 8.2
Okataina	0.97	0.95	$< 2.2 \times 10^{-16}$	9.61	4.56	4.2, 4.9
Oporoa	0.97	0.93	$< 2.2 \times 10^{-16}$	8.19	24.45	22.3, 26.6
Ponui	0.88	0.76	3.5×10^{-10}	7.12	25.64	20.1, 31.1
Rototoa	0.90	0.81	1.4×10^{-15}	4.80	5.43	4.6, 6.3

Figure 4.1. *Linear Regressions of Chlorophyll-a Concentrations Analysed by High-Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD) and Spectrophotometry both With and Without Acidification.*

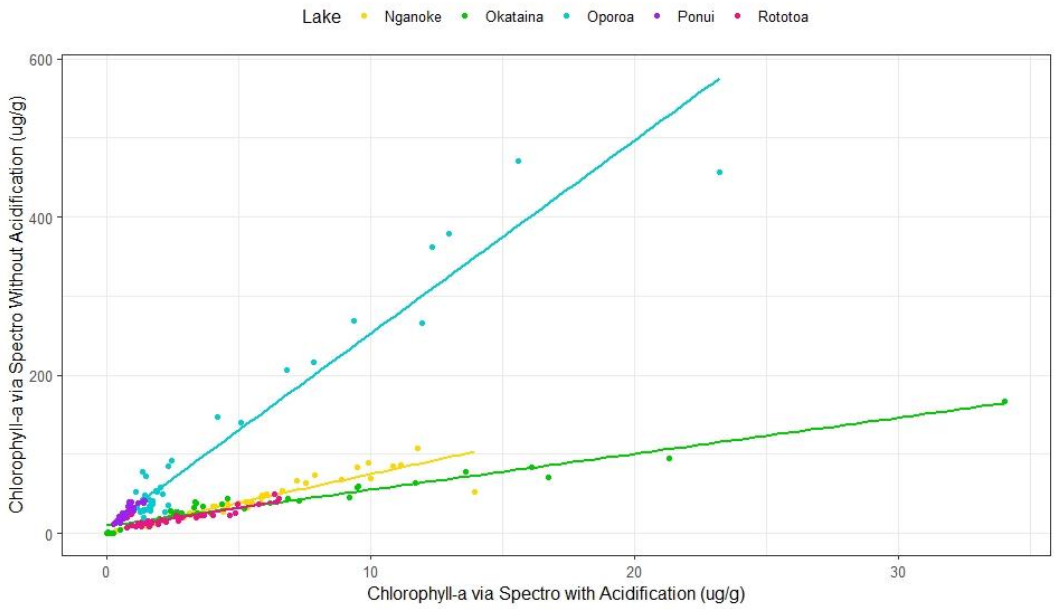
A) *HPLC-DAD (µg/g) to spectrophotometry (spectro) with acidification (µg/g).* ***B)*** *HPLC-DAD (µg/g) to spectrophotometry (spectro) without acidification (µg/g).* ***C)*** *Spectrophotometry (spectro) with acidification (µg/g) to spectrophotometry (spectro) without acidification (µg/g).*



A)



B)



C)

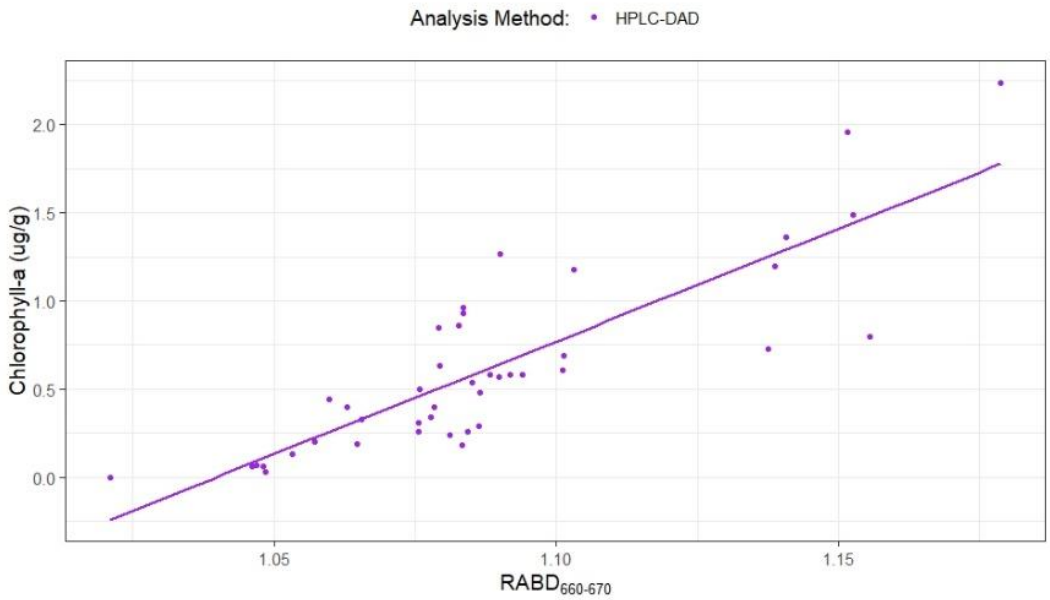
To better evaluate the three methods of chl-a analyses to determine if one method is more optimum for a calibration to $RABD_{660\text{-}670}$ within lake core sediment samples, a comparison was made between linear regressions for $RABD_{660\text{-}670}$ to chl-a via HPLC-DAD and to chl-a via spectrophotometry both with and without acidification (Table 4.2). These regressions showed that for two of the five lakes, Nganoke (Figure 4.2), and Rototoa (Figure 4.5) the slope for the HPLC-DAD analysis possessed a better fit as indicated by the r^2. Additionally, $RABD_{660\text{-}670}$ to chl-a via spectrophotometry with acidification displayed higher effect sizes for Oporoa (Figure 4.4) and Ponui (Figure 4.6). Whereas only Okataina (Figure 4.3) demonstrated a higher effect size for $RABD_{660\text{-}670}$ to chl-a via spectrophotometry without acidification. When the three chl-a analysis methods (HPLC-DAD and spectrophotometry both with and without acidification) were regressed against the $RABD_{660\text{-}670}$ signal with the data from all five lakes combined the r^2 values were higher for chl-a analysed via spectrophotometry without acidification (0.61).

This reveals that the two methods of calibrating $RABD_{660\text{-}670}$ to either chl-a via HPLC-DAD or chl-a via spectrophotometry both with and without acidification, all three produce statistically significant correlations (p-value <0.05) and r^2 values above 0.27. However, chl-a analyses via spectrophotometry without acidification produces more consistent results.

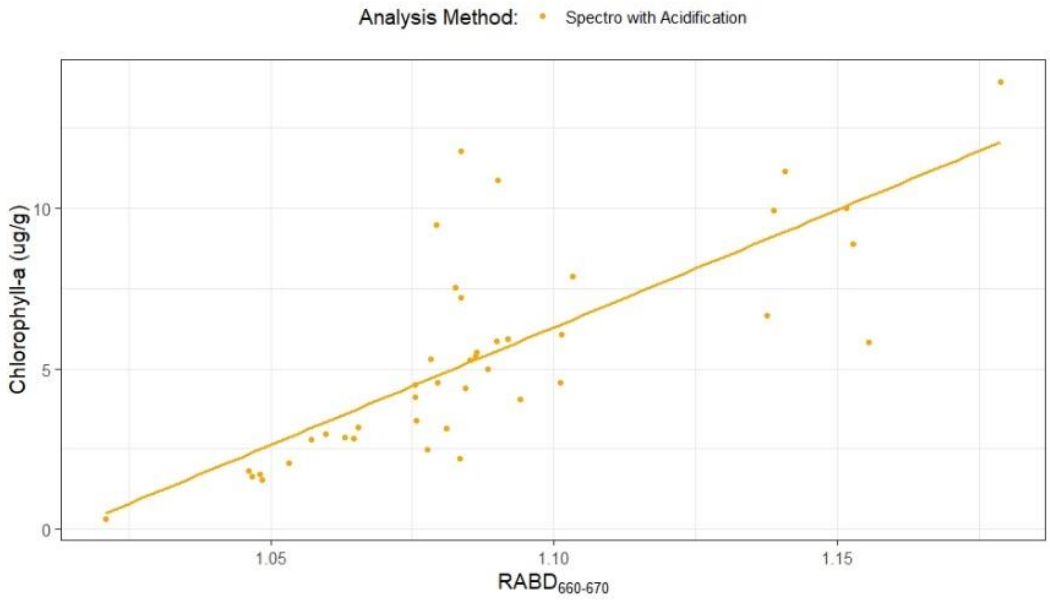
Table 4.2. *Comparison of $RABD_{660-670}$ Calibration to Chlorophyll-a Concentrations Analysed via High-Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD) and Spectrophotometry (Spectro) Both with and Without Acidification.*

Lake	Analysis Method	r^2	p-value	Intercept	Gradient	95% CI
Nganoke	**HPLC-DAD**	**0.71**	**1.9×10^{-12}**	**-13.3**	**12.8**	**10.2, 15.4**
	Spectro (acidification)	0.58	3.2×10^{-9}	-74.3	73.3	53.7, 92.9
	Spectro (non-acidification)	0.36	1.5×10^{-5}	-461.7	462.5	272.4, 652.6
Okataina	HPLC-DAD	0.41	2.8×10^{-6}	-32.0	30.6	19.3, 42.0
	Spectro (acidification)	0.51	7.1×10^{-8}	-81.5	79.3	55.0, 103.7
	Spectro (non-acidification)	**0.62**	**3.4×10^{-10}**	**-413.7**	**408.9**	**308.9, 508.8**
Oporoa	HPLC-DAD	0.59	3.9×10^{-9}	-9.2	7.9	5.8, 10.1
	Spectro (acidification)	**0.80**	**5.3×10^{-15}**	**-40.2**	**33.3**	**27.9, 38.8**
	Spectro (non-acidification)	0.72	3.6×10^{-12}	-956.7	801.0	638.5, 963.6
Ponui	HPLC-DAD	0.27	2.4×10^{-3}	-34.4	33.2	12.8, 53.5
	Spectro (acidification)	**0.38**	**2.4×10^{-4}**	**-14.9**	**14.9**	**7.7, 22.1**
	Spectro (non-acidification)	0.33	6.3×10^{-4}	-379.3	384.3	180.4, 588.3
Rototoa	**HPLC-DAD**	**0.86**	**$<2.2 \times 10^{-16}$**	**-11.6**	**11.2**	**9.7, 12.6**
	Spectro (acidification)	0.57	1.2×10^{-8}	-33.5	33.7	24.2, 43.1
	Spectro (non-acidification)	0.62	9.8×10^{-10}	-207.6	211.1	158.0, 264.2
Combined Lakes	HPLC-DAD	0.33	$<2.2 \times 10^{-16}$	-5.2	5.3	4.22, 6.36
	Spectro (acidification)	0.24	5.4×10^{-13}	-13.2	14.8	11.1, 18.6
	Spectro (non-acidification)	**0.61**	**$< 2.2 \times 10^{-16}$**	**-437.5**	**427.7**	**379.2, 476.2**

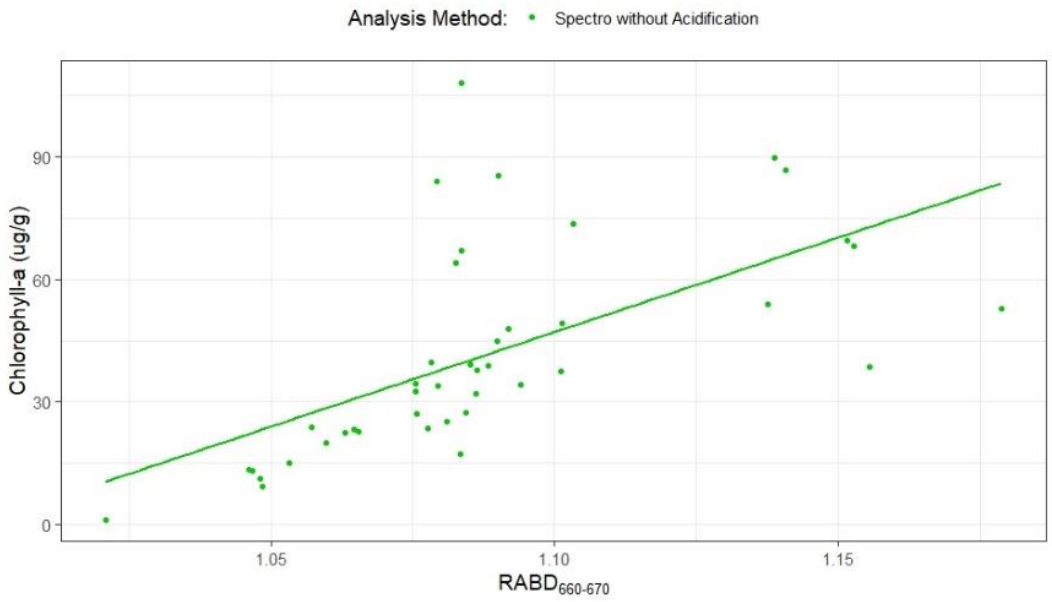
Figure 4.2. *Lake Nganoke Chlorophyll-a (chl-a) Analysis Comparison.*
A) *$RABD_{660-670}$ to chl-a via High-Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD),* ***B)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry with Acidification,* ***C)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry without Acidification*



A)

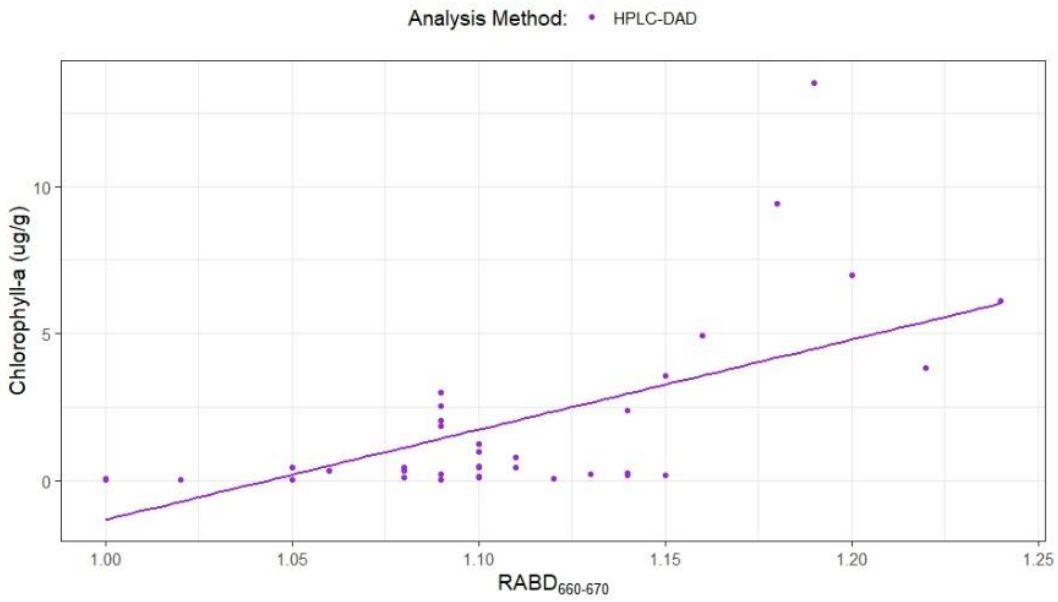


B)

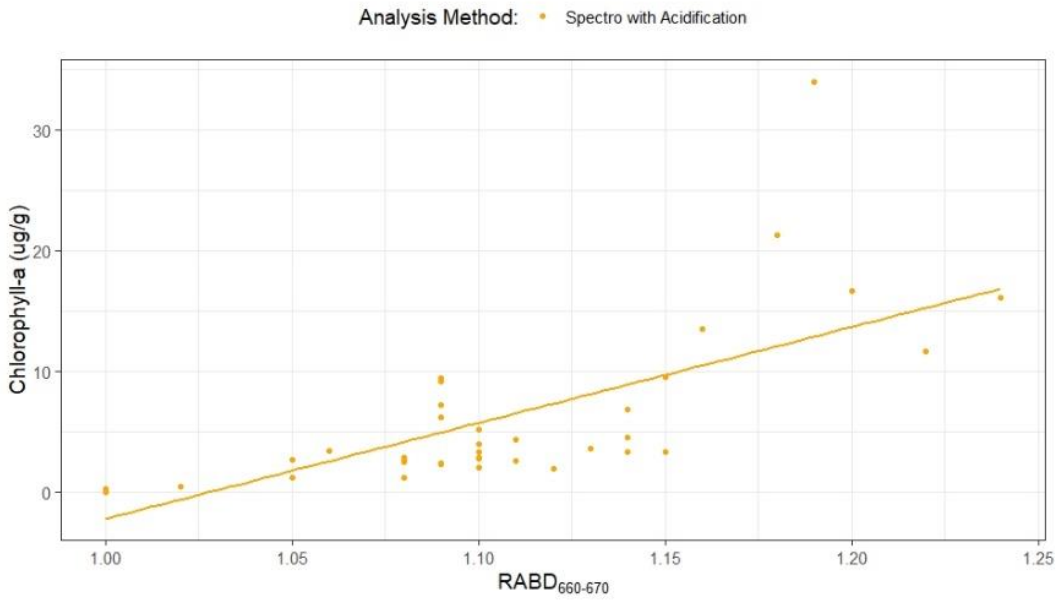

Analysis Method:
Spectro without Acidification
Chlorophyll-a (ug/g)
90
60
30
0
1.05
1.10
1.15
RABD660-670

C)

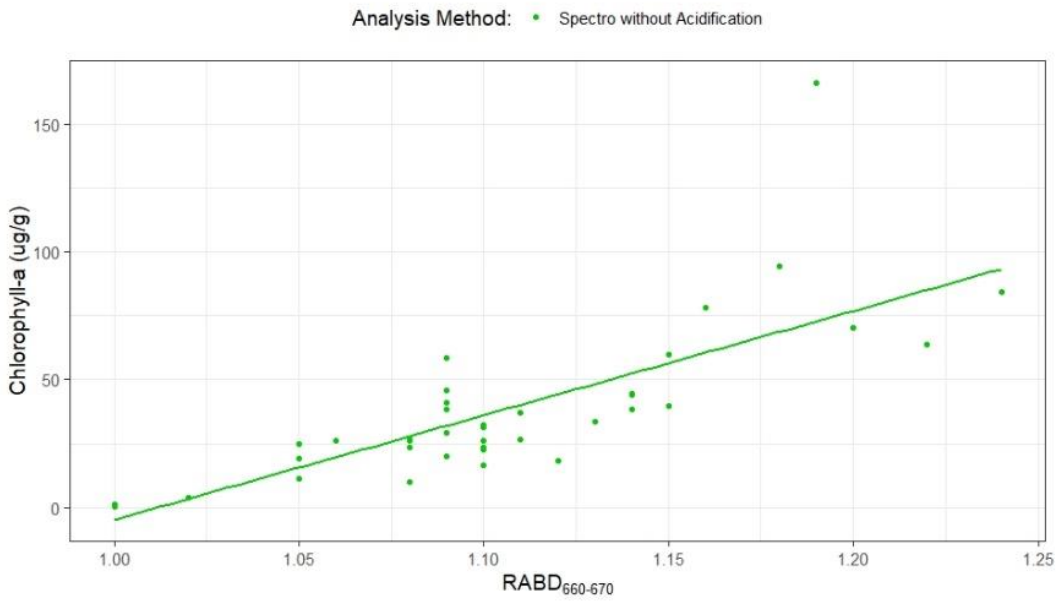
Figure 4.3. *Lake Okataina Chlorophyll-a (chl-a) Analysis Comparison.*
A) *$RABD_{660-670}$ to chl-a via High-Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD),* ***B)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry with Acidification,* ***C)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry without Acidification*



A)

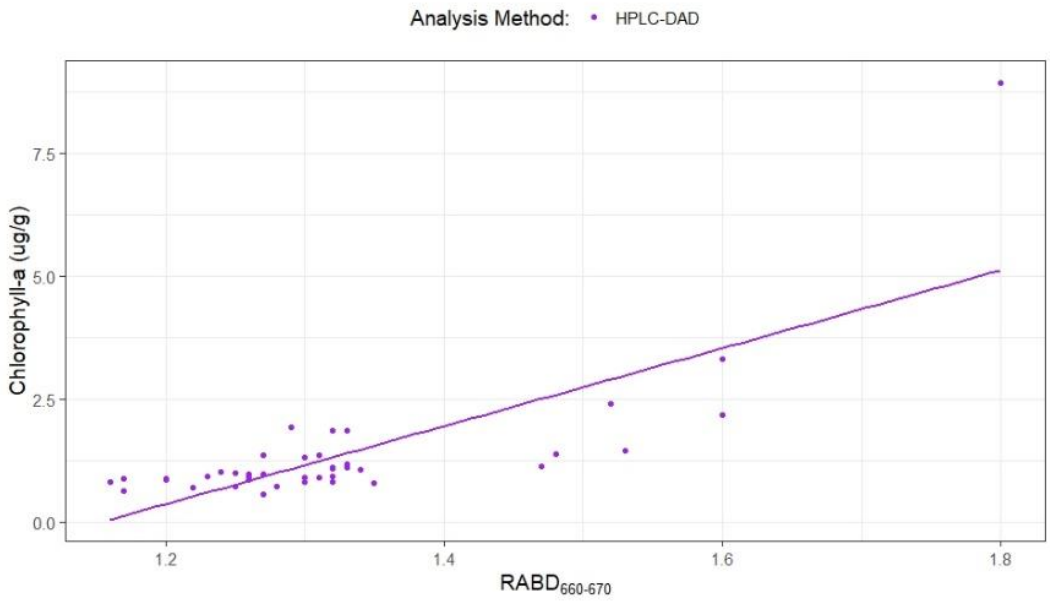


B)

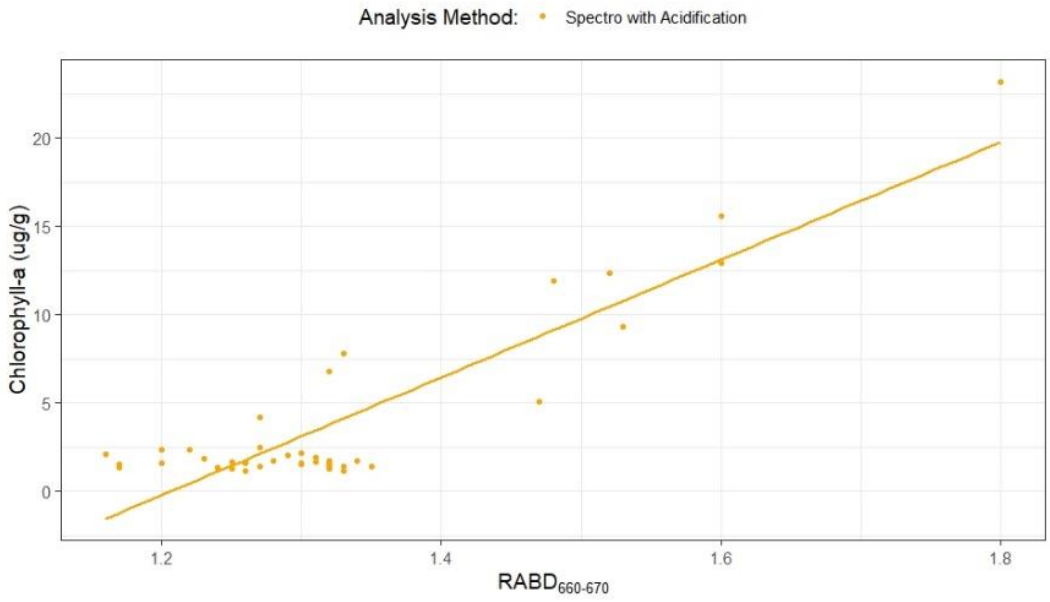

Analysis Method:
Spectro without Acidification
Chlorophyll-a (ug/g)
150
100
50
0
1.00
1.05
1.10
1.15
1.20
1.25
RABD660-670

C)

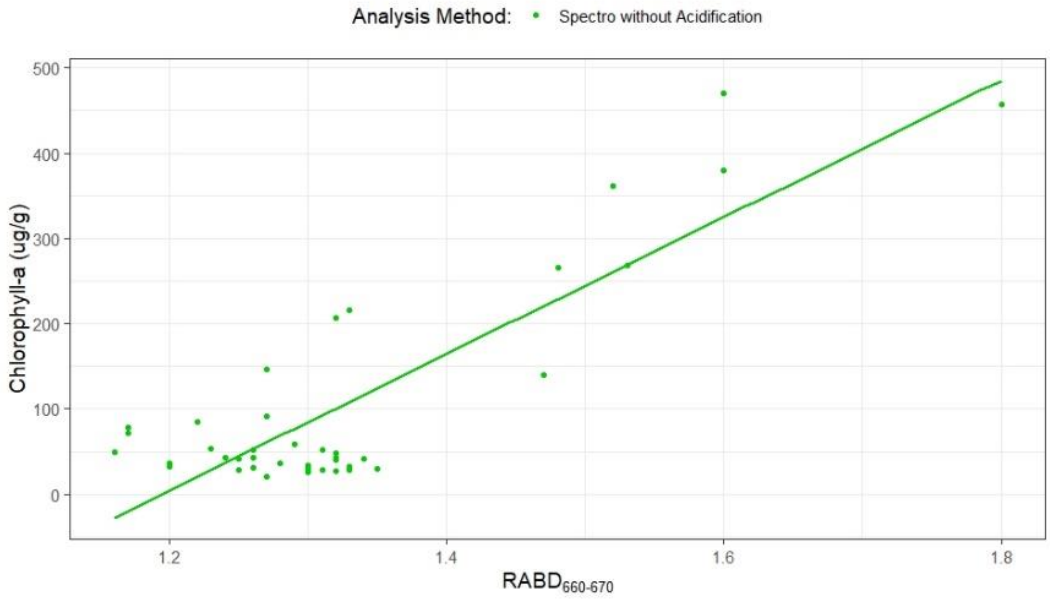
Figure 4.4. *Lake Oporoa Chlorophyll-a (chl-a) Analysis Comparison.*
A) *$RABD_{660-670}$ to chl-a via High-Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD),* ***B)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry with Acidification,* ***C)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry without Acidification*



A)

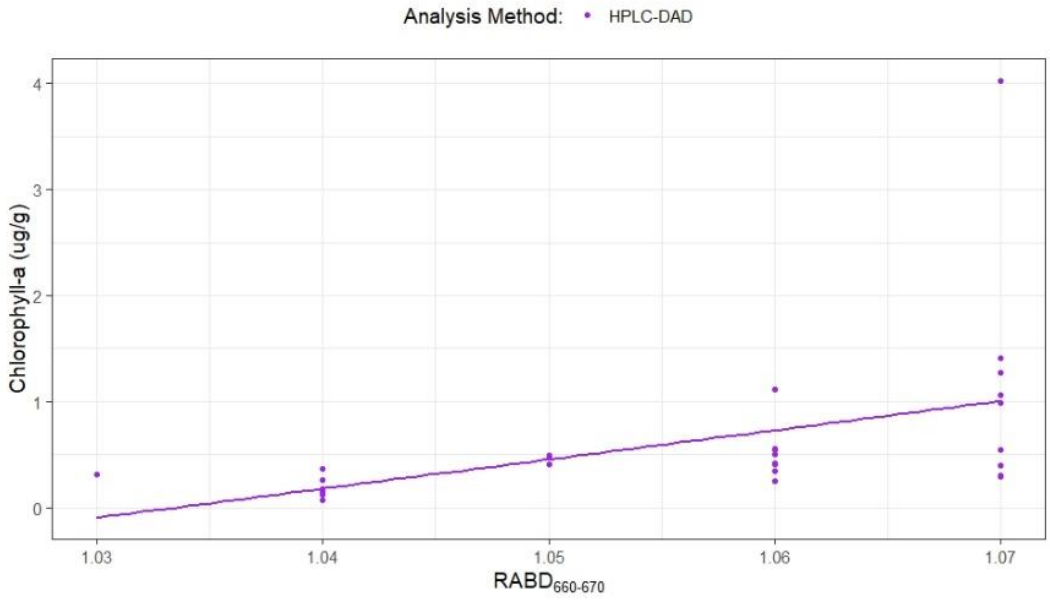


B)

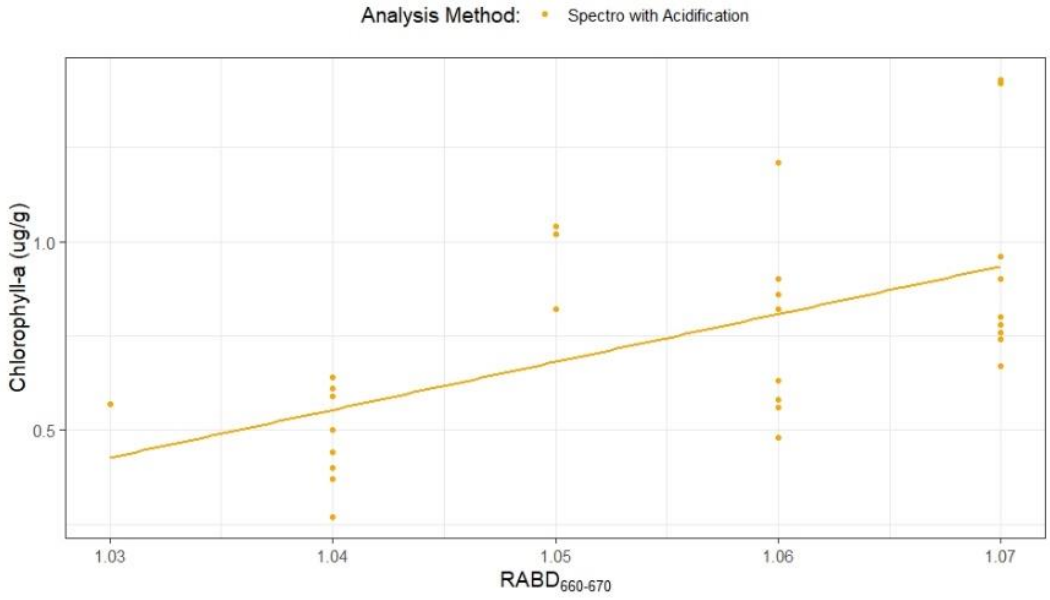

Analysis Method:
Spectro without Acidification
Chlorophyll-a (ug/g)
500
400
300
200
100
0
1.2
1.4
1.6
1.8
RABD660-670

C)

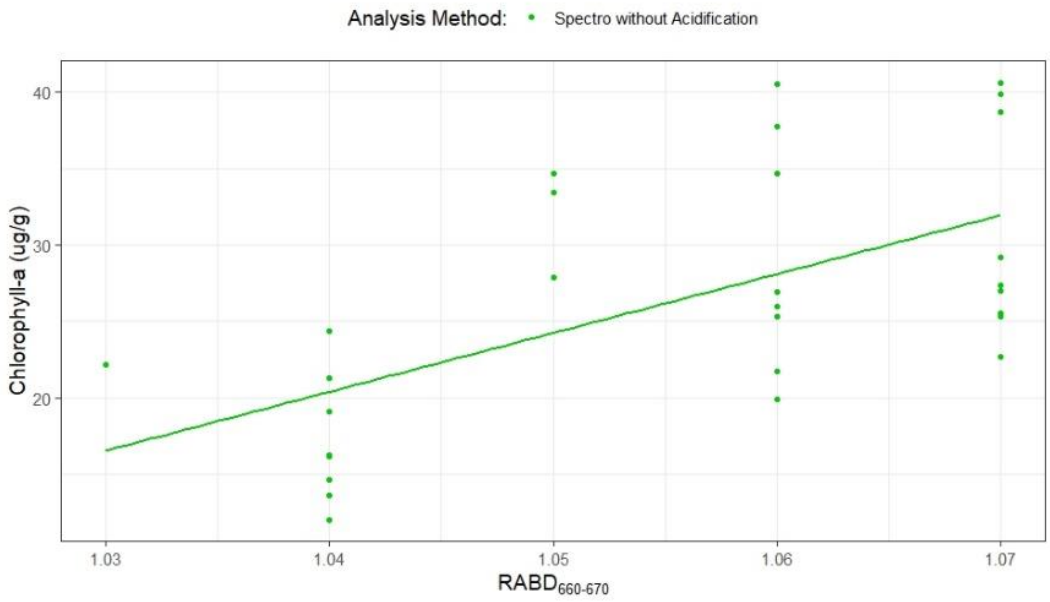
Figure 4.5. *Lake Ponui Chlorophyll-a (chl-a) Analysis Comparison.*
A) *$RABD_{660-670}$ to chl-a via High-Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD),* ***B)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry with Acidification,* ***C)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry without Acidification*



A)

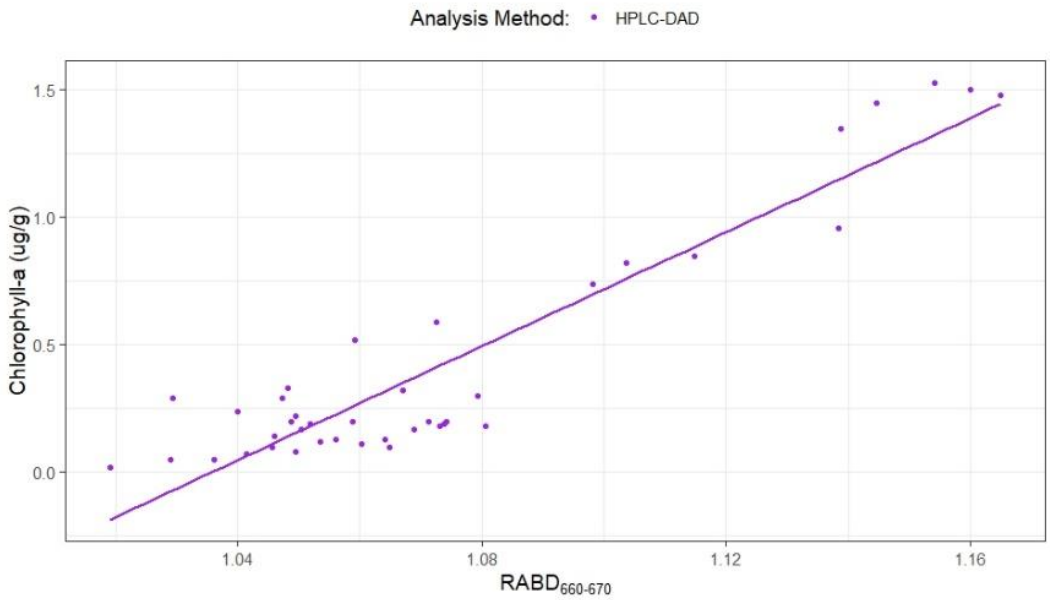


B)

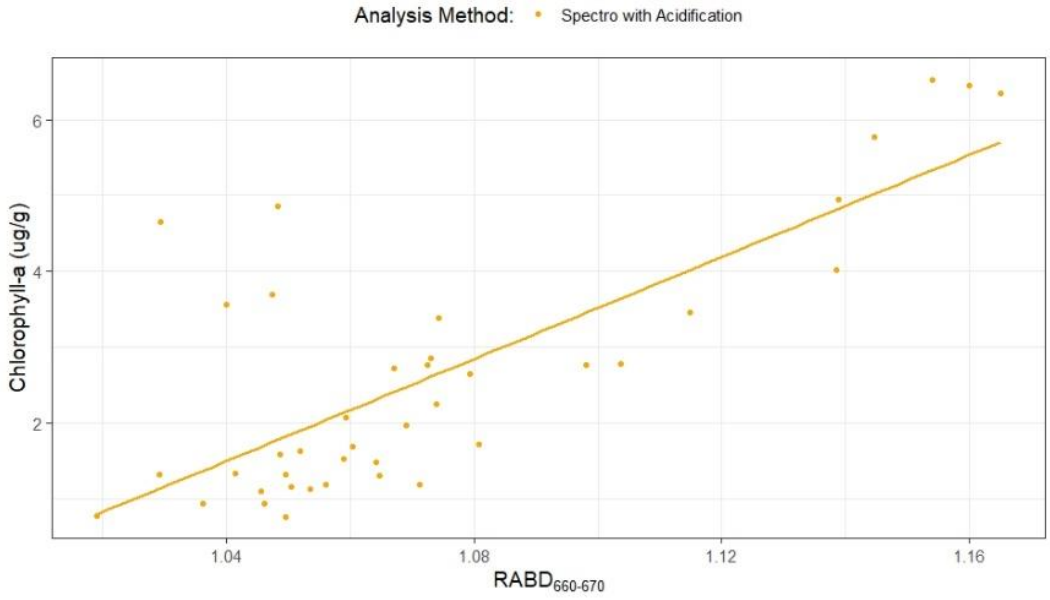

Analysis Method:
Spectro without Acidification
Chlorophyll-a (ug/g)
40
30
20
1.03
1.04
1.05
1.06
1.07
RABD660-670

C)

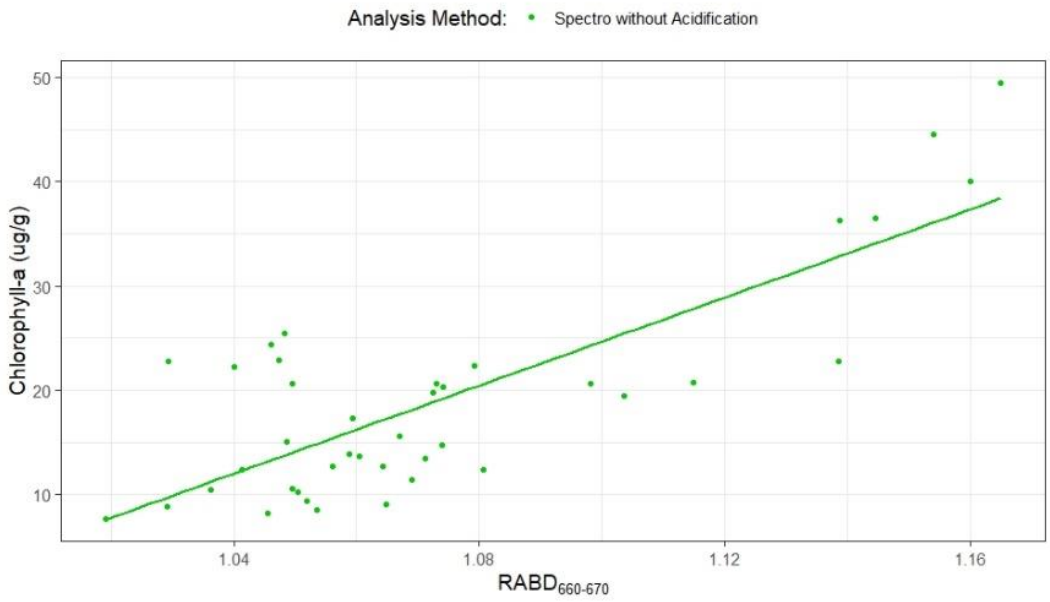
Figure 4.6. *Lake Rototoa Chlorophyll-a (chl-a) Analysis Comparison.*
A) *$RABD_{660-670}$ to chl-a via High-Performance Liquid Chromatography with Diode Array Detection (HPLC-DAD),* ***B)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry with Acidification,* ***C)*** *$RABD_{660-670}$ to chl-a via Spectrophotometry without Acidification*



A)



B)

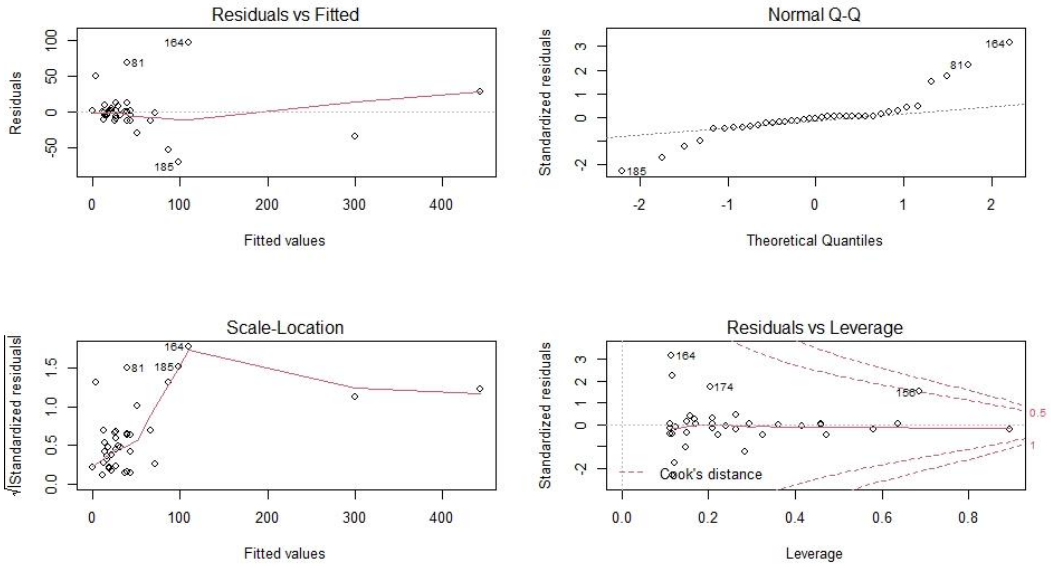

Analysis Method:
Spectro without Acidification
Chlorophyll-a (ug/g)
50
40
30
20
10
1.04
1.08
1.12
1.16
RABD660-670

C)

4.1.1.2 Calibration of RABD$_{660\text{-}670}$ to Chlorophyll-a Analysed via Spectrophotometry Without Acidification.

To validate the correlation of RABD$_{660\text{-}670}$ to chl-a analysed via spectrophotometry without acidification, linear regression diagnostic plots were assessed for linearity and outliers. Looking at the linear regression plots for the validation data from all lakes (Figure 4.7) the plots suggest a linear relationship with normal distribution in the data. However, one data point fell some distance from the rest of the data. This same data point possessed high leverage, but it did not fall outside of the Cooks' distance dashed line so was not treated as an outlier.

Figure 4.7. *Linear Regression Plot for the Combined Data of All Lakes RABD$_{660\text{-}670}$ to Chlorophyll-a Analysed via Spectrophotometry Without Acidification.*



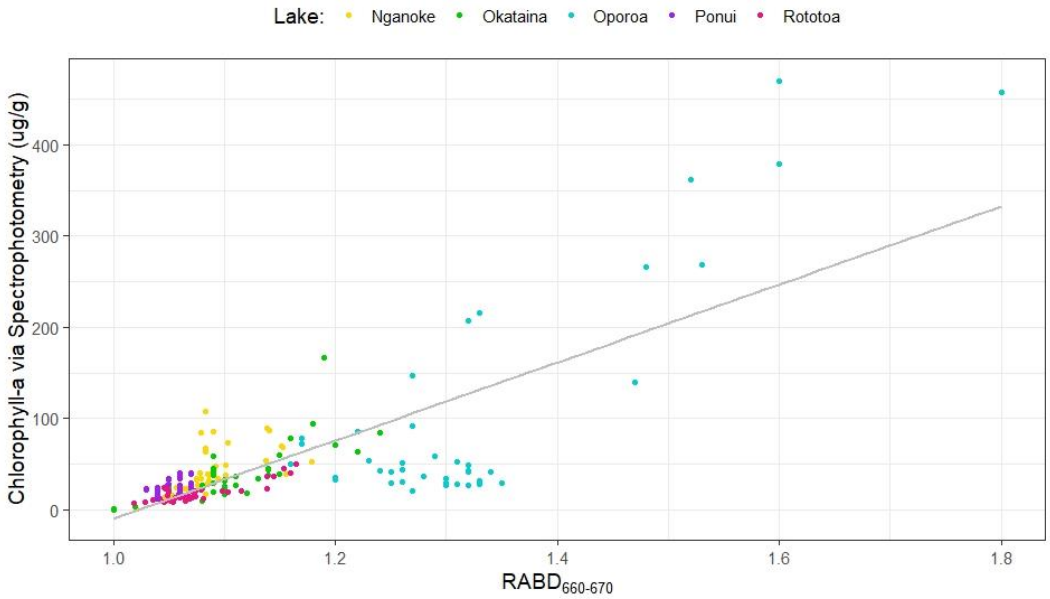
To further determine whether a calibration between the RABD$_{660\text{-}670}$ signal and the chl-a concentrations analysed via spectrophotometry without acidification required a linear regression, the combined data for the five lake cores were regressed quadratically. Both the linear and the quadratic regressions showed statistically significant p-values (Table 4.3). However, the r^2 value and the Akaike's Information Criteria (AIC) was slightly in favour of the quadratic regression (higher r^2, lower AIC). Despite this, the differences between the linear regression and the quadratic regression for the model of RABD$_{660\text{-}670}$ to chl-a via spectrophotometry without acidification were small, indicating little difference between the two regressions. Therefore, the calibration of RABD$_{660\text{-}670}$ to chl-a via spectrophotometry without acidification can be conducted using either linear or quadratic regression. However, non-linearity may not be detected within this dataset due to the noise of the data. The calibration of RABD$_{660\text{-}670}$ to chl-a via spectrophotometry without

acidification was validated on a linear regression (Figure 4.8). Figure 4.8 shows the linear regression for all five lakes; of note is the data from Lake Oporoa which deviates from the linear regression.

Table 4.3. *$RABD_{660\text{-}670}$ to Chlorophyll-a Concentrations Analysed via Spectrophotometry Without Acidification Calibration Linear Regression Compared to Quadratic Regression.*

	Linear Regression	Quadratic Regression
r^2	0.61	0.77
AIC	1909.7	1900.4
p-value	$< 2.2 \times 10^{-16}$	$< 2.2 \times 10^{-16}$

Figure 4.8. *Data from all Lakes $RABD_{660\text{-}670}$ Calibration to Chlorophyll-a Concentrations Analysed via Spectrophotometry Without Acidification.*
Model for a linear regression for all five lakes with a r^2 of 0.61, p-value of $< 2.2 \times 10^{-16}$, an intercept of -437.5, a gradient of 427.7, and CI of 379.23, 476.16.



To assess the accuracy of the linear models for all five lakes, a cross validation approach was undertaken for each lake where four validation methods were employed. The validation approaches involved a leave one out cross validation, a k-fold cross validation (10-fold), a repeated k-fold cross validation (10-fold, ten times), and a bootstrap with 100 resamples for each lake. The RMSE (root mean squared error), MAE (mean absolute error), and r^2 (coefficient of determination) results are presented in Table 4.4. The RMSE % was calculated by dividing the mean of the predictor variable ($RABD_{660\text{-}607}$) by the RMSE value. These values show that the models for all the lakes result in an uncertainty

of less than 8%. By returning small RMSE values demonstrates that the models are returning closely predicted data points to the observed data points. Additionally, the small MAE values indicate a close fit of the predicted data points to the observed data points. The r^2 values also indicate that the regression models fit the observed data relatively closely. Additionally, Table 4.4 shows that Lake Oporoa possess worse RMSE and MAE values than the other lakes. While the exact reasons for the variations in Oporoa's regression and validation results is unknown, it is worth noting that the lake has much higher chl-a concentrations within its sediments than the other lakes. These higher chl-a concentrations could be contributing to the unreliability of Lake Oporoa's results.

Table 4.4. *$RABD_{660-670}$ Calibration to Chlorophyll-a Concentrations Analysed via Spectrophotometry Without Acidification Model Validation Results.*
Validation methods performed included a leave one out cross validation (LOOCV), a k-fold cross validation (10-fold), a repeated k-fold cross validation with 10-folds repeated ten times (k-fold), and a bootstrap with 100 resamples for each lake.

Lake	Model Validation Method	RMSE	r^2	MAE
Nganoke	10-fold	0.02 (2.4%)	0.57	0.02
	k-fold	0.02 (2.5%)	0.61	0.02
	bootstrap	0.03 (2.8%)	0.41	0.02
	LOOCV	0.03 (2.8%)	0.36	0.02
Okataina	10-fold	0.04 (3.9%)	0.79	0.03
	k-fold	0.04 (4.0%)	0.80	0.03
	bootstrap	0.04 (4.4%)	0.72	0.03
	LOOCV	0.05 (4.6%)	0.48	0.03
Oporoa	10-fold	0.07 (6.9%)	0.61	0.06
	k-fold	0.07 (7.0 %)	0.71	0.06
	bootstrap	0.07 (7.3%)	0.70	0.06
	LOOCV	0.07 (7.4%)	0.68	0.06
Ponui	10-fold	0.01 (1.1%)	0.48	0.01
	k-fold	0.01 (1.0%)	0.63	0.01
	bootstrap	0.01 (1.1%)	0.37	0.01
	LOOCV	0.01 (1.1%)	0.29	0.01
Rototoa	10-fold	0.02 (2.3%)	0.56	0.02
	k-fold	0.02 (2.2%)	0.62	0.02
	bootstrap	0.02 (2.4%)	0.59	0.02
	LOOCV	0.02 (2.4%)	0.60	0.02

4.1.1.3 Universal Calibration

To assess the feasibility of a universal equation, data from all five lakes were combined. The linear regression relationship was statistically significant ($p = < 2.2 \times 10^{-16}$) and the r^2 value ($r^2 = 0.61$) indicated a regression comparable to the individual lake calibrations. When comparing the individual lake regressions on the same plot (Figure 4.8) the differences can be observed.

The $RABD_{660\text{-}670}$ values for Lake Oporoa are higher than the other four lakes, as well as the model intercept being different. However, the slope of the regression was comparable to the other lakes. Additionally, the regression line for Lake Ponui does not span as far through the chl-a via spectrophotometry without acidification concentrations as the other four lakes. Nevertheless, looking at the intercept and slope data (Table 4.5) it is feasible that data from Lake Oporoa is an outlier. However, when Lake Oporoa was removed from the dataset (Figure 4.9) the regression line for a universal calibration possessed a poorer fit to the data with a lower r^2. When Lake Ponui was removed from the dataset, there was no change to the r^2 value.

Table 4.5. *$RABD_{660\text{-}670}$ to Chlorophyll-a via Spectrophotometry Linear Regression r^2 Values. r^2 values for all five lakes individually, as well as, for the linear regression of all five lakes together, the regression of four lakes with Oporoa excluded, and the regression of four lakes with Ponui excluded.*

Lake	r^2
Nganoke	0.36
Okataina	0.62
Oporoa	0.72
Ponui	0.33
Rototoa	0.62
All 5 Lakes Model	0.61
4 Lakes Model excluding Oporoa	0.48
4 Lakes Model excluding Ponui	0.61

Figure 4.9. *Data from Four Lakes with Oporoa Excluded $RABD_{660-670}$ Calibration to Chlorophyll-a Concentrations Analysed via Spectrophotometry*
Model for a linear regression with a r^2 of 0.48, p-value of < 2.2 x 10^{-16}, an intercept of -373.4, a gradient of 374.2, and CI of 312.5, 436.0.

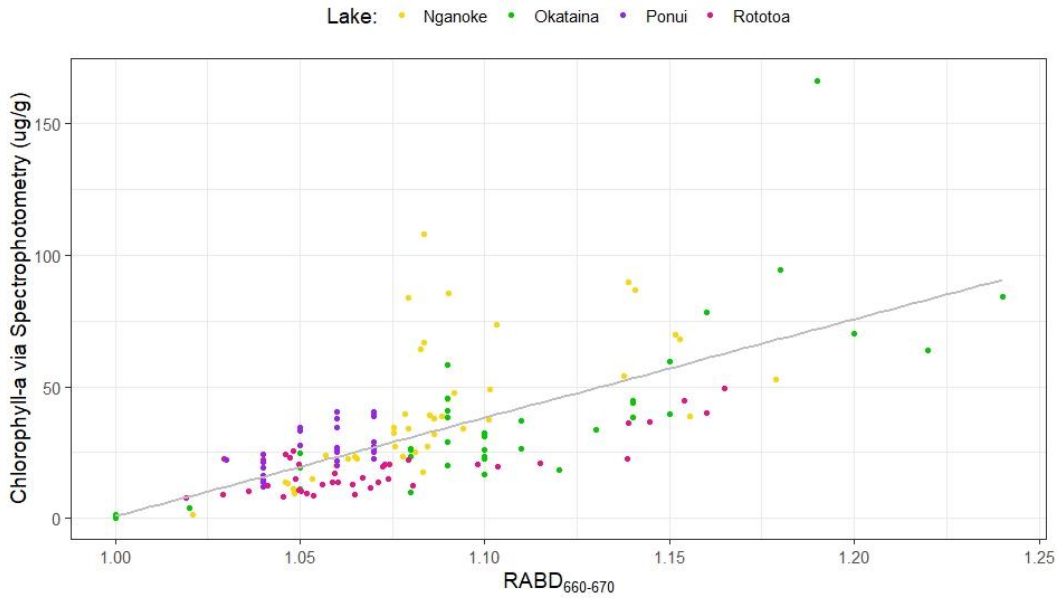
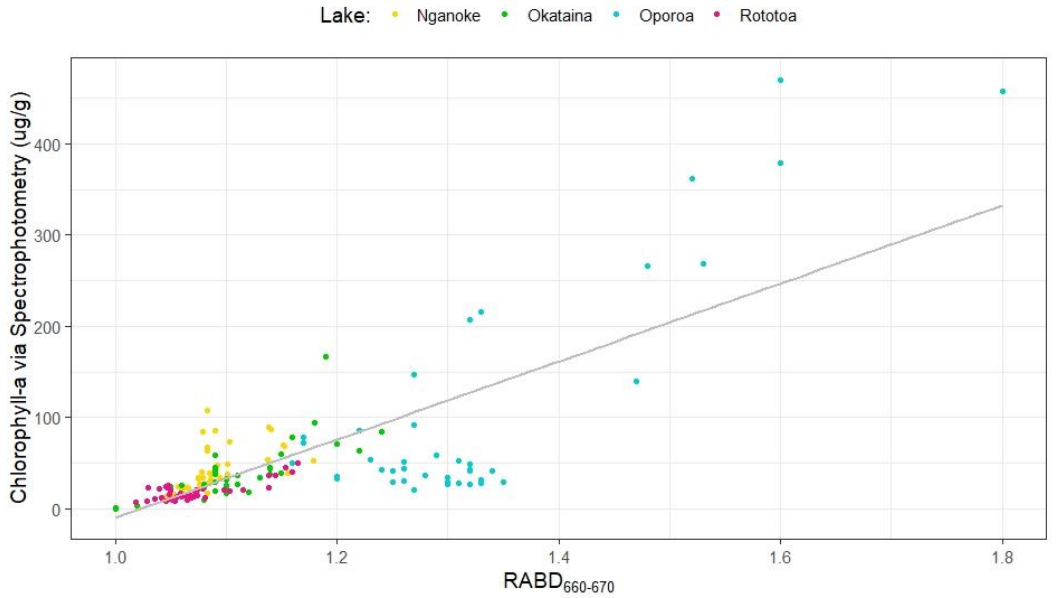


Figure 4.10. *Data from Four Lakes with Ponui Excluded $RABD_{660-670}$ Calibration to Chlorophyll-a Concentrations Analysed via Spectrophotometry*
Model for a linear regression with a r^2 of 0.61, p-value of < 2.2 x 10^{-16}, an intercept of -453.5, a gradient of 439.8, and CI of 385.7, 493.8.



4.1.2 Phycocyanin Calibration

When analysing the samples from four lake cores, a total of 161 samples were analysed for phycocyanin via fluorescence: 42 from Lake Nganoke, 42 from Lake Okataina, 40 from Lake Oporoa, and 37 from Lake Rototoa. From these 161 samples, no sample returned a concentration of phycocyanin above the level of detection (0.1 µg/mL). Therefore, no quantifiable amounts of phycocyanin could be detected within the lake core sediment samples through the fluorescence method used in this analysis.

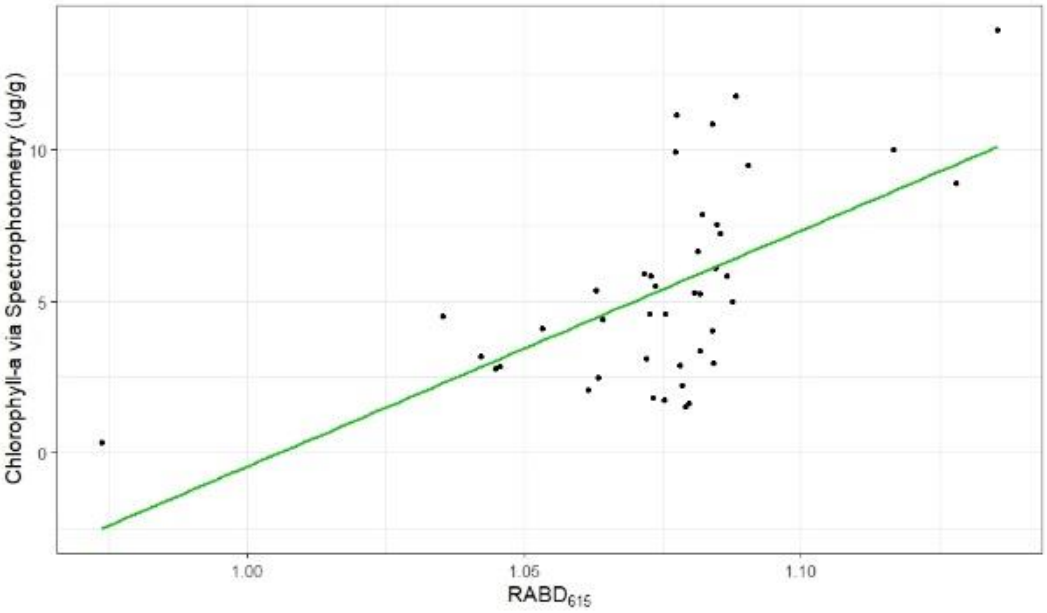
Although phycocyanin was not detected via the fluorescence method, there is a clear signal detected from the $RABD_{615}$ band. This signal correlates to the signal detected at the $RABD_{660\text{-}670}$ for chl-a. When $RABD_{615}$ was compared to chl-a via spectrophotometry concentrations using a linear regression (Figure 4.11) with a Pearson correlation, there was a statistically significant correlation (p-values between 6.4×10^{-5} to 6.2×10^{-12}) for all lakes with r^2 values between 0.33 and 0.74 (Table 4.6).

Table 4.6. *$RABD_{615}$ to Chlorophyll-a via Spectrophotometry Linear Regression Relationships.*

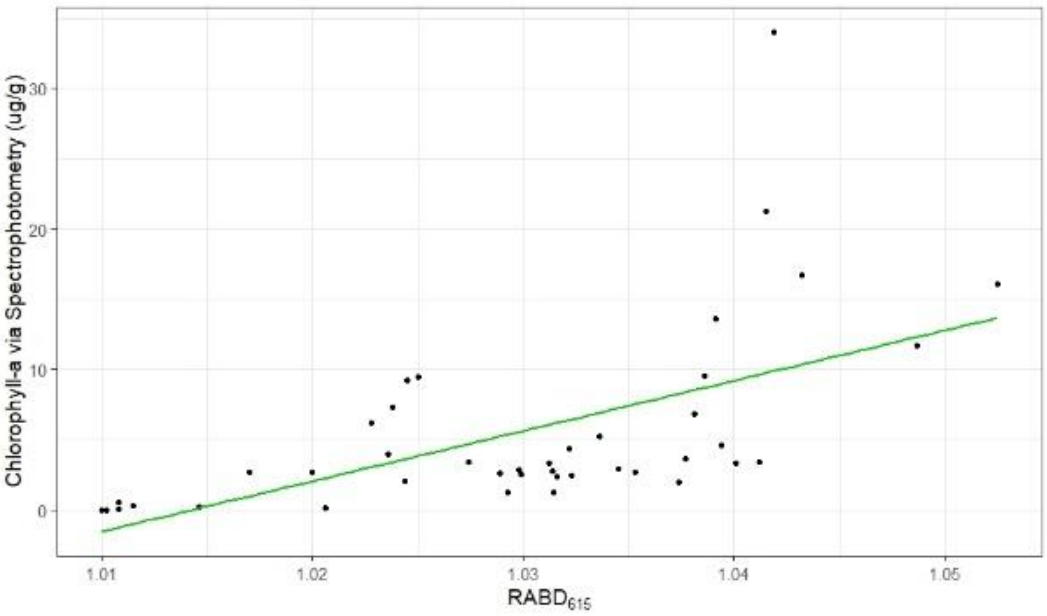
Lake	r^2	P-value	Intercept	Gradient	CI
Nganoke	0.35	2.1×10^{-5}	-78.4	77.9	45.2, 110.6
Okataina	0.32	6.4×10^{-5}	-363.3	358.2	196.1, 520.3
Oporoa	0.71	6.2×10^{-12}	-182.2	174.1	138.1, 210.1
Rototoa	0.74	6.6×10^{-12}	-477.1	480.9	384.2, 577.6

Figure 4.11. *$RABD_{615}$ to Chlorophyll-a Concentrations Analysed via Spectrophotometry Linear Regressions.*

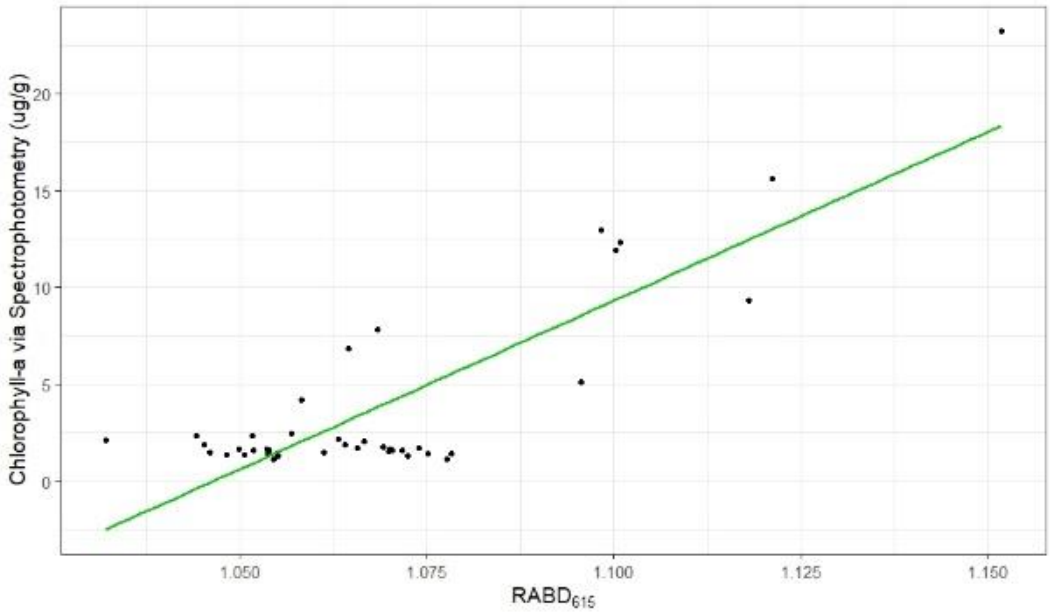
A) Lake Nganoke; B) Lake Okataina; C) Lake Oporoa; D) Lake Rototoa.



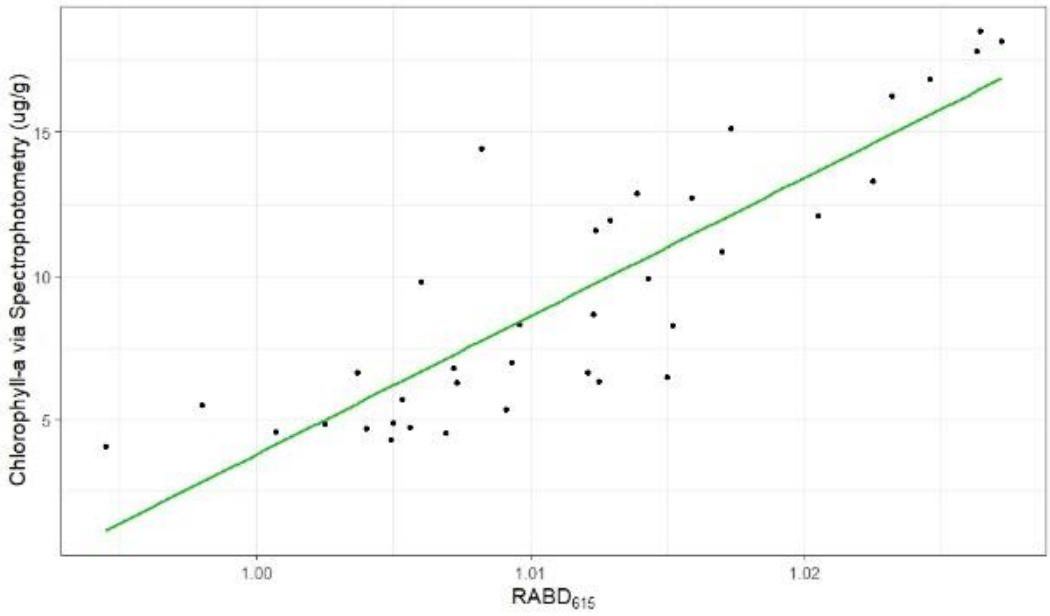
A)



B)


Chlorophyll-a via Spectrophotometry (ug/g)
20
15
10
5
0
1.050
1.075
1.100
1.125
1.150
RABD615

C)


Chlorophyll-a via Spectrophotometry (ug/g)
15
10
5
1.00
1.01
1.02
RABD615

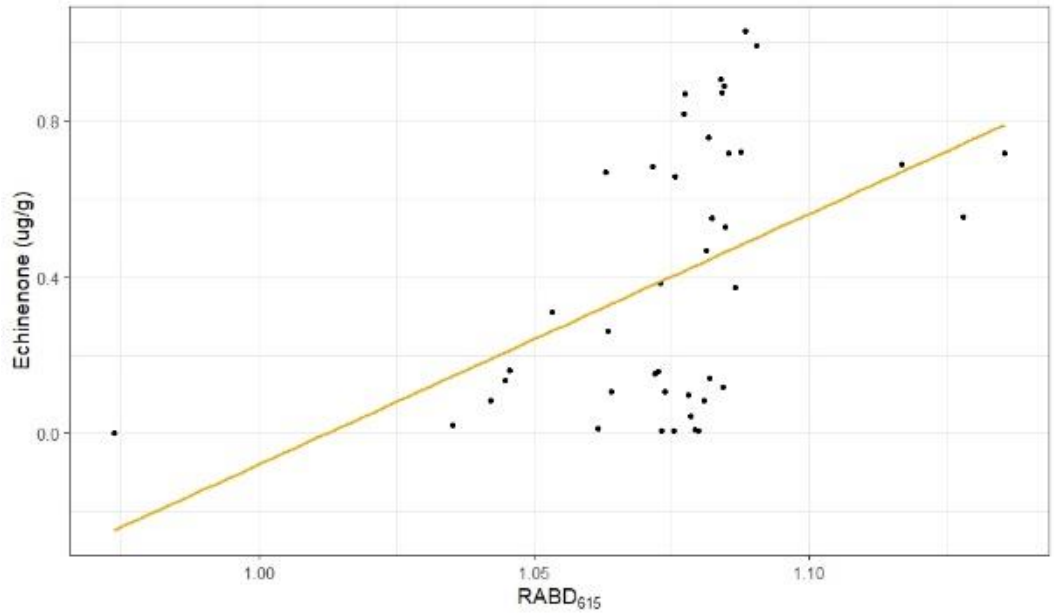
D)

Additionally, when $RABD_{615}$ levels for the sediment core samples of the four lake cores were compared to the concentrations of two carotenoid pigment indicators for cyanobacteria; echinenone and zeaxanthin (quantified via HPLC-DAD) (Figure 4.12 to 3.15, and Table 4.7), similar correlations between both the echinenone and the zeaxanthin to the $RABD_{615}$ bandwidth were observed. However, for both echinenone and zeaxanthin the correlation was not as strong as it was for the chl-a via spectrophotometry concentrations (Table 4.6). Additionally, when $RABD_{615}$ was compared with echinenone and zeaxanthin pigments, three of the four lakes resulted in positive correlations. However, Lake Oporoa did not return a relationship, with weak r^2 values and non-significant p-values. Although this result follows the trend of $RABD_{615}$ returning stronger and more significant correlations when regressed against chl-a.

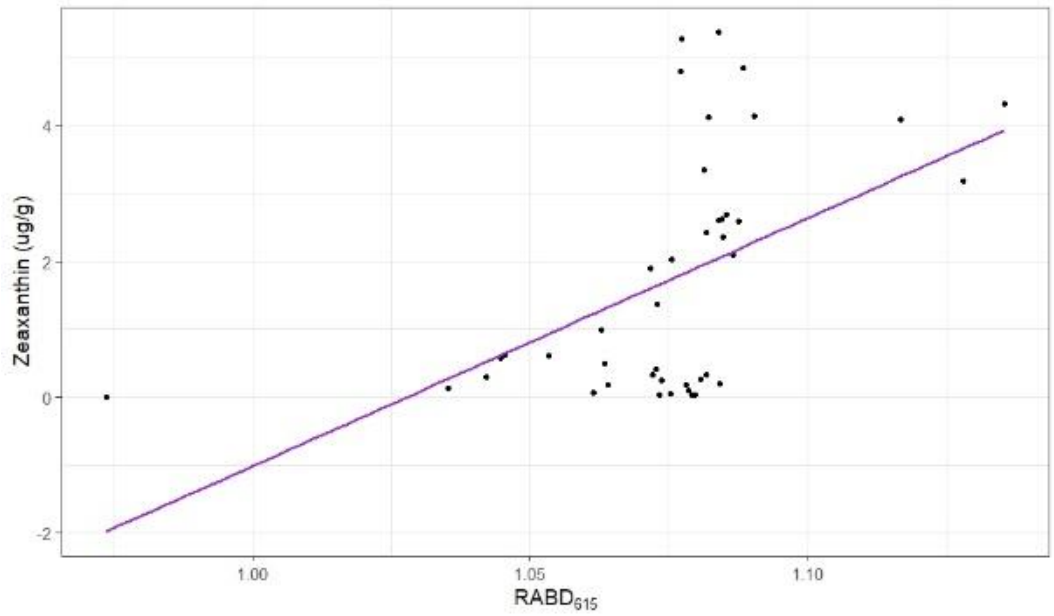
Table 4.7. *$RABD_{615}$ to Echinenone and Zeaxanthin Linear Regressions.*

Lake	Pigment	r^2	p-value	Intercept	Gradient	CI
Nganoke	Echinenone	0.20	1.6×10^{-3}	-6.5	6.4	2.6, 10.2
	Zeaxanthin	0.25	4.0×10^{-4}	-37.5	36.5	17.4, 55.6
Okataina	Echinenone	0.27	2.5×10^{-4}	-3.1	3.0	1.5, 4.5
	Zeaxanthin	0.34	2.7×10^{-5}	-15.3	15.1	8.6, 21.5
Oporoa	Echinenone	<0.01	0.7	138.5	-102.6	-667.8, 462.6
	Zeaxanthin	<0.01	0.5	1169.4	-950.8	-2314.7, -4208.1
Rototoa	Echinenone	0.58	2.3×10^{-8}	-3.4	3.4	2.4, 4.3
	Zeaxanthin	0.63	3.3×10^{-9}	-29.9	29.7	22.0, 37.4

Figure 4.12. *$RABD_{615}$ to either Echinenone or Zeaxanthin Linear Regressions for Lake Nganoke. A) $RABD_{615}$ vs. echinenone, B) $RABD_{615}$ vs. zeaxanthin.*

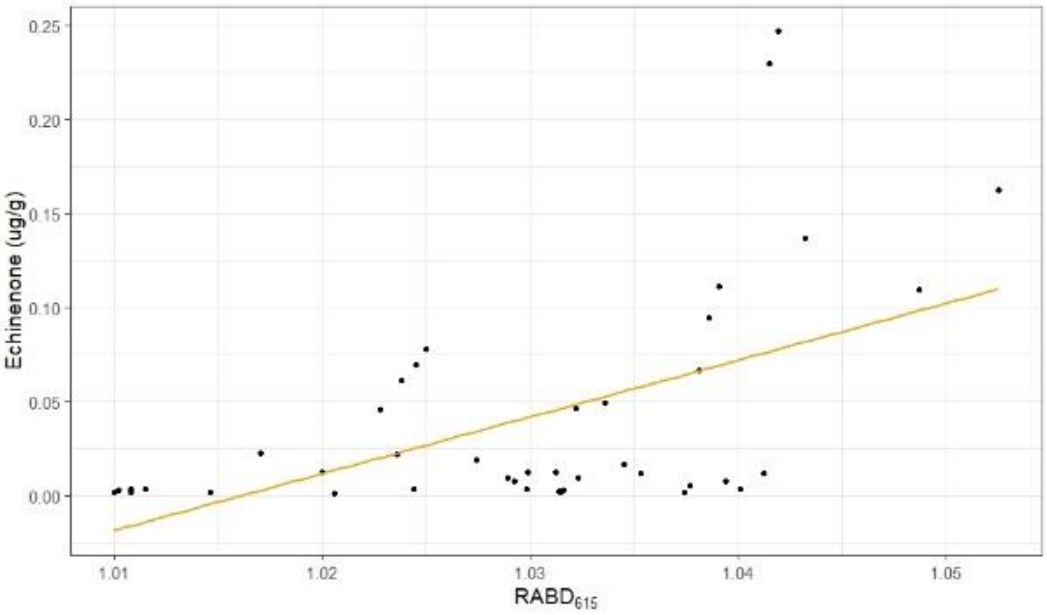


A)

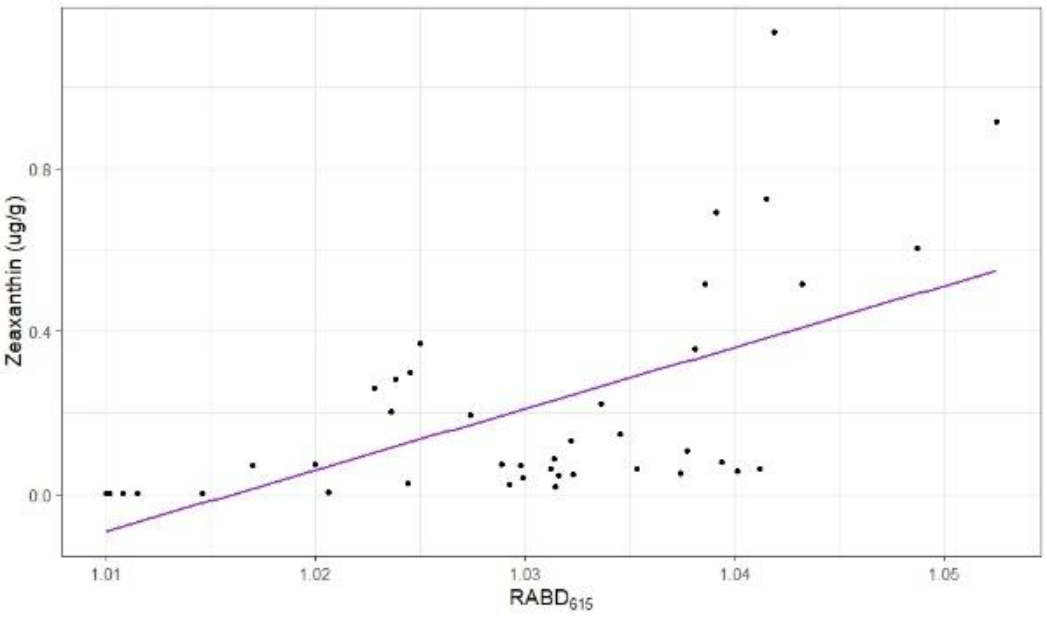


B)

Figure 4.13. *$RABD_{615}$ to either Echinenone or Zeaxanthin Linear Regressions for Lake Okataina. A) $RABD_{615}$ vs. echinenone, B) $RABD_{615}$ vs. zeaxanthin.*

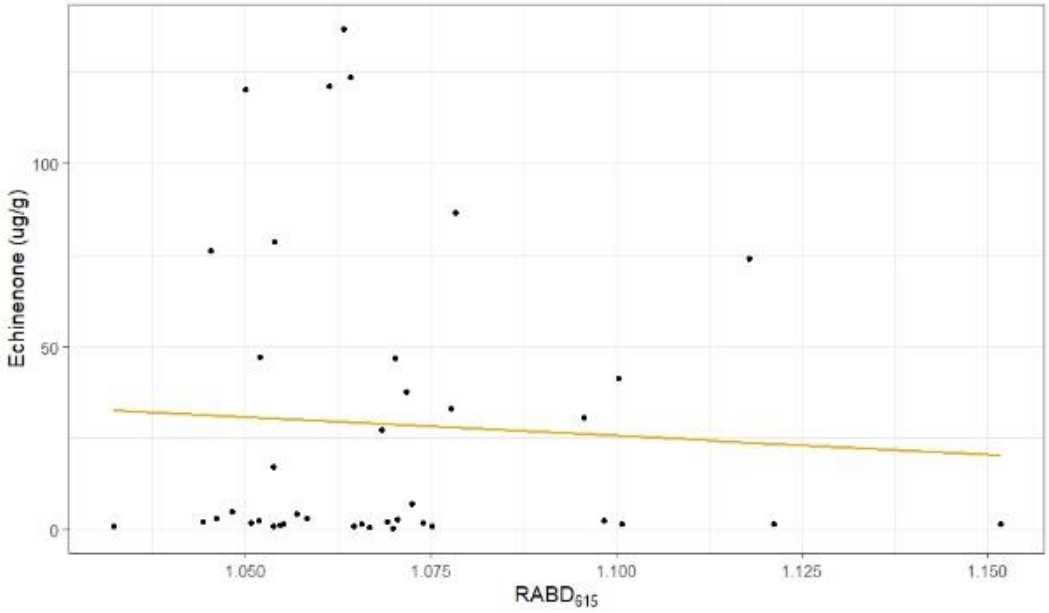


A)

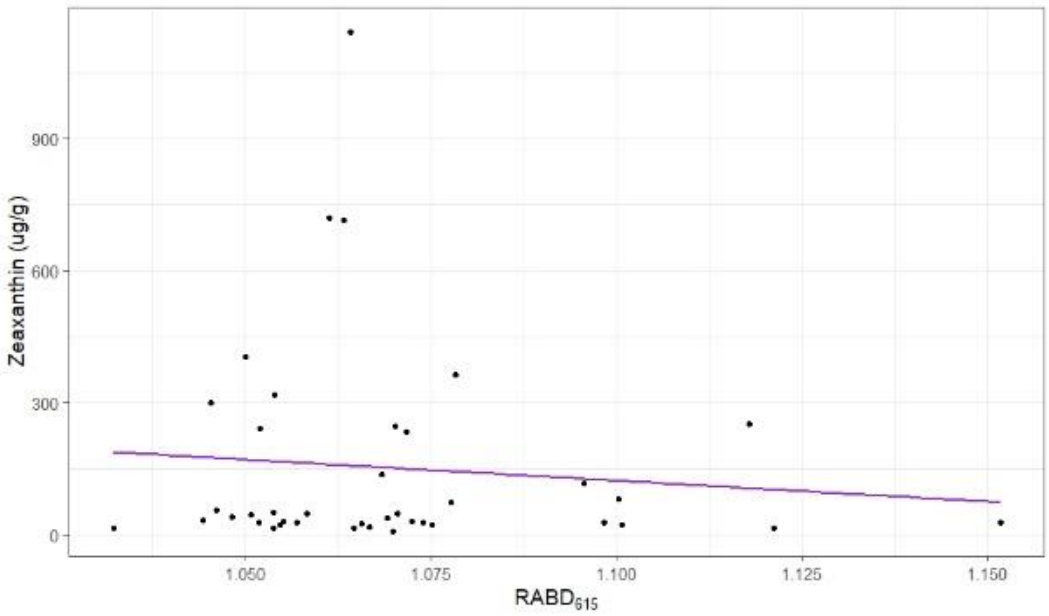


B)

Figure 4.14. *$RABD_{615}$ to either Echinenone or Zeaxanthin Linear Regressions for Lake Oporoa. A) $RABD_{615}$ vs. echinenone, B) $RABD_{615}$ vs. zeaxanthin.*

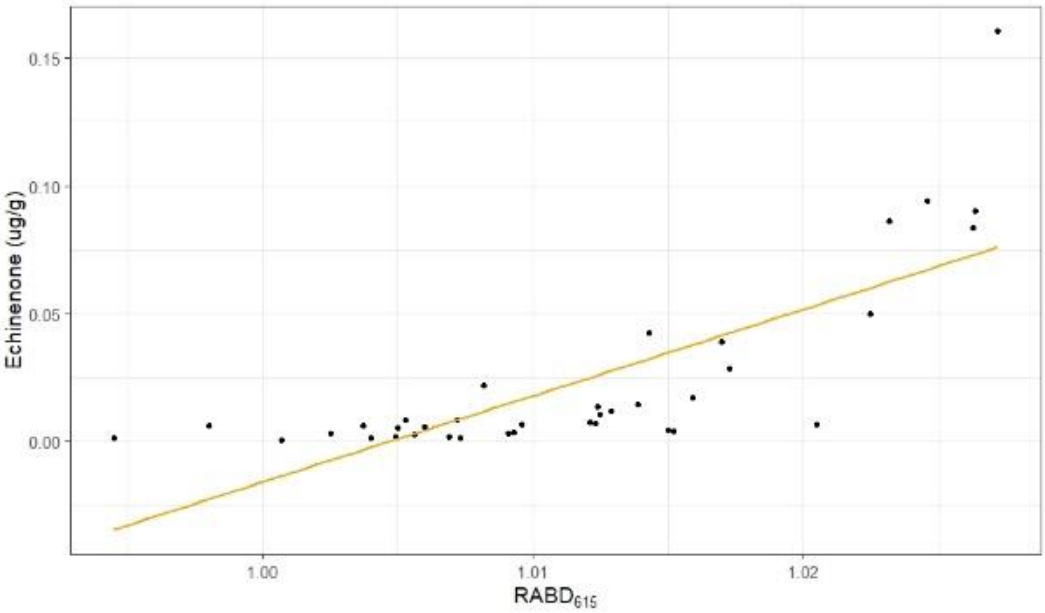


A)

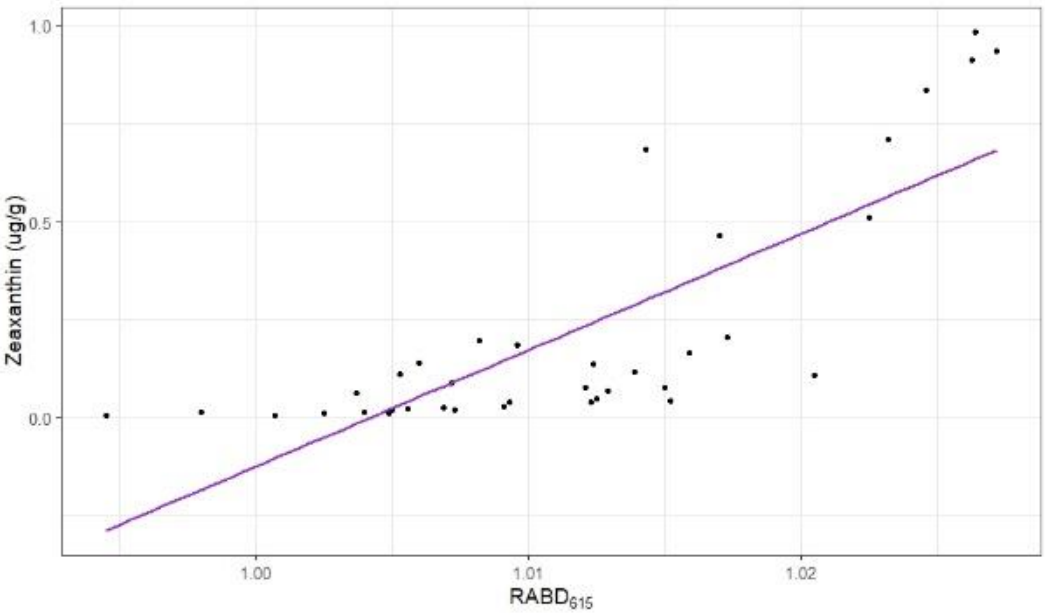


B)

Figure 4.15. *$RABD_{615}$ to either Echinenone or Zeaxanthin Linear Regressions for Lake Rototoa. A) $RABD_{615}$ vs. echinenone, B) $RABD_{615}$ vs. zeaxanthin.*



A)



B)

4.2 Predicting Lake Water Quality Using Surface Sediments

The data set of lakes used for the creation of the machine learning models consisted of 95 lakes for which the Lakes 380 team had calculated the trophic level index (TLI) from monitoring data provided by regional councils. This dataset includes three coastal shoreline/lagoons, 27 glacial lakes, 16 riverine lakes, four swamp/wetland lakes, two of tectonic/landslide origin, ten volcanic lakes, and 33 wind/aeolian formed lakes. These lakes ranged from low-land lakes (< 10 m a.s.l.) to high-altitude alpine lakes (up to 826 m a.s.l.). The TLI of these lakes ranged from 1.2 (Microtrophic) to 7.2 (Supertrophic).

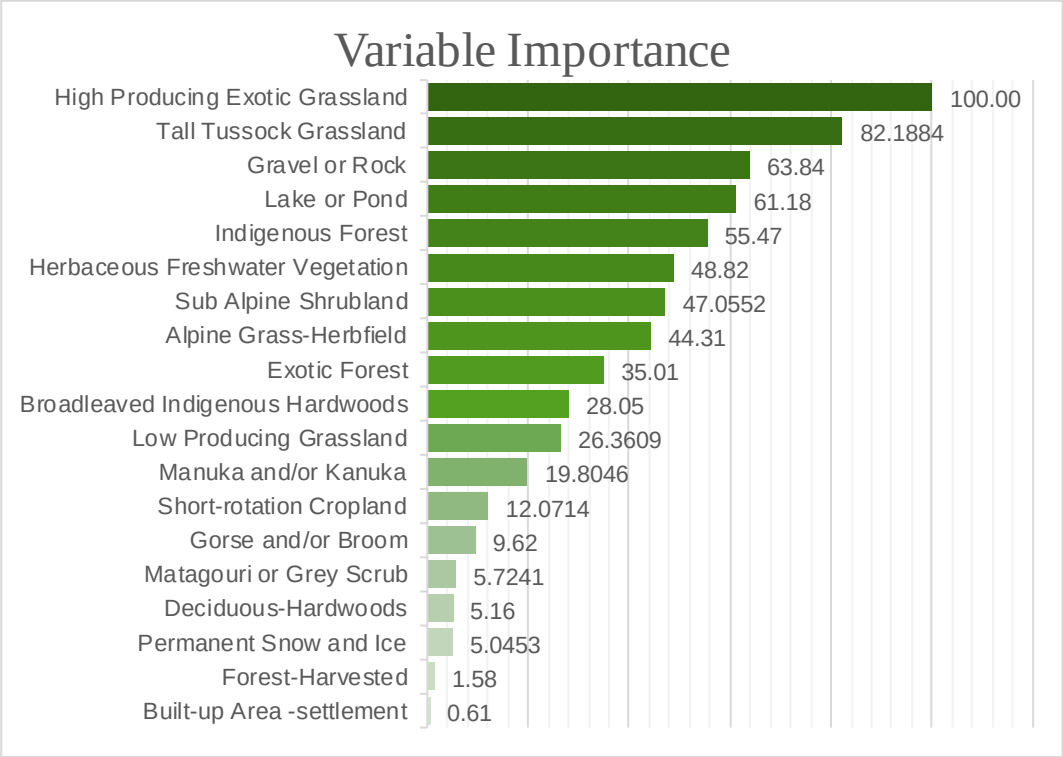
4.2.1 Predicting Lake Trophic Levels from Catchment Land Use

To assess the effects of land use on the trophic levels of lakes in New Zealand/in this study, a regression tree was fitted using the 33 land classifications from the 'LCDB v5.0 GIS raster'. None of these variables were transformed, as the TLI values possessed a normal distribution (Appendix D), and the land use variables were calculated as relative abundance.

The regression tree was then tuned to find the best minimum number of data points required to create a terminal node after it is forced to attempt a split (resulting in 13 splits as optimum), and the maximum number of internal nodes between the terminal nodes and the root node (resulting in 15 nodes being optimum). The optimum complexity parameter (cp) which controls the size of the regression tree was found to be .0.01. The final bagged tree returned a RMSE of 1.04 (25.8%) after predicting against the data split for the testing.

Figure 4.16 indicates that 19 of the 33 variables influenced the regression tree to varying degrees. The most influential land-use category was the high producing exotic grassland, followed by the tall tussock grassland category. The land-use categories which did not influence the trophic level of the lake were, depleted grassland, fern land, flax land, landslide, mixed exotic shrubland, orchard/vineyard or other perennial crop, river, sand or gravel, surface mine or dump, transport infrastructure, and urban parkland/open space. The categories not represented within the dataset were estuarine open water, herbaceous saline vegetation, and mangrove.

Figure 4.16. *Variable Importance for the Variables in the Regression Tree for Predicting Lake Trophic Levels Using the Land Use Type Ratio.*



The land use categories which did not influence the regression tree possess low frequency with depleted grassland occurring in 10% of the catchments, fern land 12%, flax land 7%, landslide 16%, mixed exotic shrubland 13%, orchard/vineyard or other perennial crop 9%, river 12%, sand or gravel 14%, surface mine or dump 10%, transport infrastructure 6%, and urban parkland/open space 10%. However, short rotation cropland occurred in 11% of the catchments and permanent snow and ice in 7%, yet these two variables were influential in the regression tree.

4.2.2 Predicting Lake Trophic Levels from Phytoplankton Pigments Within Surface Sediment

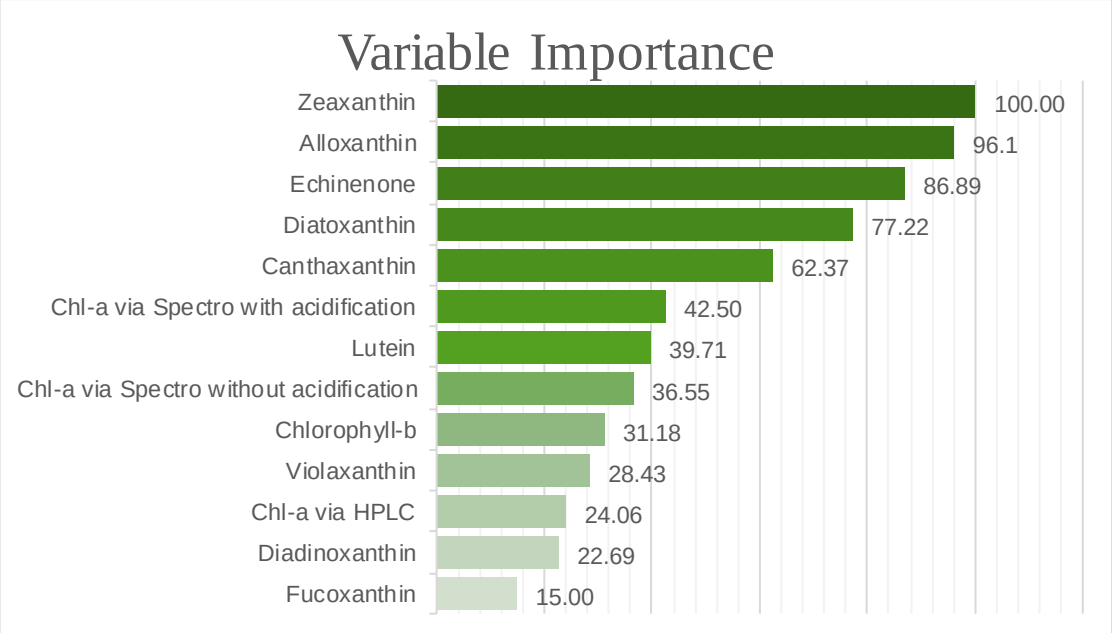
Relationships between lake TLI and phytoplankton pigment concentrations within lake surface sediment were assessed using a regression tree and a random forest model. The regression tree and random forest model was created using twelve phytoplankton pigments (with three different measurements of chl-a): twelve measured by HPLC-DAD (Alloxanthin, chl-a via HPLC, Chlorophyll-b, Canthaxanthin, Diadinoxanthin, Diatoxanthin, Echinenone, Fucoxanthin, Lutein, Peridinin, Violaxanthin, and Zeaxanthin); and chl-a by spectrophotometry both with and without acidification

4.2.2.1 Regression Tree

The regression tree was tuned to 11 splits as the optimum amount of data points, 13 as returning the optimum number of internal nodes between the terminal nodes and the root node, and a cp of 0.04. The final bagged tree returned a RMSE of 1.13 (27.8%) after testing against the predicted data split. However, the land use proportion regression tree performed slightly better (with a lower RMSE).

The variable importance values for the variables in the bagged regression tree indicates that thirteen of the fourteen variables influenced the regression tree to varying degrees (Figure 4.17). The most influential phytoplankton pigment was zeaxanthin, followed by alloxanthin and echinenone. The phytoplankton pigment which did not influence the trophic level of the lake was peridinin.

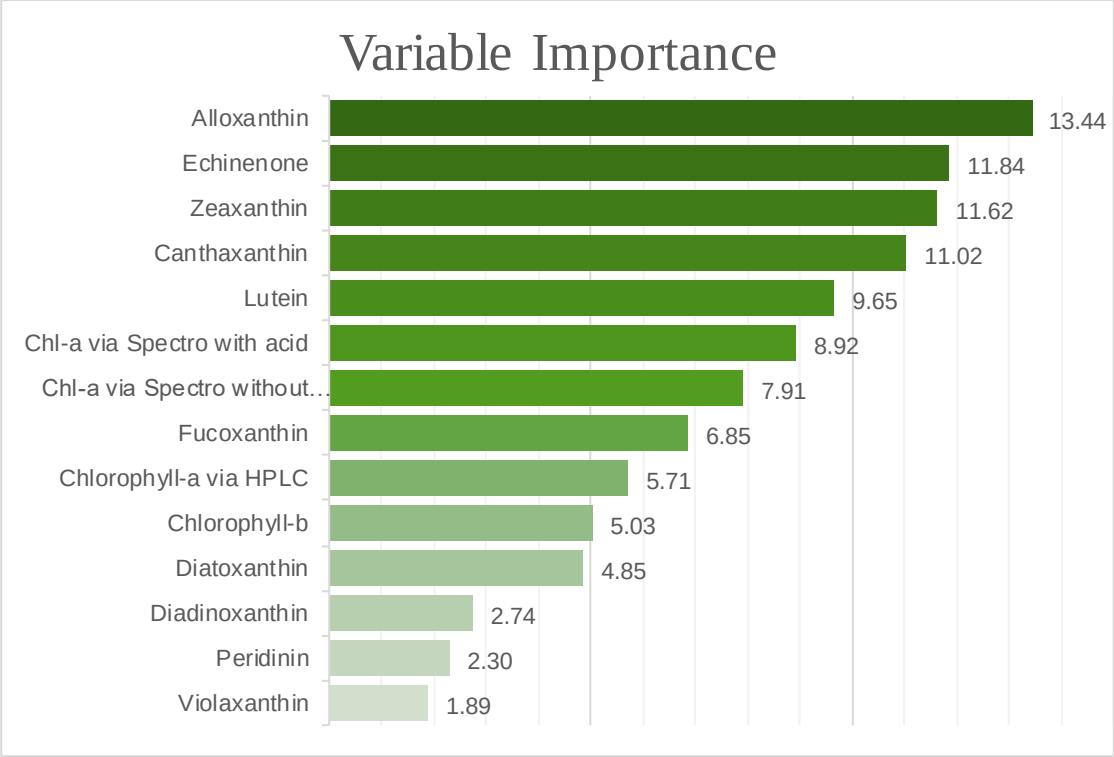
Figure 4.17. *Variable Importance for the Variables in the Regression Tree for Predicting Lake Trophic Levels Using the Phytoplankton Pigment Concentrations in Lake Sediment.*



4.2.2.2 Random Forest

A random forest model was created and tuned (mtry = 5, minimum node size = 9), on a model consisting of 500 trees, to achieve the lowest RMSE of 0.49 (11.9%). This improved on the regression by more than twice as much as the regression tree. Of the fourteen predictors in the model, all fourteen influenced the random forest model (Figure 4.18). Variable importance for the random forest tree was measured by the program by logging the reduction in MSE each time a variable was used as a node split. The most influential variables were alloxanthin, echinenone, and zeaxanthin.

Figure 4.18. *Variable Importance for the Variables in the Random Forest Model for Predicting Lake Trophic Levels Using the Phytoplankton Pigment Concentrations in Lake Sediment.*



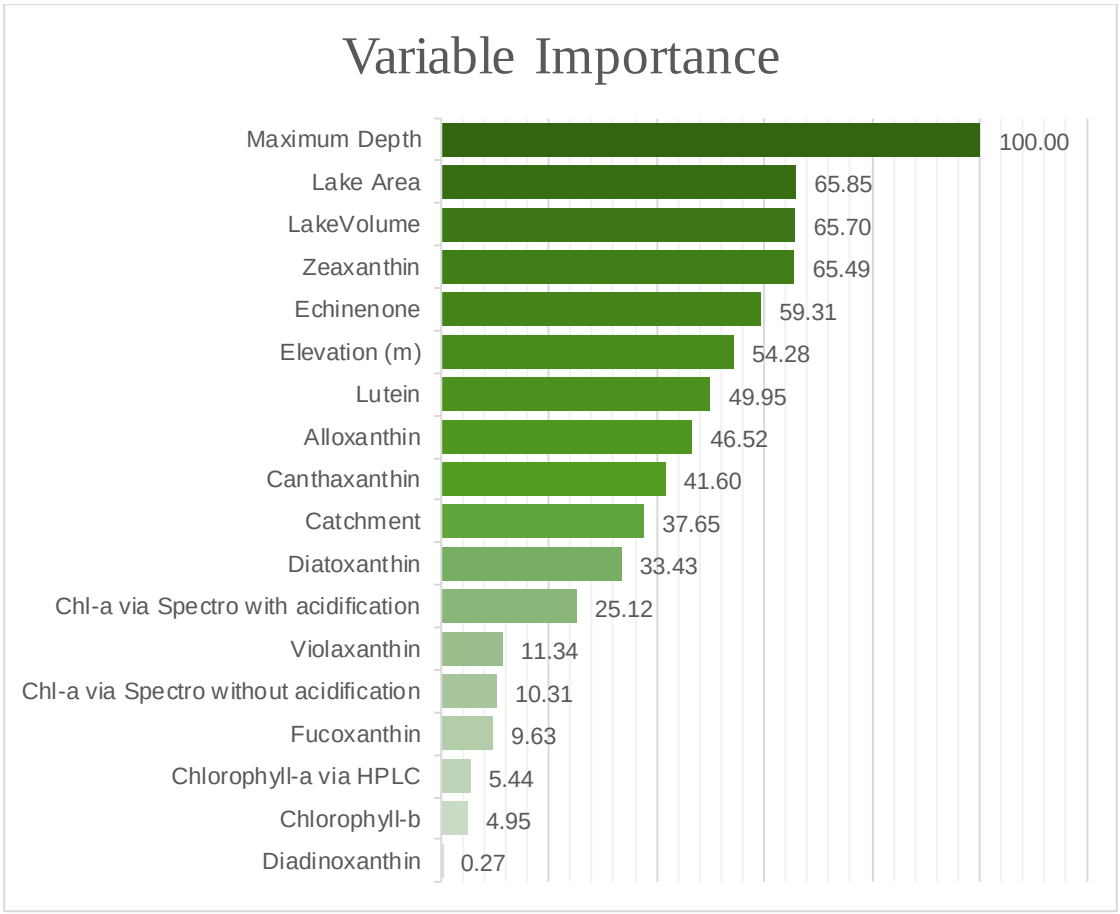
4.2.3 Creating a Predictive Model for Lake Trophic Status

A predictive model for use on the down-core samples was created using twelve phytoplankton pigments (with three different measurements of chl-a): twelve measured by HPLC-DAD (Alloxanthin, chl-a via HPLC, Chlorophyll-b, Canthaxanthin, Diadinoxanthin, Diatoxanthin, Echinenone, Fucoxanthin, Lutein, Peridinin, Violaxanthin, and Zeaxanthin); and chl-a by spectrophotometry both with and without acidification. Five physical lake characteristics that would not change through time, or would only change very slowly (Elevation, Lake Area, Lake Volume, Maximum Depth, and Catchment Area) were also included in the model. This model was termed the 'pigments +' model. These 18 variables were normalised by transforming them using a natural logarithm of the given value plus one.

4.2.3.1 Regression Tree

A regression tree was created then tuned to 16 splits as the optimum amount of data points, 14 as the optimum number of internal nodes between the terminal nodes and the root node, and a cp of 0.02. The final bagged tree returned a RMSE of 0.96 (24.6%) after predicting against the data split for the testing. Of the 19 variables used to create the regression tree 18 influenced the prediction. The predictor wihich did not influence the regression tree was peridinin (Figure 4.19). The most influential variables were those related to the size of the lake (maximum depth of the lake, lake area, and lake volume) and the most influential phytoplankton pigment concentrations were zeaxanthin and echinenone.

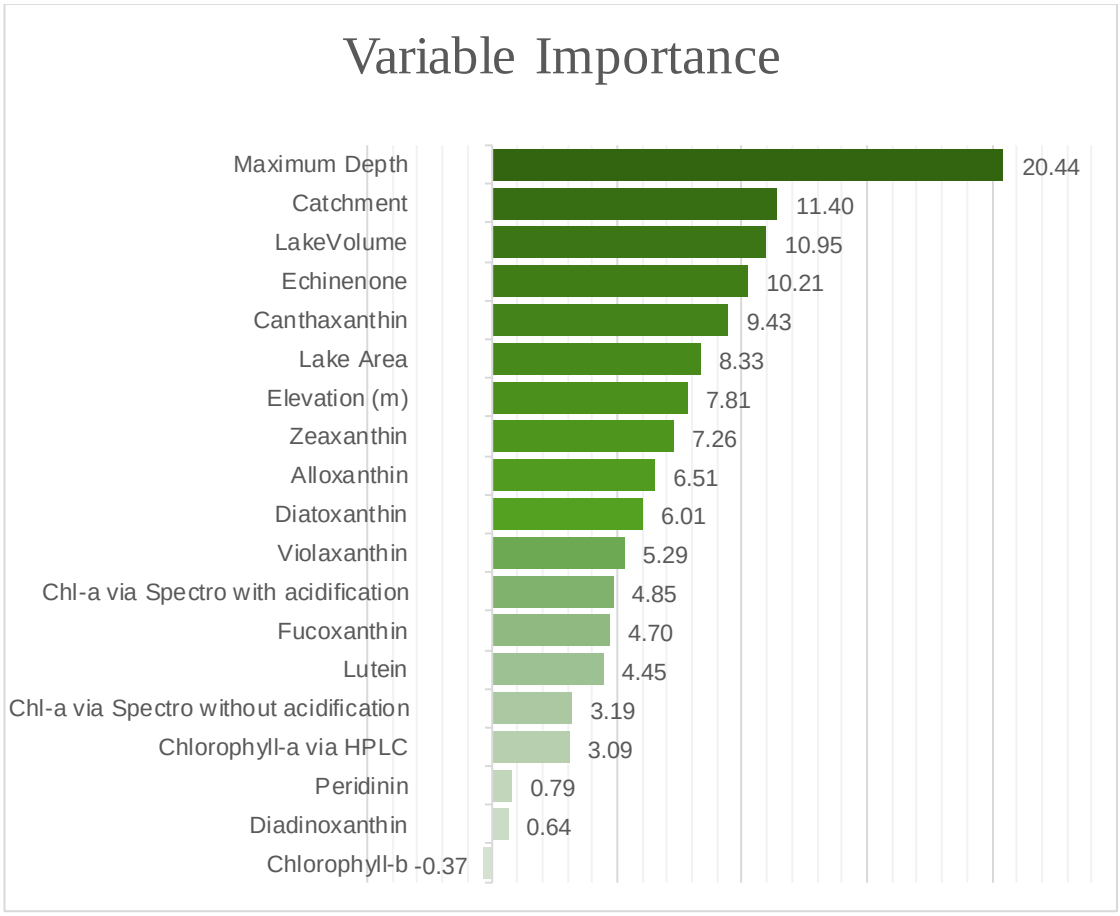
Figure 4.19. *Variable Importance for the Variables in the Bagged Regression Tree for Predicting Lake Trophic Levels Using the Phytoplankton Pigment Concentrations in Lake Sediment Plus Five Physical Characteristics of the Lake.*



4.2.3.2 Random Forest

A random forest model was created then tuned (mtry = 9, minimum node size = 3), resulting in a RMSE of 0.39 (10.0%). Once again, the random forest model improved on the regression tree more than twice as much and was the best performing TLI predictive model within this research. Of the 19 predictors in the model, all 19 influenced the random forest model (Figure 4.20). Interestingly chlorophyll-b had a negative influence on the random forest model. Variable importance for the random forest tree was measured by the program by logging the reduction in MSE each time a variable was used as a node split. The most influential variables were those related to the size of the lake (maximum depth of the lake, lake catchment size, and lake volume), with the most influential phytoplankton pigments being echinenone and canthaxanthin.

Figure 4.20. *Variable Importance for the Variables in the Random Forest Model for Predicting Lake Trophic Levels Using the Phytoplankton Pigment Concentrations in Lake Sediment Plus Five Physical Characteristics of the Lake.*

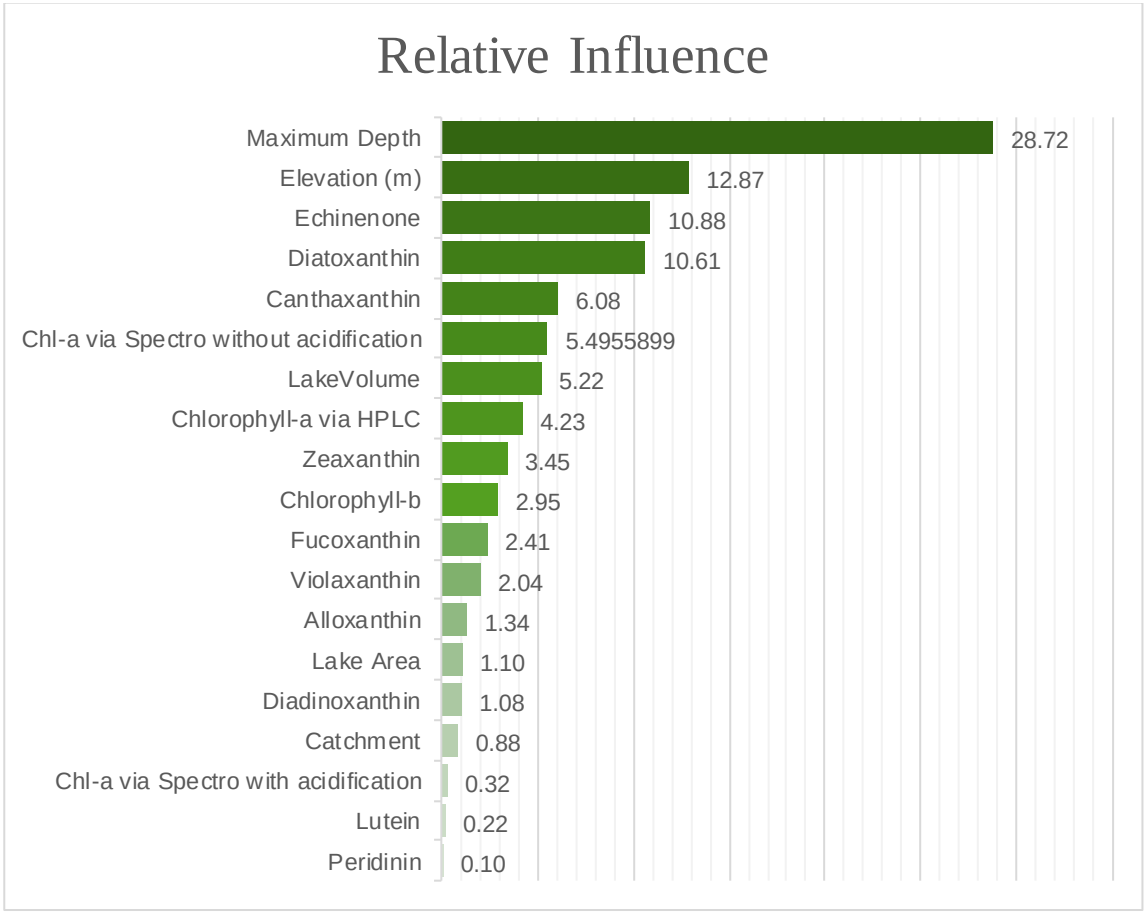


4.2.3.3 Boosted Regression Tree

A boosted regression tree was created then tuned to find the optimal parameters. The resulting model consisted of 4,000 trees within the model, with a depth of 3, and a learning rate of 0.25. Additionally, the minimum number of observations in the trees terminal nodes was set to 0.25, and the stochastic gradient descent was set to 0.75, the final tree returned a RMSE of 0.94 (24.2%).

Of the 19 variables, all 19 influenced the prediction (Figure 4.21). Variable importance for the random forest tree was measured by the program where for each split in each tree, the improvement in the MSE for regression is computed and presented as a percentage. The most influential variables of this model were the maximum depth of the lake and elevation of the lake, with the most influential phytoplankton pigments being echinenone and diatoxanthin.

Figure 4.21. *Relative Influence of the Variables in the Boosted Regression Tree for Predicting Lake Trophic Levels Using the Phytoplankton Pigment Concentrations in Lake Sediment Plus Five Physical Characteristics of the Lake.*

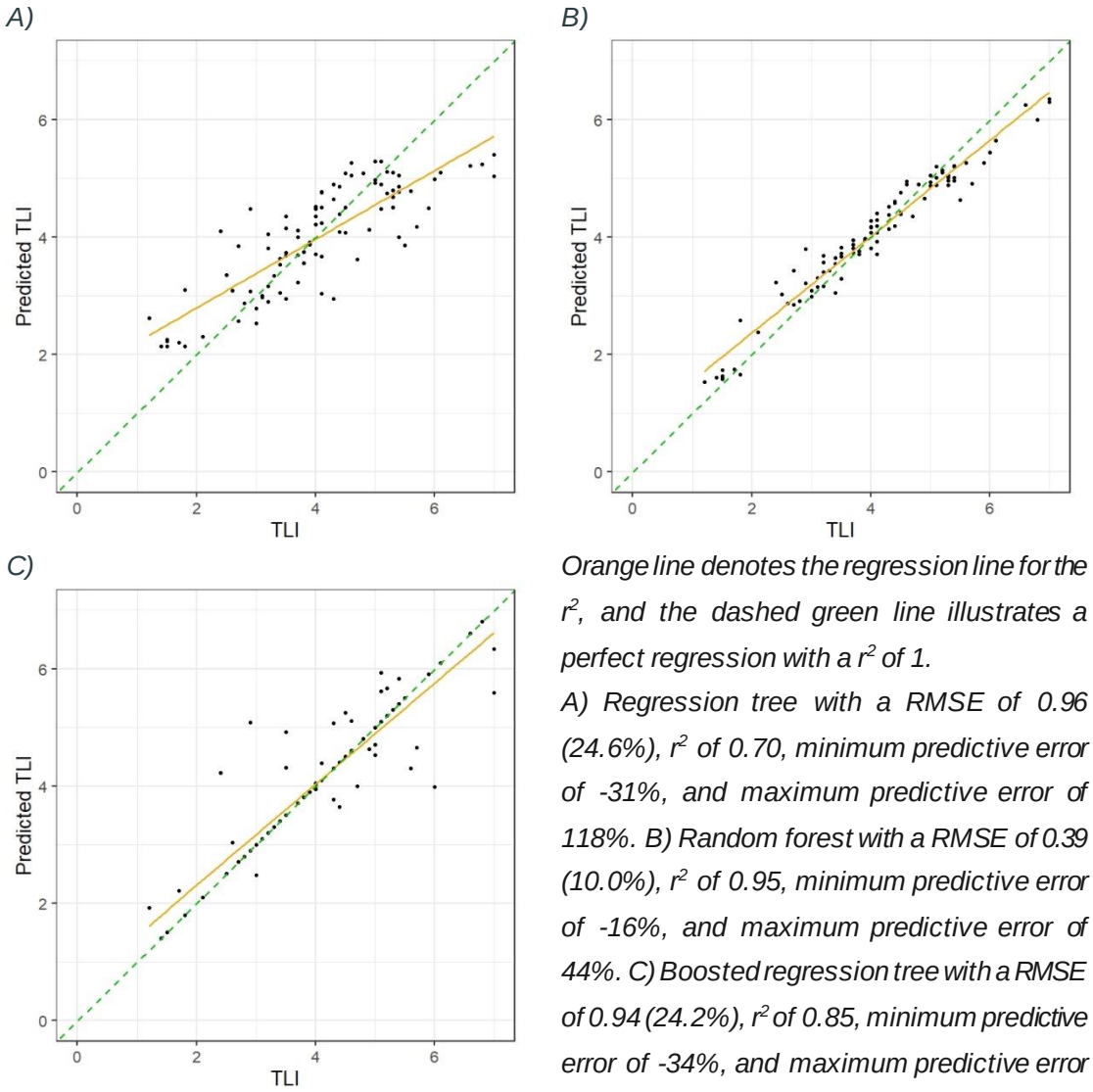


4.2.4 Preferred Model for Predictions

Of the three predictive 'pigments +' models the best performing one, with the lowest RMSE was the random forest model (Figure 4.22). While a boosted regression tree can provide smaller RMSE values than random forests, they tend to overfit models if the data contains a lot of noise (Vezhnevets & Barinova, 2007). The linear correlation of the Lake Trophic Levels to the Lake Trophic Levels Predicted by the Random Forest Model returned a r^2 of 0.95.

Within the models there are three different measurements for chl-a. By using the 'pigments +' model (the best performing model) and creating three versions with one chl-a measurement in each (Figure 4.23) we can isolate which chl-a analysis provides more accurate predictions. There is very little difference in the type of analysis utilised for chl-a analysis, with a difference in error rates of 0.5% between the three analyses. Utilising all three analyses in the random forest model appears to overfit the model, resulting in a higher error (10.0%). However, there is less than 1% difference between the random forest model utilising all three analyses and the best performing single chl-a analysis model. For subsequent analyses in this thesis, the 'pigments +' random forest model will utilise the chl-a via spectrophotometry with acidification analysis.

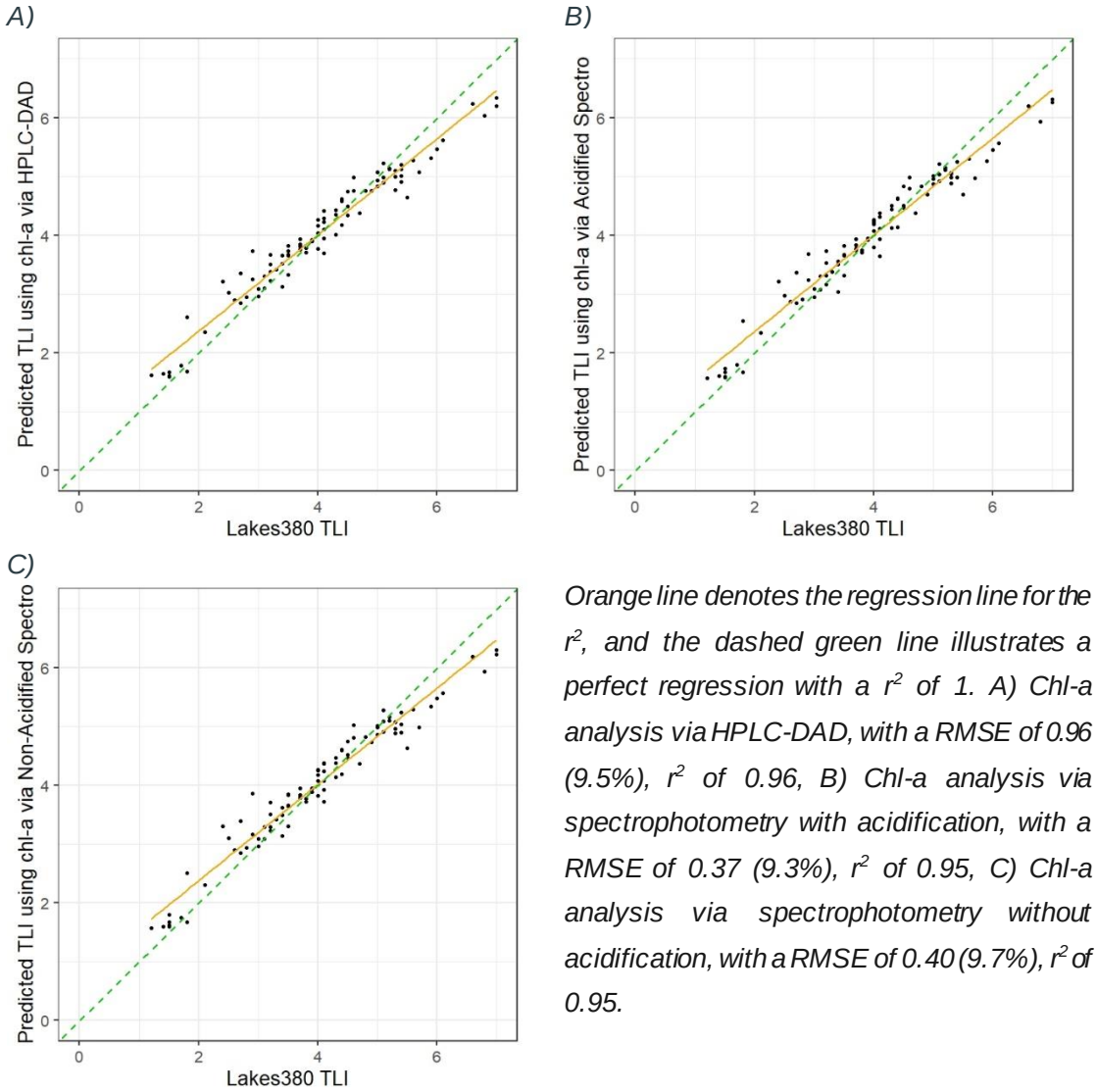
Figure 4.22. *Correlation of Measured Lake Trophic Levels to the Lake Trophic Levels Predicted by the Various 'Pigments +' Models.*



Orange line denotes the regression line for the r^2, and the dashed green line illustrates a perfect regression with a r^2 of 1.

A) Regression tree with a RMSE of 0.96 (24.6%), r^2 of 0.70, minimum predictive error of -31%, and maximum predictive error of 118%. B) Random forest with a RMSE of 0.39 (10.0%), r^2 of 0.95, minimum predictive error of -16%, and maximum predictive error of 44%. C) Boosted regression tree with a RMSE of 0.94 (24.2%), r^2 of 0.85, minimum predictive error of -34%, and maximum predictive error of 76%.

Figure 4.23. *Correlation of Measured Lake Trophic Levels to the Lake Trophic Levels Predicted by the Random Forest Pigment Plus Models with Varying Chlorophyll-a (chl-a) Analyses.*



Orange line denotes the regression line for the r^2, and the dashed green line illustrates a perfect regression with a r^2 of 1. A) Chl-a analysis via HPLC-DAD, with a RMSE of 0.96 (9.5%), r^2 of 0.96, B) Chl-a analysis via spectrophotometry with acidification, with a RMSE of 0.37 (9.3%), r^2 of 0.95, C) Chl-a analysis via spectrophotometry without acidification, with a RMSE of 0.40 (9.7%), r^2 of 0.95.

4.2.5 Modelling the Lake Trophic Status Throughout New Zealand

By applying the random forest 'pigments +' model to the lakes within the original dataset of 232 lakes without known TLI's, predictions of their current lake health could be made. Mapping out both the current and predicted TLI's (Figures 4.25 and 4.26) show the distribution of these lakes throughout New Zealand. Histograms of the current and predicted TLI's (Figure 4.24) show that the model is not predicting out to the limits of the range of the TLI (both low and high). Rather, the predicted TLIs seem to be constrained to the more central values.

Figure 4.24. *Mirrored Histogram of the Current and Predicted TLI's for New Zealand Lakes. Top histogram (green) displays the current TLI's of New Zealand Lakes used to create the model. Bottom (mirrored) histogram displays the TLI's predicted on the same lakes by the random forest 'pigments +' model.*

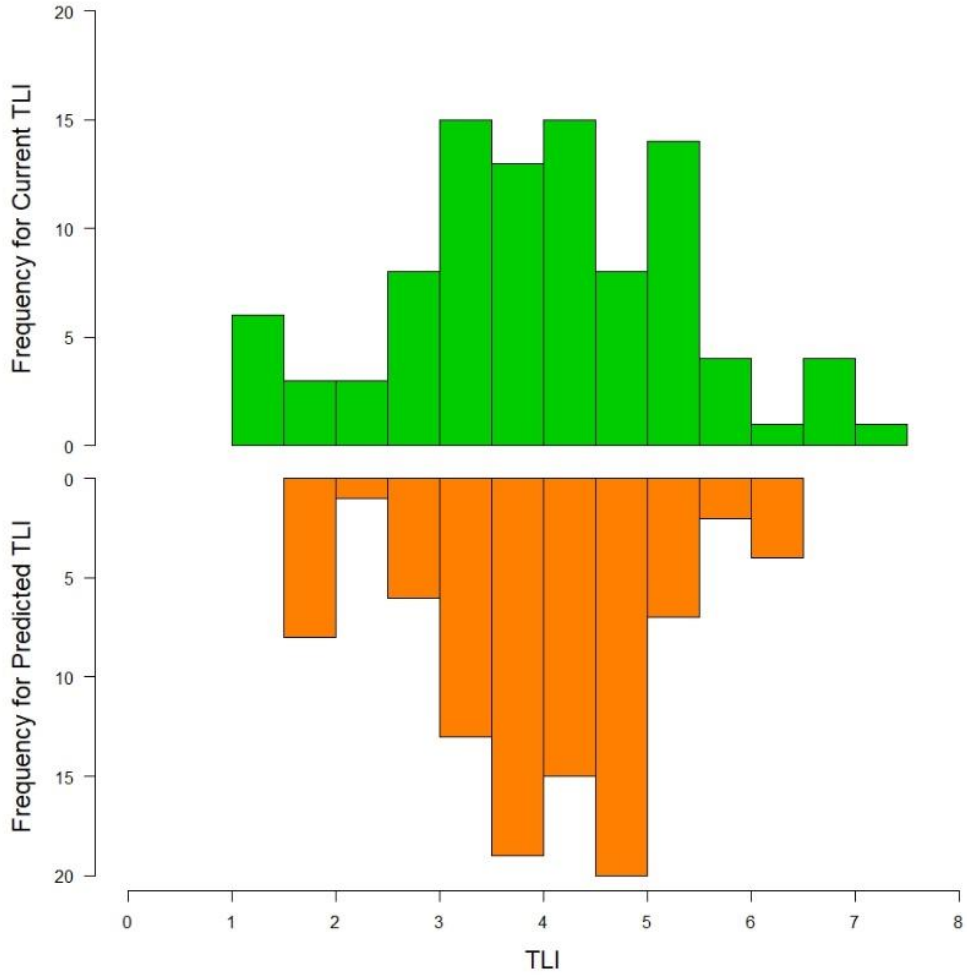


Figure 4.25. *Map of the Current Lake TLI's Within the Dataset Throughout New Zealand. Data from these 95 lakes were used to create the predictive model.*

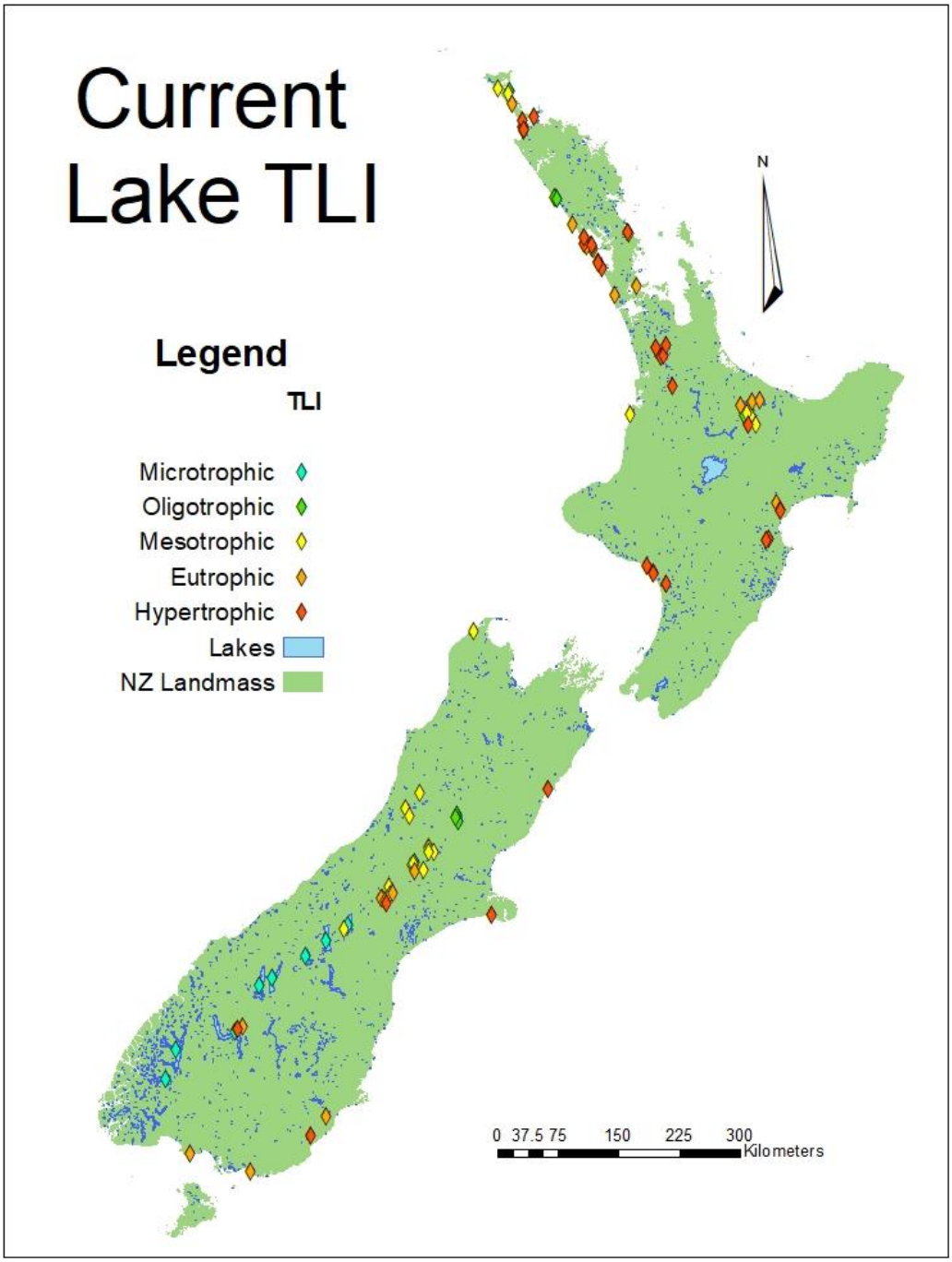
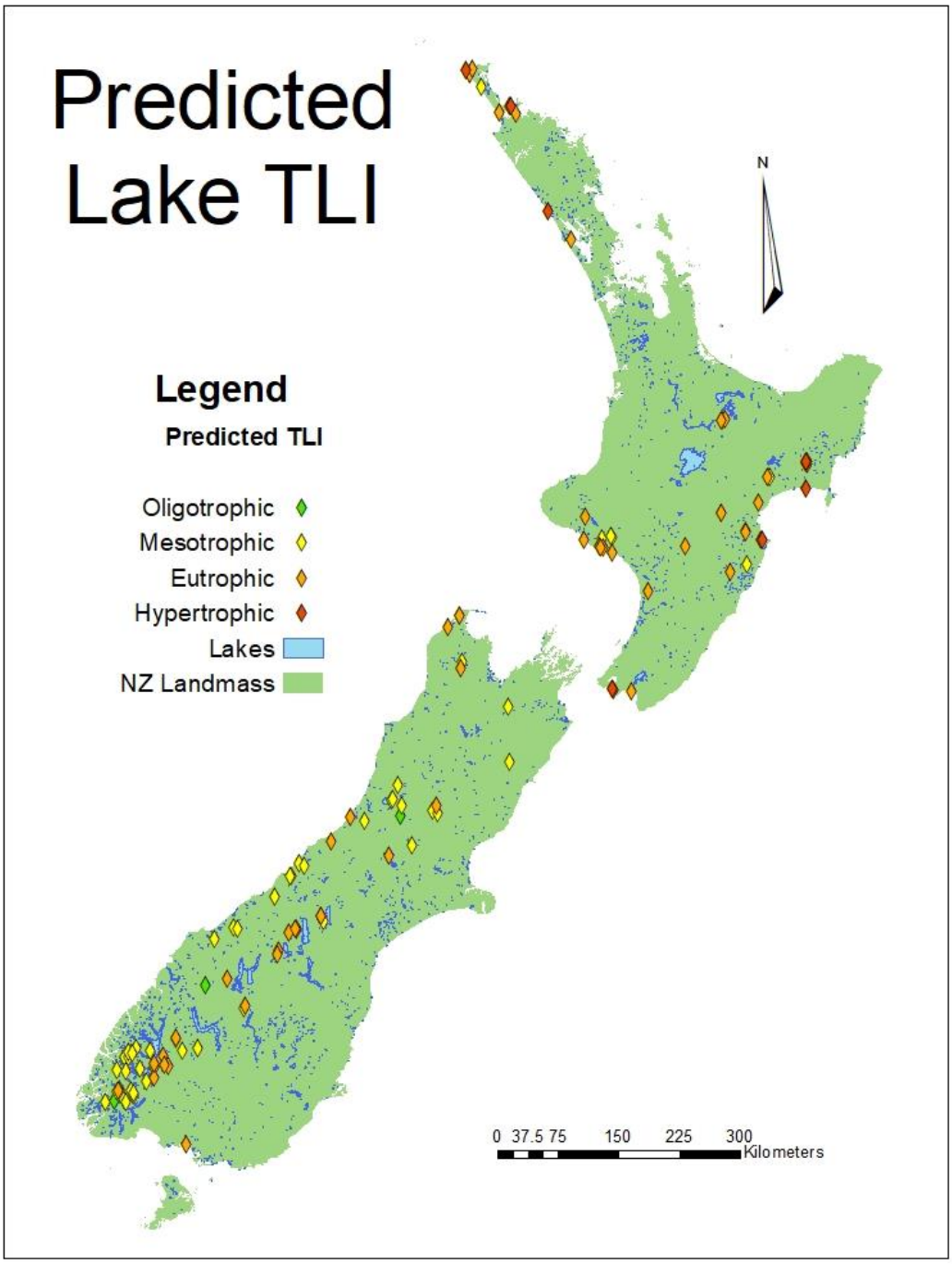


Figure 4.26. *Map of the Predicted Lake TLI's Within the Dataset Throughout New Zealand. Lakes within this dataset do not have monitoring data so no TLI could be calculated. The TLI was predicted using the random forest 'pigments +' model with an error rate of 8.8%*



4.2.6 Modelling the Lake Trophic Status Through Time

By applying the random forest 'pigments +' model to five different lake cores the range of TLI values indicating lake health through time was investigated. The retrodicted results were assessed as to whether they presented a realistic scenario of historical lake health. To contrast the results from the random forest models and the effects of the five static lake physical characteristic on the lake, the pigments-only model was performed alongside the 'pigments +' model. The approximate length of these sediment records spanned 837 years (Lake Nganoke), 455 years (Okataina), 893 years r (Oporoa), 844 years (Ponui), and 1,000 years (Rototoa). Ages of the lake core sediment samples were provided by GNS Science, Avalon, Lower Hutt.

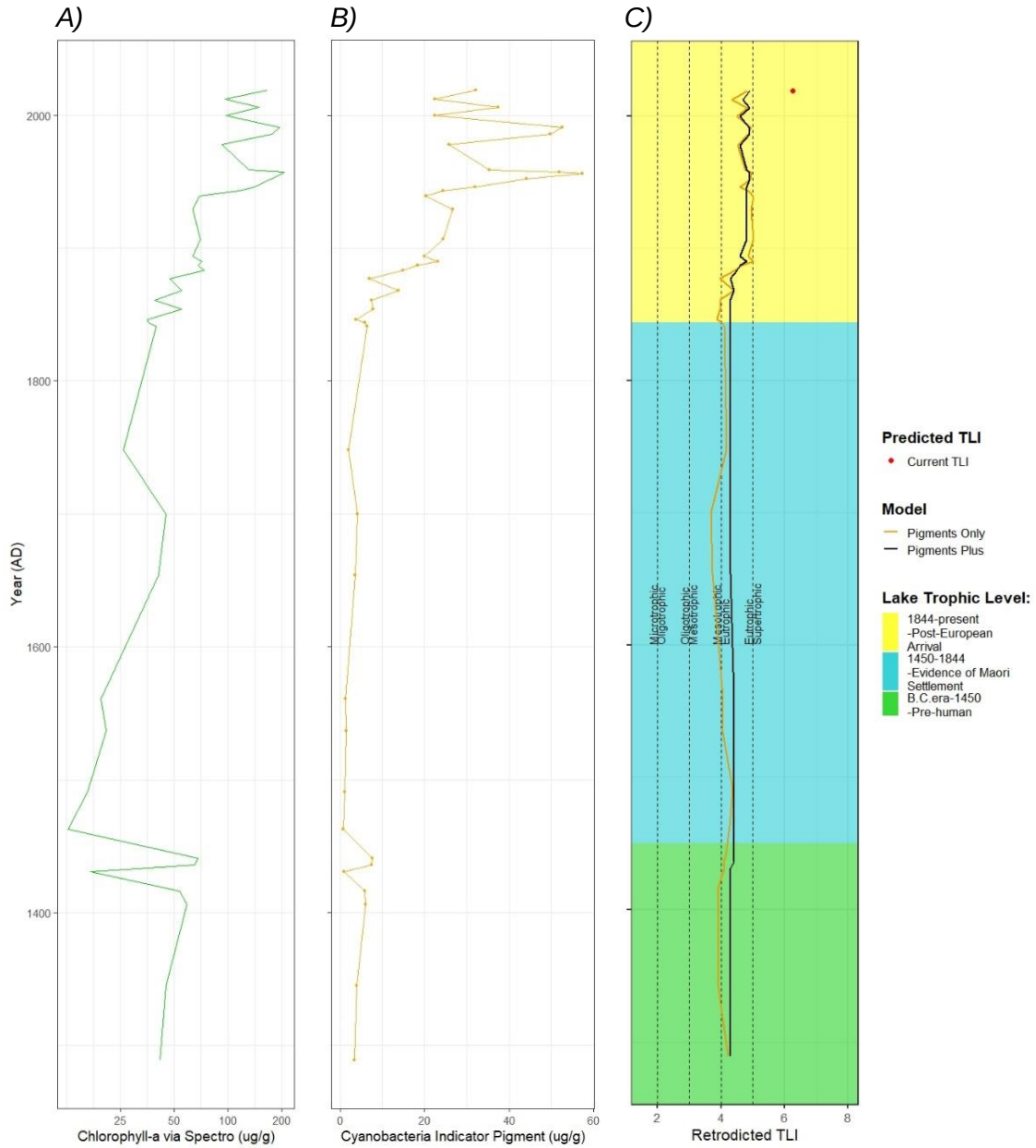
4.2.6.1 Lake Nganoke

For Lake Nganoke the random forest model could not predict a current TLI as no surface sediment sample was included in the original dataset of 232 lakes. However, a spot TLI was calculated using grab samples from the lake water when the lake was cored, this was calculated to be 6.2. The top of the lake core was estimated to be deposited in 2019, resulting in a TLI predicted by the 'pigments +' model of 4.9. In this situation the 'pigments +' model underestimated the TLI by 21%. For both models (pigments-only and pigments +) retrodicted TLI for Nganoke showed a trend towards increasing TLI values following the arrival of Europeans to the area (Figure 4.27C). In the 'pigments +' model, retrodicted values move from eutrophic to almost super-eutrophic after the arrival of Europeans to the area. Values retrodicted by the pigments-only model span a wider range, with the lake moving from the mesotrophic-eutrophic border to a eutrophic-supertrophic state in recent times. The retrodicted TLI's for both the 'pigments +' and pigments-only models conform with one another regarding the TLI of Lake Nganoke since the arrival of Europeans to the area.

The retrodicted TLI's for Nganoke showed a visual correlation through the patterns of peaks and troughs with the chl-a via spectrophotometry concentrations (Fig. 3.27A & C). A similar correspondence is observed between the phytoplankton pigment concentrations for cyanobacteria indicators (canthaxanthin, echinenone, myxoxanthophyll, and zeaxanthin) and the retrodicted TLI (Fig 3.27B & C). The only significant departure between the chl-a and cyanobacteria indicator pigment curves, and the two TLI curves occurs just prior to the approximate time where there was evidence of Māori settlement, when there is a peak in the chl-a via spectrophotometry and cyanobacteria indicator pigments which is not mirrored in the TLI values for either model.

Figure 4.27. *Lake Nganoke Retrodicted TLI and Comparison Charts.*

A) Chlorophyll-a via spectrophotometry (spectro) concentrations on a logarithmic x-axis, B) Cyanobacteria indicator pigment concentrations. Concentration is a combined total of four indicator pigments (Canthaxanthin, Echinenone, Myxoxanthophyll, and Zeaxanthin), C) Human settlement overlaid by predicted TLI. Black line denotes the 'pigments +' model, orange line denotes pigments-only model, black dashed vertical lines denote the transition from one trophic level to another, and red point denotes current TLI.



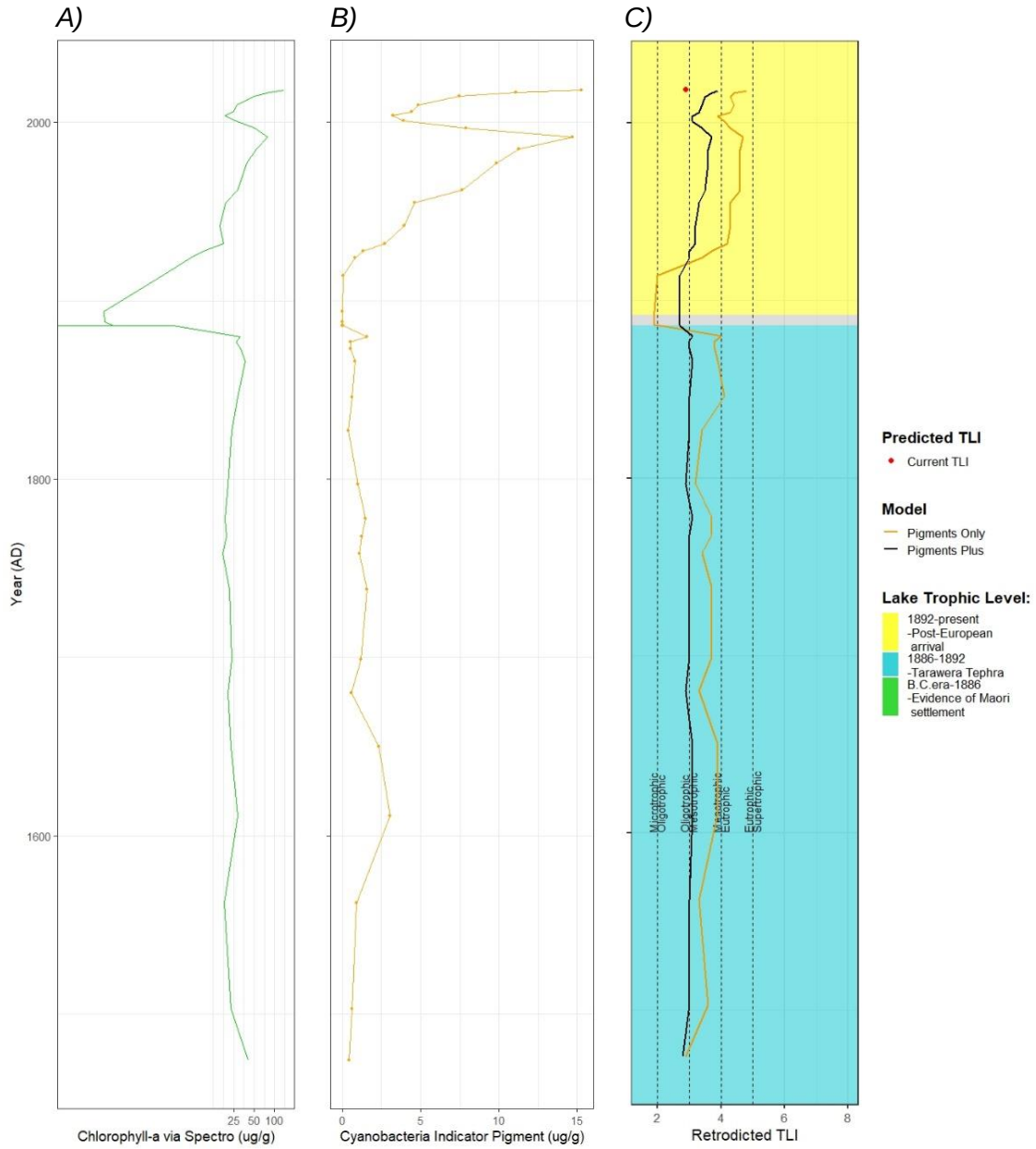
4.2.6.2 Lake Okataina

For Lake Okataina the random forest 'pigments +' model overestimated the modern TLI by 7%, where it predicted a current TLI of 3.1, whereas the TLI for that lake is currently 2.9. The retrodicted values for Okataina produced by the 'pigments +' model (Figure 4.28C) indicates that even before the arrival of humans to New Zealand, Okataina was sitting on the oligotrophic -mesotrophic border. However, since the arrival of Europeans to the area the retrodicted TLI's show an increasing trend, with the lake moving from the oligotrophic-mesotrophic border into mesotrophic during this period. As per Lake Nganoke, the values generated by the 'pigments +' model span a narrower range than those of the pigments-only model. In addition, the pigments-only model predicts a higher current TLI for Okataina from the surface sediments than the 'pigments +' model. The pigments-only model shows Okataina in a mesotrophic state before the arrival of humans to the area, then increasing to a eutrophic state after the arrival of Europeans.

A noticeable event in the history of Okataina which coincides with the arrival of Europeans to the area is the 1886 eruption of Mt Tarawera, located 13km to the southeast of the lake. Retrodicted TLI values for both the models drop suddenly, coincident with the eruption and remain low for approximately 20 years before returning to higher levels (Figure 4.28C). Cyanobacteria indicator pigments are low for most of the record and only really start to increase within the top layers of the core. The retrodicted TLI for Okataina from both models also shows high visual correlations through the patterns of peaks and troughs between the chl-a via spectrophotometry concentrations and the predicted TLI (Figure 4.28A & C). Additionally, there are also strong visual correlations through the patterns of peaks and troughs between the cyanobacteria indicator pigment concentrations (Figure 4.28B & C).

Figure 4.28. *Lake Okataina Retrodicted TLI and Comparison Charts.*

A) Chlorophyll-a via spectrophotometry (spectro) concentrations on a logarithmic x-axis, B) Cyanobacteria indicator pigment concentrations. Concentration is a combined total of four indicator pigments (Canthaxanthin, Echinenone, Myxoxanthophyll, and Zeaxanthin), C) Human settlement overlaid by predicted TLI. Black line denotes the 'pigments +' model, orange line denotes pigments-only model, black dashed vertical lines denote the transition from one trophic level to another, and red point denotes current TLI.



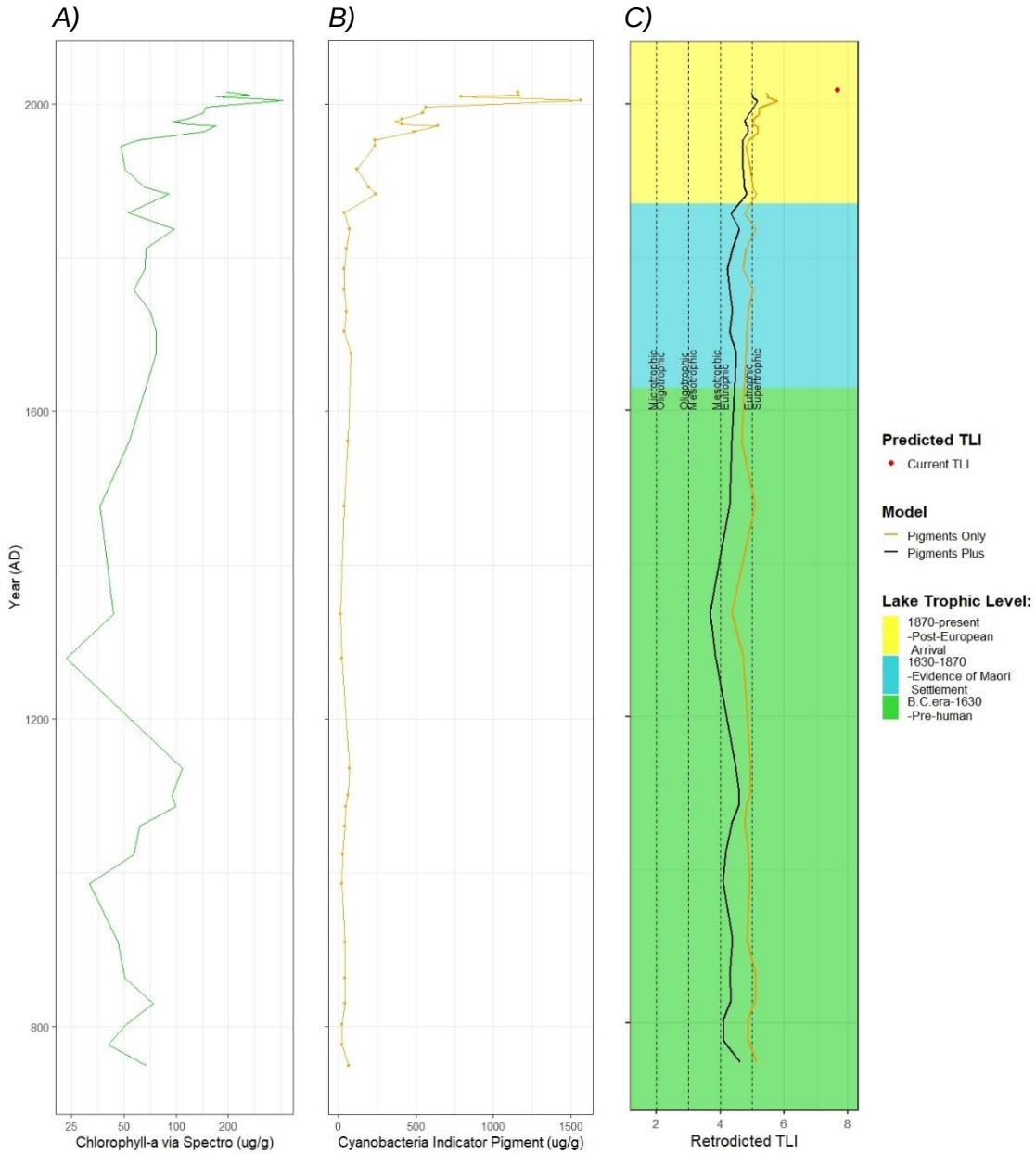
4.2.6.3 Lake Oporoa

For Lake Oporoa the random forest model predicted a TLI of 4.7 from the surface sediment samples collected from the dataset of 232 lakes. Unfortunately, there was no lake monitoring data to calculate a current TLI. However, a spot TLI was calculated using grab samples from the lake water when the lake was cored, this was calculated to be 7.7. The model underestimated the TLI of this prediction by 39%. The retrodicted TLI's produced using the 'pigments +' model for Oporoa (Figure 4.29C) showed an increasing trend since before the arrival of humans to the area. The retrodicted TLI is eutrophic in the pre-human era with a slight drop to mesotrophic around the 1300's before climbing to current super-eutrophic levels. The pigments-only model produces higher values of TLI, more eutrophic-supertrophic rather than eutrophic in the prehuman/evidence of Māori settlement eras before climbing firmly into the supertrophic level after the arrival of Europeans.

Retrodicted TLI's for Oporoa for both the 'pigments +' and pigments-only models showed visual correlations through the patterns of peaks and troughs between the chl-a via spectrophotometry concentrations (Figure 4.29A & C) and the cyanobacteria indicator pigments (Figure 4.29B & C), although not as obviously as Nganoke and Okataina. As in the case with Lake Okataina, cyanobacteria indicator pigments only really start to appear within the top layers of the core.

Figure 4.29. *Lake Oporoa Retrodicted TLI and Comparison Charts.*

A) Chlorophyll-a via spectrophotometry (spectro) concentrations on a logarithmic x-axis, B) Cyanobacteria indicator pigment concentrations. Concentration is a combined total of four indicator pigments (Canthaxanthin, Echinenone, Myxoxanthophyll, and Zeaxanthin), C) Human settlement overlaid by predicted TLI. Black line denotes the 'pigments +' model, orange line denotes pigments-only model, black dashed vertical lines denote the transition from one trophic level to another, and red point denotes current TLI.



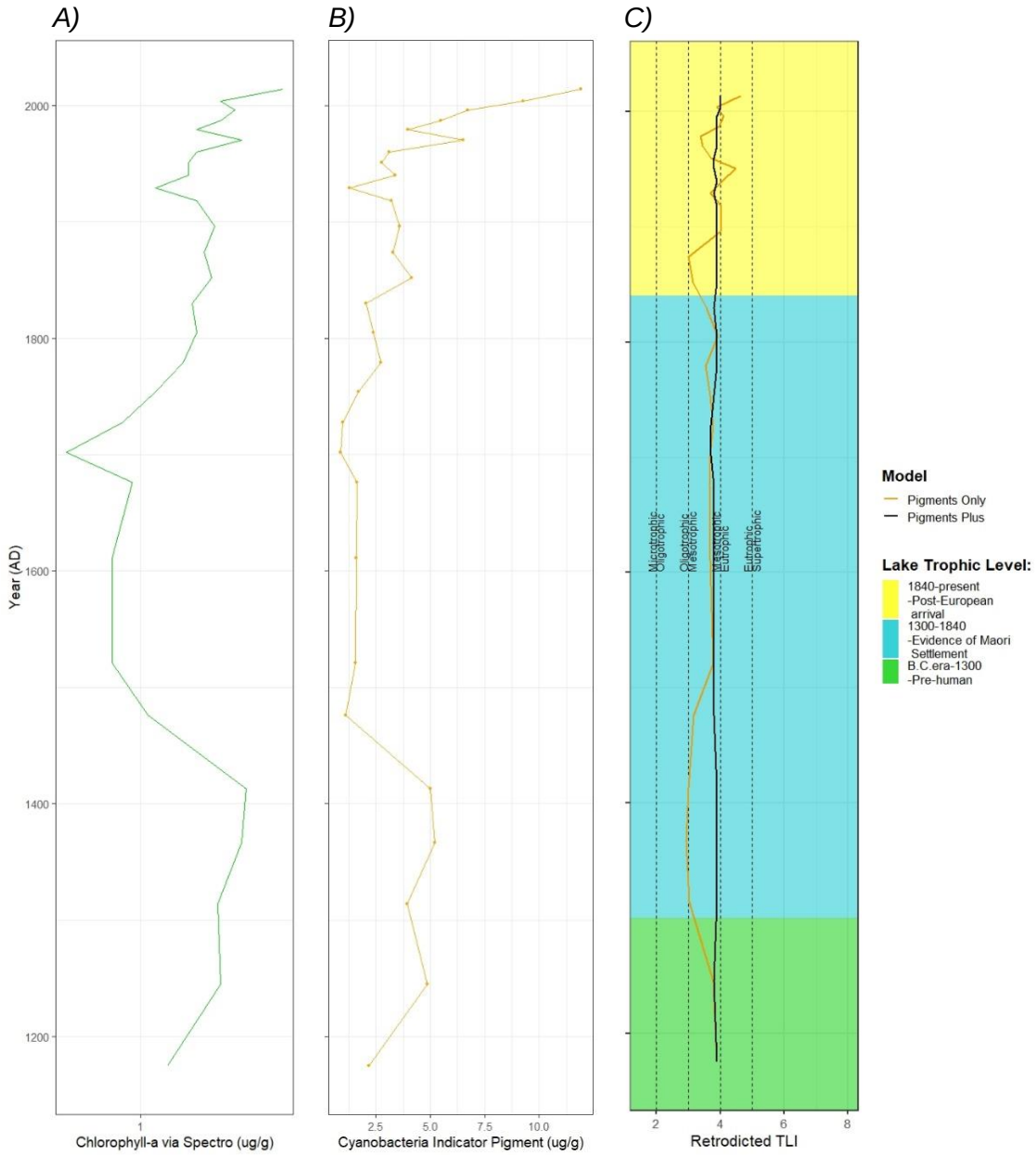
4.2.6.4 Lake Ponui

For Lake Ponui the 'pigments +' model could not predict a current TLI as no surface pigments sediment sample was included in the original dataset of 232 lakes when surface sediment samples were collected, and no other methods have been employed to determine the current TLI for this lake. The top of the lake core was estimated to be deposited in 2014, resulting in a TLI predicted by the 'pigments +' model of 4.0. The retrodicted TLI for Ponui (Figure 4.30C) using the 'pigments +' model indicate that during the pre-human/ evidence of Māori settlement eras, Ponui was in a mesotrophic state. After the arrival of Europeans to the area there was no major influence on the trophic status of the lake until approx. 20 years ago where the TLI increased, moving from mesotrophic to eutrophic. As is the case in the other lakes, the TLI values retrodicted by the pigments-only model span a wider range than those from the retrodicted TLI's with Ponui fluctuating within the mesotrophic level during the pre-human era.

Retrodicted TLI's for Ponui showed visual correlations through the patterns of peaks and troughs between the chl-a via spectrophotometry concentrations (Figure 4.30A & C) and the cyanobacteria indicator pigments (Figure 4.30B & C) within the last century. Prior to this, the peaks and troughs of the pigments do not correlate well with the TLI fluctuation. The downcore analysis showed that even before the arrival of humans to the area, lakes sediments for Ponui contained cyanobacteria indicator pigments.

Figure 4.30. *Lake Ponui Retrodicted TLI and Comparison Charts.*

A) Chlorophyll-a via spectrophotometry (spectro) concentrations on a logarithmic x-axis, B) Cyanobacteria indicator pigment concentrations. Concentration is a combined total of four indicator pigments (Canthaxanthin, Echinenone, Myxoxanthophyll, and Zeaxanthin), C) Human settlement overlaid by predicted TLI. Black line denotes the 'pigments +' model, orange line denotes pigments-only model, black dashed vertical lines denote the transition from one trophic level to another, and red point denotes current TLI.



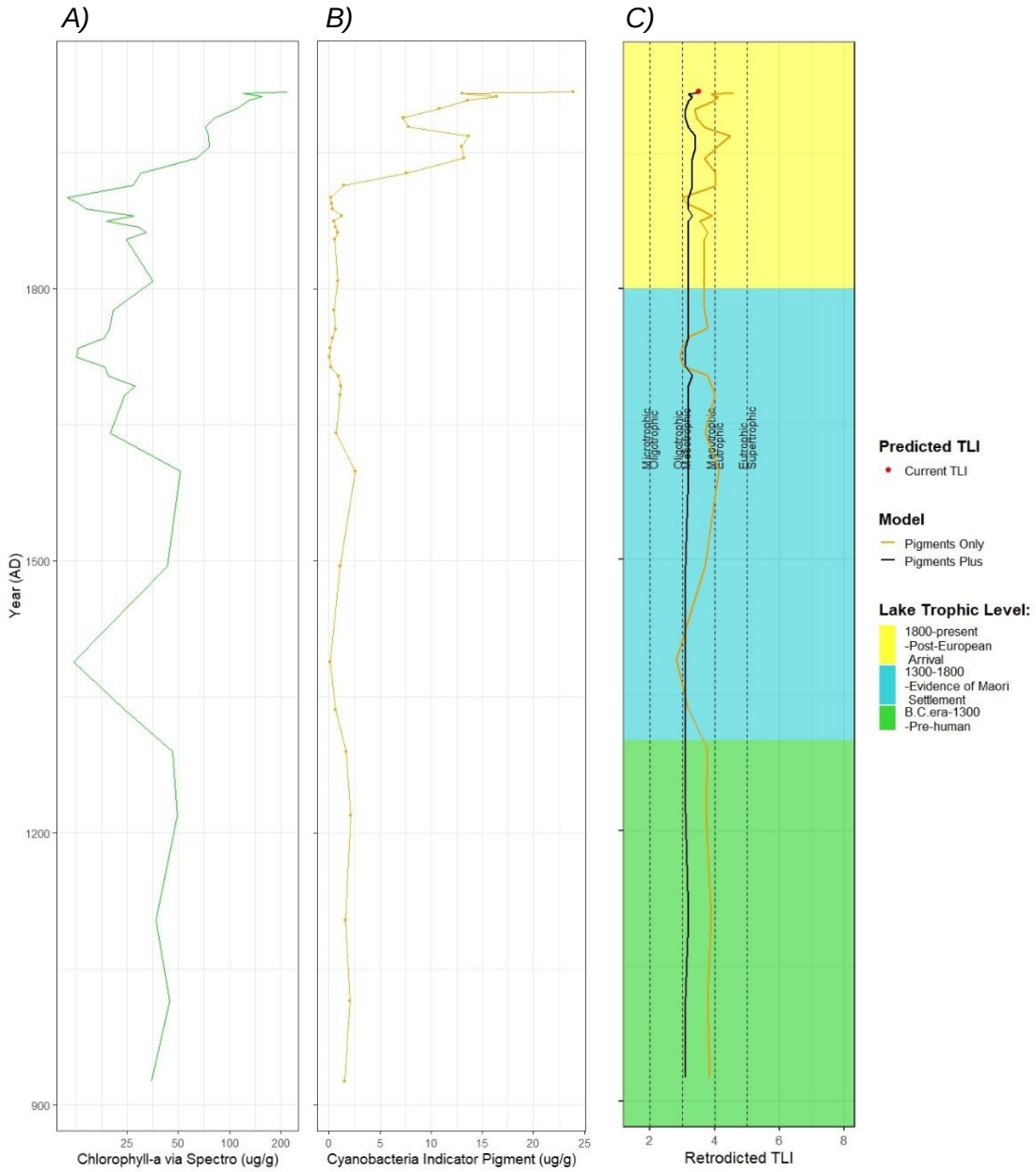
4.2.6.5 Lake Rototoa

For Lake Rototoa the random forest 'pigments +' model overestimated the TLI by only 3%, where it predicted a current TLI of 3.6, whereas the TLI for that lake is currently 3.5. The retrodicted TLI for Rototoa using the 'pigments +' model (Figure 4.31C) varies within a limited range (0.4) over the core record and has remained in a mesotrophic state since the arrival of humans to New Zealand. Additionally, cyanobacteria indicator pigments (Figure 4.31B) only really start to appear within the top layers of the core. Contrasting the 'pigments +' model to the pigments-only model shows less constraint in the retrodicted TLI's. However, the pigments-only model overshoots the present day TLI by 1.1. The retrodicted TLI's for the pigments-only model shows Rototoa fluctuating through the mesotrophic level before increasing to the eutrophic level.

The retrodicted TLI for Rototoa for both models show visual correlations through the patterns of peaks and troughs between the chl-a via spectrophotometry concentrations (Figure 4.31A & C) and the cyanobacteria indicator pigments (Figure 4.31B & C). The chl-a via spectrophotometry concentrations and the cyanobacteria indicator pigments show a recent increasing trend which does not coincide with the arrival of Europeans to the area. However, this is not obvious in the retrodicted TLI's of both models due to the fluctuations.

Figure 4.31. *Lake Rototoa Retrodicted TLI and Comparison Charts.*

A) Chlorophyll-a via spectrophotometry (spectro) concentrations on a logarithmic x-axis, B) Cyanobacteria indicator pigment concentrations. Concentration is a combined total of four indicator pigments (Canthaxanthin, Echinenone, Myxoxanthophyll, and Zeaxanthin), C) Human settlement overlaid by predicted TLI. Black line denotes the 'pigments +' model, orange line denotes pigments-only model, black dashed vertical lines denote the transition from one trophic level to another, and red point denotes current TLI.



Chapter 5 Discussion

This research was conducted with the objective of assessing the potential of phytoplankton pigments preserved/captured in sediments as indicators of lake water quality in both neoliminological (modern) and paleolimnological (ancient) settings in New Zealand lakes. In addition, HSI was evaluated as a method for detecting phytoplankton pigments within sediments. This aim has two distinct threads, 1) comparing sediment pigment measurements made by HSI against measurements made by the 'traditional' methods of HPLC and spectrophotometry; and 2) using pigment analyses from neoliminological sediments to build a predictive/retrodictive model for lake trophic status, and then applying that to the paleolimnological sediments.

5.1 Hyperspectral Calibration

This section of the research aimed to evaluate the limitations of utilising HSI to analyse lake core sediment samples. The evaluation was split into two parts. First, an investigation was conducted on whether the calibration of chl-a detected by HSI in lake core sediment samples required the use of HPLC-DAD or could spectrophotometry produce comparable results. Second, the ability of HSI to detect phycocyanin within lake core sediment samples was assessed.

5.1.1 Chlorophyll-a Calibration

The three methods of calibrating $RABD_{660\text{-}670}$ to chl-a analyses (HPLC-DAD or spectrophotometry both with and without acidification) showed that all three produce statistically significant correlations (p-value <0.05). The linear regressions revealed that chl-a via spectrophotometry without acidification showed more consistent linear regressions among the five lakes, with a r^2 falling within a narrower range (0.39) and the deviation of the gradient differing by 73.6%, as opposed to 81.2% for the chl-a via spectrophotometry with acidification. When regressed against data from all five lakes combined, the r^2 values were higher for chl-a analysed via spectrophotometry without acidification (0.61), and the 95% confidence interval of the slope was also more constrained comparatively, than the other chl-a analyses methods (HPLC-DAD and spectrophotometry with acidification).

Statistically there was little difference between the different methods. However, with spectrophotometry analyses, the absorbance of light is measured, and without acidification it will also measure chl-a diagenetic products, and other compounds absorbing light in that particular wavelength. Conversely, HSI measures the inverse of this technique, the reflection of light. As HSI measures samples in a comparable manner but using the reflected wavelengths of light rather than the absorbed wavelengths it is

more comparable to calibrate the $RABD_{660-670}$ signal measurements with spectrophotometric analysis without acidification. Additionally, chl-a and its diagenetic products provide useful information for paleolimnological analysis (Cohen, 2003), as they can be used to indicate the past presence of chlorophyll.

Additionally, two of the three methods of measuring chl-a are measuring different things. While HPLC-DAD and spectrophotometry with acidification are specific measurements of the chl-a compound, spectrophotometry without acidification analyses are measuring absorbed light from the sample at the specific wavelength of chl-a. As such spectrophotometry without acidification also measures chl-a diagenesis products, and other compounds absorbing light in that wavelength. HSI measures samples in a comparable manner, only it uses the reflected wavelengths of light rather than the absorbed wavelengths. Therefore, it is more appropriate to calibrate the $RABD_{660-670}$ signal measurements with spectrophotometric analysis.

Within existing research, the use of HSI to assess chl-a concentrations via HPLC using the $RABD_{660-670}$ signal has been restricted to the calibration of lake cores from a single lake. Butz et al. (2017); Grosjean et al. (2014); Makri et al. (2020); Sanchini and Grosjean (2020); Schneider et al. (2018). Spectral properties of the lake sediments creating variations in spectral strengths are theorised to interfere with the HSI measurements (Butz et al., 2015) suggesting that lake-specific calibrations need to be developed. However, the use of chl-a analysis via spectrophotometry without acidification to calibrate the $RABD_{660-670}$ signal returned an uncertainty of less than 8% for combined regression of all five lakes (Lakes Nganoke, Okataina, Oporoa, Ponui and Rototoa). Therefore, the utilisation of spectrophotometry measurements without acidification for HSI calibration opens the possibility for a universal equation.

The calibration of $RABD_{660-670}$ to chl-a concentrations measured by spectrophotometry without acidification was evaluated using linear regression. These values show that the models for all the lakes result in an uncertainty of less than 7.5%. This result was comparable to those achieved by others calibrating HSI data using HPLC chl-a analyses. Butz et al. (2017) achieved an error of 10-15%, Makri et al. (2020) achieved a calibration within 9%, and Zander et al. (2021) calibrated to within 5.4%. The majority of existing work follows the methods described by Butz et al. (2017) where the concentration of chl-a and its digenetic products are measured by HPLC. However, Butz et al. (2015) did suggest that spectrophotometric measurements on fresh sediments would be suitable for calibration of $RABD_{660-670}$ to chl-a.

5.1.1.1 *Limitations of the calibrations*

While HSI provides a two-dimensional image of the surface of a sediment core, HPLC-DAD and spectrophotometry analyses provide a representative analyses of the sample (Butz et al., 2015). Additionally, HSI has limitations in the shorter wavelengths (400 – 470 nm) which affect the detection of potentially important substances for ecology, such as carotenes and xanthophylls (Butz et al., 2015). Additionally, HSI possesses inherent uncertainties when assessing a target pigment connected to the potential interference of reflected spectral properties of the lake sediments (Butz et al., 2015; Yackulic, 2017). This is not necessarily limited to chl-a diagenetic products, as lake sediments contain a variety of other substances which could cause spectral variations. This interference could potentially transfer to the spectrophotometric calibration due to the method being less specific than HPLC-DAD analysis. Conversely, HSI provides a higher spatial resolution of information than analytical methods (HPLC-DAD and spectrophotometry both with and without acidification), as these analytical methods are limited to the resolution of the subsampling used to acquire the samples. Additionally, analytical methods can be limited in their sensitivity when small quantities of sample are available for extraction and analysis.

HPLC-DAD possesses its own limitations, particularly as it is highly compound specific. However, the low interference HPLC-DAD is subject to when analysing pigments balances this limitation. It is also more precise than some other traditional limnology methods, as it can distinguish and separate diverse phytoplankton pigments and their degradation products (Dos Santos et al., 2003). Spectrophotometry without acidification, on the other hand, is less compound specific than HPLC-DAD and spectrophotometry with acidification, which can result in an overestimation of chl-a in presence of its diagenetic compounds. As the diagenetic products will be measured and calculated into the chl-a concentration. However, the combined values of chl-a and its diagenesis products will still provide useful relative information for paleolimnological analysis (Cohen, 2003), particularly for evaluations of lake primary productivity.

5.1.1.2 *Universal Calibration*

The linear regression relationship for a universal calibration of the $RABD_{660\text{-}670}$ HSI signal using spectrophotometry without acidification determined chl-a concentrations was statistically significant ($p < 0.05$). The combination of five lakes resulted in a statistically significant ($p = < 2.2 \times 10^{-16}$) linear regression relationship with a r^2 value ($r^2 = 0.61$) comparable to the individual lake calibrations. However, when Lakes Nganoke, Okataina, Ponui and Rototoa were combined (without Lake Oporoa), the fit of the regression line (indicated by the r^2) did not improve. Nevertheless, while a universal

calibration between $RABD_{660-670}$ and chl-a concentrations analysed via spectrophotometry without acidification appears feasible more work needs to be conducted to confirm this.

5.1.1.3 Future Work

The combined dataset of lakes achieving a chl-a spectrophotometry without acidification determined calibration r^2 of 0.61 opens the possibility for a universal calibration equation. However, this universal calibration should be validated with a larger dataset of lakes to understand whether some lake possess significantly differing calibrations. Additionally, more work needs to be undertaken to investigate the potential outliers and lakes which differ in their results, to determine which aspects within these lakes are causing the differing results. If a consistent cause can be identified, different calibrations for different lake categories might be required.

5.1.2 Phycocyanin Detection

It has been noted by Yackulic (2017), Favot et al. (2020), and Sorrel et al. (2021) that the $RABD_{615}$ HSI signal might reflect the presence of the phycobiliprotein phycocyanin in sediment core samples. To determine if this observation is correct, phycocyanin analysis via fluorescence was undertaken on sediment core samples. When sediment core samples were analysed, no phycocyanin was detected above the level of detection (0.1 µg/mL in the extract or 0.05 µg/g of wet sediment, if 1 g of sample was extracted). Therefore, no quantifiable amounts of phycocyanin could be detected through the fluorescence method in lake core sediment samples, despite the lake cores showing detectable signals at $RABD_{615}$. However, during the development of the phycocyanin analysis via fluorescence quantifiable concentrations of phycocyanin were able to be detected within cyanobacterial cultures extracts.

Of the study lakes within this analysis, Lake Nganoke (Parrish, 2020) and Oporoa (Roygard et al., 2019), have confirmed levels of cyanobacteria. However, the topmost sample of sediment core dated 2019 AD (Lake Nganoke) and 2015 AD (Lake Oporoa) did not return a phycocyanin concentration above the level of detection. This suggests that within the study lakes, phycocyanin is not integrated into the sediment layer. Yacobi et al. (2015) noted that if phycocyanin concentration was below 10 mg/m^3 or 0.01 µg/mL and/or the phycocyanin to chlorophyll ratio was below 0.5, then the phycocyanin will be masked by the optical activity of the chlorophyll pigments. Yacobi et al. (2015) noted that phycocyanin cannot be quantified by remote sensing techniques unless these conditions had been met to avoid the phycocyanin being masked by the chlorophyll.

During the preliminary assessment of phycocyanin (via fluorescence platereader), phycocyanin concentrations were measured in 16 cyanobacteria cultures. This

confirmed that phycocyanin extracted from cyanobacterial cells is detectible using this method. However, this method could not detect quantifiable levels of phycocyanin within lake core sediment samples. This indicates that the $RABD_{615}$ signal is more likely to be measuring chl-a within the lake core rather than phycocyanin. A linear regression of $RABD_{615}$ to chl-a concentrations (via spectrophotometry with acidification) resulted in a statistically significant correlation ($p < 0.05$) for all lakes with r^2 values between 0.33 and 0.74. Additionally, similar correlations between the cyanobacterial indicator pigments echinenone and the zeaxanthin to the $RABD_{615}$ bandwidth were observed. However, for both echinenone and zeaxanthin the correlation was not as strong as it was for the chl-a concentrations (via spectrophotometry with acidification).

Several studies have assessed the remote sensing of phycocyanin within water bodies. Two of which included: Li et al. (2010) where HSI was calibrated to phycocyanin concentrations resulting in a r^2 of 0.79 for a mesotrophic reservoir (Geist Reservoir, Indiana, USA), and Kwon et al. (2020), who calibrated phycocyanin detection with drone based HSI with $r^2 > 0.8$. However, these studies have focused on the optically active nature of phycocyanin using HSI, where the distinctive reflectance of phycocyanin on the surface of the water itself was detected. With reference to paleolimnology, phycocyanin has not been widely utilised, as it is assumed that due to the high water solubility of the pigment, it will not be deposited out of the water column and into the sediment (Leavitt & Hodgson, 2002; Sanger, 1988). This appears to be correct, as phycocyanin was not detected in the sediment core samples analysed during this study.

While the remote sensing of phycocyanin within a water body appears to be promising for the detection of cyanobacterial blooms, phycocyanin was not detected in lake sediment core samples within this study. Stronger correlations observed between the $RABD_{615}$ HSI signal and chl-a concentration, compared to cyanobacterial indicator pigment concentrations, suggest that it is measuring a chl-a side peak rather than phycocyanin (as suggested in Yacobi et al. (2015)).

5.2 Predicting Lake Water Quality Using Surface Sediments

This second research component investigated the use of surface sediments to infer lake water quality using key phytoplankton pigments. The first part of the research assessed whether the quantification of key phytoplankton pigments within surface sediments could be used to create an accurate predictive model for lake trophic level indicators (TLIs). The second part of this section assessed whether the predictive model created for the first part of the section could be used to retrodict historical TLIs from lake sediment cores.

5.2.1 Predictive Lake Health Model

To assess the effectiveness of key phytoplankton pigments within surface sediments to create a predictive lake health model using TLI measurements, this section created several predictive models using a machine learning approach (regression trees, random forest models, and boosted gradient models): The use of land use proportions within the catchment to extrapolate the TLI of the lake; the use of key phytoplankton pigments to create a TLI predictive model; and the use of key phytoplankton pigments in addition to other lake characteristics to construct a TLI predictive model. These models were created using data at a nationwide scale.

5.2.1.1 The Effect of Land Use Proportions on Lake TLI Prediction

The regression tree using catchment land use to predict lake TLI returned a RMSE of 1.04 (25.8%). This tree highlighted the key driver utilised by this model as the proportion of high-producing exotic grassland within the catchment. This model was built to assess the feasibility of creating a predictive model for lake TLIs using the relationship between land use and TLIs. Previously, within New Zealand, Galbraith and Burns (2007) was able to link concentrations of nutrients within the lake water, and thus the TLI, to the extent of modification of the catchment. Additionally, Paul et al. (2012) correlated the land use in lake catchments to the trophic status of eleven New Zealand lakes located in the Rotorua lake district. Noting that lakes within catchments with higher nutrient loads from pastoral land use possessed a higher TLI and cyanobacterial dominance. Conversely, lakes with lower nutrient loads and a higher proportion of native land use within the catchment possessed a lower TLI and a chlorophyta dominance. This was supported by the catchment land use regression tree which found the highest driver of lake TLI to be high producing exotic grassland which will deliver a higher nutrient load to the lake. Paul et al. (2012), Peterson et al. (2007), and Jeppesen et al. (2005), also demonstrated relationships between the phytoplankton, and therefore the pigments laid down in the sediment, within the lakes and the corresponding land use of the lake catchment.

5.2.1.2 The Effect of Phytoplankton Pigments on Lake TLI Prediction

A regression tree between the lake TLI and the phytoplankton pigments with the surface sediments of the lake returned a RMSE of 1.13 (27.8%). While not as accurate as the regression tree created using land use proportions, the phytoplankton pigment model highlighted the key pigment drivers utilised by this model as zeaxanthin (a cyanobacteria indicator pigment), followed by alloxanthin (a pigment of the phytoplankton cryptophyta and chlorophyta) and echinenone (also a cyanobacteria indicator pigment). Additionally, a random forest model using twelve phytoplankton pigments (with three different measurements of chl-a) within the surface sediments of the lake was developed. The resulting pigment model possessed a RMSE of 0.49 (11.9%).

This tree highlighted the key phytoplankton pigments contributing to this model as the same as the regression tree, but with differing importance of the pigments. The primary pigment was alloxanthin, followed by echinenone and zeaxanthin. Additionally, alloxanthin, echinenone and zeaxanthin are some of the more stable carotenoids found in lake core sediments, isolated in sediments up to and over 5,000 years old (Watts & Maxwell, 1977; Züllig, 1986). The model placing importance on alloxanthin shows that it is also placing importance on low trophic levels as alloxanthin is indicative of chlorophyta, associated with lower TLIs in New Zealand Lakes (Paul et al., 2012). In addition to echinenone and zeaxanthin, which are indicative of cyanobacteria, associated with higher lake TLIs in New Zealand Lakes (Paul et al., 2012). This importance resulting from the model also reinforces the importance of the cyanobacterial indicators echinenone and zeaxanthin when it comes to determining lake health.

5.2.1.3 Creating a Predictive Model for Lake Trophic Status

Three machine learning models were created (regression tree, random forest, and boosted regression tree) for use on the sediment core samples to predict lake TLI using twelve phytoplankton pigments (with three different measurements of chl-a), plus five static lake physical characteristics (described hereafter as the 'pigments +' model).

The random forest predictive model provided the most accurate predictive power with a RMSE of 0.39 (10.0%). As opposed to the 'pigments +' regression tree (RMSE of 0.96 (24.6%)) and the 'pigments +' boosted regression tree (0.94 (24.2%)). The 'pigments +' regression tree identified the key drivers utilised as those related to the size of the lake (maximum depth of the lake, lake area, and lake volume). This aligns with research conducted by Hamill and Lew (2006), who identified lake depth as being closely associated with TLI within New Zealand lakes. Showing that lakes with a higher TLI were more frequently shallow lakes, and lakes with a lower TLI were more frequently deeper lakes. Additionally, the TLI association with lake depth corresponded with shallow lakes having a lower capability to assimilate nutrient loads because of their lower water volumes (Hamill & Lew, 2006).

The most influential phytoplankton pigments were identified as zeaxanthin and echinenone (both cyanobacterial indicator pigments). Additionally, the random forest identified the key drivers as the maximum depth of the lake, the lake catchment size, and the lake volume, indicating that the larger these variables, the lower the TLI would be predicted. The most influential phytoplankton pigments being echinenone and canthaxanthin (both also cyanobacterial indicator pigments). Furthermore, the 'pigments +' boosted regression tree identified the key drivers as the maximum depth of the lake and elevation of the lake, with the most influential phytoplankton pigment being

echinenone. The 'pigments +' random forest model was the best preforming model returning a r^2 of 0.95 when the TLI was regressed against the predicted TLI.

As the 'pigments +' models contained three different analyses for chl-a three versions of the random forest model were created with one chl-a measurement in each. Chl-a analysis via spectrophotometry with acidification, was the best preforming model with a RMSE of 0.37 (9.3%) and r^2 of 0.95. The improvement of the predicted error rate indicates that the combination of the three chl-a analyses within one model were overfitting the model.

When using the 'pigments +' random forest model to predict lake TLI on surface sediment samples it was noted that the predicted TLIs were constrained to the more central values. The model was not predicting out to the limits of the range of the known TLIs (both low and high). This constraint could be indicative of the model trying to predict a single variable from such a dynamic system that the model is regressing towards the average as this is the most straightforward means of prediction. Additionally, the constraint seen in this model could also indicate variables not included in this model which influence the peripheral values.

5.2.2 Retrodictive Lake Health Model

To contrast and assess the accuracy of the two best performing models (the 'pigments +' random forest model and the pigments-only random forest) they were used to retrodict lake trophic states and modelled side by side. Additionally, as previously noted, the 'pigments +' model was constrained (either due to the complexity of the lake ecosystem not being captured in the model or variables not included in this model which influence the peripheral values.) with its predictions, therefore, the pigments-only random forest was modelled alongside to provide a contrast and assess the accuracy of the retrodictions. However, the inclusion of the physical variables that would not change through time or would only change very slowly was included as this bought the predicted TLIs closer to an accurate estimate. With regards to the retrodicted TLIs these physical variables do not permit the model to fluctuate to the extremes unless a strong pigment concentration is detected. The 'pigments +' model allows is more accurate for modelling accurate current lake TILs, while the pigments model is better suited for modelling TLI changes over time.

The two models were applied to lake sediment cores from five New Zealand lakes, with sediment records spanning: ≈ 800 years for Lake Nganoke, ≈ 450 years for Lake Okataina, ≈ 900 years for Lake Oporoa, ≈ 850 years for Lake Ponui, and ≈ 1,000 years for Lake Rototoa.

5.2.2.1 *Lake Nganoke*

The retrodicted TLI for Lake Nganoke from the 'pigments +' random forest model suggests that even before the arrival of humans to New Zealand, Nganoke was in a eutrophic state. The retrodicted TLI fluctuated very little (4.3 to 4.9) throughout the ≈ 800 years that the core spanned. Land use within the catchment for Nganoke has undergone heavy modification since the arrival of humans to the area, with only 5% of the remaining flora being native. The remaining vegetation is high producing grassland (84%) and exotic forest (8%). The indication given by the regression tree created to assess the effect of land use type proportions on trophic levels of lakes is that the proportion of various land use types (i.e., high producing exotic grassland) will have a noticeable effect on the lake TLI. This is not observed in the Nganoke 'pigments +' random forest model. Unfortunately, a spatial comparison (comparison of another lake with comparable attributes) of the lake is not feasible. Of the other lakes within the dataset compiled for this research, with four possessed a comparable catchment size to Lake Nganoke, in addition to being small shallow lakes. However, Lake Nganoke possesses the largest proportion of native flora within the catchment of the comparable lakes. Therefore, a comparison of Nganoke to an equivalent relatively unmodified lake catchment to compare the retrodicted TLI's is not practical.

A study conducted by Parrish (2020) used bacterial metabarcoding to retrodict the water quality of Nganoke. This study found that during the pre-human era and during the time of Māori settlement there was relatively low diversity in bacterial taxonomic composition, suggesting that the lake was microtrophic to oligotrophic. After the arrival of Europeans, the algal diversity in Lake Nganoke increased noticeably, suggesting that the lake moved up to the eutrophic level or higher. Neither the 'pigments +' model nor the pigments model reflected Nganoke as low as oligotrophic. However, both models reflected the increase to higher TLI after the arrival of Europeans.

5.2.2.2 *Lake Okataina*

Lake Okataina's retrodicted TLI's indicates that even before the arrival of humans to New Zealand, Lake Okataina was sitting on the boarder of oligotrophic and mesotrophic. Apparent within the lake sediment core was visual evidence of the 1886 Tarawera eruption which coincides with the arrival of Europeans to the area (as seen via the tephra layer within the core). Additionally, cyanobacteria indicator pigments only really start to appear within the top layers of the core, shortly after the lake recovered from the eruption. Indicating that they were only recently incorporated into the lake sediment. Like Lake Nganoke, Lake Okataina is unable to be spatially compared to other similar lakes as it possesses the highest proportion of native flora from the comparable lakes (72%).

However, the retrodictive TLI from the 'pigments +' random forest model for Lake Okataina highlights a possible limitation of the model. Following the eruption of Tarawera all phytoplankton and their respective pigments should have been decimated. The refore, a pronounced drop in the retrodicted TLI would be expected. Within the tephra from Tarawera in the lake sediment the retrodicted TLI from the 'pigments +' model drops to its lowest value at 2.7. The possible reason for this is the static lake physical characteristics used in building the model. While these parameters allow for a relatively accurate predictive model between lakes, they could be introducing a bias within the model when it is used on a core of the same lake. Nonetheless, the small variation within the 'pigments +' model retrodicted TLI's (2.7 to 3.9) is plausible as the catchment is relatively unmodified.

A recent study conducted by Caird (2021) noted that after the Tarawera eruption and European settlement the lake possessed similar levels of primary productivity as the pre-eruption levels. However, the bacterial community has altered post-eruption with an increase in bacterial diversity. This trophic level change appears to be reflected in the 'pigments +' model with Lake Okataina possessing oligotrophic-mesotrophic waters pre-eruption. While post-eruption modelled trophic levels have shifted slightly higher to mesotrophic levels.

5.2.2.3 Lake Oporoa

The retrodicted TLIs for Lake Oporoa showed an increasing trend since before the arrival of humans to the area (both European and Māori). The predicted downcore TLI is eutrophic in the pre-human era with a slight drop to mesotrophic around the ≈ 1300's before climbing to current super-eutrophic levels. Spatially Lake Oporoa is geomorphically comparable to Lake Marion, with similar depths, volumes, and catchment size. However, Lake Marion sits at ≈. 680 m elevation while Lake Oporoa at ≈. 360 m. Lake Oporoa is a highly modified catchment with no native flora present (0%), whereas Lake Marion is relatively unmodified (87% native flora) and has a current TLI of 3.5 (mesotrophic), as opposed to Lake Oporoa's 7.7 TLI (supertrophic). The lowest retrodicted TLI for Lake Oporoa was 3.7 during pre-human times, which is comparable to Lake Marion's 3.5. The average pre-human retrodicted TLI for Lake Oporoa was 4.3, while slightly higher (eutrophic as opposed to mesotrophic) the elevation difference between the lakes makes this value conceivable.

5.2.2.4 Lake Ponui

Lake Ponui's retrodicted TLIs showed a recent sharp increase (in the past 2 decades), moving from mesotrophic to eutrophic. Prior to this Europeans had arrived within the area without influencing the trophic status of the lake significantly. However, the

retrodicted downcore TLI indicates that even before the arrival of humans to the area, Lake Ponui was in a mesotrophic state. Spatially Lake Ponui is geomorphically comparable to Lake Wiritoa, with similar elevation, depth, volume, and catchment size. However, Lake Ponui has a relatively unmodified catchment (84% native flora), whereas Lake Wiritoa is highly modified (2% native flora). Lake Wiritoa has a higher current TLI of 5.3, reflecting the catchment modification, while Lake Ponui's retrodicted recent TLI is 4.0.

While it is plausible that Lake Ponui possess a smaller TLI than Lake Wiritoa due to the catchment modification. However, Lake Ponui's retrodicted TLI is high for a relatively unmodified catchment. This could possibly be due to the presence of Perch within the lake (Jellyman, 1980) as zooplankton grazing by perch has been linked via a trophic cascade to cyanobacteria blooms (Smith & Lester, 2007).

5.2.2.5 Lake Rototoa

The retrodicted TLI's for Lake Rototoa show a consistent TLI throughout history, varying only 0.4 during the last millennium. However, cyanobacteria indicator pigments only really start to appear within the lake sediments within the last century (≈ 1920's), ≈120 years after the arrival of Europeans, but ≈ 50 years before the intensification of pine plantations to the catchment (Holloway, 2020). Spatially Lake Rototoa is geomorphically comparable to Lady Lake, with similar elevation, depth, and volume. However, Lady Lake has a catchment three times the size of Lake Rototoa. Nonetheless, Lady Lake possesses a relatively unmodified catchment (79% native flora), as opposed to Lake Rototoa (32% native flora). The current TLI of Lady Lake (3.4) is comparable to Lake Rototoa's (3.5). Therefore, the retrodicted TLI for Lake Rototoa, varying from 3.1 - 3.5, with a current TLI predicted by the model of 3.6 is a conceivable figure.

5.2.3 Comparison of Retrodicted Lakes

Within the dataset Lakes Nganoke and Oporoa are comparable small, riverine lakes, with small highly modified catchments. The TLIs modelled with the 'pigments +' model was high for both lakes. Neither of these lakes possess council monitoring data, so current TLIs were estimated from 'spot' samples (water samples collected at the time of coring). Lakes Nganoke and Oporoa show retrodicted TLI fluctuations to be less constrained than the other 'pigments +' models, through the eutrophic level, with Lake Oporoa climbing into the supertrophic level in recent times. For Lakes Nganoke and Oporoa the TLI retrodicted by the 'pigments +' model on the lake core sediment samples appears to be underestimated by the model when compared to the 'spot' TLIs (4.9 as opposed to 6.2, and 4.9 as opposed to 7.7, respectively). Additionally, the 'pigments +' model for the surface sediment sample for Lake Oporoa underestimated the lake TLI

(4.7 as opposed to 7.7) when compared to the 'spot' TLI. However, both the 'pigments +' models for the surface sediment and the lake core sediment samples for Lake Oporoa align closely (4.7 and 4.9 respectively). No surface sediment sample was collected for Lake Nganoke.

Lakes Okataina and Rototoa were also comparable deep lakes with large catchments and mesotrophic TLIs retrodicted with the 'pigments +' model. However, Okataina is a volcanic lake with a large area, while Lake Rototoa is a wind/aeolian lake with an area just over 100 ha. Lakes Okataina and Rototoa possess council monitoring data to calculate the current TLIs. The 'pigments +' model closely predicted the TLIs using the surface sediment samples at 3.1 for Lake Okataina (TLI of 2.9) and 3.6 for Lake Rototoa (TLI of 3.5). The retrodicted TLI for Lake Okataina using the lake core sediment sample was overestimated by the model (3.9) at the surface of the lake core, and it appears that pre-human era TLI's were overestimated by the pigments-only model for Lake Rototoa.

Lake Pounui, like Lake Oporoa, is a deep small lake, and like Lake Rototoa, it is a volcanic lake. Additionally, like Lakes Okataina and Rototoa, Lake Ponui has a large catchment and a retrodicted TLI as mesotrophic. However, Lake Ponui does not possess council monitoring data, nor was a surface sediment sample taken from the lake. Therefore, a comparison of the TLI retrodicted by the 'pigments +' model using the lake core sediment sample cannot be compared other TLIs for the lake.

By taking the lake TLIs retrodicted by the 'pigments +' model applied to the lake core sediment samples, presents an interesting situation. As the assumption of many degraded lakes throughout New Zealand, is that this is of anthropogenic origin. However, the retrodicted TLI's of these five lakes suggests, that while anthropogenic influence is exacerbating the degradation of the lakes since the arrival of humans to the area, before this the trophic levels of these lakes have not drastically increased.

5.2.4 Detection of Fossilised Cyanobacteria Indicator Pigments

Apparent in all five lake cores are the higher levels of cyanobacteria indicator pigments within the most recent lake sediments. This could be an indication of degradation of the pigments within the sediments with depth. However, Lakes Nganoke and Rototoa show the proportional changes with the cyanobacterial indicators are greater than the proportional changes in chl-a. This indicates that if degradation of the cyanobacterial indicators was a factor, we could expect to see less degradation than this. Additionally, higher levels of the cyanobacteria indicator pigments also coincide with the arrival of Europeans to the catchments of most of the lakes (Lakes Oporoa and Rototoa, and heightened levels in Lakes Nganoke and Okataina), therefore, a rise in cyanobacteria at

this point would be expected, due to the impacts of catchment disturbance. Lakes Nganoke, Ponui, and to some extent Okataina show cyanobacteria indicator pigments recovered from the lake sediments as old as ≈ 400 to 700 years old.

Existing literature has demonstrated the resilience of certain pigments in Lake sediment samples, for example, Züllig (1986) isolated echinenone from 14,000-year-old lake sediment (Lake or pond Lobsigen, Switzerland), while Watts and Maxwell (1977) isolated echinenone, zeaxanthin, and canthaxanthin in samples up to 5,000 years old in marine sediment samples from the Cariaco Trench, before the echinenone and zeaxanthin showed degradation. Within the same core, canthaxanthin was detected in sediments 340,000 years old. These studies suggest that decrease of cyanobacteria indicator pigments down the older lake core sediment samples is not connected to the degradation of the pigments, rather it is linked to the arrival of Europeans to the area.

5.2.5 Limitations of the 'Pigments +' Model

The predictive 'pigments +' model developed is potentially limited as the measurements within this research have not been corrected for variations in sedimentation rates. Additionally, the retrodicted models were not corrected for possible pigment degradation down the lake core. Additionally, the 'pigments +' model appears to have some limitations as illustrated by Lakes Okataina and Rototoa where the five physical lake characteristics that would not change through time, or would only change very slowly (Elevation, Lake Area, Lake Volume, Maximum Depth, and Catchment Area) appear to bias the retrodictions of the model within the same lake core. For the initial building of the model these physical lake characteristics aided in lowering the model error. However, when applied on a single lake, these physical lake characteristics appear to be biasing the model toward higher TLIs. In addition to constraining the 'pigments +' retrodicted TLIs (no extreme fluctuations observed) when they are compared to the pigments-only random forest model.

The 'pigments +' model also appears to display bias due to the calibration dataset used when creating the predictive models. While the calibration dataset contained a comprehensive number of lakes, the representation of lakes from different TLI categories was uneven, with nine lakes microtrophic, nine lakes oligotrophic, 25 lakes mesotrophic, 25 lakes eutrophic, and 27 lakes supertrophic. However, this could also reflect the proportion of lake TLIs throughout New Zealand. This could have contributed to model bias, with the model to predicting more lakes as mesotrophic to supertrophic.

Additionally, the TLI itself cannot completely account for all the catchment dynamics and influences, as there are always exceptions to the rule. For example, some lakes do not

adhere to the 'less modified catchment, the lower the lake TLI' scenario, Lake Te Kahika is a mesotrophic lake (TLI 2.7) with a highly modified catchment (<1% native flora and 90% exotic forest), whereas Lake Wainamu is a eutrophic lake (TLI 4.1) with a relatively unmodified catchment (92% native flora and 4% high producing exotic grassland). Indicating that factors other than the catchment modification are driving the lake TLI. Conversely, this could indicate that a low catchment modification of high producing grassland is more detrimental to the lake's health than a large exotic forest modification. However, it is apparent that exotic forest has a different impact on lake health than native forest. The reason for this was not investigated.

Furthermore, there is a lack of an appropriate validation dataset to compare the retrodicted TLI values, e.g., models of retrodicted lake TLIs created from proxies such as diatom abundance, nutrient concentrations (nitrogen and phosphorus), and/or fossilised DNA to reconstruct past phytoplankton communities. Contrasting these proxies against the 'pigments +' model would provide a better understanding of the accuracy of the model. Comparisons were drawn with the chl-a and cyanobacteria indicator pigments. However, while a comparison of the retrodicted TLI to common trophic status indicator pigments (chl-a and cyanobacteria indicators) aids in assessing the execution of the model, it needs to be considered that these parameters were included in the construction of the model. Therefore, it is logical that the results would be comparative.

5.2.6 Place in Ecology for this model

The analysis of phytoplankton pigments within the lake sediment core samples provides a step toward the reconstruction of both neoliminological and paleolimnological water quality. The neoliminological analysis provides a predictive tool to access lake TLI using a single surface sediment sample, rather than requiring monitoring data. Particularly, as the surface sediment of lakes captures the more recent (last couple of decades depending on the sedimentation rate of the lake) changes to the lakes TLI in a single sample. By utilising surface sediment sample as a proxy for lake TLI indications seasonal and annual variations (i.e., El Niño and La Niña events) will be averaged out into the samples and not result in strong fluctuations providing a more consistent estimation of the lake's health. This can be a valuable tool for regional councils to aid in the identification of degraded lakes. Allowing for prioritisation of rehabilitation projects. The land use model, if refined, may also provide an idea of how to balance the various land uses within a catchment to not exert too much pressure on the lake.

The retrodictive 'pigments +' model provides a paleolimnological reconstruction of the past lake TLI through the analysis of phytoplankton pigments. This reconstruction will aid

in the understanding of past lake conditions particularly if a degraded lake has historically been in this state, then rehabilitation is unlikely to be successful. On the same note, if a degraded lake has not always been in that state, rehabilitation efforts can be prioritised here.

5.2.7 Future Directions for Modelling Lake TLI Using Phytoplankton Pigments in the Lake Sediment.

There is room for refinement and expansion of the 'pigments +' predictive model. While the 'pigments +' model resulted in a relatively accurate estimate of TLI (within 10% when evaluating the surface sediment dataset), this model could be refined by investigating the feasibility of constructing a Bayesian network. A Bayesian network provides a more dynamic predictive model and can overcome some of the limitations seen in other ecological predictive models. For example, a Bayesian network can compensate for missing variables in addition to negating statistical inference when highly correlated covariates are within the dataset (Ramazi et al., 2021). Additionally, to further refine the retrodicted TLI model, suitable validation data needs to be acquired. To completely validate the accuracy of the retrodictions they need to be contrasted to other proxies for more understanding, for example, diatom analysis retrodictions of TLI. Additionally, the validation set should be based on lakes with common components, i.e., similar static lake characteristics. Therefore, they would reflect the gradients of limnological conditions that could be expected in the retrodicted TLI's (Smol, 2009). This model could be further refined by incorporating additional paleolimnological proxies into the model, if available, such as but not limited to, diatom abundance, nutrient concentrations (nitrogen and phosphorus), and/or fossilised DNA to reconstruct past phytoplankton communities.

Chapter 6 Conclusion

This research aimed to assess the potential of phytoplankton pigments preserved/captured within sediments as indicators of lake water quality in both neoliminological and paleolimnological lake sediment from New Zealand lakes. In addition to testing HSI as a method for detecting phytoplankton pigments within sediments.

When comparing sediment pigment measurements made by HSI against measurements made by the 'traditional' methods of HPLC and spectrophotometry. It was observed that the three methods of calibrating $RABD_{660\text{-}670}$ to chl-a analyses (HPLC-DAD or spectrophotometry both with and without acidification) produce statistically significant correlations (p-value <0.05). With the chl-a via spectrophotometry without acidification providing a more consistent linear regression, with a linear regression relationship comparable to the individual lake calibrations. This research showed that not only does the $RABD_{660\text{-}670}$ HSI signal from lake sediment cores provide accurate and reliable measurement of chl-a compared to those achieved through traditional analytical methods, but a universal calibration between $RABD_{660\text{-}670}$ and chl-a concentrations analysed via spectrophotometry without acidification appears feasible. However, more work with a larger dataset needs to be conducted to confirm this.

This research could find no quantifiable amounts of phycocyanin detectable through the fluorescence method in lake core sediment samples for this study, despite the lake cores showing detectable signals at $RABD_{615}$. This suggests that phycocyanin is not integrated into the sediment layer of the study lakes. However, it is plausible that the phycocyanin, if present, may have been masked by the optical activity of the chlorophyll pigments present within the sediment.

The second part of the research aimed to analyse lake surface sediments from sample set of 223 New Zealand Lakes (≈ 6% of the lakes in New Zealand) for key phytoplankton pigments to infer the water quality. Creating a random forest model using the 'pigments +' data resulted in a prediction RMSE of 0.37 (9.3%) and r2 of 0.95.

This predictive model was then used to retrodict lake trophic states from downcore samples from five New Zealand Lakes. However, due to limitations of the model and a lack of an appropriate validation dataset to compare the retrodicted TLI values could not be sufficiently assessed for accuracy. The analysis of cyanobacteria indicator pigments (canthaxanthin, echinenone, myxoxanthophyll, and zeaxanthin) suggested that the decrease of cyanobacteria indicator pigments down the older lake core sediment samples is not connected to the degradation of the pigments, rather it is linked to the

arrival of Europeans to the area. Nevertheless, by taking the lake TLIs retrodicted by the 'pigments +' model presented an interesting situation. Suggesting that while anthropogenic influence is exacerbating the degradation of the lakes since the arrival of humans to the area, before this, the trophic levels of these lakes have not drastically increased.

The predictive model of lake TLI provides a tool to access lake TLI using a single surface sediment sample, rather than requiring lake monitoring data. This can be a valuable tool for regional councils to aid in the identification of degraded lakes. Allowing for prioritisation of rehabilitation projects. The land use model, if refined, may also provide an idea of how to balance the various land uses within a catchment to not exert too much pressure on the lake. The retrodictive 'pigments +' model provides a paleolimnological reconstruction of the past lake TLI through the analysis of phytoplankton pigments aiding to better understand past lake conditions and refine rehabilitation efforts for the lakes. Particularly, if a degraded lake has historically been in this state, then rehabilitation is unlikely to be successful. On the same note, if a degraded lake has not always been in that state, rehabilitation efforts can be prioritised here.

Future directions for this research include the validation of the universal calibration of HSI signal $RABD_{660\text{-}670}$ to chl-a concentrations analysed via spectrophotometry without acidification with a larger dataset of lakes to understand whether some lake possess significantly differing calibrations. Additionally, more work needs to be undertaken to investigate the potential outliers and lakes which differ in their results, and to understand which aspects within these lakes are causing the differing results. Additionally, to further refine the retrodicted TLI model, suitable validation data needs to be acquired so they can be contrasted to other proxies (such as diatom analysis retrodictions of TLI) for a better understanding of the model. This model could be further refined by incorporating additional paleolimnological proxies into the model, if available, such as, nutrient concentrations (nitrogen and phosphorus), and/or fossilised DNA to reconstruct past phytoplankton communities.

References

Barsanti, L., & Gualtieri, P. (2014). *Algae: anatomy, biochemistry, and biotechnology*: CRC press.

Battarbee, R. W., Jones, V. J., Flower, R. J., Cameron, N. G., Bennion, H., Carvalho, L., & Juggins, S. (2002). Diatoms. In *Tracking environmental change using lake sediments* (pp. 155-202): Springer.

Bellinger, E. G., & Sigee, D. C. (2015). *Freshwater algae : identification, enumeration and use as bioindicators* (Second edition ed.): Wiley Blackwell.

Bennett, A., & Bogorad, L. (1973). Complementary chromatic adaptation in a filamentous blue-green alga. *Journal of cell biology, 58*(2), 419-435.

Bennett, K. D., & Willis, K. J. (2002). Pollen. In *Tracking environmental change using lake sediments* (pp. 5-32): Springer.

Birks, H. H. (2002). Plant macrofossils. In *Tracking environmental change using lake sediments* (pp. 49-74): Springer.

Boehmke, B. (2018a). Gradient Boosting Machines. *UC Business Analytics R Programming Guide.* Retrieved from http://uc-r.github.io/gbm_regression#gbm

Boehmke, B. (2018b). Random Forests. *UC Business Analytics R Programming Guide.* Retrieved from https://uc-r.github.io/random_forests

Boehmke, B. (2018c). Regression Trees. *UC Business Analytics R Programming Guide.* Retrieved from https://uc-r.github.io/regression_trees

Britton, G., Liaaen-Jensen, S., & Pfander, H. (1995). *Carotenoids: Vol. 1A: Isolation and Analysis*. Boston, MA: Birkhäuser Verlag.

Buchaca, T., & Catalan, J. (2007). Factors influencing the variability of pigments in the surface sediments of mountain lakes. *Freshwater Biology, 52*(7), 1365-1379.

Burns, N., Bryers, G., & Bowman, E. (2000). *Protocol for monitoring trophic levels of New Zealand lakes and reservoirs*: Lakes Consulting, Report.

Butz, C., Grosjean, M., Fischer, D., Wunderle, S., Tylmann, W., & Rein, B. (2015). *Hyperspectral imaging spectroscopy: a promising method for the biogeochemical analysis of lake sediments* (Vol. 9): SPIE.

Butz, C., Grosjean, M., Goslar, T., & Tylmann, W. (2017). Hyperspectral imaging of sedimentary bacterial pigments: a 1700-year history of meromixis from varved Lake Jaczno, northeast Poland. *Journal of Paleolimnology, 58*(1), 57-72. doi:10.1007/s10933-017-9955-1

Caird, S. (2021). *The Bio-Physical and Ecological Response of Lake Okataina, Bay of Plenty, New Zealand Following the Natural Disturbance of the 1886 Eruption of Mt Tarawera.* (Master of Science in Environmental Science). Victoria University of Wellington, Wellington. (ENSC 591)

Carlson, R. E. (1977). A trophic state index for lakes 1. *Limnology and Oceanography, 22*(2), 361-369.

Carpenter, S. R. (1981). Submersed vegetation: an internal factor in lake ecosystem succession. *The American Naturalist, 118*(3), 372-383.

Chapra, S. C., & Dobson, H. F. (1981). Quantification of the lake trophic typologies of Naumann (surface quality) and Thienemann (oxygen) with special reference to the Great Lakes. *Journal of Great Lakes Research, 7*(2), 182-193.

Chen, M., Schliep, M., Willows, R. D., Cai, Z.-L., Neilan, B. A., & Scheer, H. (2010). A Red-Shifted Chlorophyll. *Science, 329*(5997), 1318-1319. Retrieved from http://www.jstor.org.ezproxy.massey.ac.nz/stable/41075801

Cohen, A. S. (2003). *Paleolimnology: the history and evolution of lake systems*: Oxford University Press.

Conley, D. J., & Schelske, C. L. (2002). Biogenic silica. In *Tracking environmental change using lake sediments* (pp. 281-293): Springer.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research (Version 6.0-88) [R package]: InterJournal. Retrieved from https://igraph.org

Damsté, J. S., & Koopmans, M. P. (1997). The fate of carotenoids in sediments: an overview. *Pure and applied chemistry, 69*(10), 2067-2074.
Das, B., Vinebrooke, R. D., Sanchez-Azofeifa, A., Rivard, B., & Wolfe, A. P. (2005). Inferring sedimentary chlorophyll concentrations with reflectance spectroscopy: a novel approach to reconstructing historical changes in the trophic status of mountain lakes. *Canadian Journal of Fisheries and Aquatic Sciences, 62*(5), 1067-1078.
Dianto, A., Subehi, L., Ridwansyah, I., & Hantoro, W. (2019). *Clay minerals in the sediments as useful paleoclimate proxy: Lake Sentarum case study, West Kalimantan, Indonesia.* Paper presented at the IOP Conference Series: Earth and Environmental Science.
Dodson, S. I., Arnott, S. E., & Cottingham, K. L. (2000). The relationship in lake communities between primary productivity and species richness. *Ecology, 81*(10), 2662-2679.
Dos Santos, A., Calijuri, M. d. C., Moraes, E., Adorno, M., Falco, P., Carvalho, D., . . . Benassi, S. (2003). Comparison of three methods for Chlorophyll determination: Spectrophotometry and Fluorimetry in samples containing pigment mixtures and spectrophotometry in samples with separate pigments through High Performance Liquid Chromatography. *Acta Limnol. Bras, 15*(3), 7-18.
Ekpenyong, N. (2013). *Hyperspectral imaging: Calibration and applications with natural scenes.* The University of Manchester (United Kingdom),
Esri. (2016, 5/5/2016). How To: Create a watershed model using the Hydrology toolset. Retrieved from https://support.esri.com/en/technical-article/000012346
Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. A., Schofield, O., & Taylor, F. (2004). The evolution of modern eukaryotic phytoplankton. *Science, 305*(5682), 354-360.
Favot, E. J., Hadley, K. R., Paterson, A. M., Michelutti, N., Watson, S. B., Zastepa, A., . . . Smol, J. P. (2020). Using visible near-infrared reflectance spectroscopy (VNIRS) of lake sediments to estimate historical changes in cyanobacterial production: potential and challenges. *Journal of Paleolimnology, 64*(4), 335-345.
Flint, E. (1977). Phytoplankton in seven monomictic lakes near Rotorua, New Zealand. *New Zealand journal of botany, 15*(2), 197-208.
Galbraith, L. M., & Burns, C. W. (2007). Linking Land-use, Water Body Type and Water Quality in Southern New Zealand. *Landscape Ecology, 22*(2), 231-241. doi:10.1007/s10980-006-9018-x
Goetz, A. F. H., Vane, G., Solomon, J. E., & Rock, B. N. (1985). Imaging Spectrometry for Earth Remote Sensing. *Science, 228*(4704), 1147-1153. doi:10.1126/science.228.4704.1147
Gorham, E. (1960). Chlorophyll Derivatives in Surface Muds from the English Lakes. *Limnology and Oceanography, 5*(1), 29-33. doi:10.4319/lo.1960.5.1.0029
Gorham, E., Lund, J. W. G., Sanger, J. E., & Dean, W. E. (1974). Some relationships between algal standing crop, water chemistry, and sediment chemistry in the English Lakes. *Limnology and Oceanography, 19*(4), 601-617. doi:10.4319/lo.1974.19.4.0601
Greenwell, B., Boehmke, B., Cunningham, J., & GBM Developers. (2020). gbm: Generalized Boosted Regression Models. (Version 2.1.8) [R package]. Retrieved from https://CRAN.R-project.org/package=gbm
Greenwell, B. M. (2017). pdp: An R Package for Constructing Partial Dependence Plots. (Version 9(1), 421--436) [R package]: The R Journal. Retrieved from https://journal.r-project.org/archive/2017/RJ-2017-016/index.html
Grosjean, M., Amann, B. J.-F., Butz, C. F., Rein, B., & Tylmann, W. (2014). Hyperspectral imaging: a novel, non-destructive method for investigating sub-annual sediment structures and composition. *PAGES news, 22*(1), 10-11.

Guilizzoni, P., Lami, A., & Marchetto, A. (1992). Plant pigment ratios from lakes sediments as indicators of recent acidification in alpine lakes. *Limnology and Oceanography, 37*(7), 1565-1569. doi:10.4319/lo.1992.37.7.1565
Hamill, K., & Lew, D. (2006). *Snapshot of lake water quality in New Zealand*: Ministry for the Environment.
Hodgson, D. A., Wright, S. W., & Davies, N. (1997). Mass spectrometry and reverse phase HPLC techniques for the identification of degraded fossil pigments in lake sediments and their application in palaeolimnology. *Journal of Paleolimnology, 18*(4), 335-350.
Holloway, M. (2020). Auckland's Lakes. Retrieved from https://lakes380.com/rototoa-narrative/#section-7a39d8d
Jellyman, D. J. (1980). Age, growth, and reproduction of perch, Perca fluviatilis L., in Lake Pounui. *New Zealand Journal of Marine and Freshwater Research, 14*(4), 391-400.
Jeppesen, E., Søndergaard, M., Jensen, J. P., Havens, K. E., Anneville, O., Carvalho, L., . . . Foy, B. (2005). Lake responses to reduced nutrient loading–an analysis of contemporary long-term data from 35 case studies. *Freshwater Biology, 50*(10), 1747-1771.
Katsiapi, M., Moustaka-Gouni, M., & Sommer, U. (2016). Assessing ecological water quality of freshwaters: PhyCoI—a new phytoplankton community Index. *31*, 22-29. doi:10.1016/j.ecoinf.2015.11.004
Kuhn, M. (2021). caret: Classification and Regression Training (Version 6.0-88) [R package]. Retrieved from https://CRAN.R-project.org/package=caret
Kumar, H. D., & Singh, H. N. (1979). *Textbook on Algae*: Macmillan International Higher Education.
Kwon, Y. S., Pyo, J., Kwon, Y.-H., Duan, H., Cho, K. H., & Park, Y. (2020). Drone-based hyperspectral remote sensing of cyanobacteria using vertical cumulative pigment concentration in a deep reservoir. *Remote Sensing of Environment, 236*, 111517.
Lami, A., Turner, S., Musazzi, S., Gerli, S., Guilizzoni, P., Rose, N. L., . . . Yang, R. (2010). Sedimentary evidence for recent increases in production in Tibetan plateau lakes. *Hydrobiologia, 648*, 175-187. doi:10.1007/s10750-010-0263-2
Landcare Research. (2018a). NZDEM North Island 25 metre. Retrieved from https://lris.scinfo.org.nz/layer/48127-nzdem-north-island-25-metre/
Landcare Research. (2018b). NZDEM South Island 25 metre. Retrieved from https://lris.scinfo.org.nz/layer/48127-nzdem-south-island-25-metre/
Landcare Research. (2019). LCDB v5.0 - Land Cover Database version 5.0, Mainland New Zealand. Retrieved from https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/
Larkum, A. W., & Kühl, M. (2005). Chlorophyll d: the puzzle resolved. *Trends in plant science, 10*(8), 355-357.
Larned, S., Snelder, T., Whitehead, A., & Fraser, C. (2019). Water quality state and trends in New Zealand lakes: Analyses of national data ending in 2017. In: NIWA Client Report 2018359CH. Retrieved from https://www. mfe. govt. nz ….
Leavitt, P. R. (1993). A review of factors that regulate carotenoid and chlorophyll deposition and fossil pigment abundance. *Journal of Paleolimnology, 9*(2), 109-127.
Leavitt, P. R., & Hodgson, D. A. (2002). Sedimentary pigments. In J. P. Smol, J. B. Birks, & W. M. Last (Eds.), *Tracking environmental change using lake sediments* (Vol. Vol. 3, pp. 295-325): Springer.
Li, L., Sengpiel, R. E., Pascual, D. L., Tedesco, L. P., Wilson, J. S., & Soyeux, E. (2010). Using hyperspectral remote sensing to estimate chlorophyll-a and phycocyanin in a mesotrophic reservoir. *International Journal of Remote Sensing, 31*(15), 4147-4162.

Li, Y., Scales, N., Blankenship, R. E., Willows, R. D., & Chen, M. (2012). Extinction coefficient for red-shifted chlorophylls: chlorophyll d and chlorophyll f. *Biochimica et Biophysica Acta (BBA)-Bioenergetics, 1817*(8), 1292-1298.
Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. (Version 2(3), 18-22.) [R package]: R News. Retrieved from https://CRAN.R-project.org/doc/Rnews/
Lichtenthaler, H. K., & Buschmann, C. (2001). Chlorophylls and Carotenoids: Measurement and Characterization by UV-VIS Spectroscopy. *Current Protocols in Food Analytical Chemistry, 1*(1), F4.3.1-F4.3.8. doi:10.1002/0471142913.faf0403s01
Makri, S., Rey, F., Gobet, E., Gilli, A., Tinner, W., & Grosjean, M. (2020). Early human impact in a 15,000-year high-resolution hyperspectral imaging record of paleoproduction and anoxia from a varved lake in Switzerland. *Quaternary Science Reviews, 239*, 106335.
McGowan, S. (2007). Pigments in sediments of aquatic environments. *Encyclopedia of quaternary sciences. Elsevier, Amsterdam*, 2062-2074.
Milborrow, S. (2021). rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'. (Version 3.1.0) [R package]. Retrieved from https://CRAN.R-project.org/package=rpart.plot
Nürnberg, D. J., Morton, J., Santabarbara, S., Telfer, A., Joliot, P., Antonaru, L. A., . . . Boussac, A. (2018). Photochemistry beyond the red limit in chlorophyll f–containing photosystems. *Science, 360*(6394), 1210-1213.
Nygaard, G. (1949). Hydrobiological studies on some Danish ponds and lakes. *Kongl. Dansk Vid. Selskab., Biol. Skr. Kobenhavn, 7*(1), 1-293.
Paerl, H. W., & Paul, V. J. (2012). Climate change: links to global expansion of harmful cyanobacteria. *Water Research, 46*(5), 1349-1363.
Pal, R., & Choudhury, A. K. (2014). *An introduction to Phytoplanktons: diversity and ecology*: Springer.
Parrish, J. (2020). A paleolimnological investigation of agricultural intensification, water quality and ecosystem change at Lake Nganoke, southern Wairarapa, NZ.
Paul, W. J., Hamilton, D. P., Ostrovsky, I., Miller, S. D., Zhang, A., & Muraoka, K. (2012). Catchment land use and trophic state impacts on phytoplankton composition: a case study from the Rotorua lakes' district, New Zealand. In *Phytoplankton responses to human impacts at different scales* (pp. 133-146): Springer.
Peters, A., & Hothorn, T. (2021). ipred: Improved Predictors (Version 0.9-12) [R package]. Retrieved from https://CRAN.R-project.org/package=ipred
Peterson, G. S., Sierszen, M. E., Yurista, P. M., & Kelly, J. R. (2007). Stable nitrogen isotopes of plankton and benthos reflect a landscape-level influence on Great Lakes coastal ecosystems. *Journal of Great Lakes Research, 33*, 27-41.
Ptacnik, R., Lepistö, L., Willén, E., Brettum, P., Andersen, T., Rekolainen, S., . . . Carvalho, L. (2008). Quantitative responses of lake phytoplankton to eutrophication in Northern Europe. *Aquatic Ecology, 42*(2), 227-236. doi:10.1007/s10452-008-9181-z
Puddick, J., Marshall, S., & D' Souza, H. (2018). *Analysis of chlorophyll-a extracts using a 96-well plate reader with an automated liquid dispenser.* . Retrieved from
Puddick, J., Naeher, S., Pearman, J. K., Page, C. D., Romanazzi, D., Schallenberg, L. A., . . . Wood, S. A. (2021). *Using carotenoid ratios to detect shifts in freshwater cyanobacteria communities from lake sediment cores.* Journal of Paleolimnology.
Puddick, J., Nishikawa, N., & D' Souza, H. (2019). *Analysis of chlorophyll-a in periphyton samples using a 96-well plate reader and multi-channel pipette.* . Retrieved from

Ramazi, P., Kunegel-Lion, M., Greiner, R., & Lewis, M. A. (2021). Exploiting the full potential of Bayesian networks in predictive ecology. *Methods in Ecology and Evolution, 12*(1), 135-149.
Randolph, K., Wilson, J., Tedesco, L., Li, L., Pascual, D. L., & Soyeux, E. (2008). Hyperspectral remote sensing of cyanobacteria in turbid productive water using optically active pigments, chlorophyll a and phycocyanin. *Remote Sensing of Environment, 112*(11), 4009-4019. doi:https://doi.org/10.1016/j.rse.2008.06.002
Rawson, D. S. (1956). Algal Indicators of Trophic Lake Types. *Limnology and Oceanography, 1*(1), 18-25. doi:10.4319/lo.1956.1.1.0018
Rein, B., & Sirocko, F. (2002). In-situ reflectance spectroscopy–analysing techniques for high-resolution pigment logging in sediment cores. *International Journal of Earth Sciences, 91*(5), 950-954.
Reynolds, C. (1990). Temporal scales of variability in pelagic environments and the response of phytoplankton. *Freshwater Biology, 23*(1), 25-53.
Rhodes, L., Smith, K. F., MacKenzie, L., Wood, S. A., Ponikla, K., Harwood, D. T., . . . Munday, R. (2016). The Cawthron Institute Culture Collection of Micro-algae: a significant national collection. *New Zealand Journal of Marine and Freshwater Research, 50*(2), 291-316. doi:10.1080/00288330.2015.1116450
Richardson, L. L. (1996). Remote Sensing of Algal Bloom Dynamics. *BioScience, 46*(7), 492-501. doi:10.2307/1312927
Rott, E. (1984). Phytoplankton as biological parameter for the trophic characterization of lakes: With 2 figures and 2 tables in the text. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen, 22*(2), 1078-1085.
Roy, S., Llewellyn, C. A., Egeland, E. S., & Johnsen, G. (2011). *Phytoplankton pigments: characterization, chemotaxonomy and applications in oceanography*: Cambridge University Press.
Roygard, J., Kamke, J., Daly, L., Gilliland, B., Matthews, A., & Brown, L. (2019). *Lakes Management Update*. Retrieved from http://www.horizons.govt.nz/HRC/media/Media/Agenda-Reports/Strategy-Policy-Committee-2019-12-03/1929%20Lakes%20Management%20Update.pdf
RStudio Team. (2021). RStudio: Integrated Development Environment for R. (Version 1.4.1). Boston, MA: RStudio, PBC. Retrieved from http://www.rstudio.com/
Sanchini, A., & Grosjean, M. (2020). Quantification of chlorophyll a, chlorophyll b and pheopigments a in lake sediments through deconvolution of bulk UV–VIS absorption spectra. *Journal of Paleolimnology, 64*, 243-256.
Sanger, J. E. (1988). Fossil pigments in paleoecology and paleolimnology. *Palaeogeography, Palaeoclimatology, Palaeoecology, 62*(1-4), 343-359.
Sanz, N., García-Blanco, A., Gavalás-Olea, A., Loures, P., & Garrido, J. L. (2015). Phytoplankton pigment biomarkers: HPLC separation using a pentafluorophenyloctadecyl silica column. *Methods in Ecology and Evolution, 6*(10), 1199-1209. doi:10.1111/2041-210x.12406
Sartory, D., & Grobbelaar, J. (1984). Extraction of chlorophyll a from freshwater phytoplankton for spectrophotometric analysis. *Hydrobiologia, 114*(3), 177-187.
Schallenberg, M., & van der Zon, K.-A. (2019). Review of the Lake Trophic Level Index.
Schlesinger, W. H., & Bernhardt, E. S. (2020). Chapter 8 - Inland Waters. In W. H. Schlesinger & E. S. Bernhardt (Eds.), *Biogeochemistry (Fourth Edition)* (pp. 293-360): Academic Press.
Schneider, T., Rimer, D., Butz, C., & Grosjean, M. (2018). A high-resolution pigment and productivity record from the varved Ponte Tresa basin (Lake Lugano, Switzerland) since 1919: insight from an approach that combines hyperspectral imaging and high-performance liquid chromatography. *Journal of Paleolimnology, 60*(3), 381-398.
Silge, J., Chow, F., Kuhn, M., & Wickham, H. (2021). rsample: General Resampling Infrastructure (Version 0.1.0) [R package]. Retrieved from https://CRAN.R-project.org/package=rsample

Smith, K. F., & Lester, P. J. (2007). Trophic interactions promote dominance by cyanobacteria (Anabaena spp.) in the pelagic zone of Lower Karori Reservoir, Wellington, New Zealand. *New Zealand Journal of Marine and Freshwater Research, 41*(2), 143-155.
Smith, V. H., & Schindler, D. W. (2009). Eutrophication science: where do we go from here? *Trends in Ecology & Evolution, 24*(4), 201-207. doi:https://doi.org/10.1016/j.tree.2008.11.009
Smol, J. P. (2009). Paleolimnology. In G. E. Likens (Ed.), *Encyclopedia of inland waters*: Elsevier.
Smol, J. P., Birks, H., & Last, W. M. (2002). *Tracking environmental change using lake sediments: Volume 3: Terrestrial, algal, and siliceous indicators* (Vol. 3): Springer Science & Business Media.
Sorrel, P., Jacq, K., Van Exem, A., Escarguel, G., Dietre, B., Debret, M., . . . Oberhänsli, H. (2021). Evidence for centennial-scale Mid-Holocene episodes of hypolimnetic anoxia in a high-altitude lake system from central Tian Shan (Kyrgyzstan). *Quaternary Science Reviews, 252*, 106748.
Stich, H., & Brinker, A. (2004). Less is better: Uncorrected versus pheopigment-corrected photometric chlorophyll-a estimation. *Archiv fur Hydrobiologie, 162*(1), 111-120.
Stockner, J. G. (1972). Paleolimnology as a means of assessing eutrophication: With 7 figures and 3 tables in the text. *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen, 18*(2), 1018-1030.
Takaichi, S. (2011). Carotenoids in algae: distributions, biosyntheses and functions. *Marine drugs, 9*(6), 1101-1118.
Therneau, T., & Atkinson, B. (2019). rpart: Recursive Partitioning and Regression Trees (Version 4.1-15) [R package]. Retrieved from https://CRAN.R-project.org/package=rpart
Vallentyne, J. R. (1957a). Carotenoids in A 20,000-year-old sediment from Searles Lake, California. *Archives of biochemistry and biophysics, 70*(1), 29-34.
Vallentyne, J. R. (1957b). The Molecular Nature of Organic Matter in Lakes and Oceans, with Lesser Reference to Sewage and Terrestrial Soils. *J. Fish. Res., 14*(1), 33-82. doi:10.1139/f57-002
Vezhnevets, A., & Barinova, O. (2007). *Avoiding boosting overfitting by removing confusing samples.* Paper presented at the European Conference on Machine Learning.
Vincent, R. K., Qin, X., McKay, R. M. L., Miner, J., Czajkowski, K., Savino, J., & Bridgeman, T. (2004). Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. *Remote Sensing of Environment, 89*(3), 381-392. doi:https://doi.org/10.1016/j.rse.2003.10.014
Watson, S. B., McCauley, E., & Downing, J. A. (1997). Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. *Limnology and Oceanography, 42*(3), 487-495.
Watts, C., & Maxwell, J. (1977). Carotenoid diagenesis in a marine sediment. *Geochimica et Cosmochimica Acta, 41*(4), 493-497.
Whitlock, C., & Larsen, C. (2002). Charcoal as a fire proxy. In *Tracking environmental change using lake sediments* (pp. 75-97): Springer.
Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Version 3.3.5) [R package]. New York: Springer-Verlag. Retrieved from https://ggplot2.tidyverse.org
Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A Grammar of Data Manipulation (Version 1.0.7) [R package]. Retrieved from https://CRAN.R-project.org/package=dplyr
Willén, E. (2000). Phytoplankton in water quality assessment–an indicator concept. *Hydrological and limnological aspects of lake monitoring, 2*, 58-80.

Wolfe, A. P., Vinebrooke, R. D., Michelutti, N., Rivard, B., & Das, B. (2006). Experimental calibration of lake-sediment spectral reflectance to chlorophyll a concentrations: methodology and paleolimnological validation. *Journal of Paleolimnology, 36*(1), 91-100. doi:10.1007/s10933-006-0006-6

Wright, M. N., & Ziegler, A. (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. (Version 77(1), 1-17) [R package]: Journal of Statistical Software.

Wright, S. W., Jeffrey, S. W., Mantoura, R. F. C., Llewellyn, C. A., Bjørnland, T., Repeta, D., & Welschmeyer, N. (1991). Improved HPLC method for the analysis of chlorophylls and carotenoids from marine phytoplankton. *Marine Ecology Progress Series, 77*(2/3), 183. Retrieved from http://ezproxy.massey.ac.nz/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.24826571&site=eds-live&scope=site

Yackulic, E. (2017). *Productivity and Temperature Variability Over the Past 15000 Years at a Small Alpine Lake in the Southern San Juan Mountains, Colorado.* (Master of Science in Environmental Science and Policy). Northern Arizona University, Northern Arizona University.

Yacobi, Y., Köhler, J., Leunert, F., & Gitelson, A. (2015). Phycocyanin-specific absorption coefficient: Eliminating the effect of chlorophylls absorption. *Limnology and Oceanography: Methods, 13*(4), 157-168.

Zamyadi, A., McQuaid, N., Dorner, S., Bird, D. F., Burch, M., Baker, P., . . . Prévost, M. (2012). Cyanobacterial detection using in vivo fluorescence probes: Managing interferences for improved decision-making. *Journal (American Water Works Association), 104*(8), E466. Retrieved from http://ezproxy.massey.ac.nz/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.jamewatworass.104.8.E466&site=eds-live&scope=site

Zander, P. D., Żarczyński, M., Vogel, H., Tylmann, W., Wacnik, A., Sanchini, A., & Grosjean, M. (2021). A high-resolution record of Holocene primary productivity and water-column mixing from the varved sediments of Lake Żabińskie, Poland. *Science of the total environment, 755*, 143713.

Zeeb, B. A., & Smol, J. P. (2002). Chrysophyte scales and cysts. In *Tracking environmental change using lake sediments* (pp. 203-223): Springer.

Züllig, H. (1986). Carotenoids from plankton and photosynthetic bacteria in sediments as indicators of trophic changes in Lake Lobsigen during the last 14000 years. *Hydrobiologia, 143*(1), 315-319.

Appendix

Appendix A. Chemical solution preparation

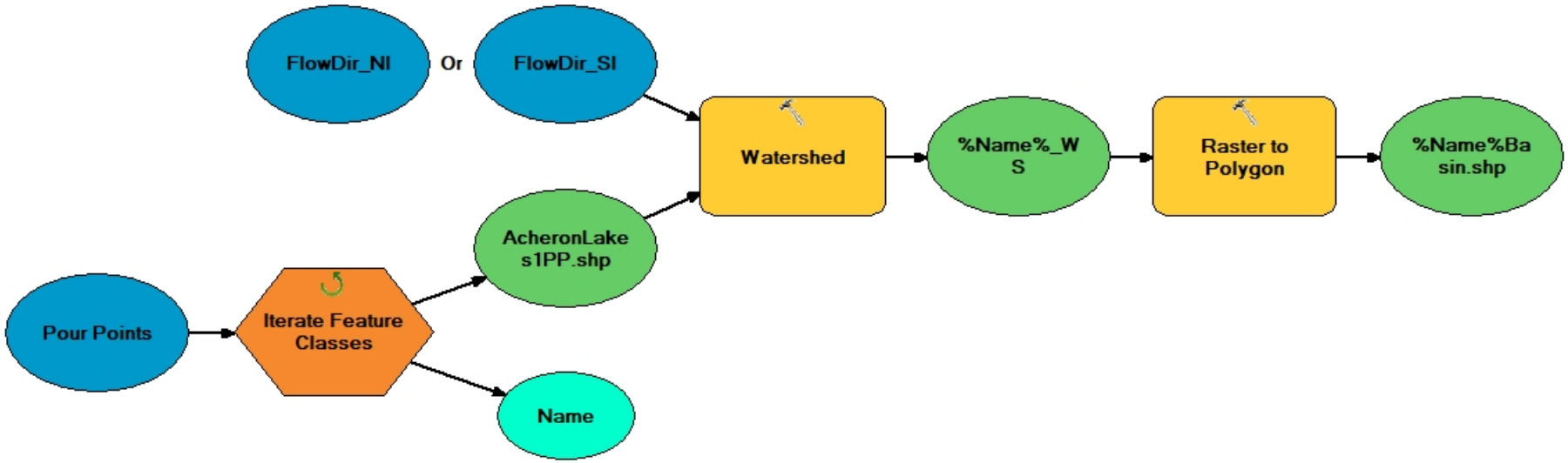
Details on how to prepare solutions used for pigment analysis.

Solution	Preparation
Phosphate Buffer (50 mM; pH 7.2)	Add 8.92 g of sodium phosphate dibasic heptahydrate ($Na_2HPO_4 \cdot 7H_2O$) and 2.31 g of sodium phosphate monobasic monohydrate ($NaH_2PO_4 \cdot H_2O$) to 800 mL of Milli-Q® ultrapure water. Correct the solution to pH 7.2 using either 1 M hydrochloric acid (HCl) to decrease pH or 1 M sodium hydroxide (NaOH) to increase. Top up the adjusted solution to 1 L with Milli-Q® ultrapure water.
90% Ethanol	In a fume hood, mix 450 mL of 100% ethanol with 50 mL of Milli-Q® ultrapure water.
1M Hydrochloric Acid (HCl)	In a fume hood, add an ampule of standardised hydrochloric acid (0.5 mol/L 1N ampoule) to partially filled 1L volumetric flask, then fill to 1 L with Milli-Q® ultrapure water.
1N Sulphuric Acid (H_2SO4)	In a fume hood, add an ampule of standardised sulfuric acid (1N H_2SO_4) to partially filled 1L volumetric flask, then fill to 1 L with Milli-Q® ultrapure water.
1% H_2SO_4	In a fume hood, add 9 mL of 1N H_2SO_4 to a 250 mL partially filled volumetric flask, then fill to 250 mL with Milli-Q® ultrapure water.
20 mg/mL $CuSO_4$	Add 2 g of $CuSO_4 \cdot 5H_2O$ to a 100 mL volumetric flask, then 50 mL of 1% H2SO4. Mix or sonicate until the $CuSO_4$ dissolves. Top up to 100 mL with 1% H_2SO_4. Compare the absorbance of the new solution (at 665 nm and 750 nm) with the previous solution using a 1 cm cuvette in the spectrophotometer which was previously blanked with 1% H_2SO_4.
methanol + 0.1% triethylamine	Mix 2 mL triethylamine to 2 L 100% methanol
40:60 methanol/ isopropyl alcohol + 0.1% triethylamine	Mix 2 mL triethylamine with 800 mL 100% methanol and 1,200 mL 100% isopropyl alcohol.

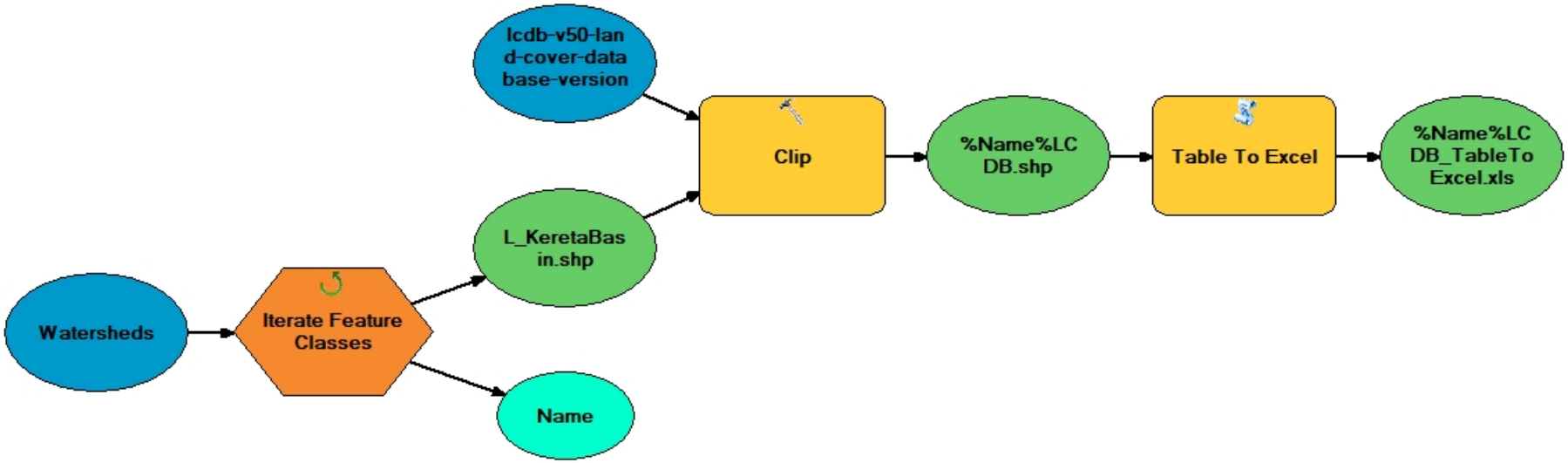
Appendix B. ArcGIS Models

ArcGIS models are workflows that create a sequence of geoprocessing tools automatically providing the output of one process into another tool as the selected input. Models are efficient ways of automating high throughput analyses such as delineating watersheds for 232 lakes and extracting data relevant data. Initial work steps were followed from Esri (2016) then adapted for the model.

Watershed Creation Model



ArcGIS model for clipping LCDB



Appendix C. Specific Characteristics of the Lakes Analysed

Characteristics of the dataset of 232 lakes utilised within this research.

Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Acheron Lakes 1	Southland	168.09508	-45.162173		Glacial	1215.02	14.84	15.10	745061.00	2,020,843.87
Acheron Lakes 2	Southland	168.099994	-45.172416		Glacial	1215.59	12.66	25.20	1062840.00	3,789,301.14
Alpine Lake	West Coast	170.137758	-43.287304		Glacial	95.36	24.07	57.20	4587404.00	1,656,579.85
Alta Tarn/Little McKay	Otago	169.189836	-44.878914		Glacial	1714.93	26.05	1.40	124305.00	247,578.40
Big Lagoon	Southland	168.176044	-46.352698		Coastal Shoreline/ Lagoon	15.34	2.90	16.50	160101.00	NA
Dumb-bell Lake	Canterbury	169.760098	-44.241908		Glacial	1557.18	21.99	14.50	1063610.00	1,124,038.47
Duncan Stream 1	Canterbury	170.043707	-44.011167		Glacial	1839.10	16.32	2.90	158932.00	525,255.41
Duncan Stream 2	Canterbury	170.052239	-44.017164		Glacial	1698.82	21.99	1.70	125984.00	503,830.25
Duncan Stream 3	Canterbury	170.059998	-44.032518		Glacial	1336.03	13.56	11.30	510992.00	7,822,438.12
Hartley Tarn	Canterbury	170.439594	-43.881374		Glacial	858.61	18.17	5.90	355542.00	582,714.33
Horseshoe	Hawkes Bay	176.760833	-39.92		Tectonic/ Landslide	175.89	23.09	31.00	2384056.00	781,690.54
Horseshoe Lake	Canterbury	172.521944	-42.597778		Riverine	456.65	17.75	4.32	255327.47	NA
Hurimoana Swamp	Hawkes Bay	176.719924	-39.554087		Swamp/ Wetland	38.09	11.91	20.20	800772.00	3,579,944.64
Irishman stream	Canterbury	169.940542	-44.052604		Glacial	1795.46	17.39	1.20	66884.00	36,366,567.92
Island Lake	Southland	167.365662	-45.779249		Glacial	715.18	31.01	41.80	4326328.00	3,399,898.88
Jane Lake	Southland	167.110974	-45.712428		Glacial	855.73	26.05	8.60	748513.00	1,453,243.29
Kaihoka Lakes 1	Tasman	172.59835	-40.553914		Coastal Shoreline/ Lagoon	52.35	15.99	5.30	281261.00	294,068.44
Kaihoka Lakes 2	Tasman	172.602592	-40.554992		Coastal Shoreline/ Lagoon	37.47	11.50	6.80	259070.00	562,771.83
Kaikereru	Hawkes Bay	177.549444	-38.755278		Volcanic	298.16	23.99	6.90	552029.00	252,184.58

Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Kaikiore	Hawkes Bay	177.541111	-38.767778		Volcanic	265.15	24.15	5.60	449195.00	3,432,258.26
Kaikura	Taranaki	174.404495	-39.70007		Coastal Shoreline/ Lagoon	58.13	5.54	5.90	109286.00	8,457,127.78
Kangaroo Lake	West Coast	171.548724	-42.613492		Glacial	96.80	50.63	125.50	21172991.00	6,173,111.08
Kaurapataka	West Coast	171.696405	-42.78562		Tectonic/ landslide	408.42	48.52	34.90	5644693.00	3,632,017.37
Kautuku Swamp	Hawkes Bay	176.732865	-39.573314		Swamp/ Wetland	39.78	11.54	10.40	398613.00	2,322,554.73
Kawau	Taranaki	175.351585	-40.262325		Riverine	71.28	13.70	1.00	45767.00	1,574,237.86
Kaweka 1	Hawkes Bay	176.3625	-39.361389		Man-made	677.25	18.05	8.60	517607.00	1,868,216.77
Kaweka 2	Hawkes Bay	176.365278	-39.363056		Man-made	677.90	18.27	5.10	313441.00	339,423.76
Killanery	Tasman	172.808308	-40.852047			11	11.8	0.6		57,851.00
L 15929	Taranaki	174.816663	-39.841709		Wind/Aeolian	59.22	15.66	6.30	328534.00	11,311,148.21
L 24423	Northland	173.320882	-34.876688		Wind/Aeolian	7.74	2.50	23.50	196089.00	2,305,954.06
L 45434	Canterbury	169.747966	-44.285174		Glacial	1516.66	14.89	5.90	292422.00	2,480,827.04
L 46774	West Coast	169.737734	-43.640618		Glacial	1193.70	26.06	10.00	867390.00	525,987.90
L 50335	Northland	174.027104	-36.29181		Swamp/ Wetland	11.29		2.50	50937.00	1,614,754.05
L 51931	Southland	167.706476	-45.581972		Riverine	237.93	12.04	2.60	104474.00	106,772.87
L 52064	Southland	167.716506	-45.471914		Glacial	275.68	24.60	2.80	229546.00	174,910.91
L 52117	Southland	167.895873	-45.453998		Swamp/ Wetland	376.96	0.50	8.70	747569.00	3,148,565.61
L 52487	Southland	167.870808	-45.346308		Riverine	354.31	12.88	2.00	85013.00	249,616.57
L 54655	Otago	168.952186	-44.533008		Glacial	1413.81	27.64	3.10	285713.00	1,214,463.58
Lady Lake	West Coast	171.573965	-42.601819		Glacial	111.09	23.00	140.80	10790978.00	15,997,825.39
Lake 34599	Hawkes Bay	176.963611	-39.647222		Coastal Shoreline/ Lagoon	42.29	13.06	6.00	261487.00	265,898.59
Lake 36343	Hawkes Bay	176.955833	-39.651111		Coastal Shoreline/ Lagoon	55.22	15.20	6.90	351252.00	1,247,155.34

Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Lake Alexandrina	Canterbury	170.451473	-43.937539	3.0	Glacial	711.09	27.00	645.70	58109414.00	47,246,367.41
Lake Bright	Southland	167.142116	-45.691707		Glacial	812.93	25.47	18.10	1533204.00	5,600,045.31
Lake Brunner / Kōtukuwhakaoka	West Coast	171.447500	-42.615833	3.4	Glacial	76.40	109.00	4060.97	1475485331.00	910,431,342.97
Lake Camp	Canterbury	171.055247	-43.614686	3.3	Glacial	675.94	13.00	43.80	1895901.00	6,102,972.61
Lake Carrot	Northland	173.187516	-35.022290	4.4	Wind/Aeolian	54.98	3.00	2.10	20633.00	166,526.91
Lake Catherine	Canterbury	171.560048	-43.221696	2.5	Glacial	667.12	18.63	18.40	1142957.00	2,724,633.10
Lake Cecil	Southland	167.384201	-45.304816		Glacial	633.06	28.45	14.00	1330511.00	11,723,303.11
Lake Chalice	Marlborough	173.31	-41.568056		Landslide	775.71	35.69	42.50	5056408.90	14,028,791.95
Lake Clearwater	Canterbury	171.046680	-43.604066	4.0	Glacial	675.00	18.00	196.60	11798938.00	32,800,697.13
Lake Denny	Canterbury	171.122007	-43.670566	5.5	Riverine	677.85	15.09	5.40	270752.00	15,958,017.21
Lake Douglas	West Coast	169.0911	-43.967272		Glacial	627.75	43.53	64.00	9283409.00	22,161,531.76
Lake Dudding	Manawatu Wanganui	175.279000	-40.100300	5.7	Wind/Aeolian	91.85	19.72	7.84	515105.85	1,155,879.52
Lake Eggeling	West Coast	169.145945	-43.985353		Glacial	980.00	27.64	10.00	918527.00	5,513,550.57
Lake Emily	Canterbury	171.227277	-43.551066	4.1	Riverine	674.31	17.31	19.30	1116175.00	2,350,205.12
Lake Emma	Canterbury	171.106261	-43.636279	4.5	Glacial	639.76	3.00	166.80	1667881.00	29,878,344.88
Lake Eva	Southland	167.337171	-45.286146		Glacial	909.79	27.41	13.40	1219933.00	2,905,513.00
Lake Evelyn	Canterbury	171.538898	-43.254488	3.2	Glacial	592.29	23.01	17.60	1352575.00	49,495,471.49
Lake Eyles	Southland	167.468845	-45.25051		Glacial	1004.46	27.46	41.90	3832533.00	2,654,754.94
Lake Forsyth (Lake Wairewa)	Canterbury	172.736957	-43.805443	6.1	Coastal Shoreline/ Lagoon	2.63	4.00	558.80	7450066.00	111,147,560.30
Lake Freeman	Southland	167.337057	-45.343426		Glacial	734.82	26.29	10.50	923470.00	11,156,205.28
Lake Gault	West Coast	169.984172	-43.430438		Glacial	333.23	42.10	31.80	4460964.00	1,415,809.57
Lake George	Southland	167.858439	-46.356639	4.5	Coastal Shoreline/ Lagoon	10.47	2.00	90.80	605401.00	43,303,580.12
Lake Georgina	Canterbury	171.568705	-43.317001	4.0	Glacial	536.66		17.50	581716.00	6,109,568.24

Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Lake Grasmere	Canterbury	171.774469	-43.063871	3.4	Riverine	589.18	15.00	62.40	3117796.00	11,675,366.35
Lake Greaney	West Coast	168.784653	-44.092708		Glacial	1183.03	21.84	15.90	1157343.00	1,558,654.17
Lake Hakanoa	Waikato	175.167500	-37.552222	6.8	Riverine	15.50	2.50	56.39	469952.14	5,746,410.34
Lake Harihari	Waikato	174.722778	-38.210833	3.9	Wind/Aeolian	16.04	8.13	18.39	498605.40	1,251,509.28
Lake Haupiri	West Coast	171.659782	-42.445279	3.5	Glacial	176.03	43.58	223.80	32514025.00	38,642,987.05
Lake Hawden	Canterbury	171.849859	-43.104203	3.5	Glacial	571.06	4.00	35.40	471496.00	3,858,596.08
Lake Hawea	Otago	169.307122	-44.456417	1.2	Glacial	283.16	384.00	15177.00	19426539520.00	1,370,386,978.14
Lake Hay	Southland	167.047436	-45.806496		Glacial	704.37	34.38	26.50	3040254.00	2,049,333.91
Lake Hayes	Otago	168.811058	-44.974953	4.7	Glacial	324.78	35.00	273.82	31945596.67	24,876,866.05
Lake Heather	Northland	173.193598	-35.050958	5.0	Wind/Aeolian	37.85	5.60	8.30	166575.00	517,889.82
Lake Henrietta	Canterbury	171.500517	-43.229210		Glacial	571.75	18.67	4.30	270664.00	8,845,922.09
Lake Henry	Southland	167.717397	-45.431004		Glacial	218.16	21.38	3.50	251240.00	185,930.20
Lake Herengawe	Taranaki	174.640591	-39.792905		Wind/Aeolian	54.90	18.60	14.30	887649.00	5,059,256.60
Lake Heron	Canterbury	171.169529	-43.483910	3.2	Riverine	692.39	37.00	694.50	85658441.00	183,575,216.00
Lake Hochstetter	West Coast	171.659782	-42.445279		Glacial	252.28	42.80	500.20	71369317.00	25,424,189.43
Lake Humuhumu	Northland	174.121361	-36.327457	3.7	Wind/Aeolian	51.05	15.00	139.60	6978376.00	8,942,973.56
Lake Ianthe	West Coast	170.622189	-43.055498		Riverine	9.48	9.11	470.80	14303850.00	16,085,303.20
Lake Ida	Canterbury	171.538249	-43.235059	3.2	Glacial	675.65	9.00	10.00	301251.00	4,093,053.41
Lake Ione	Southland	167.33763	-45.279517		Glacial	914.82	24.75	17.70	1462855.00	2,178,542.24
Lake Jaquiery	Southland	167.241653	-45.834946		Glacial	689.17	27.41	10.40	952418.00	5,190,373.16
Lake Johnson	Otago	168.732361	-45.007715	5.4	Glacial	391.20	27.00	25.47	2292303.94	1,224,359.10
Lake Kahuparere	Northland	174.158237	-36.369540	4.0	Wind/Aeolian	57.50	7.50	7.30	183458.00	17,893,638.35
Lake Kaiiwi	Northland	173.653697	-35.815094	2.9	Wind/Aeolian	78.30	16.00	26.80	1427544.00	5,316,148.16
Lake Kaniere	West Coast	171.145806	-42.829775		Tectonic/ Landslide	108.09	54.24	1473.80	266478692.00	54,703,914.05
Lake Kanono	Northland	174.143699	-36.363631	4.4	Wind/Aeolian	57.29	14.00	77.10	3597560.00	17,274,786.10
Lake Kapoai	Northland	173.833139	-36.045759		Wind/Aeolian	37.72	12.00	2.50	74695.00	2,657,533.90

Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Lake Karaka	Northland	174.039526	-36.313304	5.0	Wind/Aeolian	5.92	5.00	10.60	175839.00	8,072,388.95
Lake Karangata	Hawkes Bay	177.5425	-38.772778		Volcanic	259.94	33.68	11.04	1240073.95	760,165.33
Lake Katawich	Northland	173.164009	-34.943646		Wind/Aeolian	35.85	11.30	6.90	261419.00	1,531,294.01
Lake Kereta	Auckland	174.280097	-36.592355	5.2	Wind/Aeolian	17.07	5.00	23.60	117947.00	8,389,920.41
Lake Kohangapiripiri	Wellington	174.857033	-41.361741		Coastal Shoreline/ Lagoon	15.71	1.80	10.80	275937.00	3,774,185.95
Lake Kohangatera	Wellington	174.867192	-41.368318		Coastal Shoreline/ Lagoon	14.27	2.50	21.30	542232.00	20,565,642.79
Lake Kuwakatai	Auckland	174.237147	-36.527718	5.3	Wind/Aeolian	56.30	19.00	27.80	1760235.00	9,837,275.54
Lake Lochnager	Otago	168.603678	-44.590747		Glacial	1056.40	45.03	303.60	45576048.00	21,700,318.84
Lake Lockett	Tasman	172.624877	-41.07948		Glacial	1285.44	26.28	27.40	2398479.00	1,481,212.88
Lake Lucille	Southland	167.262574	-45.496883		Glacial	804.21	26.29	12.30	1076648.00	1,612,478.34
Lake Lyndon	Canterbury	171.693466	-43.306460	3.1	Glacial	825.96	28.00	88.40	8249652.00	15,250,255.29
Lake Mahinapua	West Coast	170.919334	-42.793725		Glacial	13.83	10.00	393.80	13125625.00	38,487,198.22
Lake Mangawhio	Taranaki	174.794396	-39.654043		Tectonic/ Landslide	119.95	17.24	9.20	529832.00	5,399,320.42
Lake Mapourika	West Coast	170.204077	-43.313652		Glacial	52.28	40.48	889.70	120051931.00	65,036,709.63
Lake Marion	Canterbury	172.231667	-42.678889		Glacial	677.43	18.63	9.98	619729.59	824,204.61
Lake Mason	Canterbury	172.168056	-42.733889		Glacial	671.89	25.46	52.60	4467988.00	8,638,633.05
Lake Mcgregor	Canterbury	170.470924	-43.936175		Glacial	709.56	25.47	37.10	3148391.00	49,339,324.14
Lake Mckay	Otago	169.212716	-44.850345		Glacial	1695.19	12.83	2.90	122358.00	910,009.58
Lake Mike	Southland	166.902314	-45.826198		Glacial	486.50	43.31	47.40	6849086.00	8,955,581.47
Lake Minerva	Southland	167.344749	-45.326148		Glacial	779.92	25.95	16.90	1462986.00	6,034,898.55
Lake Moawhitu (Durville)	Marlborough	173.810666	-40.809941				13.0			2,091,989.89
Lake Moeraki	West Coast	169.275915	-43.728283		Glacial	-19.34	12.75	241.80	10272174.00	99,088,431.54
Lake Mokeno	Northland	174.063316	-36.351529	4.6	Wind/Aeolian	8.31	5.50	159.20	2919345.00	8,066,051.64
Lake Morehurehu	Northland	172.997073	-34.642364	3.7	Wind/Aeolian	16.97	14.10	35.90	1676633.00	3,026,983.80

Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Lake Morgan	West Coast	171.707229	-42.671405		Glacial	1230.03	26.05	11.30	979894.00	683,154.59
Lake Moumahaki	Taranaki	174.665746	-39.69149		Riverine	95.30	37.32	32.90	4086180.00	20,560,850.41
Lake Ned	Otago	168.433194	-45.286903		Glacial	844.86	20.15	11.70	784467.00	35,390,649.84
Lake Ngahewa	Waikato	176.373945	-38.315366		Volcanic	412.97	23.18	8.40	650819.00	7,374,785.26
Lake Ngakapua	Northland	173.192614	-35.020496	4.1	Swamp/ Wetland	56.19	4.90	6.10	105005.00	443,435.82
Lake Ngakeketa	Northland	172.772042	-34.518375		Wind/Aeolian	38.14	8.70	11.40	329330.00	5,769,816.53
Lake Nganoke	Wellington	175.18696	-41.3555		Riverine	19.91	7.00	3.10	72263.00	1,484,231.92
Lake Ngapouri	Bay of Plenty	176.335309	-38.338764		Volcanic	477.39	19.01	21.30	1349818.00	7,471,876.08
Lake Ngatu	Northland	173.197639	-35.031680	4.0	Wind/Aeolian	37.40	6.50	51.70	1119982.00	2,652,062.12
Lake Nigel	Otago	168.427815	-45.295682		Glacial	861.35	17.67	16.60	975013.00	33,257,252.55
Lake Norma	Southland	167.273963	-45.336915		Riverine	405.70	30.97	17.60	1812463.00	18,491,302.58
Lake Ohau	Canterbury	169.852524	-44.224624	1.7	Glacial	455.14	129.00	5926.77	2548511100.00	1,145,648,090.76
Lake Okareka	Bay of Plenty	176.362101	-38.170668	3.4	Volcanic	344.39	33.50	334.10	37308073.00	19,785,560.89
Lake Okaro	Bay of Plenty	176.394741	-38.298901	5.1	Volcanic	419.72	18.00	30.10	1808142.00	4,733,799.50
Lake Okataina	Bay of Plenty	176.407460	-38.128860	2.9	Volcanic	302.04	78.50	1072.80	280713462.00	63,457,468.28
Lake Onoke	Wellington	175.131667	-41.380278		Coastal Shoreline/ Lagoon	0.18	7.70	622.38	15965758.43	3,430,288,608.03
Lake Oporoa	Manawatu Wanganui	175.867549	-39.751437	4.4	Riverine	357.55	27.06	7.10	638026.00	400,127.52
Lake Opouahi	Hawkes Bay	176.835840	-39.146529		Riverine	485.30	18.26	5.80	355443.00	600,734.83
Lake Orakai	Hawkes Bay	176.888152	-39.236089		Tectonic/ Landslide	168.35	23.30	3.30	256405.00	7,842,548.96
Lake Orbell	Southland	167.669139	-45.292665	3.5	Glacial	880.67	40.71	34.50	4676367.00	5,175,322.05
Lake Otuhie	Tasman	172.420242	-40.68533		Tectonic/ Landslide	4.75	9.10	84.70	2566636.00	15,351,177.65
Lake Oturi	Taranaki	174.62159	-39.779442		Wind/Aeolian	59.01	12.14	10.90	442281.00	3,583,431.51

Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Lake Paringa	West Coast	169.406752	-43.719989		Glacial	-1.68	18.46	474.60	29205220.00	79,668,906.98
Lake Pauri	Manawatu Wanganui	175.100556	-39.978056	5.4	Wind/Aeolian	57.58	7.85	19.25	503840.32	7,059,125.63
Lake Pearson	Canterbury	171.779191	-43.103656	3.8	Glacial	603.15	17.00	201.90	11443202.00	41,533,544.73
Lake Peel	Tasman	172.602653	-41.145754		Glacial	1353.06	26.05	4.70	408922.00	645,042.74
Lake Poerua	West Coast	171.493084	-42.706604	3.7	Riverine	95.37	7.75	212.70	5494733.00	22,566,735.46
Lake Pounui	Wellington	175.115342	-41.344251		Tectonic/ Landslide	14.08	16.25	45.95	2489671.67	667,311.24
Lake Pukaki	Canterbury	170.165528	-44.060630		Glacial	477.64	70.00	17273.59	4030504426.67	6,880,527.95
Lake Pupuke	Auckland	174.764815	-36.780636	1.5	Volcanic	14.17	57.00	103.80	19730959.00	1,358,619,973.90
Lake Pyramid (Burnt Lake)	Southland	167.364793	-45.747901	4.1	Glacial	730.74	21.84	13.30	971544.00	1,859,904.54
Lake Rakatu	Southland	167.59316	-45.624662		Glacial	174.90	35.13	130.30	15255359.00	2,324,264.82
Lake Rerewhakaaitu	Bay of Plenty	176.503369	-38.296111		Volcanic	438.74	15.80	517.00	27228195.00	106,553,230.70
Lake Richter	Southland	167.505453	-45.485206	3.8	Glacial	311.61	24.95	14.10	1174620.00	31,821,297.31
Lake Roe	Southland	167.152298	-45.70719		Glacial	973.83	18.47	19.20	1179307.00	448,391.15
Lake Rotoehu	Bay of Plenty	176.533117	-38.023584		Volcanic	296.87	13.50	790.10	35554381.00	678,724.35
Lake Rotoiti	Bay of Plenty	176.433276	-38.038078	4.9	Volcanic	270.40	126.00	3369.10	1415024352.00	75,203,981.18
Lake Rotokaha	Hawkes Bay	177.558611	-38.775556	4.1	Volcanic	255.87	25.85	9.44	813481.62	704,470,618.40
Lake Rotokare	Taranaki	174.410367	-39.45291		Volcanic	192.22	24.48	15.70	1278222.00	5,274,749.83
Lake Rotokawau (Aupouri)	Northland	173.205751	-35.018884		Swamp/ Wetland	50.54	3.00	15.20	606136.00	1,982,696.92
Lake Rotokawau (Karikari)	Northland	173.308949	-34.870514	4.6	Wind/Aeolian	7.86	2.50	66.30	552712.00	394,947.46
Lake Rotokawau (Potu/Poutu)	Northland	174.147309	-36.347070		Wind/Aeolian	59.99	11.00	25.70	942203.00	5,035,617.56
Lake Rotomanuka	Waikato	175.315278	-37.925833	3.5	Swamp/ Wetland	58.28	8.70	13.64	218225.00	1,649,720.87
Lake Rotongaio	Hawkes Bay	177.011592	-38.943619	4.9	Volcanic	394.50	22.66	9.30	704646.00	NA

Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Lake Rotonuiaha	Hawkes Bay	177.039308	-38.947486		Volcanic	278.49	26.82	43.80	3917914.00	404,752.91
Lake Rotootuauru/Swan	Northland	174.139627	-36.331229		Wind/Aeolian	36.98	5.50	17.20	614425.00	6,793,507.65
Lake Rotopiko / Lake Serpentine N	Waikato	175.318611	-37.941667	5.4	Swamp/ Wetland	53.53	11.76	4.64	181730.38	10,441,919.25
Lake Rotopokaka (Cola Lake)	Northland	173.383394	-34.949906	5.3	Wind/Aeolian	1.97	3.50	11.70	136438.00	NA
Lake Rotoroa 23681	Northland	173.196136	-35.058951		Wind/Aeolian	32.64	8.00	27.70	738252.00	3,641,625.27
Lake Rotoroa 37530	Hawkes Bay	177.029432	-38.948362	4.0	Volcanic	316.01	26.66	14.60	1293490.00	1,083,074.45
Lake Rotorua	Bay of Plenty	176.268158	-38.075540		Volcanic	270.02	44.80	8047.90	1201815791.00	4,434,723.38
Lake Rototekoiti	West Coast	169.737985	-43.640241	4.3	Glacial	1236.90	35.28	22.90	2698186.00	507,912,684.40
Lake Rototoa	Auckland	174.237926	-36.510555		Wind/Aeolian	76.31	29.00	106.60	10305165.00	576,400.21
Lake Rototuna	Northland	174.040307	-36.252092	5.2	Wind/Aeolian	118.56	5.00	8.90	147536.00	2,188,361.02
Lake Sarah	Canterbury	171.776467	-43.049118	3.2	Glacial	575.37	6.70	22.00	491431.00	22,815,254.98
Lake Selfe	Canterbury	171.518659	-43.239632	3.0	Glacial	568.66	30.00	65.40	6543730.00	6,302,596.07
Lake Story	Southland	167.172096	-45.741325		Glacial	992.84	27.65	26.50	2440490.00	1,398,447.64
Lake Sumner / Hokakura	Canterbury	172.216389	-42.700278	2.1	Glacial	504.71	134.50	1373.11	615612149.00	314,801,269.30
Lake Taharoa 21917	Northland	173.648705	-35.807090	1.8	Wind/Aeolian	78.40	37.00	204.20	25182937.00	4,395,315.38
Lake Tarawera	Bay of Plenty	176.428345	-38.195766	3.1	Volcanic	278.48	87.50	4115.40	1200331183.00	342,461,430.60
Lake Taylor	Canterbury	172.231111	-42.767500	2.6	Glacial	578.50	40.50	206.81	27919893.38	16,279,673.93
Lake Te Anau	Southland	167.750511	-45.196832	1.5	Glacial	169.44	417.00	34296.61	47672294016.00	3,110,417,779.07
Lake Te Kahika	Northland	173.001344	-34.623788	2.7	Wind/Aeolian	15.59	10.80	14.50	531127.00	4,797,463.20
Lake Tekapo	Canterbury	170.521303	-43.906406	1.4	Glacial	685.79	120.00	9659.43	3863772480.00	1,444,088,605.71
Lake Thomas	Southland	167.946409	-45.471251		Riverine	488.83	23.90	28.90	2300959.00	814,440.72
Lake Tikitapu	Bay of Plenty	176.331526	-38.195768	2.8	Volcanic	414.33	27.50	144.20	13220396.00	5,042,058.21
Lake Troup	Southland	167.131341	-45.478637		Glacial	941.46	27.64	16.20	1495448.00	803,975.20

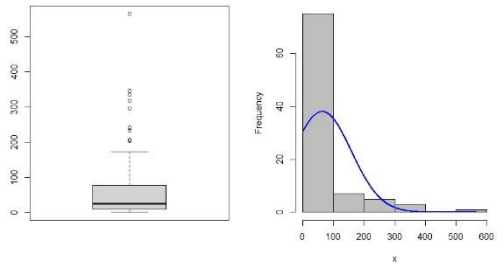
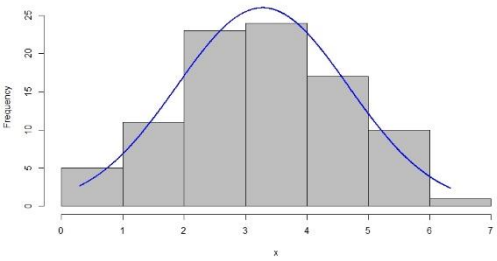
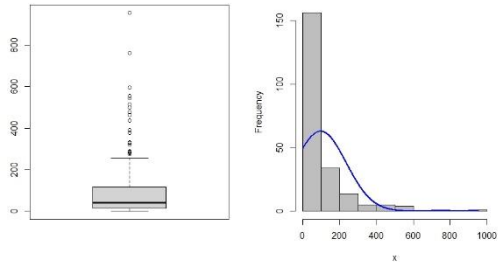
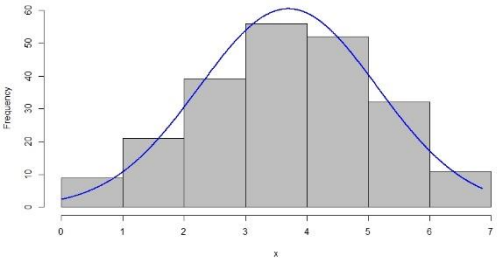
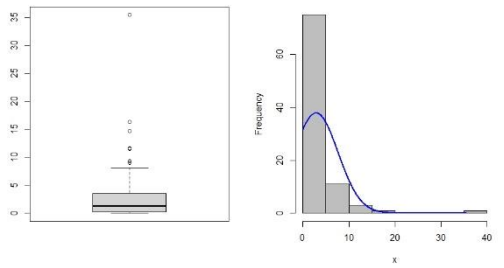
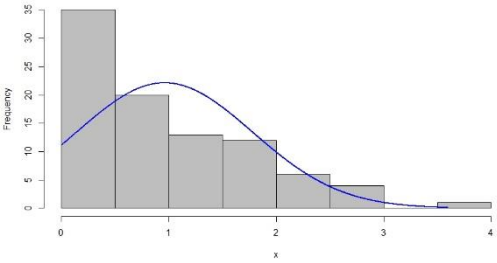
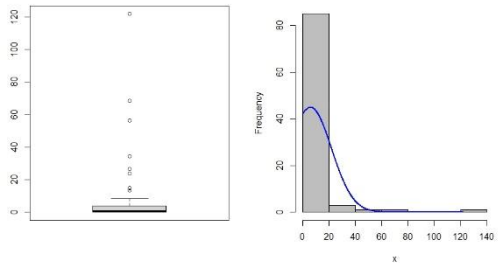
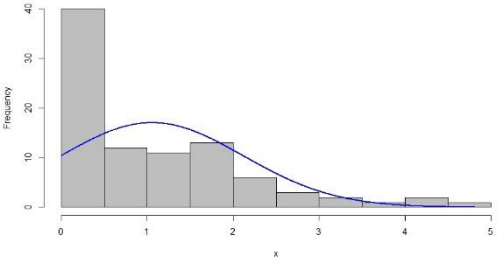
Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Lake Tuakitoto	Otago	169.824879	-46.228154	5.4	Riverine	21.03	3.00	131.74	1317384.00	144,850,512.20
Lake Tutaeinanga	Waikato	176.322376	-38.333028		Volcanic	492.45	14.59	3.10	150918.00	530,926.76
Lake Tutira	Hawkes Bay	176.892759	-39.224505	4.5	Tectonic/ Landslide	160.37	52.88	170.00	29964100.00	9,903,502.05
Lake Ursula	Southland	167.127915	-45.699764		Glacial	1017.21	11.73	2.40	93273.00	86,509.68
Lake Vincent	Southland	168.822927	-46.592383	4.3	Wind/Aeolian	19.16	5.00	17.20	286801.00	2,341,172.92
Lake Von River Valley	Southland	168.424963	-45.086926		Glacial	687.47	18.69	10.90	680091.00	NA
Lake Waahi	Waikato	175.125556	-37.559167	5.6	Riverine	13.09	5.00	444.57	7409551.67	91,854,223.12
Lake Wahakari	Northland	172.924789	-34.652031		Wind/Aeolian	53.47	12.50	84.00	3358396.00	7,792,894.14
Lake Wai Raupo	Northland	172.758029	-34.515854		Wind/Aeolian	55.96	7.00	11.20	261260.00	2,171,638.50
Lake Waihola	Otago	170.078036	-46.020350	4.8	Riverine	12.36	2.20	607.56	4455420.57	71,049,758.56
Lake Waikare	Waikato	175.200833	-37.435833		Riverine	6.76	1.80	3437.42	20624548.80	3,165,829.61
Lake Waikare	Taranaki	174.805236	-39.669440	7.0	Tectonic/ Landslide	133.96	23.76	7.40	586908.00	13,304,798,527.63
Lake Waikere	Northland	173.629502	-35.798844	2.4	Wind/Aeolian	91.04	30.00	29.70	2974459.00	1,325,543.32
Lake Waikopiro	Hawkes Bay	176.894308	-39.235158	5.3	Tectonic/ Landslide	169.21	16.00	10.00	769891.00	1,132,785.81
Lake Wainamu	Auckland	174.468258	-36.889872	4.1	Wind/Aeolian	36.36	15.00	15.00	749008.00	5,029,054.37
Lake Wainui	Northland	173.882274	-36.101340	4.0	Wind/Aeolian	40.26	11.80	3.80	151178.00	1,208,166.57
Lake Waiparera	Northland	173.181403	-34.943930	5.0	Wind/Aeolian	31.26	6.00	108.60	2171542.00	5,961,427.82
Lake Waiporohita	Northland	173.347832	-34.900907	5.1	Wind/Aeolian	14.27	3.70	6.90	68642.00	377,924.35
Lake Waitahora	Northland	172.805572	-34.455619		Wind/Aeolian	12.41	3.85	2.10	26878.00	6,181,900.31
Lake Wakatipu	Otago	168.695948	-45.031880	1.5	Glacial	237.03	380.00	29825.23	37778631253.33	3,102,950,618.14
Lake Wanaka	Otago	169.109423	-44.530971	1.8	Glacial	233.74	311.00	20399.82	21147817547.00	2,579,832,240.53
Lake Westmere	Manawatu Wanganui	174.999671	-39.895472	5.9	Wind/Aeolian	96.21	23.39	8.10	629776.00	3,676,121.74
Lake Whangape	Waikato	175.050833	-37.470556	6.6	Riverine	17.07	3.50	1078.62	12583939.67	314,528,320.62
Lake Whatumā / Lake Hatuma	Hawkes Bay	176.525278	-40.021944		Tectonic/ Landslide	132.38	3.00	152.90	14667954.00	15,280,177.14

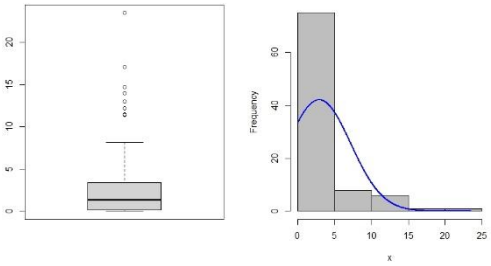
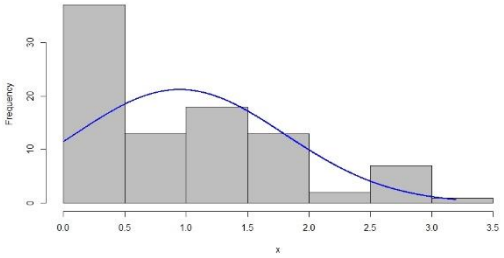
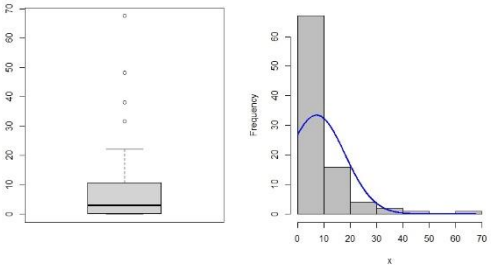
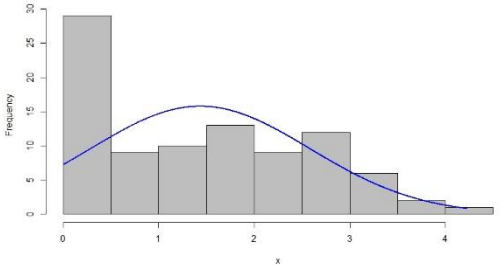
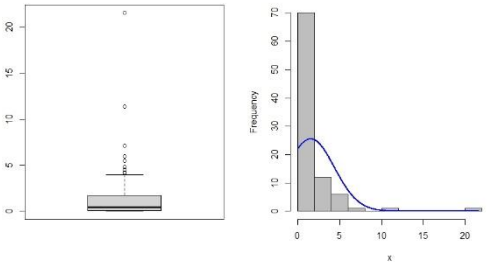
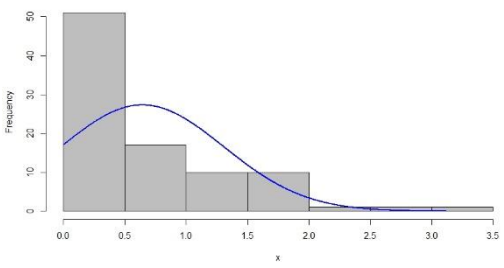
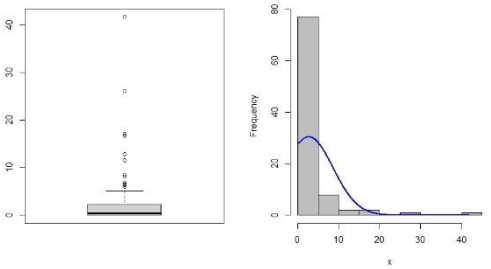
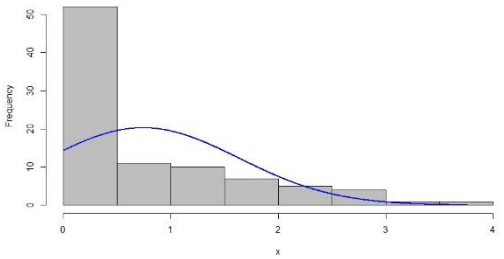
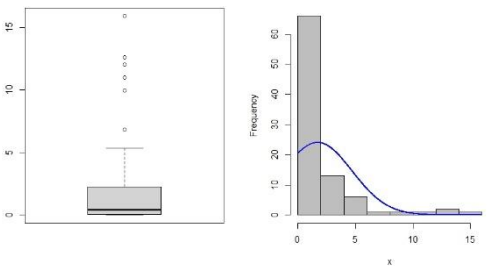
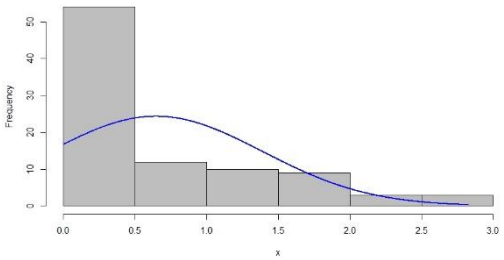
Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Lake Wiritoa	Manawatu Wanganui	175.089083	-39.974633	5.3	Wind/Aeolian	51.07	28.88	21.76	2094447.10	8,814,691.76
Loch Katrine / Waitetemoroiti	Canterbury	172.200000	-42.718333	2.7	Glacial	519.88	28.00	77.90	7270659.00	12,266,745.02
Long White Lagoon	Southland	168.15906	-46.353774		Swamp/ Wetland	15.29	15.00	1.40	13485.00	1,634,126.99
Manapouri	Southland	167.546684	-45.515162	1.5	Glacial	141.64	444.00	14177.70	20982949824.00	4,535,735,769.86
Mangarakau	Tasman	172.484985	-40.650392	3.5	Coastal Shoreline/ Lagoon	14.32	9.70	15.30	493547.00	4,352,083.25
Maori Lake E	Canterbury	171.17665	-43.571033	3.9	Riverine	626.83	17.37	9.50	549360.00	91,181,958.10
Maratoto	Waikato	175.304167	-37.886667	5.3	Swamp/ Wetland	56.08	7.00	17.80	416333.00	662,687.29
Marymere	Canterbury	171.853976	-43.116835		Glacial	615.66	12.88	24.00	1030611.00	1,599,506.62
McRae/McCrae	Canterbury	173.33	-42.183611		Glacial	883.47	41.62	65.80	9123657.00	7,311,647.14
Oingo Lake	Hawkes Bay	176.748298	-39.561721	5.1	Riverine	20.72	1.80	85.10	3915649.00	9,964,965.22
Rotorua-Kaikoura	Canterbury	173.581394	-42.404092	7.2	Riverine	27.53	3.00	43.20	432387.00	3,949,268.87
Rototai	Tasman	172.819081	-40.845441				14.0			35,287.84
Runanga Lake	Hawkes Bay	176.70521	-39.576913	7.0	Riverine	38.27	0.90	110.50	9703287.00	100,980,127.40
Shallow Lake	Southland	167.33694	-45.71276		Glacial	317.57	40.43	13.00	1757219.00	119,537,413.90
Sheppard	Canterbury	172.251111	-42.762222		Glacial	581.57	24.59	109.00	8935281.00	12,158,473.58
Skiffington Swamp	West Coast	169.996492	-43.425136		Glacial	333.52	37.28	27.60	3434889.00	3,125,009.47
South Mavora Lake	Southland	168.175172	-45.308398		Glacial	599.76	35.04	162.80	19018350.00	346,398,491.30
Spectacle Lake	Auckland	174.629987	-36.180147	6.0	Wind/Aeolian	16.00	7.00	43.80	1023022.00	6,561,352.18
Stony Tarn	Canterbury	170.445921	-43.874181		Glacial	817.20	17.27	7.10	406277.00	896,418.37
Te Ketekete	Northland	172.710728	-34.476569		Swamp/ Wetland	17.26	3.00	10.60	234840.00	27,896,655.20
Tomarata Lake	Auckland	174.650210	-36.193773	4.3	Wind/Aeolian	32.63	5.00	14.40	239265.00	1,455,718.71
Waiau	Taranaki	174.680908	-39.79448		Wind/Aeolian	56.83	16.12	24.00	1289169.00	12,227,726.11
Waihau	Hawkes Bay	177.554722	-38.755000		Volcanic	296.49	24.61	9.80	800480.00	534,177.44

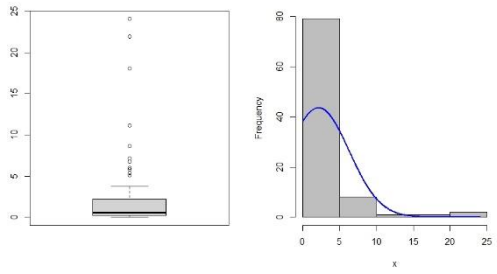
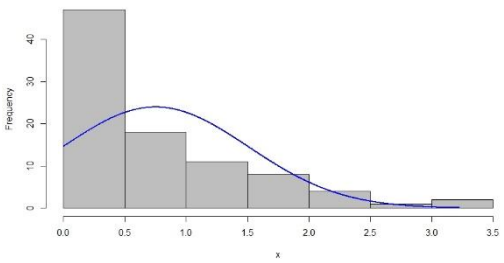
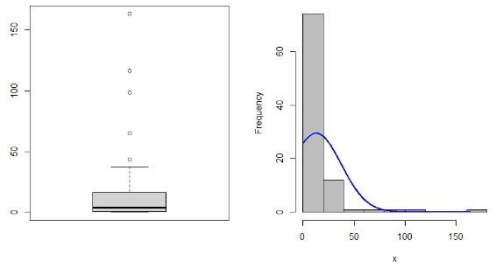
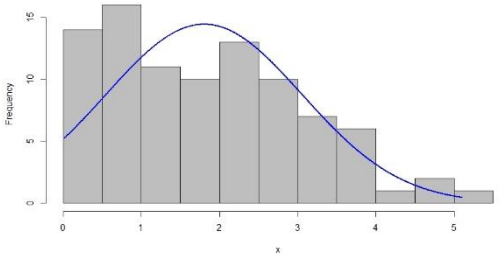
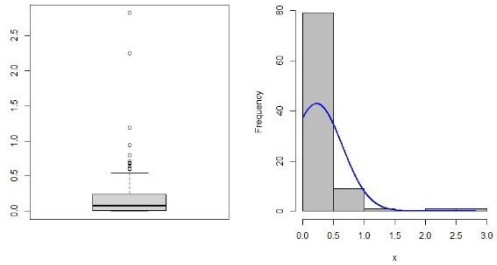
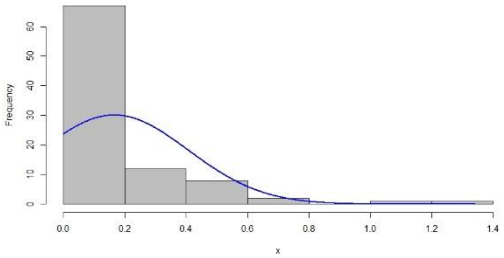
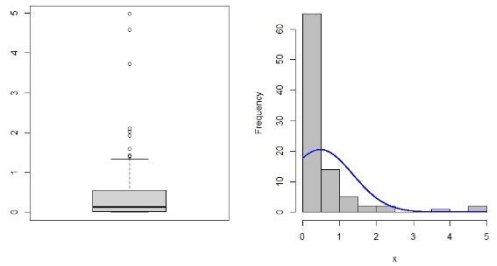
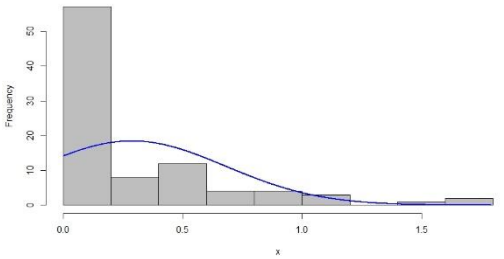
Name	Region	Longitude	Latitude	TLI	Geomorphic Class	Elevation (m)	Maximum Depth (m)	Lake Area (ha)	Lake Volume (L)	Catchment size (m^2)
Waihopo Lake	Northland	173.042826	-34.755756	4.1	Swamp/ Wetland	35.02	3.60	3.50	34765.00	1,284,110.55
Waingata	Northland	174.151118	-36.351414		Wind/Aeolian	57.28	7.00	11.00	257661.00	739,389.29
Waipara	Northland	172.854429	-34.586209	3.7	Wind/Aeolian	77.12	10.00	2.00	66115.00	485,779.38
Whakaki Lagoon	Hawkes Bay	177.554914	-39.045391		Coastal Shoreline/ Lagoon	20.48	2.75	474.90	4352717.00	13,889,979.22

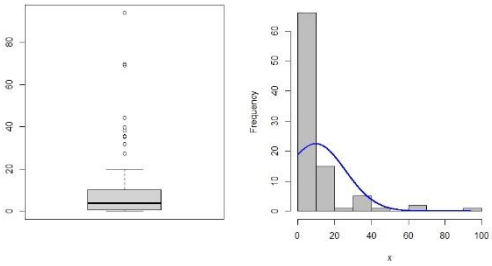
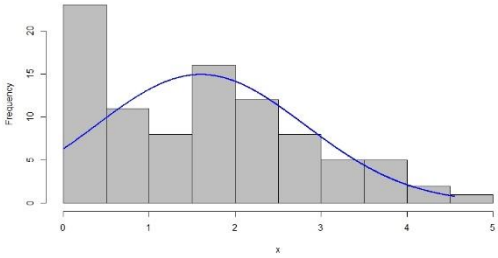
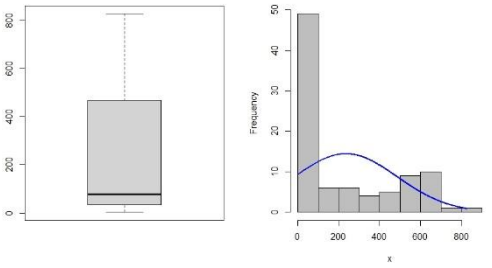
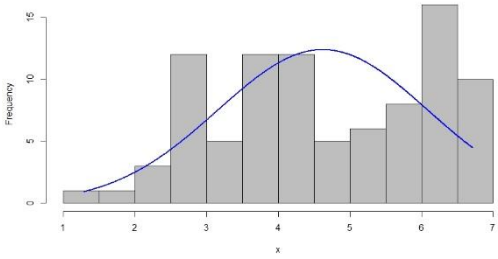
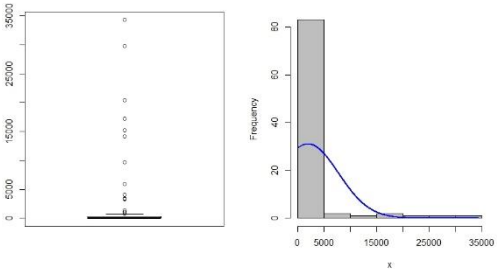
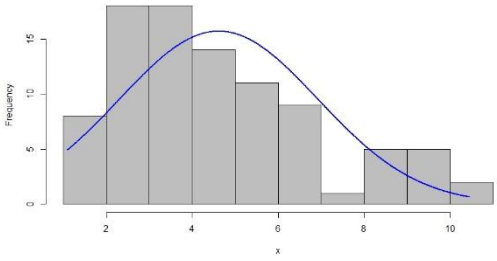
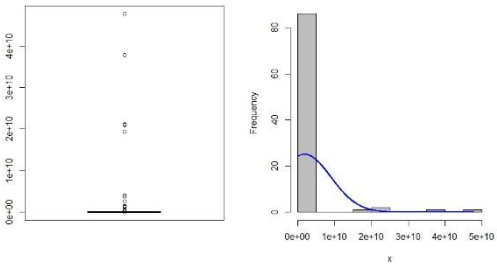
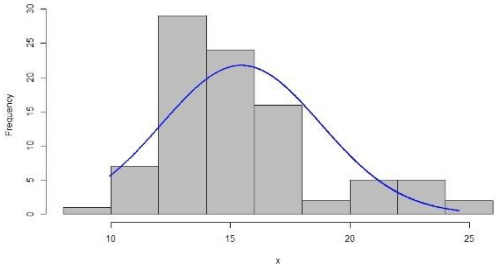
Appendix D. Distribution of Pre- and Post- Transformation of Model Variables

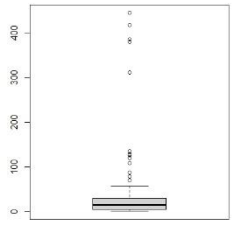
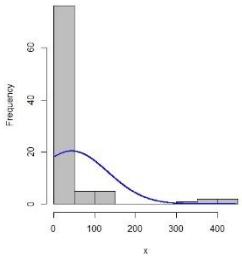
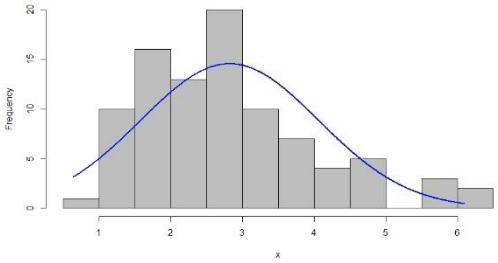
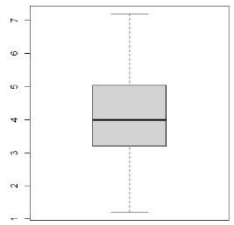
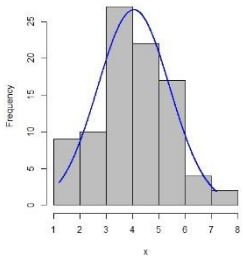
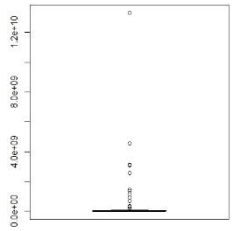
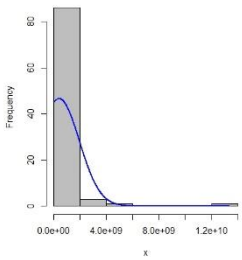
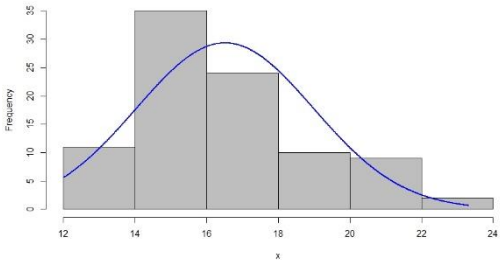
Distributions of key phytoplankton pigments and static lake characteristic used for machine learning. The land use variables were calculated as relative abundance and not transformed. Therefore, the distributions are not displayed. Distributions were normalised by transforming them using a natural logarithm of the given value plus one

Raw Data	Normalised Data
Chlorophyll-a via Spectrophotometry with Acidification	
	
Chlorophyll-a via Spectrophotometry without Acidification	
	
Alloxanthin	
	
Chlorophyll-a via HPLC	
	

Raw Data	Normalised Data
Chlorophyll-b	
	
Canthaxanthin	
	
Diadinoxanthin	
	
Diatoxanthin	
	
Echinenone	
	

Raw Data	Normalised Data
Fucoxanthin	
	
Lutein	
	
Peridinin	
	
Violaxanthin	
	

Raw Data	Normalised Data
Zeaxanthin	
	
Elevation (m)	
	
Lake Area (ha)	
	
Lake Volume (L)	
	

Raw Data		Normalised Data
Maximum Depth (m)		
		
TLI		
		Not transformed due to normal distribution.
Catchment Area (ha)		
		

Appendix E. R Code for Predictive Models

Regression Trees

```
set.seed(42)
data_split <- initial_split(data, prop = .7)
data_train <- training(data_split)
data_test  <- testing(data_split)

#Tuned Land use regression tree
optimal_tree <- rpart(
  formula = Lakes380_TLI ~ .,
  data    = data_train,
  method  = "anova",
  control = list(minsplit = 13, maxdepth = 15, cp = 0.01))

#Tuned pigments-only regression tree
optimal_tree <- rpart(
  formula = Lakes380_TLI ~ .,
  data    = data_train,
  method  = "anova",
  control = list(minsplit = 19, maxdepth = 15, cp = 0.03))

#Tuned 'pigments +' regression tree
optimal_tree <- rpart(
  formula = Lakes380_TLI ~ .,
  data    = data_train,
  method  = "anova",
  control = list(minsplit = 16, maxdepth = 14, cp = 0.02))

pred <- predict(optimal_tree, newdata = data_test)
RMSE(pred = pred, obs = data_test$Lakes380_TLI)

rpart.plot(optimal_tree)
optimal_tree

#Bagging the trees with caret
ctrl <- trainControl(method = "cv",  number = 10) #Specify 10-fold
cross validation

# CV bagged model
bagged_cv <- train(Lakes380_TLI ~ .,
  data = data_train,
  method = "treebag",
  trControl = ctrl,
  importance = TRUE)

#Predict
pred <- predict(bagged_cv, data_test)
RMSE(pred, data_test$Lakes380_TLI)
```

Random Forest Models

```
set.seed(42)
data_split <- initial_split(data, prop = .7)
data_train <- training(data_split)
data_test  <- testing(data_split)

#Tuned pigments-only random forest
rf <-randomForest(Lakes380_TLI~., data=data, mtry=11,
min.node.size = 3, importance=TRUE, ntree=500)

#Tuned 'pigments +' random forest
rf <-randomForest(Lakes380_TLI~., data=data, mtry=11,
min.node.size = 3, importance=TRUE, ntree=500)

print(rf)

#Predict
pred <- predict(rf, data_test)
RMSE(pred, data_test$Lakes380_TLI)
```

Boosted Regression Tree

```
set.seed(42)
data_split <- initial_split(data, prop = .7)
data_train <- training(data_split)
data_test  <- testing(data_split)

#Tuned boosted regerssion tree
gbm.fit.final <- gbm(
  formula = Lakes380_TLI ~ .,
  distribution = "gaussian",
  data = data_train,
  n.trees = 100,
  interaction.depth = 5,
  shrinkage = 0.1,
  n.minobsinnode = 1,
  bag.fraction = 0.3,
  train.fraction = 1,
  n.cores = NULL,
  verbose = FALSE)

print(gbm.fit.final)

#Predicting
pred <- predict(gbm.fit.final, n.trees = gbm.fit.final$n.trees,
data_test)
RMSE(pred, data_test$Lakes380_TLI)
```