Water quality state and trends in New Zealand lakes

Analyses of national lakes data ending in 2020

Prepared for Ministry for the Environment

October 2021

Prepared by:
Amy Whitehead
Caroline Fraser (LWP)
Ton Snelder (LWP)
Richard White

For any information regarding this report please contact:

Amy Whitehead
Quantitative Freshwater Ecologist
Freshwater Ecology
+64-3-343 7854
amy.whitehead@niwa.co.nz

National Institute of Water & Atmospheric Research Ltd
PO Box 8602
Riccarton
Christchurch 8011

Phone +64 3 348 8987

NIWA CLIENT REPORT No: 2021297CH
Report date: October 2021
NIWA Project: MFE21502

Revision	Description	Date
Version 1.0	Final Report	8 October 2021

Quality Assurance Statement		
	Reviewed by:	Scott Larned
	Formatting checked by:	Rachel Wright
	Approved for release by:	Phil Jellyman

Contents

Tables

Figures

Executive summary

Introduction

The New Zealand Ministry for the Environment (MfE) and Stats NZ Tatauranga Aotearoa use the results from analyses of lake water quality state and trends to inform policy development and meet their requirements for environmental reporting on the freshwater domain under the Environmental Reporting Act 2015. The data used for these analyses come from regional council state-of-the-environment (SoE) monitoring programmes. MfE have commissioned national-scale analyses of lake water quality data periodically since 2003. The current study was commissioned to analyse lake water quality state and trends for the period ending in December 2020.

The primary aim of the current study was the produce accurate estimates of recent state and temporal trends at individual lakes. The results for individual lakes have been provided to MfE as supplementary files. These lake-specific results may then be aggregated and summarised by MfE in different ways (e.g., by environmental class, region, entire nation) to meet other environmental reporting requirements and to better inform policymakers. In this study, we aggregated the results for individual lakes into lake elevation × depth classes.

The brief for this work consisted of seven major steps:

1. Compile lake water quality data from regional councils and Land Air Water Aotearoa (LAWA).
2. Organise and process the data, including error correction, application of reporting conventions and links to spatial data for each lake.
3. Assess the suitability of data for nine physical, chemical, microbial and ecological variables for statistical analyses, determine which variables have sufficient corresponding data for state and trend analyses, and apply site-inclusion rules.
4. Carry out analyses of water-quality state, including comparisons of state at lakes aggregated by elevation × depth classes.
5. Carry out trend analyses using 10-, 20- and 30-year periods ending in late 2020, including comparisons of trends in elevation × depth classes.
6. Evaluate trends for each of the water quality variables at each monitoring site for rolling windows of 10-years duration starting in 1990 and incrementing by one year (ending 31 December) to a final window ending in 2020 (i.e., a total time period of 30 years).
7. Assess water quality trends at the national scale using two approaches: categorical levels of confidence and a statistical analysis of the proportions of decreasing trends.

As an additional step, we used the water-quality state dataset to assess lakes against attribute states that are set out in the National Policy Statement for Freshwater Management 2020 (NPS-FM; New Zealand Government 2020).

Methods

Data acquisition and processing.

We used three procedures to acquire updated data for the current report: requests to Land Air Water Aotearoa (LAWA) data managers for available data from regional councils acquired for the annual LAWA refresh, interrogation of data servers operated by individual regional councils and direct requests to councils for data that were unavailable through data servers or LAWA. These data were organised into a consistent format and stored in a single RData file.

Data processing was carried out in four steps: 1) application of consistent conventions for variable names, site identifiers, date and time formats, and other data structure elements; 2) correction of errors identified using time-series plots and quantile plots; 3) exclusion of data generated using non-comparable methods; and 4) attachment of spatial information to the data for each monitoring site.

Processed data were then assessed for suitability for statistical analysis on the basis of duration and frequency of sampling. Following this assessment (and in consultation with MfE), all nine of the monitoring variables assessed were retained for analysis: Secchi depth (Seechi), concentrations of ammoniacal nitrogen (NH4N), oxidised or nitrate-nitrogen (NO3N), unfiltered total nitrogen (TN), dissolved reactive phosphorus (DRP), unfiltered total phosphorus (TP), phytoplankton biomass as chlorophyll *a* (CHLA), the bacterium *Escherichia coli* (ECOLI), and the Trophic Level Index (TLI).

State analyses.

The state dataset consisted of data for the nine variables listed above, for the 2016–2020 period, at lakes for which measurements were available in at least 80% of the years (four out of five years) and at least 80% of the seasons in the period (either 48 of 60 months, or 16 of 20 quarters). For several variables, many data were “censored”, i.e., reported as a value less than an analytical detection limit or as a value greater than a reporting limit. Censored values were replaced by imputation prior to analysis – several rules were used to make this process consistent.

For each lake × variable combination, medians were used to represent water-quality state for the lake. In addition, the state dataset was used for comparisons with attribute states that are set out in the NPS-FM. We evaluated the National Objective Framework (NOF) grades for all relevant lake attributes.

Trend analyses.

The trend analyses utilised data for the nine variables listed above for the 10-, 20- and 30-year periods ending in December 2020. Trend analyses were based on estimates of trend rate from the Sen slope estimator, and estimates of the confidence in the trend direction, derived from Kendall tests. In our previous national-scale water quality trend analyses (Larned et al. 2018), censored values in the trend datasets were replaced with imputed values, and lake × variable combinations for which more than 15% of the data consisted of censored entries were excluded. In the current study, we modified this approach to improve Sen slope and confidence interval estimation and reduce the number of excluded lakes.

The trends for all lake × variable combinations were classified into nine confidence categories on the basis of the confidence that a given trend is decreasing. The categories range from “virtually certain” (confidence 99–100%) to “exceptionally unlikely” (confidence 0-1%).

Two approaches were also used to evaluate patterns of trends at the national scale and within environmental classes. Both approaches involved aggregating multiple lakes into elevation × depth classes, and into a single spatial domain covering the entire country. The first approach was to tally the proportion of lakes in each of the nine confidence categories described above. The second approach was to estimate the proportion of decreasing trends (P_d), and the 95% confidence interval for those proportions, for all lakes in New Zealand and in each elevation × depth class.

Results

Water quality state.

Between 14 and 83 lakes met the inclusion rules for analyses of nutrient, Seechi, CHLA and TLI state. The low elevation, shallow lake class had the lowest median Seechi and the highest median CHLA, NH4N, TN, TP and TLI levels. Median Seechi was high and median nutrient, CHLA and TLI levels were generally low in the high-elevation lake classes. Median values for these variables in the low elevation, deep lake classes were intermediate. ECOLI concentrations were highest in the low-elevation, shallow lake class, and were substantially lower in the high-elevation lake classes. Grading for the NOF *E. coli* combined numeric attribute state indicated that 20% of lakes were below the national bottom line (graded D or E). These exceedances all occurred in low elevation, shallow lakes. Only one lake was below the bottom line for ammonia (median), but a greater number (23%) were below the bottom line for ammonia (maximum). A number of lakes were below the bottom line for the phytoplankton (trophic state) attribute, with 41% and 26% of sites assigned a D grade for maximum chlorophyl-a and median chlorophyl-a, respectively. Multiple lakes and 21% of sites were below the bottom line for the TN (33%) and TP (21%) attributes; all of these were low elevation lakes.

Water quality trends.

In contrast to the analyses of water-quality state, the lake elevation × depth classes did not account for much variability in trend magnitude, for any trend period. For most lake classes, only 1–7 sites were available for trend analyses, which made estimates of median trend magnitudes unreliable. For classes with more lakes, median trend magnitudes were generally less than 1% per year in the 10-year period. Although median values were low, individual lake x variable trend rates were up to 20% increasing and as low as -15% decreasing.

The 10-year P_d statistics ranged from 31–60%. For all water quality variables, the 95% confidence intervals for P_d included 50% and we could not infer widespread increasing or decreasing trends. The 20-year P_d statistics ranged from 47–70%. NH4N had a majority (i.e., P_d >50%) of decreasing 20-year trends at the 95% confidence level. For all other variables, the 95% confidence intervals for the P_d included 50%. The number of lakes that qualified for 30-year analyses of trends in each variable were small (0–15 sites), and it was unlikely that lakes used to calculate the PIT statistics were representative of all New Zealand lakes.

Discussion

The statistical power of state and trend analyses and the degree to which lakes in the analyses represented all lakes in New Zealand were limited by the small number of lakes with sufficient data. The small numbers of lakes resulted from the scarcity of lakes in council SoE monitoring networks and the exclusion of some monitored lakes due to inadequate data (although the updated procedures for handling censored values reduced the number of excluded lakes). Three general steps can be taken to alleviate problems caused by the small number of lakes in national-scale analyses: 1) alter rules about data adequacy to reduce the number of lakes excluded from analyses, as applied in this study; 2) increase the number of lakes in council monitoring networks; 3) ensure that all core water-quality variables are measured at most or all lakes in each council network.

1 Introduction

The Ministry for the Environment (MfE) and Stats NZ Tatauranga Aotearoa use analyses of lake water quality state and trends to inform policy development and meet their requirements for environmental reporting on the freshwater domain under the Environmental Reporting Act 2015. In this report, we use “lake water quality” as a general term to refer to the physical, chemical and biological variables that are included in lake state-of-environment (SoE) monitoring programmes. In a previous report for MfE, we provided water quality state and trends based on monitoring data from 155 monitored lakes; the time-series for each lake × variable combination had an ending date in December 2017 (Larned et al. 2018). In the current report, we have undertaken a new data compilation in order to report updated states and trends; the lake × variable ending dates in the current report are in December 2020.

The brief for this work consisted of seven major steps:

1. Compile lake water quality data from regional councils and Land Air Water Aotearoa (LAWA).
2. Organise and process the data, including error correction, application of reporting conventions and links to spatial data for each lake.
3. Assess the suitability of data for nine physical, chemical, microbial and ecological variables for statistical analyses, determine which variables have sufficient corresponding data for state and trend analyses, and apply site-inclusion rules.
4. Carry out analyses of water-quality state, including comparisons of state at lakes aggregated by elevation × depth classes.
5. Carry out trend analyses using 10-, 20- and 30-year periods ending in late 2020, including comparisons of trends at lakes aggregated by elevation × depth classes.
6. Carry out rolling 10-year trend analyses, including comparisons of trends at lakes aggregated by elevation × depth classes.
7. Assess water quality trends at the national scale using two approaches: categorical levels of confidence and a statistical analysis of the proportions of improving trends.

As an additional step, we used the water-quality state dataset to grade lake monitoring sites for water quality variables that are used to define numeric attribute states in the National Objectives Framework (NOF) of the National Policy Statement for Freshwater Management (NPS-FM; New Zealand Government 2020). We determined the *Escherichia coli* attribute state for individual lakes and determined the number of lakes at which the NOF bottom-lines for phytoplankton biomass, ammoniacal-nitrogen, total nitrogen and total phosphorus concentrations were exceeded.

The main components of the current report are detailed methods for data processing and analysis, summaries of water-quality state and trends at the national scale and within four contrasting land-cover classes, and supplementary files with lake-specific results and spatial data for each lake. The detailed methods and tabulated, lake-specific results will enable MfE to use the results for a wide range of purposes (e.g., mapping, inter-comparisons between environmental classes or geographic domains, estimation of reference conditions) that are all based on a single, comprehensive explanation of the methods.

The analyses in this report were aligned where possible with attributes defined by the NPS-FM. The NPS-FM requires regional councils, through their regional plans, to set freshwater objectives that provide for freshwater values, and to set limits and develop management actions to achieve those objectives. The NPS-FM identifies multiple attributes to assist regional councils in developing numeric objectives for rivers and lakes, and policies (including limits) for achieving those objectives. By expressing the current lake water quality state in terms of attribute bands, this report provides information that is relevant to management and decision-making processes.

2 Data acquisition and processing

New Zealand regional and district councils carry out SoE monitoring at approximately 150 lakes; most lakes are represented by a single monitoring site, but some lakes have 2–5 monitoring sites. For the monitoring sites used in this report, monthly or quarterly monitoring has been underway for at least five years and continues to the present. A variety of physical, chemical and biological indicators of water quality are measured at these sites.

Lake water-quality data from council monitoring programmes are periodically acquired and federated into databases for national-scale SoE reports and investigations of monitoring performance (e.g., Larned et al. 2015, 2018; Sorrell et al. 2006; Verburg et al. 2010). In the current project, the lake monitoring database used for the preceding national-scale report (Larned et al. 2018) was updated with data that had been collected between 2018 and December 2020. In this section we describe the water quality variables, data sources and organisation of the lake monitoring database, and explain the data processing procedures used to derive datasets suitable for state and trend analyses.

2.1 Water quality variables

We assessed lake water quality using nine variables that correspond to physical, chemical and biological conditions (Table 2-1). The lake water-quality variables were Secchi depth, concentrations of ammoniacal nitrogen, oxidised or nitrate-nitrogen, unfiltered total nitrogen, dissolved reactive phosphorus, unfiltered total phosphorus, and phytoplankton biomass as chlorophyll *a*, and the Trophic Level Index. Hereafter, the variables are referred to by the abbreviations listed in Table 2-1.

Table 2-1: Lake water quality variables included in this study.

Variable type	Variable	Abbreviation	Units
Physical	Secchi depth	Secchi	m
	Ammoniacal nitrogen	NH4N	mg l^{-1}
	Nitrate	NO3N	mg l^{-1}
Chemical	Total nitrogen (unfiltered)	TN	mg l^{-1}
	Dissolved reactive phosphorus	DRP	mg l^{-1}
	Total phosphorus (unfiltered)	TP	mg l^{-1}
	Chlorophyll *a*	CHLA	mg l^{-1}
Biological	*Escherichia coli*	ECOLI	cfu 100 ml^{-1}
	Trophic Level Index	TLI	unitless

Secchi depth (referred to as Seechi) is a measure of water clarity and gives an indication of the amount of light-scattering and light-absorbing particulate and dissolved matter in lakes. Seechi measures the maximum depth at which a black and white Secchi disk is visible to an observer at the lake surface.

Five different nutrient species (NO3N, NH4N, DRP, TN and TP) were included because they influence the growth of planktonic, epiphytic and benthic algae and vascular plants (macrophytes) in lakes, and because ammonia can be toxic to lake organisms at high concentrations. Nutrient enrichment can promote proliferations of planktonic algae (phytoplankton) and epiphytic algae on the surfaces of lake macrophytes. These algae can inhibit macrophyte growth by reducing light penetration. At elevated concentrations, free ammonia (NH_3) can be toxic to lake fish and invertebrates (Randall and Tsui 2002). The concentration of free ammonia and consequent risk to fish and invertebrates is determined by water temperature, pH and salinity, as well as the concentration of total ammonia ($NH_4 + NH_3$).

Chlorophyll *a* concentration (CHLA) is a measure of lake phytoplankton biomass. High chlorophyll *a* concentrations may occur during periods of high internal and/or external nutrient loading, and are the primary indicators of eutrophication. Phytoplankton chlorophyll *a* concentrations are also used to calculate Trophic Level Index scores, as described below.

The Trophic Level Index (TLI) is an indicator variable that summarises data related to lake trophic state and potential primary production. The TLI is used to classify New Zealand lakes into trophic classes (e.g., oligotrophic, eutrophic); TLI scores increase with increasing eutrophication. There are two versions of TLI in use in New Zealand, one with three variables (TLI3) and one with four variables (TLI4) (Burns et al. 2000; Verburg et al. 2010). TLI3 scores are derived from log-transformed concentrations of CHLA, TN and TP. TLI4 uses Seechi data in addition to CHLA, TN and TP concentrations. However, Seechi data were not available for all lakes in the current study. Moreover, Seechi data are strongly influenced by factors that are independent of trophic state, such as fine glacial sediment and tannins. To ensure consistent calculations, we calculated both TLI3 and TLI4 scores for all lakes in our national dataset (using the formulae given by Sorrell et al. 2006) and used these scores in lieu of TLI scores provided in council datasets.

The concentration of the bacterium *Escherichia coli* (ECOLI) is used as an indicator of human or animal faecal contamination and the risk of infectious human disease from waterborne pathogens in contact-recreation and drinking water.

NO3N and ECOLI are not core variables in all lake monitoring programmes in New Zealand, and the number of lakes for which there were sufficient NO3N and ECOLI data for statistical analysis was substantially lower than for the other variables in Table 2-1.

We used attributes for lakes that have been defined by the NPS-FM to provide context to the water quality state analyses. Five of the nine variables used in the current report are also attributes in the NPS-FM: phytoplankton (as CHLA), TN, TP, NH4N and ECOLI. The TN attribute bands distinguish between two lake classes, polymictic lakes and seasonally stratified and brackish lakes. The bands for TN and TP refer to annual medians alone. The bands for CHLA and NH4N refer to annual median and annual maximum values. The bands for ECOLI refer to median and 95th percentile values, percent of samples exceeding 260 cfu 100 ml^{-1}, and percent of samples exceeding 540 cfu 100 ml^{-1}.

2.2 Data acquisition

Lake water-quality monitoring data have been acquired periodically from regional councils for recent national scale analyses for MfE (Larned et al. 2015, 2018; Sorrell et al. 2006; Verburg et al. 2010). For each successive analysis, a database consisting of site information, sampling dates and measurements of a wide range of monitoring variables was updated. The database also contains metadata (e.g., methods, alternative variable labels, analytical detection limits). Until the current project, the data were maintained in an MS Access database; we have now shifted to storing data in an RData file.

We used three procedures to acquire updated data for the current report: requests to Land Air Water Aotearoa (LAWA) data managers for data acquired from regional councils for the annual LAWA refresh, interrogation of data servers operated by individual regional councils and NIWA (for NRWQN data) and direct requests to councils for data that were unavailable through data servers or LAWA. We used the data acquired through these three procedures to update the dataset used for the previous national-scale analysis (Larned et al. 2018). The data from each source required site-matching and verification, grid-reference conversions, and other processing to resolve inconsistencies between the two datasets, as described in the next section.

2.3 Data processing

Lake water-quality data were processed in several steps to ensure that the datasets acquired from different sources were internally consistent, that site information was complete and accurate, that consistent measurement procedures were used, and that the data were as error-free as possible.

Step 1. Reporting conventions. The water-quality data received from councils and LAWA varied in reporting formats, reporting conventions for variable names, site identifiers, date and time formats, units of measurement, and other data structure elements. We first organised data from all sources into a single format. Then we applied a consistent set of reporting conventions. Common errors included mislabelled site-names, incorrect units and data transcription errors. We applied a flagging system developed in the previous project that attaches metadata to individual data points. Flags include censored data (see Section 2.4), unit conversions, and values that were synthesised from other data (e.g., TLI scores).

Step 2. Monitoring site spatial information. The following spatial data were associated with each lake monitoring site: site name, location and regional council identifier (if available) and spatial coordinates (WGS84 latitude and longitude).

Lake monitoring sites were grouped by water surface elevation and maximum depth. Two elevation classes (0–300 m a.s.l., and > 300 m a.s.l.) and four depth classes (0–5 m, 5–15 m, 15–50 m, > 50 m) were used to define eight elevation × depth classes.

The rationale for the elevation × depth classification was: 1) elevation corresponds closely to catchment land-use and vegetation, which influence external loading to lakes; and 2) depth corresponds to lake mixing regime, which influences nutrient concentrations during summer in the surface layer, and burial of nutrients in the sediment. Lakes in the < 5 m depth class are likely to be wind-mixed frequently throughout the year, lakes in the 5–15 m depth class are likely to be mixed occasionally during summer by surface cooling, lakes in the 15–50 m and > 50 m depth classes are expected to be seasonally stratified. The same lake classification was used in the previous national-scale lake water quality report (Larned et al. 2018).

Step 3. Reducing multiple measurements into single values for each site x sampling date x variable combination. This step was used to process data from samples taken from multiple depths at each site and replicate samples taken at the same site x depth. At the sample level, we relied on water depth information in the raw dataset to inform our choice as to data suitability. For records where sample depth was explicitly specified in the raw data, we used the shallowest depth (typically 0–1 m) for which data were available. Records where sample depth was indicated only by a descriptive term clearly referring to a surface or near-surface sample (e.g., 0–25m tube, composite, epilimnion, surface, top) we accepted the data as given; samples described in other ways (including anoxic, bottom, deep, hypolimnion, middle, thermocline) were rejected. We then estimated TLI for all records with sufficient data, as described in Section 2.2, and added these to the pooled nutrient/ECOLI/Seechi data.

Step 4. Comparable field and laboratory methods. The next data processing step was to assess methodological differences for all variables. For most of the variables, two or more measurement procedures were represented in the datasets. We grouped data by procedure, then pooled data for which different procedures gave comparable results, based on assessments set out below. Data measured using the less-common and non-comparable methods were eliminated. Table 2-2 lists the most common procedures used for each variable, and the procedures corresponding to data retained for analysis.

The data produced by multiple procedures used to measure ECOLI, NO3N, NH4N, and DRP and were pooled, based on the assumption that the different procedures gave comparable results. In contrast, some procedures used to measure TN, TP are unlikely to give comparable results. Most councils use the alkaline persulfate digestion method and unfiltered water samples. A smaller group of councils uses a sulphuric acid digestion procedure to measure total Kjeldahl nitrogen (TKN) and calculates TN as TKN + NO3N. At least one council uses filtered samples for the data labelled TN and TP, although the filtered samples are more correctly labelled total dissolved nitrogen and phosphorus. The alternative methods could generate substantial differences in reported TN and TP concentrations (Horowitz 2013; Patton and Kryskalla 2003). Therefore, only TN and TP measured by the persulfate digestion method with unfiltered samples were retained for analysis. Seechi measurements made by vertically deployed Secchi discs were retained. Measurements made using water samples in horizontal clarity tubes at some shallow sites were omitted because clarity tubes due not simulate the in situ light field (Davies-Colley and Smith 2001). Laboratory CHLA measurements made by spectrophotometry were retained, and in situ measurements made using laboratory fluorometry or in situ fluorometry were omitted due to differences in the effects of other photosynthetic pigments, and for in-situ measurements, interference by dissolved organic matter (Gregor and Maršálek 2004).

Step 5. Error correction and adjustment. We manually inspected the data to correct identifiable errors (e.g., transcription errors), and to rescale data where changes in units (e.g., from mg L^{-1} to μg L^{-1}) caused scale problems. We used time-series plots and quantile plots to identify and remove gross outliers for each variable. Where necessary, values were adjusted to ensure consistent units of measurement across all datasets.

At the completion of the data processing steps, our dataset comprised 162 monitoring sites across 128 lakes, with values for some or all of the variables listed in Table 2-1.

Table 2-2: Measurement procedures for water quality variables. Procedures retained: data generated by the procedures in this column were retained for analysis in this study.

Variable	Measurement procedures	Procedures retained
ECOLI	Colilert QuantiTray 2000 Membrane filtration	Both procedures (presumed to give comparable results)
NO3N	Ion chromatography, filtered samples Cadmium reduction, filtered samples Azo dye colourimetry, filtered samples	All procedures (nitrite in cadmium-reduction and Azo-dye measurements is presumed to be negligible in unpolluted water)
NH4N	Phenol/hypochlorite colorimetry, filtered samples	Phenol/hypochlorite colorimetry, filtered samples
TN	Persulfate digestion, unfiltered samples Dissolved inorganic+organic nitrogen, filtered samples Kjeldahl digestion (TKN + NNN)	Persulfate digestion, unfiltered samples
TP	Persulfate digestion, unfiltered samples Dissolved inorganic+organic phosphorus, filtered samples	Persulfate digestion, unfiltered samples
DRP	Molybdenum blue colourimetry, unfiltered samples	Molybdenum blue colourimetry, unfiltered samples
Seechi	Secchi disk Horizontal clarity tube	Secchi disk
CHLA	Acetone pigment extraction, spectrofluorometric measurement. In situ and laboratory fluorometry	Acetone pigment extraction, spectrofluorometric measurement
TLI	Calculated from CHLA, TN, TP and Seechi	Procedures retained for CHLA, TN, TP and Seechi

2.4 Note on censored values

For several water-quality variables, some true values are too low or too high to be measured with precision. For very low values of a variable, the minimum acceptable precision corresponds to the "detection limit" for that variable; for very high values of a variable, the minimum acceptable precision corresponds to the "reporting limit" for that variable. Cases where values of variables are below the detection limit or above the reporting limit are often indicated by the data entries "<DL" and ">RL", where DL and RL are the laboratory detection limit and reporting limit, respectively. In some cases, the censored values had been replaced (by the monitoring agency) with substituted values to facilitate statistical analyses. Common substituted values are 0.5×detection limit and 1.1×reporting limit.

Water-quality datasets from New Zealand lakes often include DRP, TP and NH4N measurements that are below detection limits, and ECOLI and Seechi measurements that are above reporting limits. Although common, replacement of censored values with constant multiples of the detection and reporting limits can result in misleading results when statistical tests are subsequently applied to those data (Helsel 2012). In this study, different procedures were used to handle censored data in the state and trend analyses. The procedure used for state analyses is set out in Section 3.1.2. The procedure use for trend analyses is set out in section 3.2.

3 Analysis methods

3.1 Water quality state analyses

3.1.1 Grading of monitoring sites

Water quality state for lake monitoring sites was graded based on attributes and associated attribute state bands defined by the National Objectives Framework (NOF) of the National Policy Statement – Freshwater Management (NPS-FM; New Zealand Government 2020) (Table 3-2).

Each table of Appendix 2 of the NPS-FM (2020) represents an **attribute** that must be used to define an objective that provides for a particular environmental **value**. For example, Appendix 2A, Table 6, defines the nitrate toxicity attribute, which is defined by nitrate-nitrogen concentrations that will ensure an acceptable level of support for the “Ecosystem health (Water quality)” value. Objectives are defined by one or more **numeric attribute states** associated with each attribute. For example, for the nitrate-nitrogen attribute there are two numeric attribute states defined by the annual median and the 95^{th} percentile concentrations.

For each attribute, the NOF defines categorical attribute states in four (or five) **attribute bands**, which are designated A to D (or A to E, in the case of the *E. coli* attribute). The attribute bands represent a graduated range of support for environmental values from high (A band) to low (D or E band). The ranges for attribute states that define each attribute band are defined in Appendix 2 of the NPS-FM (2020). For most attributes, the D band represents a condition that is unacceptable (with the threshold between the C and the D band being referred to as “**bottom line”**) in any waterbody nationally. In the case of the nitrate (toxicity) and ammonia (toxicity) attributes in the 2020 NPS-FM, the C band is unacceptable, and for the DRP attribute, no bottom line is specified.

The primary aim of the attribute bands designated in the NPS-FM is to provide a basis for objective setting as part of the NOF process. The attribute bands avoid the need to discuss objectives in terms of technically complicated numeric ranges. Each band is associated with a narrative description of the outcomes for values that can be expected if that attribute band is chosen as the objective. However, it is also logical to use attribute bands to provide a grading of the current state of water quality; either as a starting point for objective setting or to track progress toward objectives.

A monitoring site can be **graded** for each attribute by assigning it to attribute bands (e.g., a site can be assigned to the A band for the nitrate toxicity attribute). The grades are referred to as ‘NOF grades’ in the results below. Site grading is done by using the numeric attribute state (e.g., annual median nitrate-nitrogen) as a **compliance statistic**. The value of the compliance statistic for a site is calculated from a record of the relevant water quality variable (e.g., the median value is calculated from the observed monthly nitrate-nitrogen concentrations). The site’s compliance statistic is then compared against the numeric ranges associated with each attribute band and a grade assigned for the site (e.g., an annual median total phosphorous concentration of 0.015 mg/l would be graded as “B-band”, because it lies in the range >0.01to ≤0.02 mg/l). Note that for attributes with more than one numeric attribute state, we have provided a grade for each numeric attribute state (e.g., for the ammonia (toxicity) attribute, grades are defined for both the median and maximum concentrations).

Table 3-1 provides a summary of the NOF attributes and numeric attribute states calculated as part of this study. In addition to the NOF attributes in Table 3-1 we also report on lake water quality state for raw (not pH adjusted) Ammoniacal Nitrogen (NH4N), Trophic level index (TLI3 and TLI4), and Nitrate (NO3N). For these variables, we report the median of the observations.

Table 3-1: **Details of the NOF attributes used to grade the state of lake monitoring sites.**

NPS-FM Reference – NOF Attribute	Calculation guidance	Numeric attribute state description	Units	Abbreviated name
A2A; Table 1 – Phytoplankton (trophic state)		Median concentration of chlorophyll-a	mg chl-a/ m^{-3}	NOF.CHLA.Med
		Maximum concentration of chlorophyll-a	mg chl-a/ m^{-3}	NOF.CHLA.Max
A2A; Table 3 – Total nitrogen (trophic state)		Median concentration of total nitrogen	mg l^{-1}	NOF.TN.Med
A2A; Table 4 – Total phosphorous (trophic state)		Median concentration of total phosphorus	mg l^{-1}	NOF.TP.Med
A2A; Table 5 – Ammonia (toxicity)	Based on temperature and pH adjusted Ammoniacal-N	Median concentration of Ammoniacal-N	mg l^{-1}	NOF.NH4N.Med
		Maximum concentration of Ammoniacal-N	mg l^{-1}	NOF.NH4N.Max
A2A; Table 9 - *Escherichia coli*	minimum of 60 samples over a maximum of 5 years,	% exceedances over 260 cfu 100 ml^{-1}	%	NOF.ECOLI.260
		% exceedances over 540 cfu 100 ml^{-1}	%	NOF.ECOLI.540
		Median concentration of *E. coli*	cfu 100 ml^{-1}	NOF.ECOLI.Med
		95th percentile concentration of *E. coli*	cfu 100 ml^{-1}	NOF.ECOLI.p95

3.1.2 Handling censored values

Censored values were replaced by imputation for the purposes of calculating the compliance statistics. Left censored values (values below the detection limit(s) were replaced with imputed values generated using ROS (Regression on Order Statistics; Helsel 2012), following the procedure described in Larned *et al.* (2015). The ROS procedure produces estimated values for the censored data that are consistent with the distribution of the uncensored values and can accommodate multiple censoring limits. When there are insufficient non-censored data to evaluate a distribution from which to estimate values for the censored observations, censored values are replaced with half of their reported value.

Censored values above the detection limit were replaced with values estimated using a procedure based on "survival analysis" (Helsel 2012). A parametric distribution is fitted to the uncensored observations and then values for the censored observations are estimated by randomly sampling values larger than the censored values from the distribution. The survival analysis requires a minimum number of observations for the distribution to be fitted; hence in the case that there were fewer than 24 observations, censored values above the detection limit were replaced with 1.1* the detection limit. The supplementary file outputs provide details about whether and how imputation was conducted for each site by attribute assessment.

3.1.3 Time period for assessments and minimum data requirements

When grading sites based on NOF attributes, it is general practice to define consistent time periods for all sites and to define the acceptable proportion of missing observations (i.e., data gaps) and how these are distributed across sample intervals so that site grades are assessed from comparable data. The time period, acceptable proportion of gaps and representation of sample intervals by observations within the time period are commonly referred to as site inclusion or filtering rules (e.g., Larned et al. 2018) but are also termed 'site screening criteria' and 'completeness criteria' (Snelder et al. 2021).

The grading assessments were based on a compliance statistic, (e.g., the median value of the observations), made for the five-year time-period to end of December 2020. The start and end dates for this period were determined by the availability of quality assured data (see Section 2), MfE reporting time periods and consideration of statistical precision of the compliance statistics used in the grading of state. The statistical precision of the compliance statistics depends on the variability in the water quality observations and the number of observations. For a given level of variability, the precision of a compliance statistic increases with the number of observations. This is particularly important for sites that are close to a threshold defined by an attribute band because the confidence that the assessment of state is 'correct' (i.e., that the site has been correctly graded) increases with the precision of the compliance statistics (and therefore with the number of observations). As a general rule, the rate of increase in the precision of compliance statistics slows for sample sizes greater than 30 (i.e., there are diminishing returns on increasing sample size with respect to precision (and therefore confidence in the assigned grade) above this number of observations; McBride 2005).

In this study, a period of five years represented a reasonable trade-off for grading assessments because it yielded a sample size of 30 or more for many sites and variable combinations). The five-year period for the state analyses is consistent with the 2013-2017 period used in the previous national water-quality state analyses (Larned et al. 2018). Because water quality data tend to fluctuate seasonally, it is also important that each season is well-represented over the period of record. In New Zealand, it is common to sample either monthly or quarterly, and in these cases, seasons are defined by months or quarters. We therefore applied a rule that restricted site × variable combinations in the state analyses to those with measurements for at least 80% of the sampling intervals in that period (at least 48 of 60 months or 16 of 20 quarters). Site × variable combinations that did not comply with these rules were excluded from the state analysis.

For grading the *E. coli* attribute the NPS-FM requires 60 observations over five years. For monthly monitoring, this requires collection of all monthly observations (i.e., no missing data). For this study, we relaxed the rule to require observations for 90% of months over the five-year period (54 observations). Both this relaxation and our default sample number are subjective choices. Therefore, within the supplementary files we provide state assessments for all sites regardless of whether they meet the filtering rules, as well as details about the number of observations and number of years with observations. This will allow MfE to apply tighter or more lenient filtering rules as required.

3.1.4 Calculation of percentiles and compliance statistics

For each lake site and variable, we characterised the current state using percentiles (5th, 20th, 25th, 50th, 75th, 80th, 95th) derived from the distribution of measured values for the period 2016 to 2020 (inclusive). All percentiles were calculated using the Hazen method.[1]

The compliance statistics specified as "Annual" (maximum, median, 95th percentile) in the NPS-FM were calculated from the data pertaining to the entire five-year state period.

When reporting state results as box and whisker graphs, the median compliance statistic across sites for each lake was used, so as to not bias the distributions towards lakes with multiple monitoring locations.

3.1.5 pH adjustment of ammonia

Ammonia is toxic to aquatic animals. When in solution, ammonia occurs in two forms: the ammonium cation (NH_4^+) and unionised ammonia (NH_3); the relative proportions of the forms are strongly dependent on pH (and temperature). Unionised ammonia is more toxic to fish than ammonium, hence the total ammonia toxicity increases with increasing pH (and/or temperature) (ANZECC & ARMCANZ 2000). The NOF attributes related to ammoniacal-N concentrations in freshwater require a correction to account for pH and temperature. Despite this requirement, the results in the current report are not temperature-corrected due to insufficient temperature data. We applied a pH correction to NH4N to adjust values to equivalent pH 8 values, following the methodology outlined in Hickey (2014). For pH values outside the range of the correction relationship (pH 6–9), the maximum (pH<6) and minimum (pH>9) correction ratios were applied. pH adjustment of ammonia was performed after imputation of censored values (Section 3.1.2). In results tables and figures adjusted ammoniacal-N is abbreviated as "NH4N (Adj.)".

3.2 Water quality trend analyses

3.2.1 Sampling dates, seasons and time periods for analyses

It is important to define the period and seasons used in trend assessment, and to determine whether the observations are adequately distributed over time, for two reasons. First, because variation in many water quality variables is associated with the time of the year or "season", the robustness of

[1] (http://www.mfe.govt.nz/publications/water/microbiological-quality-jun03/hazen-calculator.html) Note that there are many possible ways to calculate percentiles. The Hazen method produces middle-of-the-road results, whereas the method used in Excel does not (McBride 2005, chapter 8).

trend assessment is likely to be diminished if the observations are biased to certain times of the year. Second, a trend assessment will always represent a period; essentially that defined by the first and last observations. The resulting characterisation of the change in the observations over the period is likely to be diminished if the observations are not reasonably evenly distributed across the time period. For these reasons, important steps in the data compilation process include specifying the seasons, the period, and ensuring adequately distributed data.

Monitoring programmes are generally designed to sample at a fixed frequency, (e.g., monthly, quarterly). The trend analysis ‘season’ is generally specified to match this sampling frequency (e.g., seasons are months, bi-months or quarters). There is therefore generally an observation for each sample interval (i.e., each season within each year). The sampling frequency for some variables is annual. For example, annual sampling is common for biological sampling such as macro-invertebrates.

Two common deviations from the prescribed sampling regime are (1) the collection of more than one observation in a sample interval (e.g., two observations within a month) and (2) a change in sampling interval within the time period. Both of these deviations occurred in the national datasets, particularly type (2), as there were many sites with changes in sampling frequency, largely moving from lower frequency (e.g., bi-monthly or quarterly) to monthly monitoring. In our trend analyses we identified sites for which sampling intervals had changed and used the coarser sampling interval for each site to define seasons. For the part of the record with a higher frequency, the observations in each season were defined by taking the observation closest to the midpoint of the coarser season. The reason for not using the median value case is that it can induce a trend in variance, which will invalidate the null distribution of the test statistic (Helsel et al. 2020). We note that in previous national trend assessments (e.g., Larned et al. 2018) the median (rather than temporally central values) of seasons with multiple observations was used.

The trend at each site was characterised by the rate of change of the central tendency of the observations of each variable through time. Because water quality is constantly varying through time, the evaluated rate of change depends on the period over which the trend is assessed (e.g., Ballantine et al. 2010; Larned et al. 2016). Therefore, trend assessments are carried out for specified periods. In the current study, MFE requested that trends be evaluated for periods of 10, 20 and 30 years, ending in December 2020. In addition, MFE requested trends by evaluated for each of the water quality variables at each monitoring site for rolling windows of 10-years duration starting in 1990 and incrementing by one year (ending 31 December) to a final window ending in 2020 (i.e., a total time period of 30 years).

For a national study that aims to allow robust comparison of trends between sites and to provide a synoptic assessment of the whole country it is important that trends are commensurate in terms of their statistical power and representativeness of the time period. In these types of studies, it is general practice to ensure the assessed site trends are commensurate in terms of the time period by defining consistent trend durations and start dates. It is also general practice to define the acceptable proportion of gaps and how these are distributed across sample intervals so that the reported trends are assessed from data with comparable statistical power. We defined the acceptable proportion of gaps and representation of sample intervals by observations with filtering rules.

There is not a single set of agreed site filtering rules for trend assessments performed over many sites and variables such as the present study. Instead, filtering rules are generally defined for individual studies. The choice of filtering rules is based in part on the trade-off between highly restrictive rules, which increase the robustness of the individual trend analyses but generally exclude numerous sites thereby reducing spatial coverage, and highly lenient rules that retain more sites but decrease robustness. In general, this trade-off is also affected by the period duration. Steadily increasing monitoring effort in New Zealand over the last two decades means that shorter and more recent periods will generally have a larger number of eligible sites.

The application of filtering rules for variables that are measured at quarterly intervals or more frequently requires two steps. First, retain sites for which observations are available for at least *X*% of the years in the period. Second, retain sites for which observations are available for at least *Y*% of the sample intervals.

In this study, we used filtering rules applied by Larned et al. (2018), which set X and Y to 80%. Further, the definition of seasons was flexible in order to maximise the number of sites that were included. If the site failed to comply with filter rule (2) when seasons were set as months, a coarsening of the data to bi-monthly seasons was applied and the filter rule (2) was reassessed, and then repeated with seasons as quarterly if bi-monthly seasons failed to comply with filter rule (2). If the data then complied with filter rule (2), the trend results based on the coarser (i.e., bi-monthly or quarterly) seasons were retained for reporting. It is noted that this decision implies a tolerance of variable levels of statistical power and temporal representativeness across the sites that were included in the analysis. In this study, we also included bi-months as an intermediate coarseness between months and quarters, as this is a historically used sampling interval for some regional councils.

3.2.2 Handling censored values

Censored values are managed in a special way by the non-parametric trend assessment methods described in Section 3.2.2. It is therefore important that censored values are correctly identified in the data. Detection limits or reporting limits that have changed through the trend period (often due to analytical changes) can induce trends that are associated with the changing precision of the measurements rather than actual changes in the variable. This possibility needs to be accounted for in the trend analysis and this is another reason that it is important that censored values are correctly identified in the data.

We applied a “hi-censor” filter in the trend assessments to minimise biases that might be introduced due to changes in detection limits through the trend assessment period (Helsel et al. 2020). The hi-censor filter identifies the highest detection limit for each water quality variable in the trend assessment period and replaces all observations below this level with the highest detection limit and identifies these as censored values.

The water quality datasets included a small number of left censored values that were much larger than the apparent detection limit at any given time (outliers). Unsupervised application of the hi-censor filter in these circumstances can lead to the unnecessary loss of statistical power in the assessment. To avoid this problem, we employed the following approach. We expected that systematic changes in detection limit would be relatively consistent for a variable across a regional council. To explore patterns in detection level, we plotted left censored data over time by variable and regional council and used these plots to identify the occurrence of outliers. We identified a maximum realistic detection level for each variable and regional council and capped the hi-censor level at these values.

Overall, the application of the hi-censor filter generally had limited impact on the trend assessment, except for NH4N, as there was a significant shift in the detection limit, and most of the observations were generally very small (of similar magnitude to the detection limit). We note that in previous national scale assessments (e.g., Larned et al. 2018), a hi-censor filter was not applied.

3.2.3 Seasonality assessment

For many site × variable combinations, observations vary systematically by season (e.g., by month or quarter). In cases where seasons are a major source in variability, accounting for the systematic seasonal variation should increase the statistical power of the trend assessment (i.e., increase the confidence in the estimate of direction and rate of the trend). The purpose of a seasonality assessment was to identify whether seasons explain variation in the water quality variable. If this was true, then seasonal versions of the trend assessment procedures were used at the trend assessment step (Section 0).

We evaluated seasonality using the Kruskall-Wallis multi-sample test for identical populations. This is a non-parametric ANOVA that determines the extent to which season explains variation in the water quality observations. Following Hirsch et al. (1982), we identified site × variable combinations as being seasonal based on the *p*-value from the Kruskall-Wallis test with α=0.05. For these sites/variable combinations, subsequent trend assessments followed the “seasonal” variants, described in Section 0.

The choice of α is subjective and a value of 0.05 is associated with a very high level of certainty (95%) that the data exhibit a seasonal pattern. In our experience there are generally diminishing differences between the seasonal and non-seasonal trend assessments associated with the Kruskall-Wallis test for *p*-values values larger than 0.05 (Helsel et al. 2020).

3.2.4 Analysis of trends

The purpose of trend assessment is to evaluate trend direction (i.e., increasing or decreasing) and rate of the change in the central tendency of the observed water quality values over the period of analysis (i.e., the trend rate). Because the observations represent samples of the water quality over the period of analysis, there is uncertainty about the conclusions drawn from their analysis. Therefore, statistical models are used to determine the direction and rate of the trend and to evaluate the uncertainty of these determinations.

We have evaluated trends using the LWPTrends functions in the R statistical computing software. The methods are based on recently published guidance for environmental trend assessment (Snelder et al. 2021). A brief description of the theoretical basis for these functions is provided below.

Assessments of trend directions

Trend directions and the confidence in trend directions were evaluated using either the Mann Kendall assessment or the Seasonal Kendall assessment. Although the non-parametric Sen slope regression also provides information about trend direction and its confidence, the Mann Kendall assessment was used, rather than Sen slope regression, because the former more robustly handles censored values. However, Sen slope regression was used for assessing trend rates.

The Mann Kendall assessment requires no *a priori* assumptions about the distribution of the data but does require that the observations are randomly sampled and independent (no serial correlation) and that there is a sample size of ≥ 8. Both the Mann Kendall and Seasonal Kendall assessments are based on calculating the Kendall *S* statistic, which is explained diagrammatically in Figure 3-1.

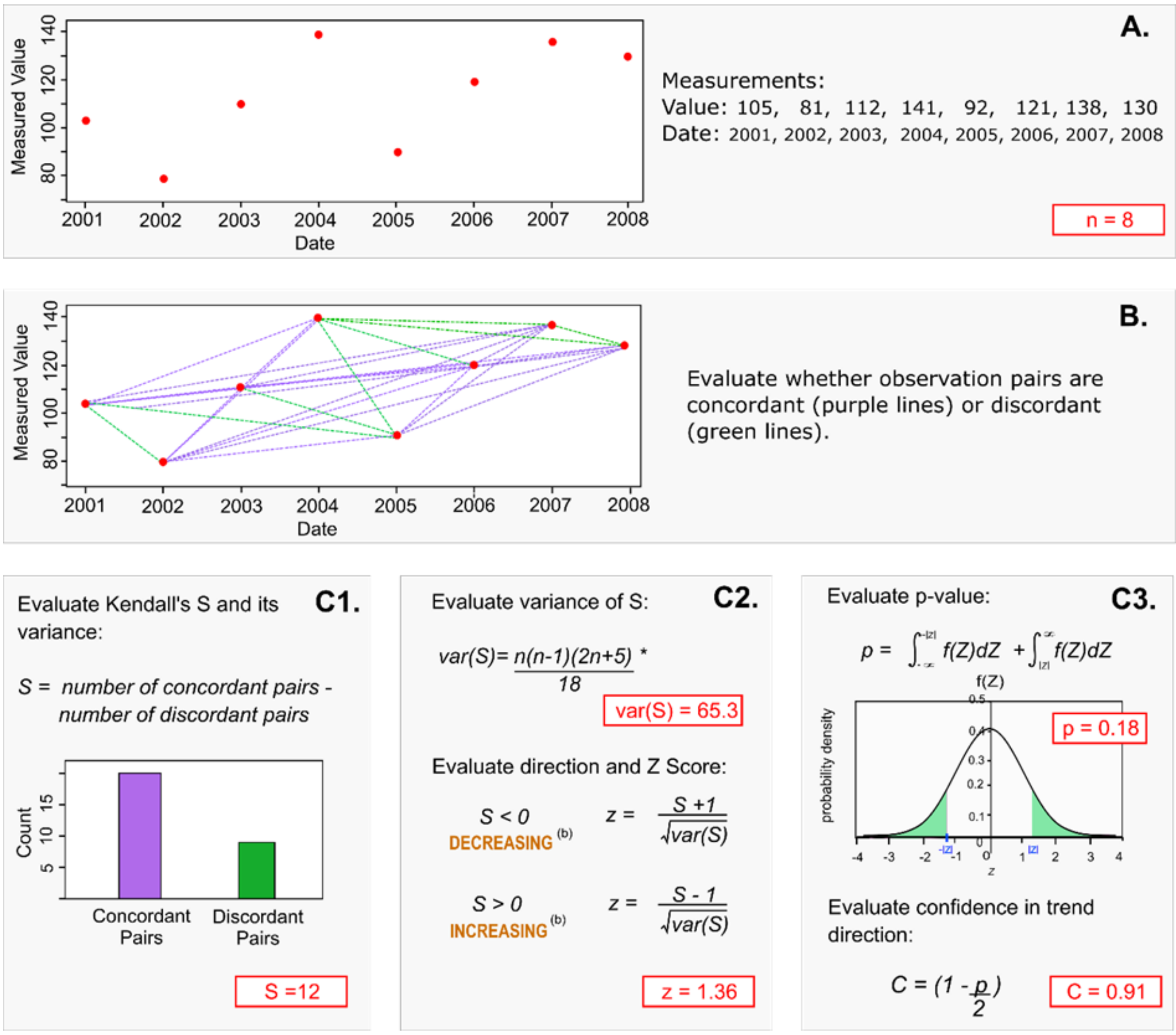

Figure 3-1: Schematic diagram demonstrating how the Kendall *S* statistic and confidence in trend direction (C) is calculated. See text for calculation details (from Snelder et al. 2021).

The Kendall *S* statistic is calculated by first evaluating the differences between all pairs of water quality observations (Figure 3-1 A and B). Positive differences are termed 'concordant' (i.e., the observations increase with increasing time) and negative differences are termed 'discordant' (i.e., the observations decrease with increasing time). The Kendall *S* statistic is the number of concordant pairs minus the number of discordant pairs (Figure 3-1, C1). The water quality trend direction is indicated by the sign of *S* with a positive or negative sign indicating an increasing or decreasing trend, respectively (Figure 3-1, C2). In the special case that the S is equal to zero, the trend is pronounced "indeterminate" (i.e., the trend direction cannot be determined).

The seasonal version of the Kendall *S* statistic *S* is calculated in two steps. First, for each season, the *S* statistic is calculated in the same manner as shown in Figure 3-1 but for data pertaining to observations in each individual season. Second, *S* is the sum of values over all seasons ($S = \sum_1^n S_i$), where S_i is the number of concordant pairs minus the number of discordant pairs in the i^{th} season and n is the number of seasons. The variance of *S* is calculated for each season and then summed over all seasons.

The sign (i.e., + or -) of the *S* statistic calculated from the sample represents the best estimate of the population trend direction but is uncertain (i.e., the direction of the population trend cannot be known with certainty). Confidence in the calculated *S* statistic in Mann's (1945) original trend test and subsequent extensions by Hirsch et al. (1982) was originally on null hypothesis significance testing (NHST). The significance of S was evaluated based on the null hypothesis of no trend (or the trend is zero). Mann (1945) showed that the *S* statistic was normally distributed, and *S* could be converted to Z-scores based on the formula shown in Panel C3 of Figure 3-1. This model describes the expected range of values of *S* if they were repeatedly calculated from many random samples, drawn from a population with no trend (i.e., the null hypothesis was true), each having the same number of observations as the actual water quality data and drawn from a population with no trend (i.e., the null hypothesis was true). The derived distribution allows the evaluation of the probability of observing a value of *S* that is as least as extreme as the observed value, if the null hypothesis was true. That probability is the *p*-value and is shown by the areas of the distribution that are cut off at the calculated value of *S* (Figure 3-1, C3). Note that for a two-tailed test, the *p*-value includes the area defined by both tails because the test is concerned with the extremity of the value and does not consider if *S* is positive or negative.

NHST produces a two-category measure of confidence in the trend assessment (i.e., significant or insignificant) based on rejection of the null hypothesis when the *p*-value is smaller than an arbitrary value known as the significance level or alpha value (α). In this study we define a continuous measure of confidence, which we call confidence in the trend direction (C). Confidence in the trend direction is calculated as:

$$C = 1 - {}^{p}/_{2}$$

where *p* is the *p*-value calculated for either Kendall S or its seasonal variant (Hirsch et al. 1982; Mann 1945).

The value C can be interpreted as the probability that the sign of the calculated value of *S* indicates the direction of the population trend (i.e., that the calculated trend direction is correct). The value C ranges between 0.5, indicating the true trend direction is equally likely to be in the opposite direction to that indicated by the sign of S, to 1, indicating complete confidence that the sign of *S* is the same as the true trend. Further discussion of the derivation of *C* and the benefits of *C* over traditional NHST significance testing is provided by Snelder et al. (2021).

As the size of the sample (i.e., the number of observations) increases, confidence in the trend direction increases. When the sample size is very large, *C* can be high, even if the trend rate is very low. It is important therefore that *C* is interpreted correctly as the confidence in direction and not as the importance of the trend. As stated at the beginning of this section; both trend direction and trend rate are relevant and important aspects of a trend assessment.

Assessments of trend rates

The method used to assess trend rate is based on non-parametric Sen slope regressions of water quality observations against time. The Sen slope estimator (SSE; Hirsch et al. 1982) is the slope parameter of a non-parametric regression of the concentration against time. SSE is calculated as the median of all possible inter-observation slopes (i.e., the difference in the measured observations divided by the time between sample dates (Figure 3-2).

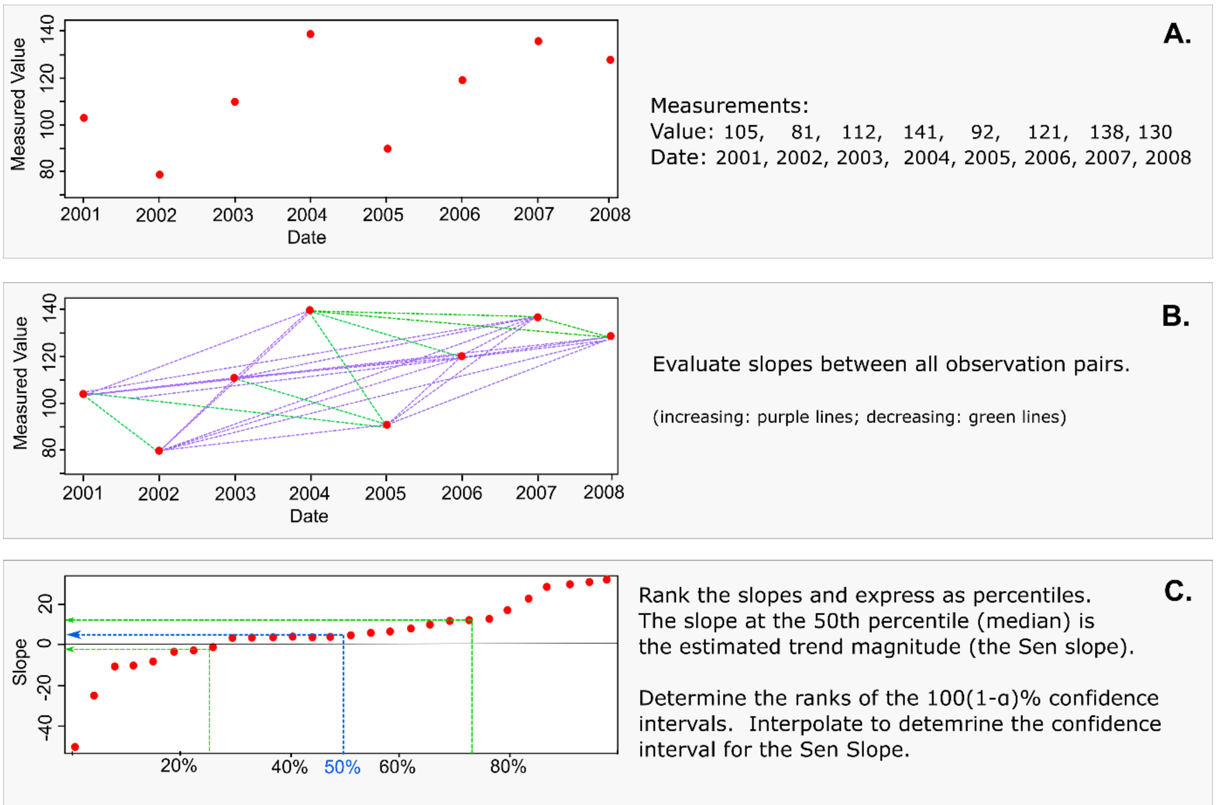

Figure 3-2: Schematic diagram of the calculation of the Sen slope, which is used to characterise trend rate (from Snelder et al. 2021).

The seasonal Sen slope estimator (SSSE) is calculated in two steps. First, for each season, the median of all possible inter-observation slopes is calculated in same manner as shown in Figure 3-2 but for data pertaining to observations in each individual season. Second, *SSSE* is the median of the seasonal values.

Uncertainty in the assessed trend rate is evaluated following a methodology outlined in Helsel et al. (2020). To calculate the 100(1-α)% two-sided symmetrical confidence interval about the fitted slope parameter, the ranks of the upper and lower confidence limits are determined, and the slopes associated with these observations are applied as the confidence intervals.

The inter-observation slope cannot be definitively calculated between any combination of observations in which either one or both observations comprise censored values. Therefore, it is usual to remove the censor sign from the reported laboratory value and use just the 'raw' numeric component (i.e., <1 becomes 1) multiplied by a factor (such as 0.5 for left-censored and 1.1 for right-censored values). This ensures that in the Sen slope calculations, any left-censored observations are always treated as values that are less than their 'raw' values and right censored observations are always treated as values that are greater than their 'raw' values. The inter-observation slopes associated with the censored values are therefore imprecise (because they are calculated from the replacements). However, because the Sen slope is the median of all the inter-observation slopes, the Sen slope is unlikely to be affected by censoring when a small proportion of observations are censored. As the proportion of censored values increase, the probability that the Sen slope is affected by censoring increases. The outputs from the trend assessment provide an 'analysis note' to identify Sen Slopes where one or both of the observations associated with the median inter-observation slope is censored.

The relative Sen Slope estimator (RSSE) is the Sen Slope divided by the median value from the observation data and expresses the trend rate as a percentage change per year.

3.2.5 Interpretation of trends

The trend assessment procedures used here allow a more nuanced inference than the categorical measure of confidence associated with NHST (i.e., significant or not significant). The confidence in direction (C) can be transformed into a continuous scale of confidence the trend was decreasing (C_d). For all trends with $S < 0$, $C_d = C$, and for all $S > 0$ a transformation is applied so that $C_d = 1\text{-}C$. C_d ranges from 0 to 1.0. When C_d is very small, a decreasing trend is highly unlikely, which because the outcomes are binary, is the same as an increasing trend is highly likely.

The approach to presenting levels of confidence of the Intergovernmental Panel on Climate Change (IPCC; Stocker et al. 2014) is one way of conveying the confidence of trend directions (Table 3-2). These same categorical levels of confidence were used to express the confidence that water quality was decreasing for each site and variable in this report. The trend for each site × variable combination was assigned a categorical level of confidence that the trend was decreasing according to its evaluated confidence, direction and the categories shown in Table 3-2.

Table 3-2: Level of confidence categories used to convey the confidence that the trend direction was decreasing. The confidence categories are used by the Intergovernmental Panel on Climate Change (IPCC; Stocker et al. 2014).

Categorical level of confidence trend was decreasing	Value of C_d (%)
Virtually certain	0.99–1.00
Extremely likely	0.95–0.99
Very likely	0.90–0.95
Likely	0.67–0.90
About as likely as not	0.33–0.67
Unlikely	0.10–0.33
Very unlikely	0.05–0.10
Extremely unlikely	0.01–0.05
Exceptionally unlikely	0.00–0.01

Some trends were classified as “not analysed” for either of two reasons:

1. When a large proportion of the values were censored (data has <5 non-censored values and/or <3 unique non-censored values). This arises because trend analysis is based on examining differences in the value of the variable under consideration between all pairs of sample occasions. When a value is censored, it cannot be compared with any other value and the comparison is treated as a “tie” (i.e., there is no change in the variable between the two sample occasions). When there are many ties there is little information content in the data and a meaningful statistic cannot be calculated.

2. When there is no, or very little, variation in the data because this also results in ties. This can occur because laboratory analysis of some variables has low precision (i.e.,

values have few or no significant figures). In this case, many samples have the same value, and this then results in ties.

3.2.6 Aggregation of site trends

Aggregating water-quality trend results from multiple sites is intended to indicate water quality changes over a domain of interest (e.g., environmental classes, regions, national). In the present study, we aggregated trend results using both trend magnitudes and trend directions.

The distributions of trend magnitude across sites were characterised using box and whisker plots of the relative Sen slope estimates (RSSE) and relative seasonal Sen slope estimates (RSSSE). Sen slopes were relativised by dividing the SSE and SSSE values by the duration of the trend period to give estimates of temporal change in % yr^{-1}. When reporting trend magnitude results as box and whisker graphs, the median RSSE or SSSE across sites for each lake was used, to not bias the distributions towards lakes with multiple monitoring locations.

We used two different approaches for aggregating trend directions. The first approach involved the calculation of the aggregate proportion of sites in each categorical level of confidence that the trend was decreasing (shown in Table 3-2) for each variable; these values were plotted as colour coded stacked bar charts. These charts provide a graphical representation of the proportions of increasing and decreasing trends at the levels of confidence indicated by the categories. We also used this approach for each of the outputs of the 10-year trends for rolling windows. In order to reduce bias towards lakes with multiple monitoring locations, the median C_d across sites was used to represent the lake C_d.

The second approach also utilises the confidence that the true trend was decreasing to provide a probabilistic estimate of the proportion of decreasing site-specific trends (P_d) within a geographic or environmental domain. Note that P_d is equivalent to the proportion of improving trends (PIT) statistic reported in Larned et al. (2018) without the additional subjective step of defining the direction that constitutes improvement for each water quality variable. For a given water-quality variable, the trends at multiple monitoring sites distributed across a domain of interest can be assumed to represent independent samples of the population of trends, for all sites within that domain. Before calculating the trend, we evaluated the median C_d for each lake, across all monitoring sites in order to reduce bias in the aggregate statistics towards lakes with more than one monitoring site.

The statistic P_d is calculated by letting the sampled sites within this domain be indexed by s, so that $s \in \{1, \dots, S\}$ and let I be a random Bernoulli distributed variable which takes the value 1 with probability $p = C_d$ and the value 0 with probability $q = 1 - C_d$ (where C_d is the confidence that the trend was decreasing, as described in Section 0). Therefore, $I_s = 1$ denotes a decreasing trend at site $s \in \{1, \dots, S\}$ when the estimated $p_s \geq 0.5$ and an increasing trend as 0 when $p_s < 0.5$. Then, the estimated proportion of sites with decreasing trends in the domain is:

$$P_d = \sum_{s=1}^{s=S} I_s / S$$

Because the variance of a random Bernoulli distributed variable is $Var(I) = p(1-p)$, and assuming the site trends are independent, the estimated variance of P_d is:

$$Var(P_d) = \frac{1}{S^2} \sum_{s=1}^{s=S} Var(I_s) = \frac{1}{S^2} \sum_{s=1}^{s=S} p_s(1-p_s)$$

P_d and its variance represent an estimate of the population proportion of decreasing trends, within a spatial or environmental domain, and the uncertainty of that estimate. It is noted that the proportion of increasing trends is the complement of the result (i.e., 1 - P_d). The estimated variance of P_d can be used to construct 95% confidence intervals[2] around the P_d statistics as follows:

$$CI_{95} = P_d \pm 1.96 \times \sqrt{Var(P_d)}$$

We calculated P_d and its confidence interval for all water quality variables and for domains of interest defined by the entire country, and by the altitude x depth classes defined in Section2.3. We also calculated P_d for each of the rolling 10-year time windows.

[2] Note that +/- 1.96 are approximately the 2.5th and 97.5th percentile of a standard normal distribution.

4 Results – lake state

Between 14 and 83 lakes (including 20 to 103 sites) met the filtering rules for the state analysis of nutrients, Seechi, CHLA and TLI; the number of qualifying lakes varied by water quality variable and by elevation × depth class (Table 4-1). The number of lake × variable combinations is higher than in the dataset used in the previous analysis of lake state (Larned et al. 2018).

The geographic distribution of lakes in the state dataset is shown in Figure 4-1. The qualifying lakes are sparsely and unevenly distributed on both the North and South Islands. Lakes with NO3N and ECOLI data were particularly sparse, with 14 and 26 qualifying lakes, respectively.

The distributions of lake-medians for the lake water quality variables for the 2016–2020 period are summarized with box-and-whisker plots in Figure 4-2. These plots indicate that some variability in the most widely measured variables is explained by the lake elevation × depth classes. The low elevation, shallow lake class had the lowest median Seechi and the highest median CHLA, ECOLI, TN, TP and TLI levels. Median Seechi was highest, and median nutrient, CHLA and TLI levels were generally low in the high-elevation lake classes. Median values for these variables in the low elevation, deep lake classes were intermediate, but generally showing better water quality with increasing lake depth. The complete set of state analysis results are provided in the supplementary file “LakeStateResults_2016to2020_v210916 csv”.

The distribution of ECOLI concentration percentiles (5th, 20th, 50th, 80th and 95th) are shown in Figure 4-3, and the distribution of the two ECOLI exceedance measures, G260 and G540 (the percentage of observations that exceeded 260 and 540 cfu 100 ml^{-1}, respectively) are shown in Figure 4-4. For each percentile class, ECOLI concentrations were highest in the low-elevation, shallow lake class, and were substantially lower in the high-elevation lake classes (Figure 4-3). The medians of lake G260 and G540 values were also highest in the low-elevation, shallow lake class (Figure 4-4).

Table 4-1: Number of lakes (and sites, in brackets) by class and water quality variable included in the state analyses. Elevation × depth classes are given as elevation range in top line, maximum-depth range on bottom line. NS: no lakes met the inclusion rule for that water quality variable. Note NH4N (Adj.) is the adjusted ammoniacal-N.

Variable	Elevation × depth class 0–300 m 0–5 m	 0–300 m 5–15 m	 0–300 m 15–50 m	 0–300 m > 50 m	 > 300 m 0–5 m	 > 300 m 5–15 m	 > 300 m 15–50 m	 > 300 m > 50 m	Total
CHLA	25 (39)	29 (31)	10 (11)	8 (11)	1	1	6	3	83 (103)
ECOLI	13 (25)	3 (5)	4	3	1	1	1	NS	26 (40)
NH4N	24 (33)	29 (31)	9 (10)	8 (11)	1	1	6	3	81 (96)
NH4N_adj	22 (29)	26 (28)	6 (7)	5 (8)	1	1	5	3	69 (82)
NO3N	5 (9)	2 (4)	4	2	NS	NS	1	NS	14 (20)
Secchi	18 (28)	27 (28)	9 (10)	7 (10)	NS	NS	6	3	70 (85)
TLI3	25 (37)	29 (31)	10 (11)	8 (11)	1	1	6	3	83 (101)
TLI4	18 (28)	27 (28)	9 (10)	6 (9)	NS	NS	6	3	69 (84)
TN	25 (37)	29 (31)	10 (11)	8 (11)	1	1	6	3	83 (101)
TP	25 (37)	29 (31)	10 (11)	8 (11)	1	1	6	3	83 (101)

Figure 4-1: Locations of lake monitoring sites used for state analyses of lake water quality.

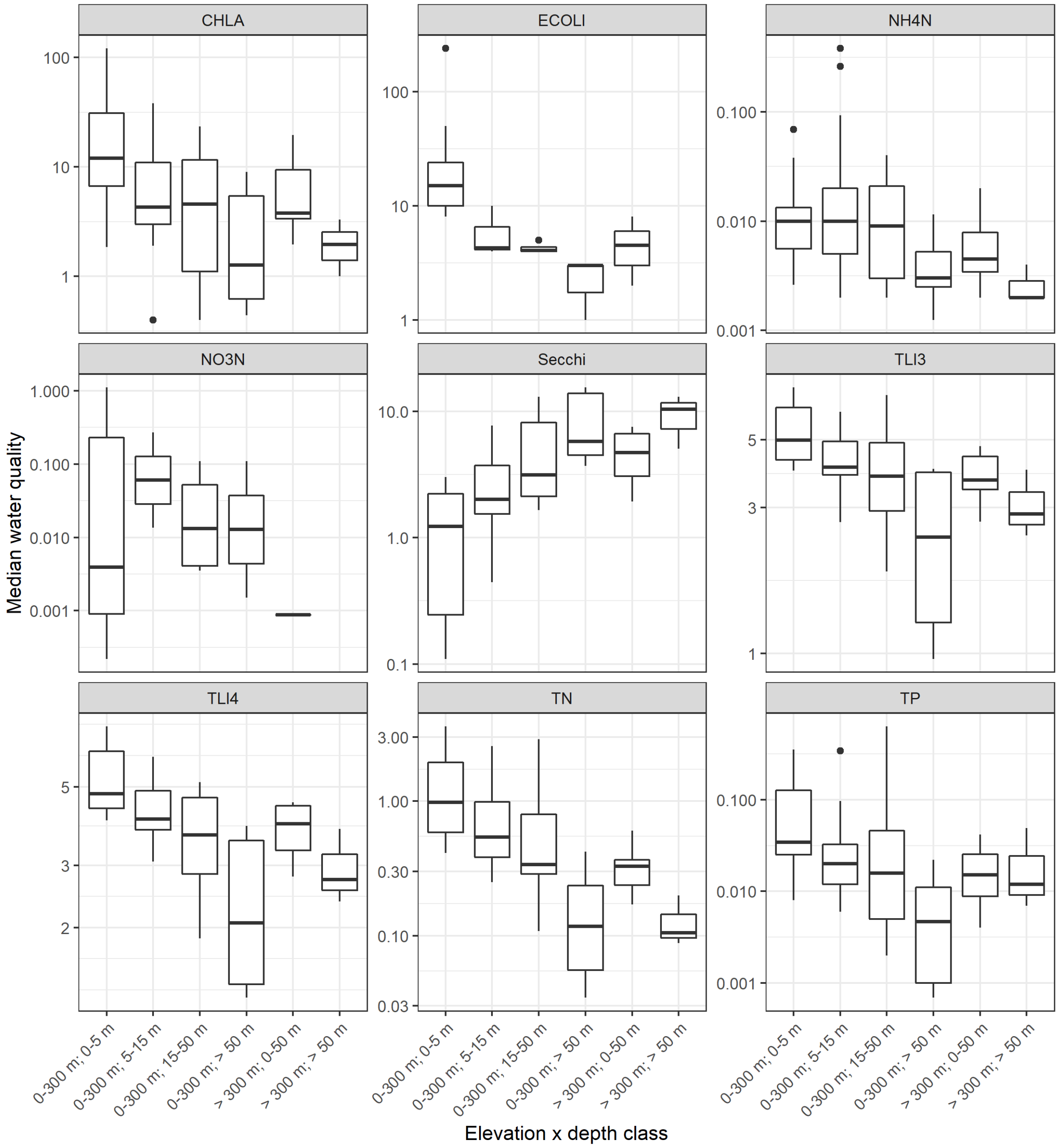

Figure 4-2: Lake water quality state in elevation × depth classes. Box-and-whisker plots show the distributions of lake medians within lake elevation × depth classes. For y-axes units of measure refer to Table 2-1. Black horizontal line in each box indicates the median of site medians, and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as black circles. Note log-scale on Y-axes.

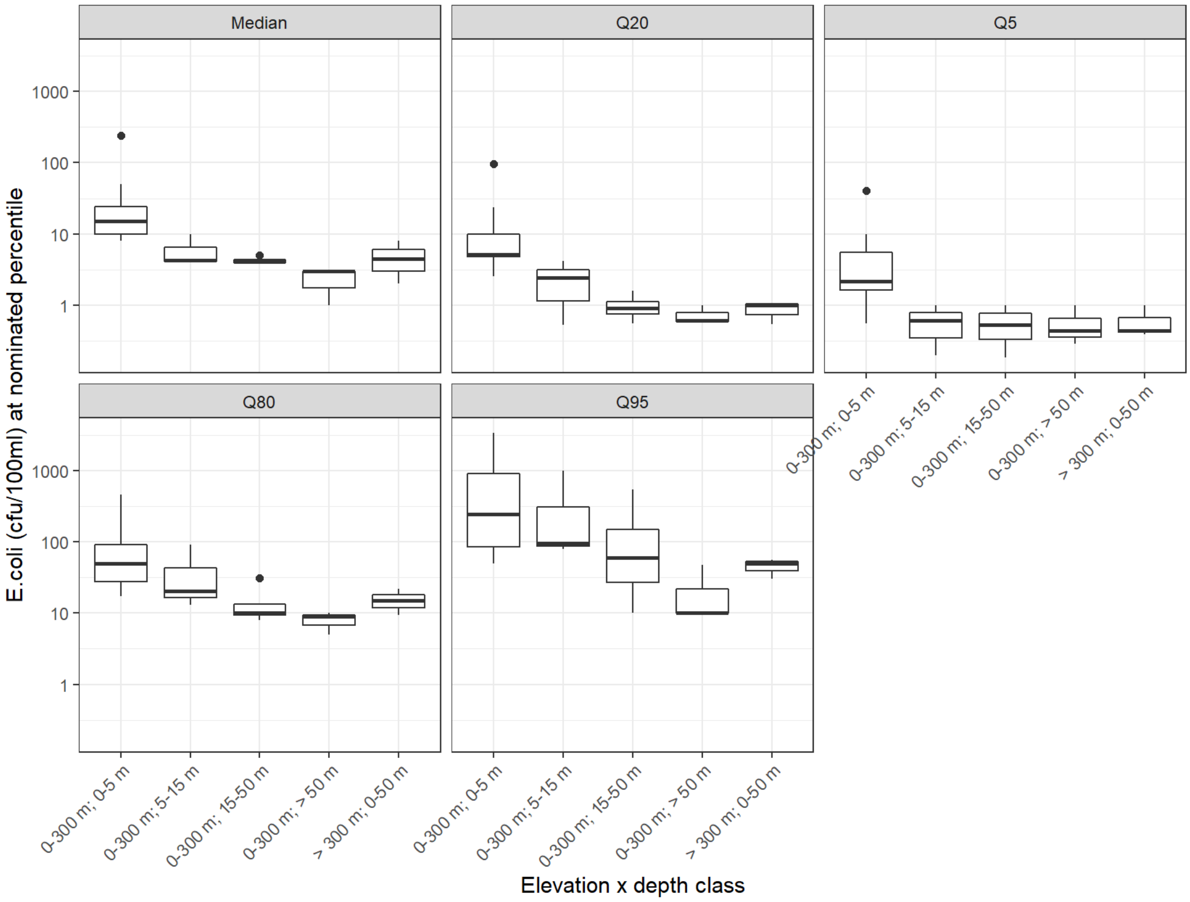

Figure 4-3: ECOLI concentrations in lake elevation × depth classes. Box-and-whisker plots show the distributions of lake percentiles within elevation x depth classes. Black horizontal line in each box indicates the median of site percentiles, and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as black circles. Note log-scale on Y-axes.

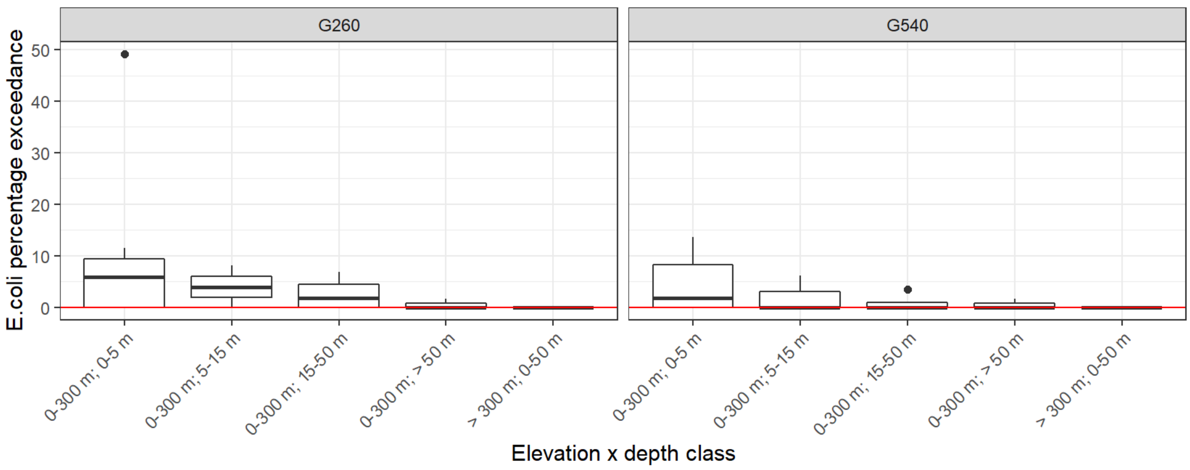

Figure 4-4: ECOLI percent exceedance in lake elevation × depth classes. Box-and-whisker plots show the distributions of percentage exceedance over 540 cfu 100 ml^{-1} (G540) and 260 cfu 100 ml^{-1} (G260) at lakes within land cover classes. Black horizontal line in each box indicates the median of percent exceedances and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as black circles.

4.1 NOF attribute state grades

A summary of lake water quality state grades for each NOF attribute, including the number and lakes and sites that are classified in each NOF grade, is provided in Table 4-2. Figure 4-5 and Figure 4-6 provide maps for each attribute showing the sites coloured by their evaluated state grade. Figure 4-7 and Figure 4-8 show the percentage of sites belonging to each grade, by elevation x depth class and variable.

For the *E. coli* combined numeric attribute, 20% of lakes were below the national bottom line for state (graded D or E). These exceedances all occurred in low elevation, shallow lakes. Only one lake was below the bottom line for ammonia toxicity (median), but a greater number (23%) were below the bottom line for ammonia toxicity (maximum). A number of lakes were below the bottom line for the phytoplankton (trophic state) attribute, with 41% and 26% of sites assigned a D grade for maximum chlorophyl-a and median chlorophyl-a, respectively. For the TN and TP attributes, 33% and 21% of sites were below the bottom line, respectively. All D grades for TP and TN were in low elevation lakes.

Table 4-2: Summary of the number of lakes and sites (in brackets, where this is different from the number of lakes) included in the state grading. Cells shown in grey indicate grades that are below the NOF national bottom line.

Numeric attribute state	NOF Grade				
	A	B	C	D	E
NOF.Chla.Max	17 (18)	17 (18)	17 (24)	36 (43)	0
NOF.Chla.Med	17 (21)	23 (25)	23 (28)	22 (29)	0
NOF.ECOLI.Combined	19 (27)	5 (6)	0	5 (6)	1
NOF.ECOLI.G260	25 (37)	1	0	2	0
NOF.ECOLI.G540	20 (28)	6 (9)	2	0	1
NOF.ECOLI.Med	25 (38)	0	0	2	0
NOF.ECOLI.p95	19 (27)	5 (6)	0	6 (7)	0
NOF.NH4N.Max	22 (24)	34 (40)	17 (18)	0	0
NOF.NH4N.Med	64 (77)	4	1	0	0
NOF.TN.Med	9	29 (34)	20 (24)	28 (34)	0
NOF.TP.Med	20 (21)	20 (21)	27 (34)	18 (25)	0

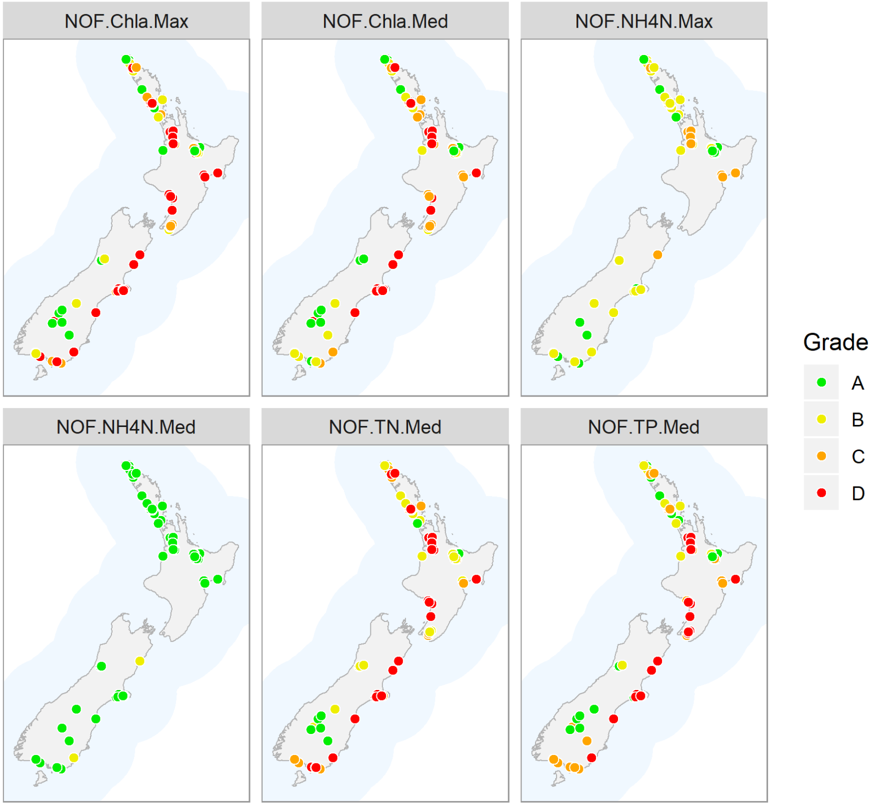

Figure 4-5: Maps showing lake states graded by NOF physico-chemical attributes.

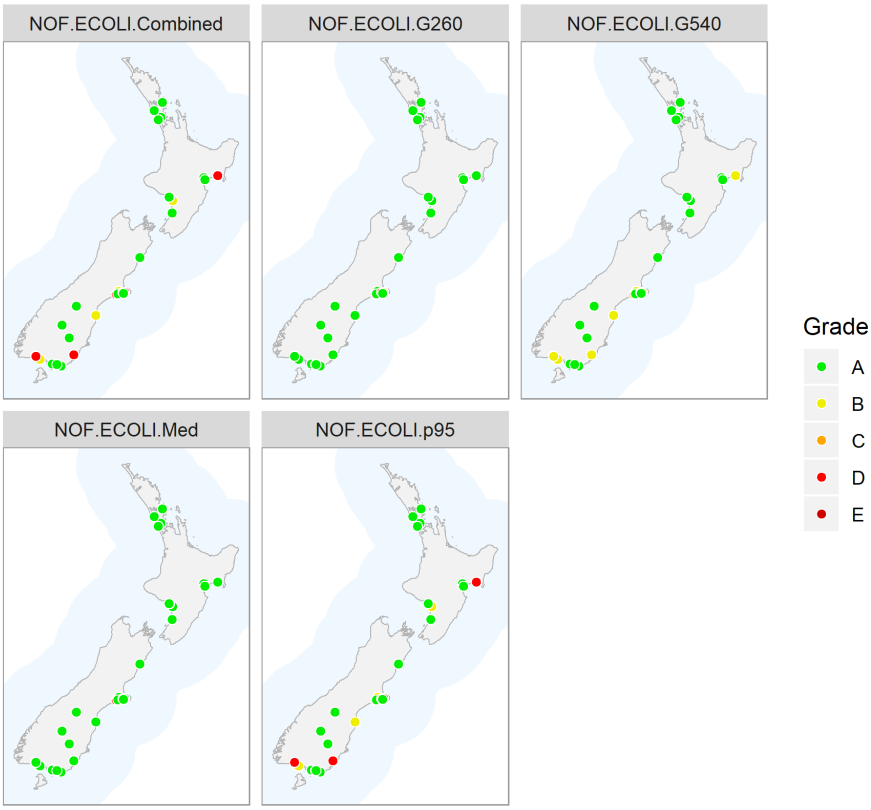

Figure 4-6: Maps showing lake states graded by the NOF *E. coli* attribute.

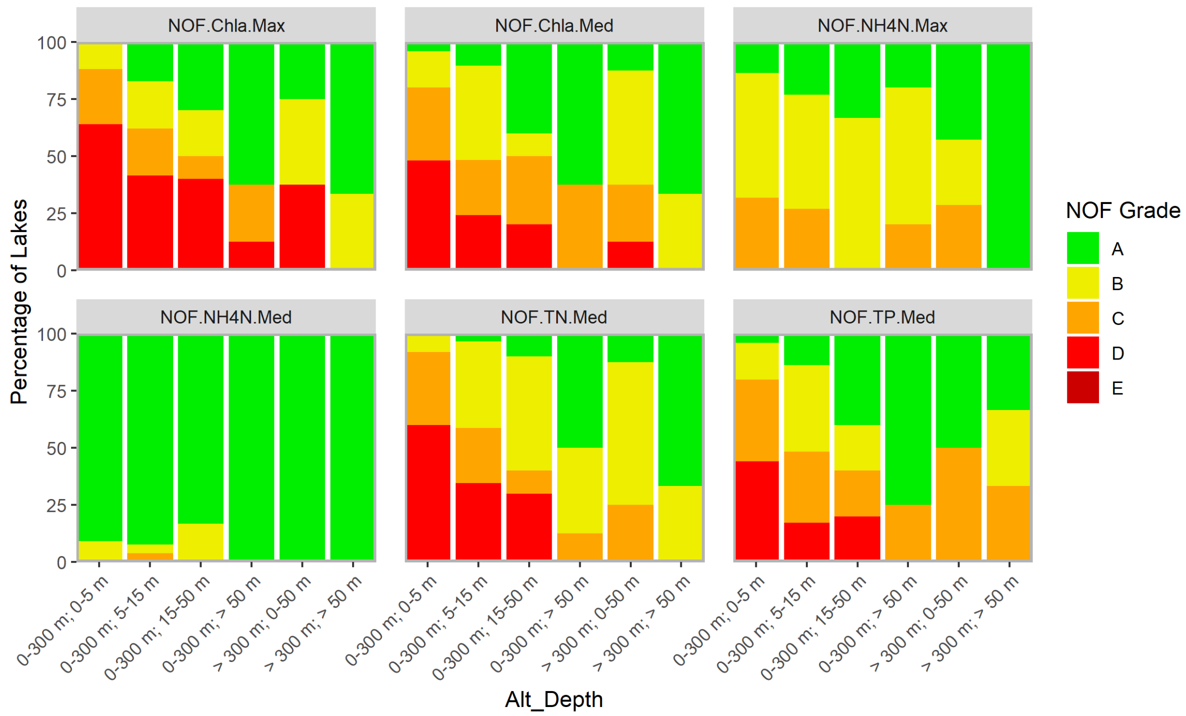

Figure 4-7: Stacked bar charts showing the percentage of lakes assigned to each NOF grade (median over sites within a lake), by land cover class for NOF physico-chemical attributes.

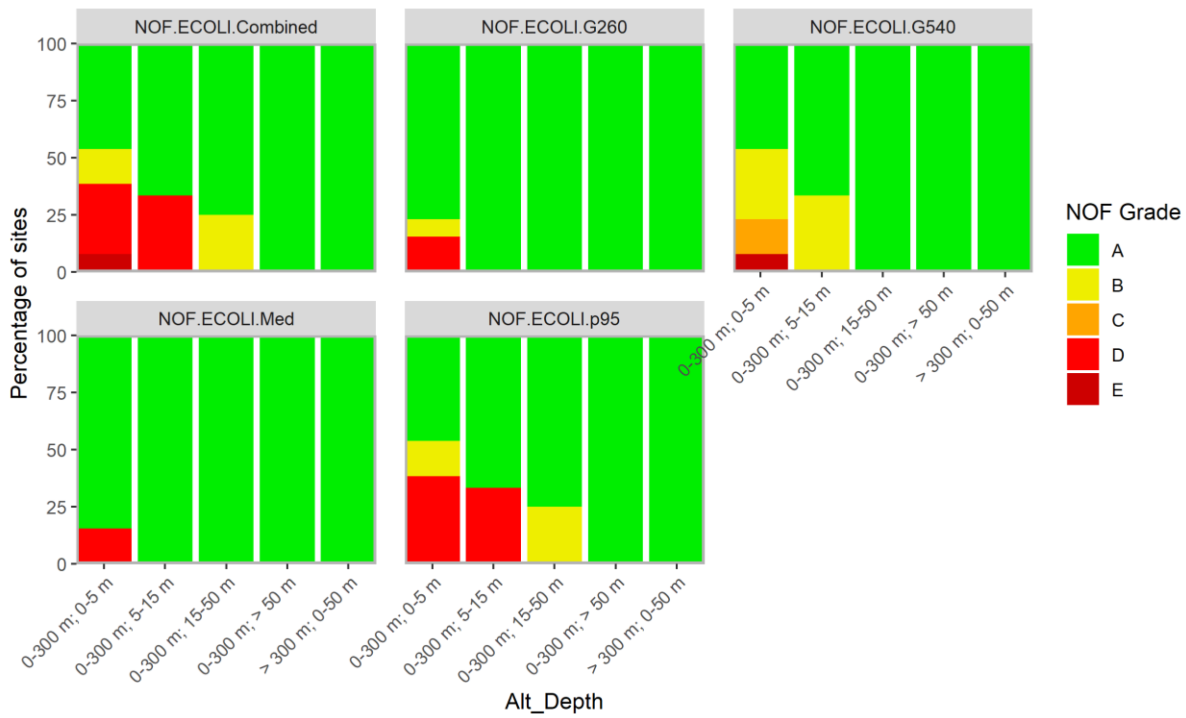

Figure 4-8: Stacked bar charts showing the percentage of lakes assigned each NOF grade (median over sites within a lake), by land cover class for NOF *E. coli* attributes.

5 Results – lake trends

5.1 Ten-year trends (2011–2020)

Between 5 and 65 lakes (and 5 and 77 sites) met the filtering rules for the 10-year trend analysis of water quality variables. The number of lakes varied widely by lake class (Table 5-1). There was only one qualifying lake in each of two high-elevation, shallow lake classes (and not for all variables), and no qualifying lake variable combinations in several other classes (particularly for NO3N). The qualifying lakes are sparsely and unevenly distributed on both the North and South Islands (Figure 5-1). All lake sites, lake classes and numbers of sampling dates are included in the supplementary file "LakeTrends_to2020_v210916.csv".

Table 5-1: Number of lakes and sites (in brackets where different from the number of lakes) in the elevation × depth lake classes included in the 10-year trend analyses. The lake and site numbers shown refer to lakes or sites that met the site inclusion requirements in Section 3.2.1 (measurements were available for at least 80% of the years and at least 80% of seasons).

Variable	Elevation × depth class								Total
	0–300 m 0–5 m	0–300 m 5–15 m	0–300 m 15–50 m	0–300 m > 50 m	> 300 m 0–5 m	> 300 m 5–15 m	> 300 m 15–50 m	> 300 m > 50 m	
CHLA	20 (29)	23	7 (8)	5 (8)	1	NS	5	3	64 (77)
ECOLI	6 (10)	NS	1	2	1	1	1	NS	12 (16)
NH4N	19 (23)	23	7 (8)	5 (8)	1	1	5	3	64 (72)
NO3N	NS	NS	2	2	NS	NS	1	NS	5
Secchi	14 (21)	22	7 (8)	4 (7)	NS	NS	5	3	55 (66)
TLI3	20 (27)	23	7 (8)	5 (8)	1	NS	5	3	64 (75)
TLI4	14 (18)	22	7	3 (6)	NS	NS	5	3	54 (61)
TN	20 (27)	23	7 (8)	5 (8)	1	1	5	3	65 (76)
TP	20 (27)	23	7 (8)	5 (8)	1	1	5	3	65 (76)

5.1.1 Trend rate

Box-and-whisker plots summarise the assessed trend rates for each lake water quality variable in each lake elevation × depth class for the 10-year period from 2011–2020 (Figure 5-2). All estimated trend rates are included in these plots, irrespective of the level of confidence in the assessment (see Section 0). The plots indicate that lake classes did not account for a substantial amount of the variation in trends for any variable; this is in contrast with the state analyses of lake variables, where water-quality state clearly varied between some lake classes (Figure 4-2, Figure 4-3, Figure 4-4).

There were only two classes with more than 7 lakes: 0–300m 0–5m; and 0–300m;5–15m (14–23 lakes, excluding for ECOLI). The small number of lakes in the other classes (1–7 lakes) makes the estimates of class median trend rates unreliable. Median absolute trend magnitudes were largest for CHLA in the 0–300m; 0–5m class (4.3%) and Secchi in the 0–300m; 5–15m class (2.9%). For other variables in these classes, median absolute trend rates were generally less than 1%. Although median values were low, individual lake x variable trend rates were up to 20% increasing and as low as -15% decreasing.

Figure 5-1: Locations of lake sites included in the 10-year trend analyses by water quality variable.

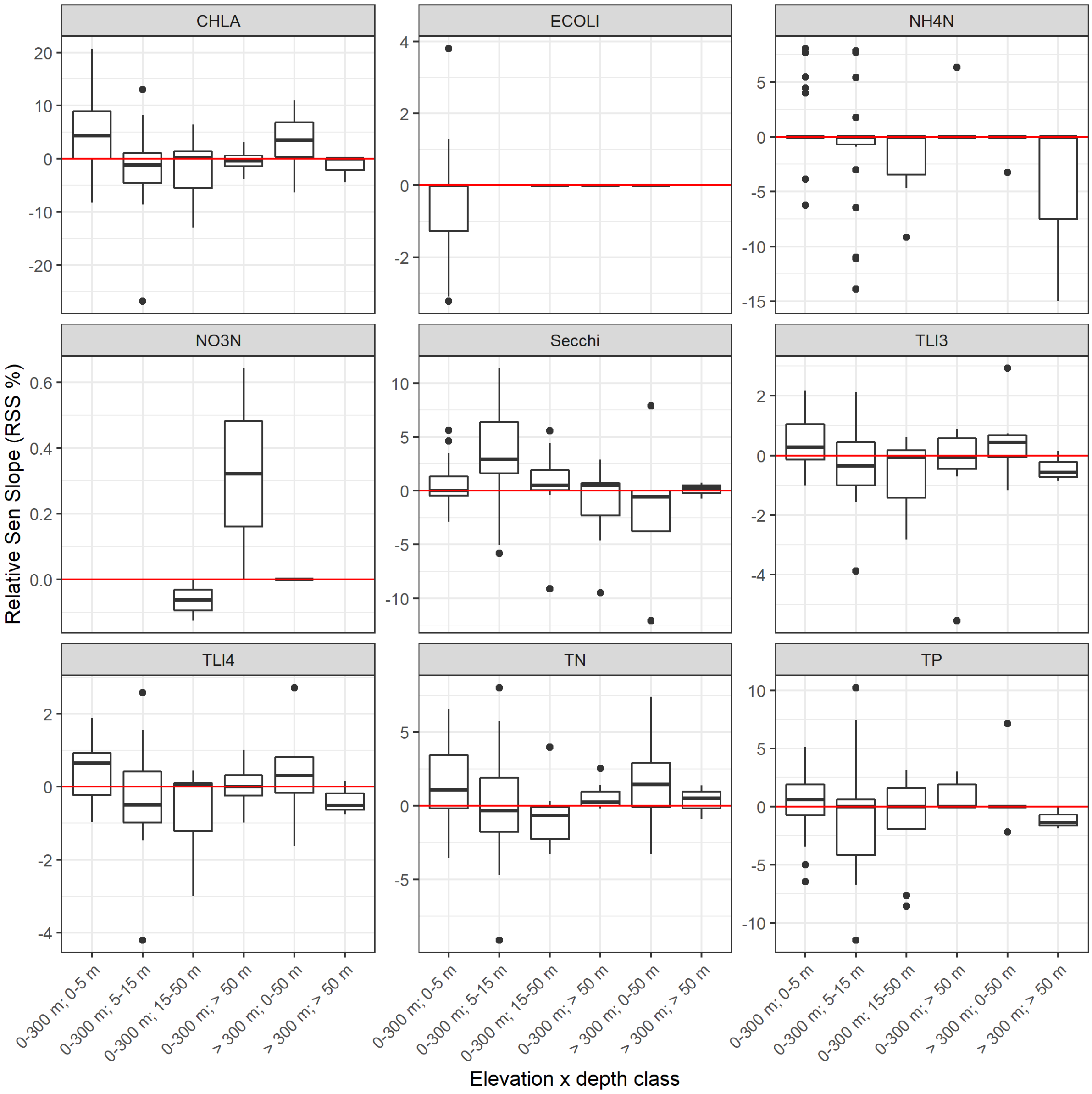

Figure 5-2: Summary of 10-year lake trend rates for the water quality variables, in lake elevation × depth classes. Box-and-whisker plots show the distributions of site trends within each class. Black horizontal line in each box indicates the median of site trends and the box indicates the inter-quartile range (IQR). The red line indicates a rate of zero. Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as black circles.

5.1.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence 10-year trends were decreasing for each site × variable combination. The spatial distributions of categorised individual sites are shown in Figure 5-3. Because confidence that a trend is decreasing is the complement of the confidence that a trend is increasing, "unlikely" decrease, could also be categorised as "likely" increase. Note that for Secchi and TLI, decreasing trends indicate degradation, whereas for all other variables decreasing trends indicate improvement.

Figure 5-3: Lake water quality monitoring sites categorised by the confidence that the 10-year trend was decreasing (C_d) for each variable. C_d is expressed using the confidence categories in Table 3-2. Only sites that met the sampling requirements outlined in Section 3.2.1 are shown in the figure.

5.1.3 Aggregate Trends

Figure 5-4 shows the proportions of sites belonging to each of the nine categorical levels of confidence for C_d defined in Table 3-2 for the 10-year trends. These plots provide a national-scale assessment of the relative proportions of lakes with decreasing versus increasing trends.

The national-scale proportions of decreasing trends (P_d) and their confidence intervals are summarised in Table 5-2. The 10-year P_d statistics ranged from 31-60%. For all variables, the 95% confidence intervals for the P_d included 50% and we could not infer widespread increases or decreases for these variables.

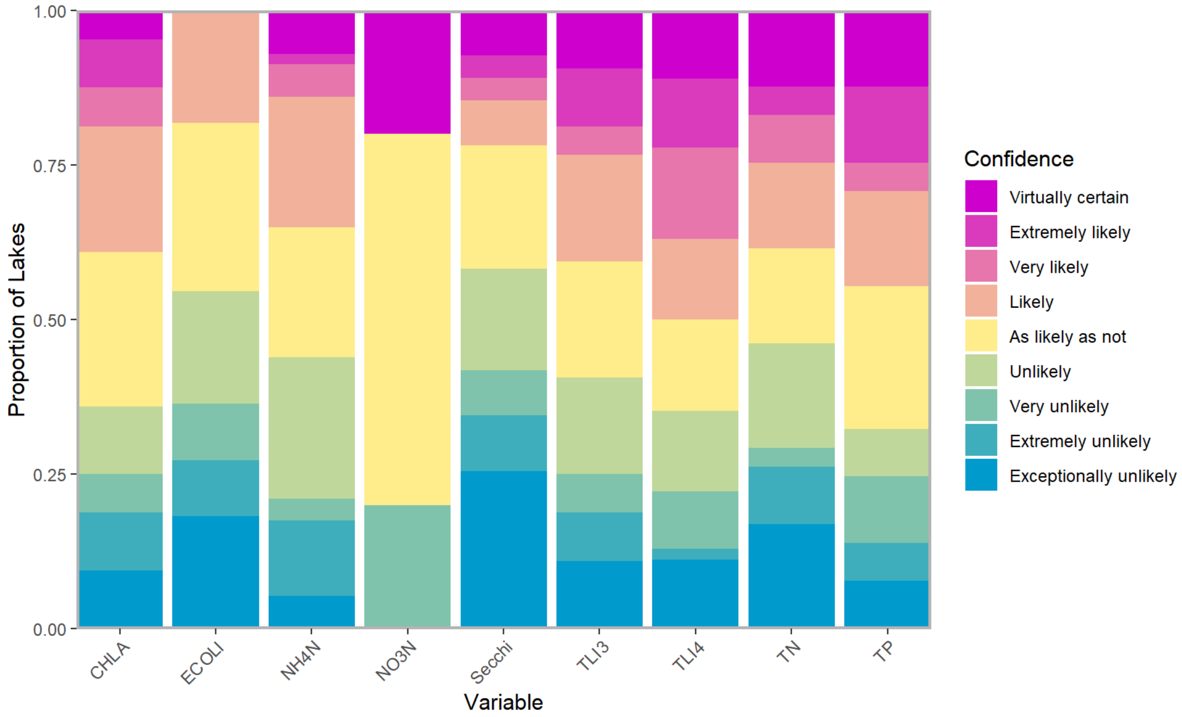

Figure 5-4: Summary plot representing the proportion of lakes with decreasing 10-year trends at each categorical level of confidence. The plot shows the proportion of sites with decreasing trends at levels of confidence defined in Table 3-2.

Table 5-2: Proportions of decreasing 10-year lake trends (P_d) by variable.

Variable	Number of lakes	P_d (%)	95% confidence interval for P_d (%)
CHLA	64	50.8	34.9–66.7
ECOLI	11	36.4	0.0–81.1
NH4N	57	49.1	31.9–66.3
NO3N	5	60	9.0–100.0
Secchi	55	30.9	10.7–51.1
TLI3	64	51.6	35.7–67.5
TLI4	54	54.6	38.1–71.1
TN	65	47.7	31.0–64.4
TP	65	55.4	40.3–70.5

5.2 Twenty-year trends (2001–2020)

Between one and 37 lakes (and one and 52 sites) met the inclusion rules for the 20-year trend analysis (Table 5-3). Only one lake met the inclusion rules for NO3N, and only six for ECOLI. The number of lakes in each variable × lake class combination was very small (0–13 lakes). The majority of qualifying lakes were in the Bay of Plenty, Waikato, Otago and Auckland Regions (Figure 5-5). All lake locations, lake classes and numbers of sampling dates are included in the supplementary file "LakeTrends_to2020_v210916.csv".

Table 5-3: Number of lakes and sites (in brackets where different from the number of lakes) in the elevation × depth lake classes included in the 20-year trend analyses. The site numbers shown refer to sites that met the site inclusion requirements in Section 3.2.1 (measurements were available for at least 80% of the years and at least 80% of seasons).

Variable	Elevation × depth class								Total
	0–300 m 0–5 m	0–300 m 5–15 m	0–300 m 15–50 m	0–300 m > 50 m	> 300 m 0–5 m	> 300 m 5–15 m	> 300 m 15–50 m	> 300 m > 50 m	
CHLA	12 (18)	5	5 (6)	9 (14)	NS	NS	NS	2	37 (49)
ECOLI	1	NS	NS	4	NS	NS	NS	NS	6
NH4N	11 (14)	5	4 (5)	9 (13)	NS	NS	NS	2	36 (44)
NO3N	NS	NS	NS	1	NS	NS	NS	NS	1
Secchi	6 (9)	5	5	4 (6)	NS	NS	NS	2	26 (31)
TLI3	10 (13)	4	4 (5)	6 (9)	NS	NS	NS	2	30 (37)
TLI4	4 (7)	3	4	2 (3)	NS	NS	NS	2	19 (23)
TN	11 (14)	6	4 (5)	6 (9)	NS	NS	NS	2	34 (41)
TP	13 (19)	7	5 (6)	9 (13)	NS	NS	NS	2	41 (52)

5.2.1 Trend rate

Box-and-whisker plots summarise the assessed trend rates for each lake water quality variable in each elevation x depth class for the for the 20-year period from 2001–2020 (Figure 5-6). All estimated trend rates are included in these plots, irrespective of the level of confidence in the assessment (see Section 3.2.5).

Figure 5-5: Locations of lake sites included in the 20-year trend analyses by water quality variable.

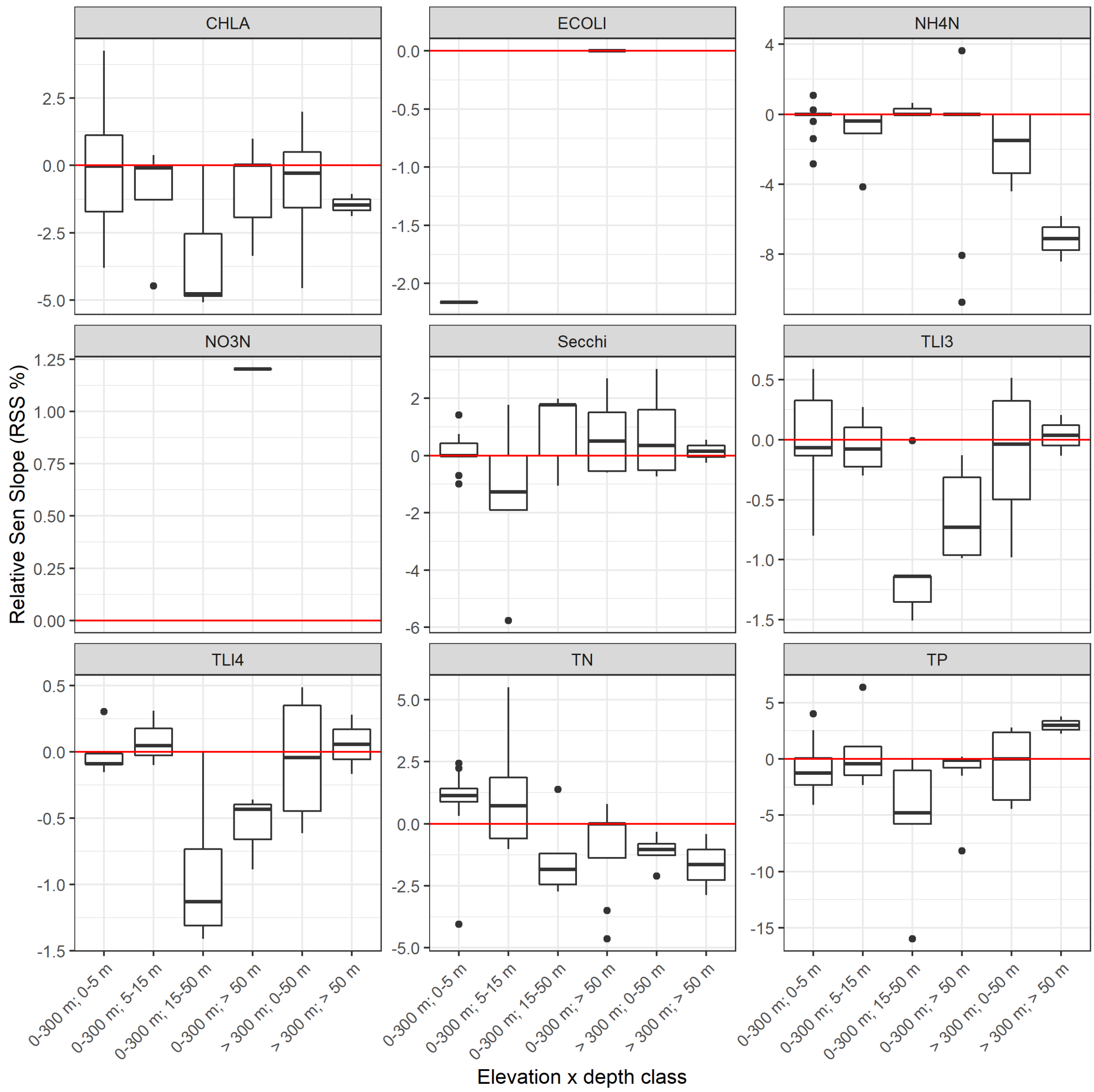

Figure 5-6: Summary of 20-year lake trend rates for the water quality variables, in lake elevation × depth classes. Black horizontal line in each box indicates the median of site trends, and the box indicates the inter-quartile range (IQR). The red line indicates a rate of zero. Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as black circles.

5.2.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence 20-year trends were decreasing for each site × variable combination. The spatial distributions of categorised individual sites are shown in Figure 5-7. Because confidence that a trend is decreasing is the complement of the confidence that a trend is increasing, “unlikely” decrease, could also be categorised as “likely” increase. Note that for Secchi and TLI, decreasing trends indicate degradation, whereas for all other variables decreasing trends indicate improvement.

Figure 5-7: Lake water quality monitoring sites categorised by the confidence that the 20-year trend was decreasing (C_d) for each variable. C_d is expressed using the confidence categories in Table 3-2. Only sites that met the sampling requirements outlined in Section 3.2.1 are shown in the figure.

5.2.3 Aggregate Trends

Figure 5-8 shows the proportions of sites belonging to each of the nine categorical levels of confidence for C_d defined in Table 3-2 for the 20-year, raw trends. These plots provide national-scale assessments of the relative proportions of lakes with decreasing versus increasing trends.

The national-scale proportions of decreasing trends (P_d) and their confidence intervals are summarised in Table 5-4. The 10-year P_d statistics ranged from 47–70%. NH4N had a majority (i.e., P_d >50%) of decreasing trends at the 95% confidence level. For all other variables, the 95% confidence intervals included 50% and we could not infer widespread increases or decreases for these variables.

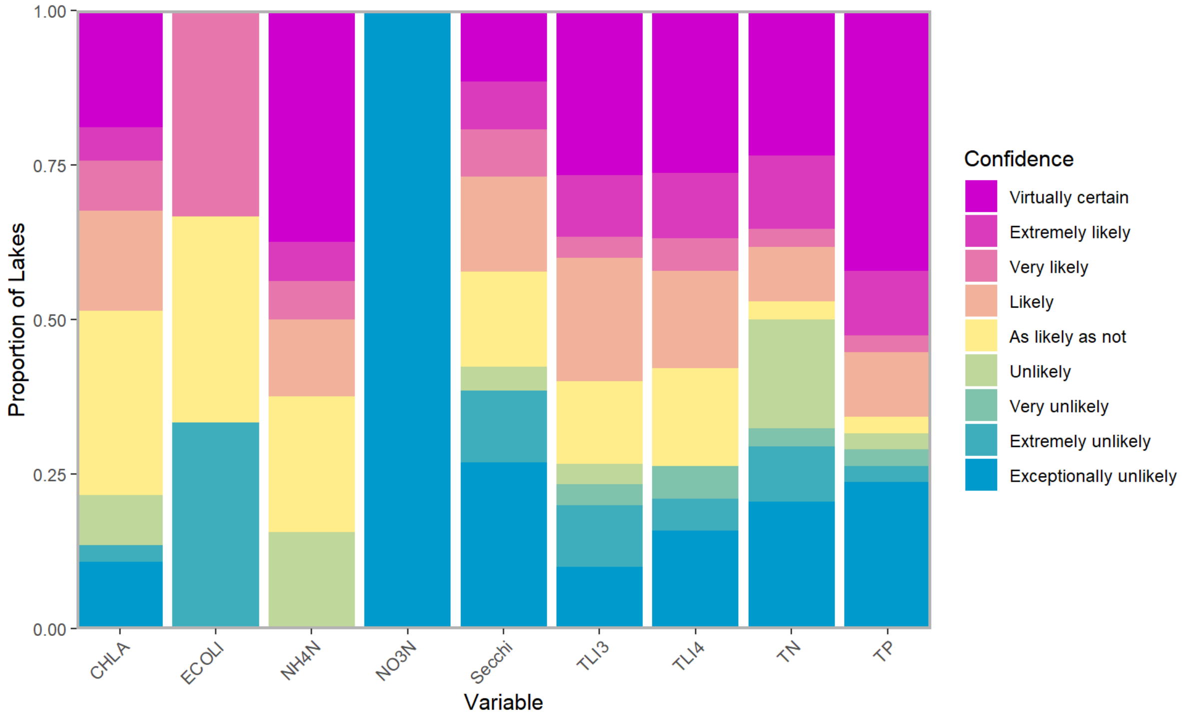

Figure 5-8: Summary plot representing the proportion of sites with decreasing 20-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with decreasing trends at levels of confidence defined in Table 3-2.

Table 5-4: Proportions of decreasing trends (P_d) for 20-year time period.

Variable	Number of sites	P_d (%)	95% confidence interval for P_d (%)
CHLA	37	64.9	47.1–82.7
NH4N	32	70.3	54.2–86.4
Secchi	26	57.7	31.4–84.0
TLI3	30	70	49.6–90.4
TLI4	19	63.2	37.1–89.3
TN	34	47.1	24.2–70.0
TP	38	65.8	47.8–83.8

5.3 Thirty-year trends (1991–2020)

Between two and 15 lakes met the inclusion rules for the 30-year trend analysis of seven of the nine water quality variables. No lakes met the inclusion rules for ECOLI or NO3N. The number of lakes varied widely by lake class (Table 5-5). The qualifying lakes were limited to three regions: Bay of Plenty (maximum of nine lakes), Waikato (two lakes), Otago (maximum five lakes) and Auckland (three lakes) (Figure 5-9). TLI is only available in Bay of Plenty. All lake locations, lake classes and numbers of sampling dates are included in the supplementary file "LakeTrends_to2020_v210916.csv".

Table 5-5: Number of lakes in the elevation × depth lake classes included in the 30-year trend analyses. The lake numbers shown refer to lakes that met the site inclusion requirements in Section 3.2.1 (measurements were available for at least 80% of the years and at least 80% of seasons). All lakes only had one site.

	Elevation × depth class								
Variable	0–300 m 0–5 m	0–300 m 5–15 m	0–300 m 15–50 m	0–300 m > 50 m	> 300 m 0–5 m	> 300 m 5–15 m	> 300 m 15–50 m	> 300 m > 50 m	Total
CHLA	2	3	1	1	NS	NS	4	1	12
ECOLI	NS	NS	NS	NS	NS	NS	NS	NS	NS
NH4N	4	2	NS	4	NS	NS	4	1	15
NO3N	NS	NS	NS	NS	NS	NS	NS	NS	NS
Secchi	2	3	1	1	NS	NS	4	1	12
TLI3	NS	NS	NS	1	NS	NS	3	1	5
TLI4	NS	NS	NS	NS	NS	NS	1	1	2
TN	1	NS	NS	1	NS	NS	4	1	7
TP	4	2	NS	2	NS	NS	4	1	13

5.3.1 Trend rate

Box-and-whisker plots were used to summarise the assessed trend rates for each of the water quality variables for the 30-year period from 1991–2020 across the lake elevation x depth classes (Figure 5-10). All estimated trend rates are included in these plots, irrespective of the level of confidence in the assessment (see Section 3.2.5).

The very low numbers of lakes included in the 30-year trend assessment means that the distributions shown in Figure 5-10 are unlikely to accurately reflect trends in the populations of lakes in each elevation x depth class.

Figure 5-9: Locations of lake sites included in the 30-year trend analyses by water quality variable.

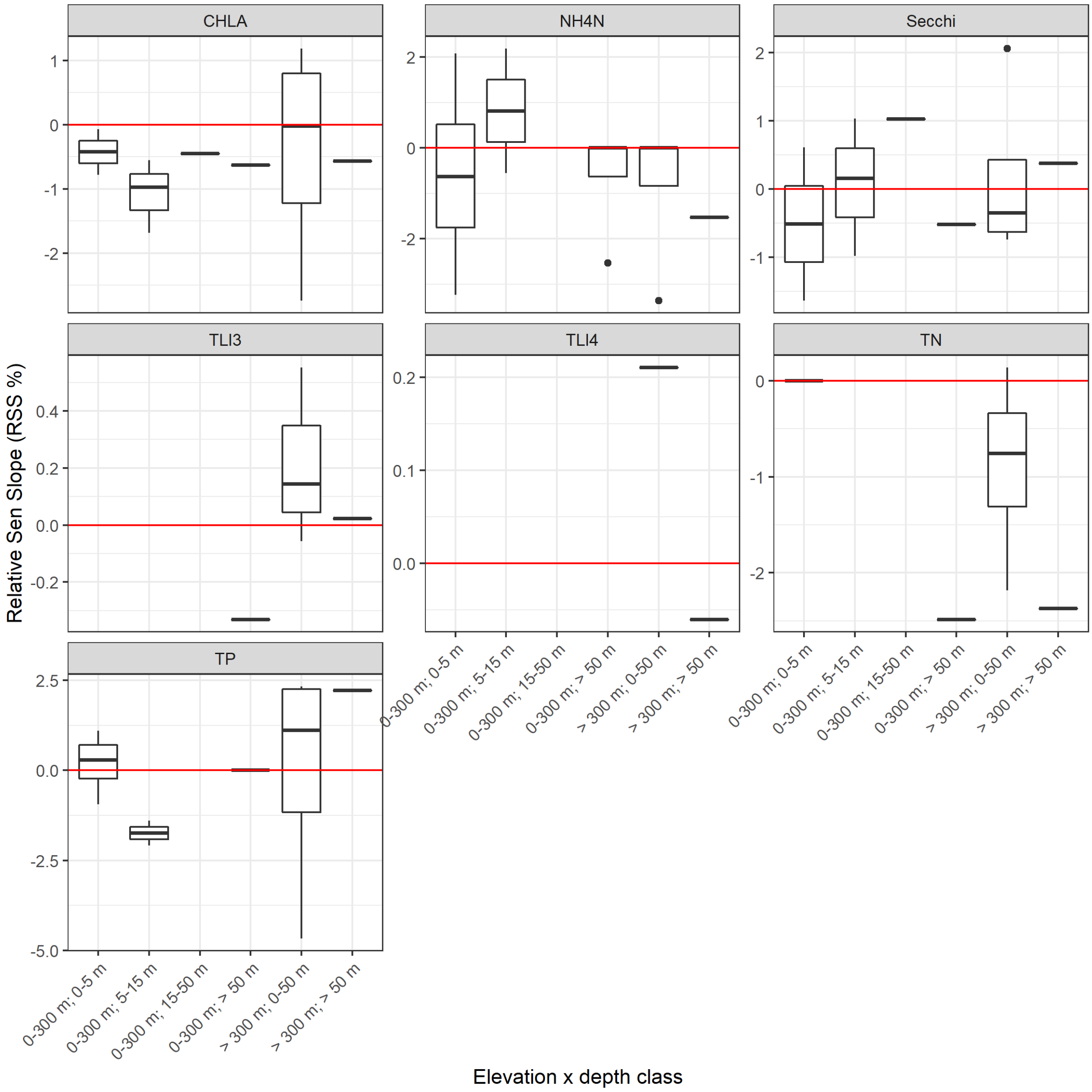

Figure 5-10: Summary of 30-year lake trend rates in water quality variables, in lake elevation × depth classes. Black horizontal line in each box indicates the median of site trends, and the box indicates the inter-quartile range (IQR). The red line indicates a rate of zero. Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as black circles.

5.3.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence 30-year trends were decreasing for each site × variable combination. The spatial distributions of categorised individual sites are shown in Figure 5-11. Because confidence that a trend is decreasing is the complement of the confidence that a trend is increasing, “unlikely” decrease, could also be categorised as “likely” increase. Note that for Secchi and TLI, decreasing trends indicate degradation, whereas for all other variables decreasing trends indicate improvement.

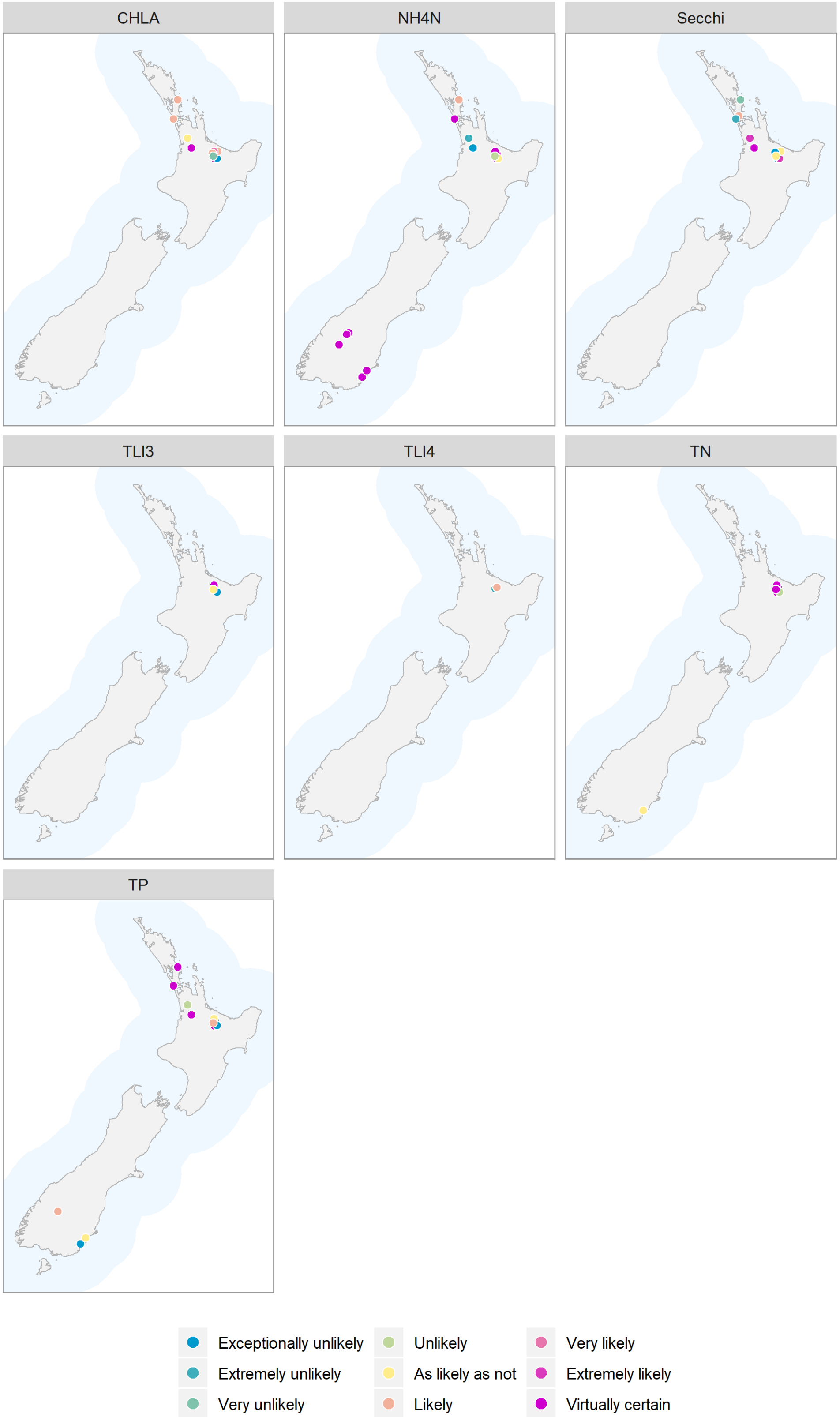

Figure 5-11: Lake water quality monitoring sites categorised by the confidence that the 30-year trend was decreasing (C_d) for each variable. C_d is expressed using the confidence categories in Table 3-2. Only sites that met the sampling requirements outlined in Section 3.2.1 are shown.

5.3.3 Aggregate Trends

Figure 5-12 shows the proportions of sites belonging to each of the nine categorical levels of confidence for C_d defined in Table 3-2 for the 30-year trends. These plots provide a national-scale assessment of the relative proportions of lakes with decreasing versus increasing trends.

The national-scale proportions of decreasing trends (P_d) and their confidence intervals are summarised in Table 5-6. The 10-year P_d statistics ranged from 40–83%. CHLA had a majority of decreasing (i.e., P_d>50%) trends, at the 95% confidence level. For all other variables, the 95% confidence intervals for the P_d included 50% and we could not infer widespread increases or decreases for these variables.

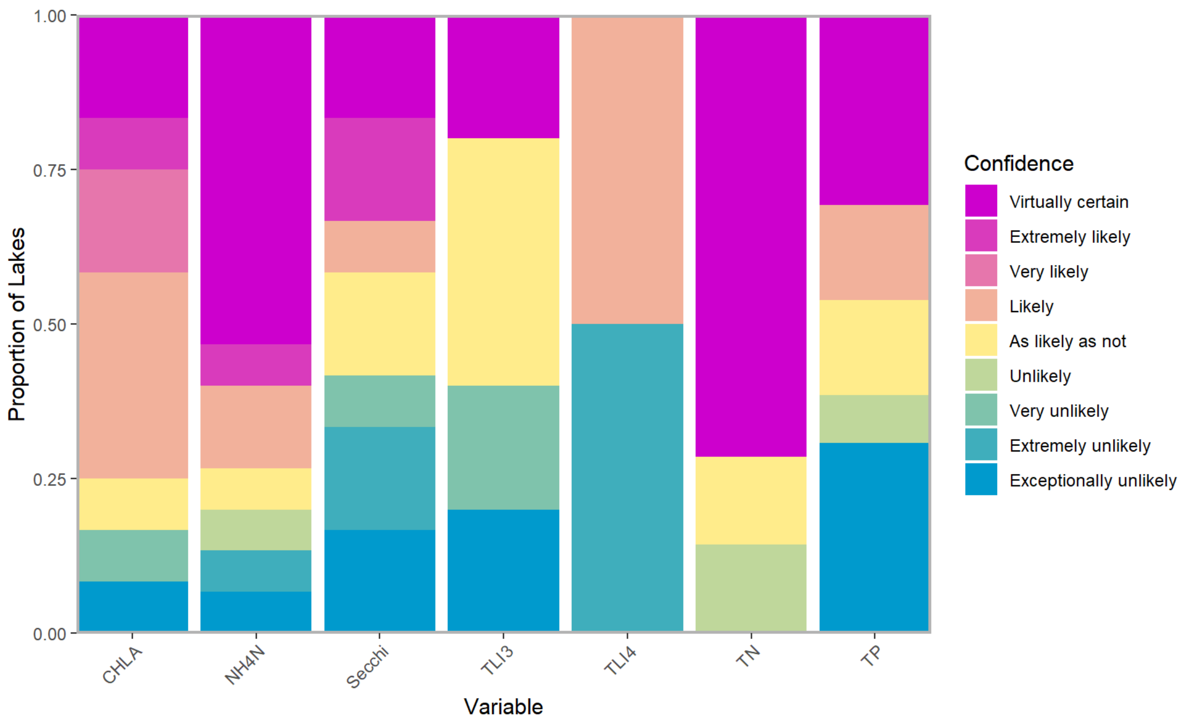

Figure 5-12: Summary plot representing the proportion of sites with decreasing 30-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with decreasing trends at levels of confidence defined in Table 3-2.

Table 5-6: Proportions of decreasing trends (P_d) for 30-year time period.

Variable	Number of sites	Pd (%)	95% confidence interval for Pd (%)
CHLA	12	83.3	55.3–100.0
NH4N	15	73.3	49.8–96.8
Secchi	12	50	11.6–88.4
TLI3	5	40	0.0–100.0
TLI4	2	50	0.0–100.0
TN	7	71.4	43.6–99.2
TP	13	53.8	17.7–89.9

5.4 Rolling ten-year trends

The number of lakes that met the filtering rules for each of the ten-year windows and variables ranged from zero to 70. The changes in the number of sites included in the rolling ten-year trend assessment over time are shown in Figure 5-13.

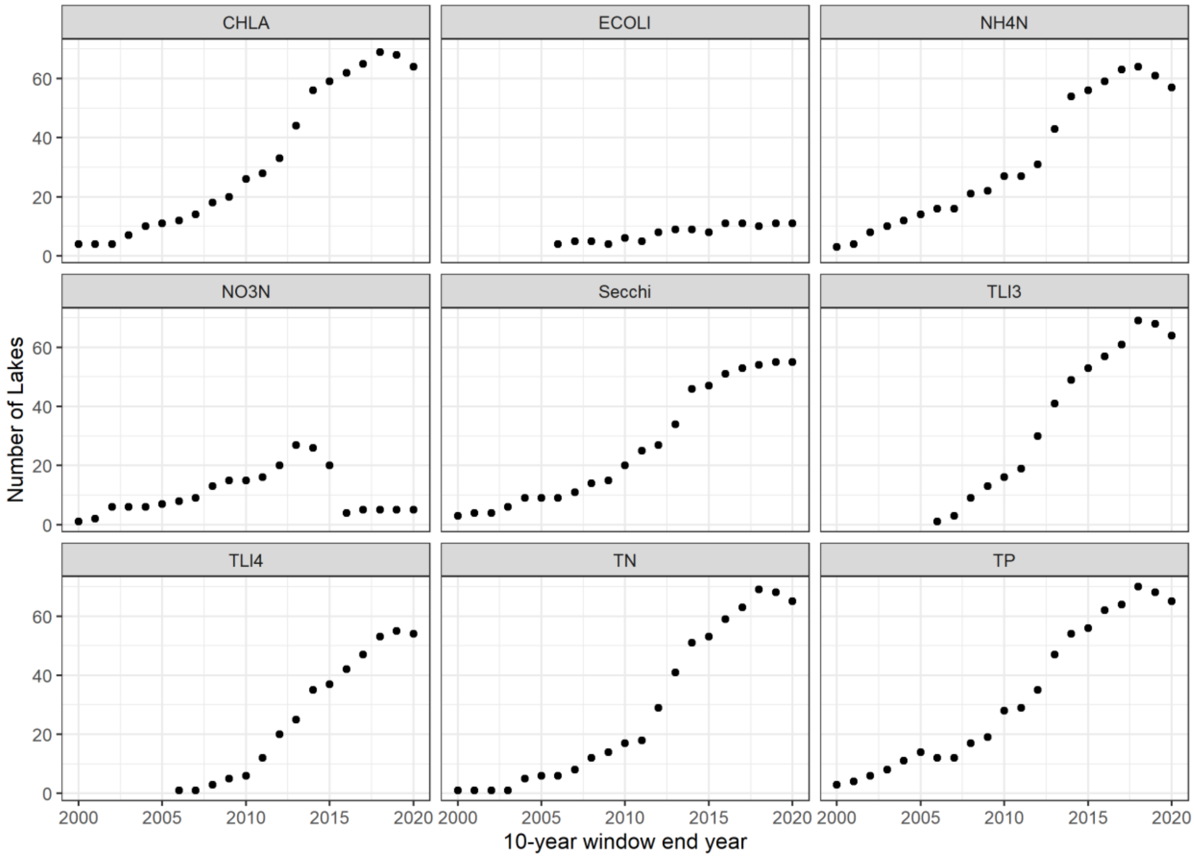

Figure 5-13: Number of lakes included for each 10-year period in the rolling trends assessment by variable.

5.4.1 Trend rate

Box-and-whisker plots were used to summarise the estimated trend rates for each of the water quality variables for each 10-year window (Figure 5-14). All estimated trend rates are included in these plots, irrespective of the levels of confidence in the assessment (see Section 0). Time windows are only shown where the sample size was at least 20 lakes. This was an arbitrary cut-off point selected to minimise bias that might be associated with a small sample size but maximise the number of time windows that were reported (note, this criterion completely excluded ECOLI from the rolling 10-year trend assessment).

The median trend rate for TN has been monotonically increasing over the past ten ten-year periods. All other variables show patterns that are quasi-periodic or random. We note that the patterns may be influenced by the number of sites included in each ten-year window.

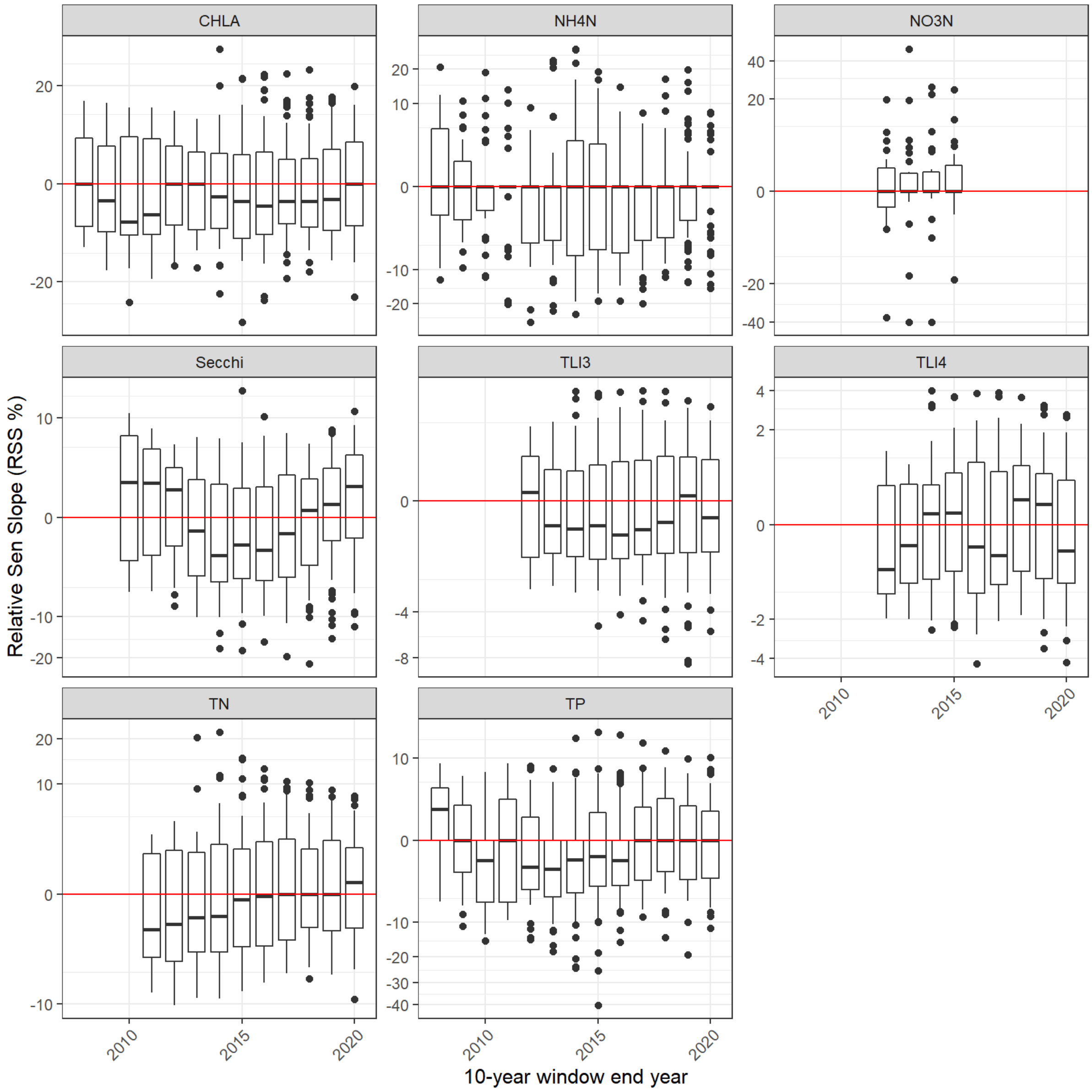

Figure 5-14: Summary of rolling window 10-year trend rates. Box-and-whisker plots show the distributions of site trend rates within each ten-year window. Black horizontal line in each box indicates the median of site trends, and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as black circles. Note, y-axis has a signed square root transformation. The red line indicates a rate of zero. Units for each variable are given in Table 2-1.

5.4.2 Aggregate trends

Figure 5-15 shows the proportions of lakes belonging to each of the nine categorical levels of confidence for C_d defined in Table 3-2 for each of the 10-year time window trends. These plots provide a national-scale assessment of the relative proportions of lakes with decreasing versus increasing trends, and how these have changed through time. Time windows are only shown where the sample size was at least 20 lakes (note, this criterion completely excluded ECOLI from the rolling 10-year trend assessment).

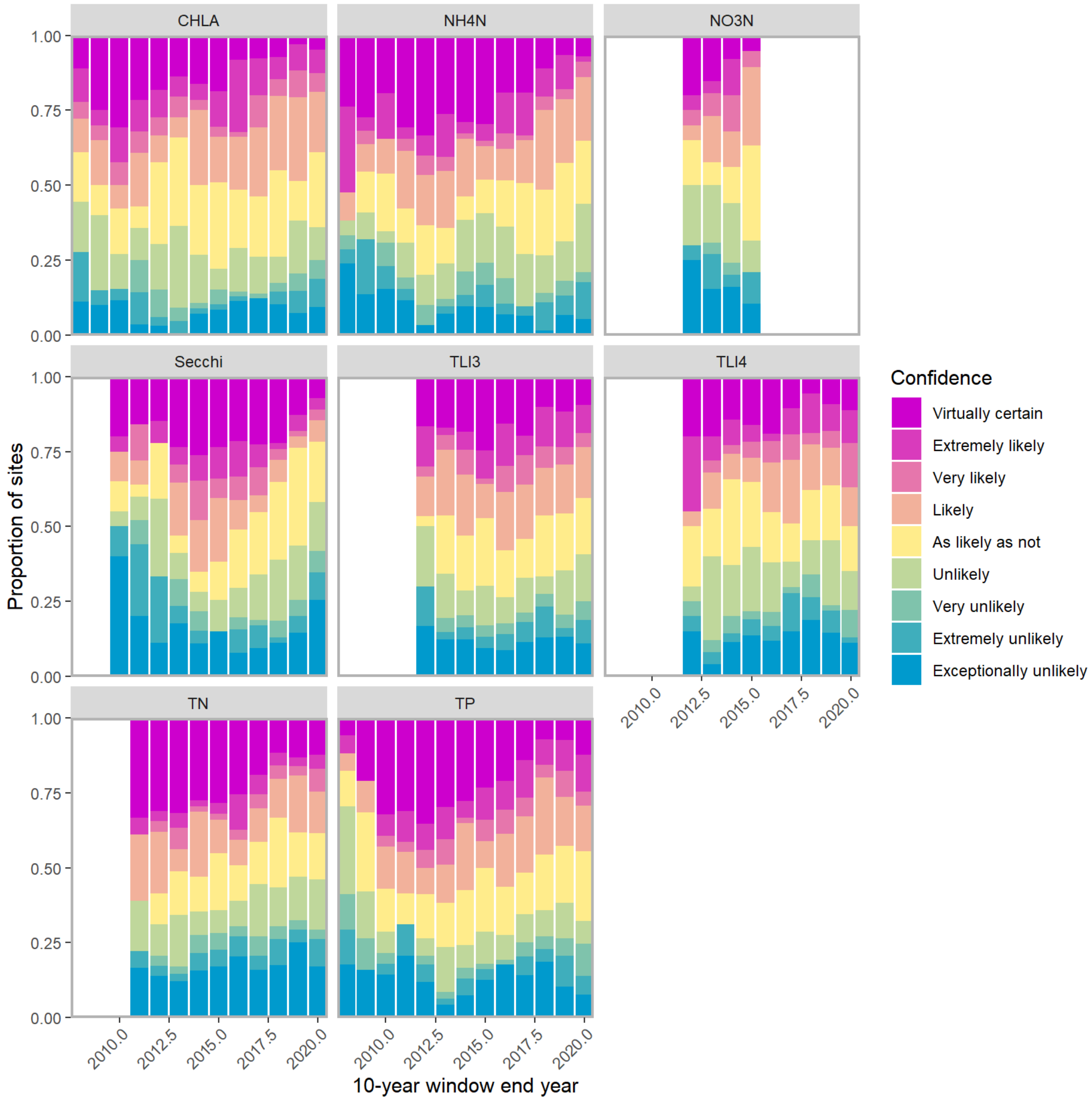

Figure 5-15: Summary plot representing the proportion of sites with decreasing 10-year time-period trends at each categorical level of confidence for each time window. The plot shows the proportion of lakes with decreasing trends at levels of confidence defined in Table 3-2.

The national-scale proportions of decreasing trends (P_d) and their confidence intervals for each time window with more than 20 sites are summarised in Figure 5-16. TN exhibits an approximately monotonic change in the P_d score, with the proportion of sites with decreasing (i.e., improving) trends reducing over time. The variation in P_d varied between lake water quality variables, with the greatest variation observed for Secchi (ranging from 31% to 69%). For CHLA and NH4N, a majority of sites had decreasing trends in each time window. However, that the majority of sites were decreasing was not established at the 95% confidence level for most time windows.

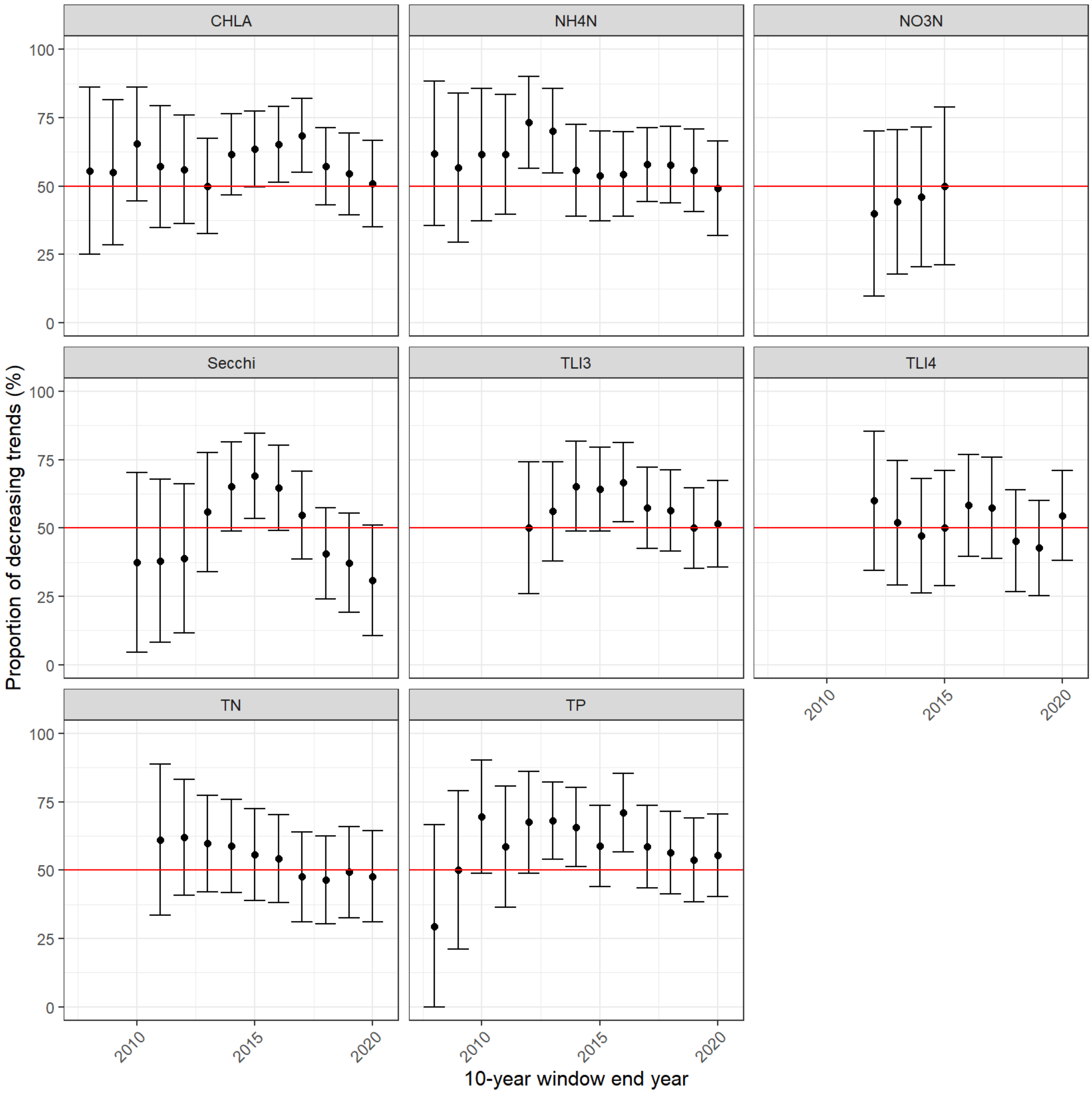

Figure 5-16: Proportions of decreasing trends (P_d) within each 10-year window. Error bars are 95% confidence intervals.

6 Discussion

The primary purposes of the state and trend analyses reported here are:

- to provide MfE with information required for reporting on the freshwater domain, and
- to support policy development.

The detailed information for each lake monitoring site is contained in the supplementary files that accompany this report. The sites and their water quality conditions can be aggregated in many ways to meet different information requirements (e.g., grouped by region or environmental class, distributed along environmental gradients). Therefore, we limited our summaries of the results to example tables and plots, and we focus this discussion on the methods used, rather than a detailed interpretation of the results.

We have used the same state assessment methodology as used in the previous national-scale water quality state analyses (Larned et al. 2018). There have been some changes in the trend assessment methodology and terminology used in the report. These changes have largely been made to align the reporting with recently published trend guidance (Helsel et al. 2020; Snelder et al. 2021). The differences are summarised below:

Changes in method:

- A hi-censor filter has been applied.
  - Previously a hi-censor filter was not applied, but using a high censor removes the possibility that the reported trend is associated with a change in censoring level rather than a change in the variable with time.
- An additional sampling frequency (bi-monthly) has been added.
  - Previously sites that were predominantly monitored on a bi-monthly frequency were evaluated based on quarterly seasons. Including the bi-monthly seasons increases the statistical power for these sites.
- When more than one observation is available within a sampling period, we use only the sample that is closest to the centre of time of the sampling period.
  - Previously, where more than one observation per sampling period existed we used the median of the sample period. However, where there are changes in sampling frequency, this averaging reduces variance in the higher frequency period, and can artificially induce trends (Helsel et al. 2020).

Changes in terminology and reporting:

- In the current report, the main measure of trend direction is C_d, the confidence that the trend was decreasing. This is the same quantity that was referred to as “P”, probability that the trend was decreasing in Larned et al. (2018). In the previous report, the complement of P was taken for variables for which decreasing trends indicated degradation (and P for all other variables), to provide a metric “probability that the trend was improving”. We have not assigned trend directions to improving or degrading categories in this report to avoid subjectivity associated with the choice of trend directions that are regarded to indicate improvement and degradation.

- In the current report, aggregate proportions of sites that are decreasing are reported as P_d. In the previous report, aggregate proportions of sites that are improving were reported as PIT. P_d and PIT are derived in the same way, with the exception that PIT used "the probability that the trend was improving", i.e., a conversion in the confidence was applied for variables where decreasing trends indicated degradation. Again, this change was to avoid subjectivity associated with the choice of trend directions that are regarded to indicate improvement and degradation.

The statistical power of state and trend analyses and the degree to which lakes in the analyses represented all lakes in New Zealand were limited by the small sample sizes (i.e., the number of lakes with sufficient data). These limitations also applied to previous analyses of national-scale state and trends in New Zealand lake water quality. The small numbers of lakes result from the scarcity of lakes in council SoE monitoring networks and the exclusion of some monitored lakes due to inadequate data, as discussed in Sections 3.1 and 3.2. The procedures used in this study to handle censored values prevented some lakes from being excluded, but the number of lakes retained for analysis was still very limited. In the current study, we initially compiled data from 162 sites across 128 lakes. This number was reduced to a maximum of 103 sites at 83 lakes (for state analysis) and 77 sites at 65 lakes (for 10-year trend analysis) after applying inclusion rules about sampling frequency and duration. In previous studies, lake numbers ranged from 112 (Verburg et al. 2010) to 155 (Larned et al. 2018), which indicates that limitations caused by small sample sizes is a long-standing issue.

Three general steps can be taken to alleviate problems caused by the small number of lakes used in national-scale analyses: 1) alter rules about data adequacy to reduce the number of lakes excluded from analyses; 2) increase the number of lakes in council monitoring networks; 3) ensure that all core water-quality variables are measured in all lakes in each council network (Larned and Unwin 2012). We have addressed the first step in the current study, although further modification of inclusion rules may be needed in subsequent studies. Increasing the number of lakes in council monitoring networks is costly, but a small number of new sites that fill major gaps in environmental or geographic coverage would be beneficial for national-scale analyses. For example, there were limited qualifying lakes in many of the eight elevation × depth classes that we used (e.g., Table 4-1 and Table 5-1), and these environments could be prioritised for new lake monitoring sites. Many other approaches can be used to identify high-priority gaps in lake monitoring networks. The increased labour costs of expanded monitoring networks could be minimised through the use of autonomous water quality sensors ('lake monitoring buoys'). Finally, the national network of monitored lakes can be effectively expanded by ensuring that all core variables are measured in all lakes in each monitoring programme. The gaps in coverage of individual variables are indicated by the uneven distribution of lake × variable combinations in Table 4-1 and Table 5-1. CHLA, NH4N, TN and TP are measured at most lakes, whereas NO3N and ECOLI are only measured at a small subset. Ensuring that NO3N, ECOLI and other core variables are measured at all lakes would greatly expand the spatial coverage of lakes for those variables, at a modest additional cost.

7 Acknowledgements

We thank Isaac Bain (Ministry for the Environment), Lance Morell and Sonja Miller (Stats NZ Tatauranga Aotearoa) for project support and feedback. Many thanks to the LAWA Refresh Team [Abi Loughnan (Otago Regional Council), Eric Goodwin (Cawthron), Jane Groom and Coral Grant (Auckland Council)] for data updates and advice, as well as support from Juliet Milne. We thank the many regional council staff who provided additional data and information about monitoring programmes.

8 Glossary of abbreviations and terms

CHLA	Phytoplankton biomass as chlorophyll *a*
DRP	Dissolved reactive phosphorus
ECOLI	*Escherichia coli*
IPCC	Intergovernmental Panel on Climate Change
LAWA	Land Air Water Aotearoa
MfE	Ministry for the Environment
NH4N	Ammoniacal nitrogen
NO3N	Oxidised or nitrate-nitrogen
NOF	National Objectives Framework
NPS-FM	National Policy Statement for Freshwater Management
RSSE	Relative Sen slope estimator
RSSSE	Relative seasonal Sen slope estimator
Seechi	Secchi disk
SoE	State of the Environment
SSE	Sen slope estimator
SSSE	Seasonal Sen slope estimator
TLI	Trophic level index
TN	Total nitrogen
TP	Total phosphorus

9 References

ANZECC & ARMCANZ (2000) 1 Australian and New Zealand Guidelines for Fresh and Marine Water Quality *Australian and New Zealand Guidelines for Fresh and Marine Water Quality: The Guidelines*. Canberra, ACT.

Ballantine, D.J., Booker, D.J., Unwin, M.J., Snelder, T.H. (2010) *Analysis of national river water quality data for the period 1998–2007*. NIWA Client Report CHC2010–038 prepared for the Ministry of the Environment. NIWA, Christchurch.

Burns, N., Bryers, G., Bowman, E. (2000) Protocol for monitoring lake trophic levels and assessing trends in trophic state. *Client Report: 99/2 prepared for Ministry for the Environment*, : 130.

Davies-Colley, R.J., Smith, D.G. (2001) Turbidity, suspended sediment, and water clarity: a review. *Journal of the American Water Resources Association*, 37: 1085–1101.

Gregor, J., Maršálek, B. (2004) Freshwater phytoplankton quantification by chlorophyll a: a comparative study of in vitro, in vivo and in situ methods. *Water Research*, 38: 517–522.

Helsel, D.R. (2012) *Statistics for Censored Environmental Data Using Minitab and R*. Second edi. Wiley.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J. (2020) *Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, Book 4, Chapter A3*. U.S. Geological Survey, 458 pp.

Hickey, C.W. (2014) *Derivation of indicative ammoniacal nitrogen guidelines for the National Objectives Framework*. NIWA Memorandum MFE13504. NIWA, Hamilton.

Hirsch, R.M., Slack, J.R., Smith, R.A. (1982) Techniques of trend analysis for monthly water quality data. *Water Resources Research*, 18: 107–121.

Horowitz, A.J. (2013) A Review of Selected Inorganic Surface Water Quality-Monitoring Practices: Are We Really Measuring What We Think, and If So, Are We Doing It Right?. *Environmental Science & Technology*, 47: 2471–2486.

Larned, S.T., Snelder, T.H., Unwin, M.J., McBride, G.B. (2016) Water quality in New Zealand rivers: current state and trends. *New Zealand Journal of Marine and Freshwater Research*, 50: 389–417.

Larned, S.T., Snelder, T.H., Unwin, M.J., McBride, G.B., Verburg, P.H., McMillan, H.K. (2015) CHC2015-03 Prepared for Ministry for the Environment *Analysis of water quality in New Zealand lakes and rivers*. NIWA Client Report CHC2015-033 prepared for Ministry for the Environment. NIWA, Christchurch.

Larned, S.T., Snelder, T.H., Whitehead, A.L., Fraser, C. (2018) *Water quality state and trends in New Zealand lakes: Analyses of national data ending in 2017*. NIWA Client Report 2018359CH prepared for the Ministry for the Environment. NIWA, Christchurch.

Larned, S.T., Unwin, M.J. (2012) *Representativeness and statistical power of the New Zealand river monitoring network. National Environmental Monitoring and Reporting:*

Network Design, Step 2. NIWA Client Report CHC2012-079 prepared for Ministry for the Environment. Christchurch.

Mann, H.B. (1945) Nonparametric Tests Against Trend. *Econometrica*, 13: 245.

McBride, G.B. (2005) *Using Statistical Methods for Water Quality Management: Issues, Problems and Solutions (Vol. 19)*. John Wiley & Sons.

New Zealand Government (2020) *National Policy Statement for Freshwater Management 2020*. Ministry for the Environment, Wellington.

Patton, C.J., Kryskalla, J.R. (2003) *Evaluation of alkaline persulfate digestion as an alternative to Kjeldahl digestion for determination of total and dissolved nitrogen and phosphorus in water*. Resources Investigations Report 3-4174.

Randall, D.J., Tsui, T.K.N. (2002) Ammonia toxicity in fish. *Marine Pollution Bulletin*, 45: 17–23.

Snelder, T.H., Fraser, C., Larned, S.T., Whitehead, A.L. (2021) *Guidance for the Analysis of Temporal Trends in Environmental Data*. NIWA Client Report 2021017WN. NIWA, Christchurch.

Sorrell, B., Unwin, M.J., Dey, K., Hurren, H. (2006) *A Snapshot of Lake Water Quality*. NIWA, Christchurch.

Stocker, T., Qin, D.Q., Plattner, G.K. (2014) *Climate Change 2013: The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Verburg, P.H., Hamill, K., Unwin, M.J., Abell, J. (2010) *Lake water quality in New Zealand 2010: Status and trends*. NIWA Client report HAM2010 prepared for Ministry for the Environment. NIWA, Hamilton.