Water quality state and trends in New Zealand lakes

Analyses of national data ending in 2017

Prepared for Ministry for the Environment

December 2018

Prepared by:
Scott Larned
Ton Snelder (LWP)
Amy Whitehead
Caroline Fraser (LWP)

For any information regarding this report please contact:

Scott Larned
Chief Scientist – Freshwater and Estuaries

+64-3-343 7834
scott.larned@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: 2018359CH
Report date: December 2018
NIWA Project: MFE18502

Quality Assurance Statement		
	Reviewed by:	Graham McBride
	Formatting checked by:	Janine Pineda-Verhoogh
	Approved for release by:	Clive Howard-Williams

Contents

Tables

Figures

Executive summary

Introduction

The New Zealand Ministry for the Environment (MfE) and Statistics New Zealand (Stats NZ) 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. MfE have commissioned national-scale analyses of lake water quality data periodically since 2006. The current study was commissioned to analyse water quality state and trend for the period ending in late 2017.

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 policy-makers. In this study, we aggregated the results for individual lakes into lake elevation × depth classes.

The brief for this work consisted of six major steps:

- Compile lake water quality data from regional councils and Land and Water Aotearoa (LAWA).
- Organise and process the data, including error correction, application of reporting conventions and links to spatial data for each lake.
- Assess the suitability of data for 11 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.
- Carry out analyses of water-quality state, including comparisons of state at lakes aggregated by elevation × depth classes.
- Carry out trend analyses using 10-, 20- and 28-year periods ending in late 2017, including comparisons of trends in elevation × depth classes.
- 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 assess lakes against attribute states that are set out in the National Policy Statement for Freshwater Management of 2014 (NPS-FM), and in the 2017 amendments to the NPS-FM (New Zealand Government 2014, 2017).

Methods

Data acquisition and processing. We used three procedures to acquire updated data: interrogation of data servers operated by individual regional councils and LAWA, requests to LAWA data managers for the most recent (2017) data, and direct requests to councils for data that were unavailable through data servers or LAWA.

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 the Ministry), nine of the 11 monitoring variables assessed were retained for analysis: Secchi depth (SECCHI), 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 2013-2017 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 determined the ECOLI attribute state for lakes in the dataset and the number of lakes in which the NPS-FM bottom-lines for CHLA, TN, and TP were exceeded.

Trend analyses. The trend analyses utilised data for the nine variables listed above for the 10-, 20- and 28-year periods ending in December 2017. Trend analyses were based on estimates of trend magnitude from the Sen slope estimator, and estimates of the confidence in the trend direction, which were made using P-values from Kendall tests. In our previous national-scale water quality trend analyses (Larned et al. 2015), 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 in two ways. The first approach was the same used in the previous report (Larned et al. 2015). This approach is conservative because improving and degrading trend categories are reserved for trends where the 90% confidence intervals exclude zero. The newer approach classified trends into nine confidence categories on basis of the probability that a given trend is improving. The categories range from "virtually certain" (probability 99-100%) to "exceptionally unlikely" (probability 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 number of lakes in each of the nine confidence categories described above. The second approach was to estimate the proportion of improving trends (PIT), 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 seven and 63 lakes met the inclusion rules for analyses of nutrient, SECCHI, CHLA and TLI state. The low elevation, shallow lake class had the lowest median SECCHI and the highest median CHLA, NH4N, TN, TP and TLI levels. Median SECCHI 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. Median ECOLI concentrations and the medians of lake G260 and G540 values were highest in the low-elevation, shallow lake class. The proportions of lakes that exceeded bottom lines for CHLA, TN, TP were generally highest in the low-elevation, shallow class. Only six lakes met the sample-size requirement for assigning ECOLI attribute states; four of these lakes were in the A band, and two were in the B band.

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 classes, only 1-6 sites were available for trend analyses, which made estimates of median trend magnitudes unreliable. For classes with more lakes, trend magnitudes were generally less than 2% per year in the 10-year period and less than 0.5% per year in the 20- and 28-year periods.

The 10-year PIT statistics ranged from 39-88%. Five variables (CHLA, DRP, NH4N, NO3N and TP) had a majority of improving (i.e., >50%) trends at the 95% confidence level. The remaining variables had 95% confidence intervals for the PIT that included 50% (ECOLI, SECCHI, TLI3, TN), and we could not infer widespread degradation or improvement in those variables. The number of lakes that qualified for 20-year and 28-year analyses of trends in each variable were small (0-21 sites), which raises the possibility that lakes used to calculate the PIT statistics were poorly representative of all New Zealand lakes.

Discussion

We recommend adopting the approaches set out in this report to increase the information yield from trend analyses, and ultimately, from regional council and national monitoring programmes. We recognise that progressive changes in data analysis methods can impede comparisons between consecutive reports. To alleviate that problem, we provided results of trend analyses using both the methods of Larned et al. (2015) and the new methods, and we recommend presenting the results in parallel as we have in the current report. Finally, we note that the current report does not represent the last word in water-quality data analysis; further advancements are inevitable and beneficial.

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. The procedures used to handle censored values prevented some lakes from being excluded, but the number of lakes retained for analysis was still very limited. 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

MfE and Stats NZ 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 SoE monitoring programmes. In a previous report for MfE, we provided water quality state and trends based on monitoring data from 156 monitored lakes; the time-series for each lake × variable combination had an ending date in December 2013 (Larned et al. 2015). 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 2017.

The brief for this work consisted of six major steps:

- Compile lake water quality data from regional councils and LAWA.
- Organise and process the data, including error correction, application of reporting conventions and links to spatial data for each lake.
- Assess the suitability of data for 11 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.
- Carry out analyses of water-quality state, including comparisons of state at lakes aggregated by elevation × depth classes.
- Carry out trend analyses using 10-, 20- and 28-year periods ending in late 2017, including comparisons of trends at lakes aggregated by elevation × depth classes.
- 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 assess lakes against attribute states that are set out in the NPS-FM of 2014, and in the 2017 amendments to the NPS-FM (New Zealand Government 2014, 2017). We determined the *Escherichia coli* attribute state for individual lakes and determined the number of lakes at which the NPS-FM bottom-lines for phytoplankton biomass, total nitrogen and total phosphorus concentrations were exceeded. There are also NPS-FM bottom-lines for planktonic cyanobacteria and pH-adjusted ammoniacal-nitrogen. However, there were insufficient data for state analyses of planktonic cyanobacteria and insufficient pH data for calculating pH-adjusted ammoniacal-nitrogen concentrations.

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 of 2014[1], and in the 2017 amendments to the NPS-FM[2]. 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.

The methods used in the current study include three advances on our previous national-scale water-quality analyses: 1) modified statistical procedures were used to determine the directions of trends (and associated confidence) and the magnitudes of trends; 2) a new approach was used to express confidence in the direction of trends and 3) a new procedure was used to make probabilistic estimates of the PIT for each variable within a given domain (e.g., an environmental class), based on the likelihood that water quality was improving for each lake representing that domain.

[1] http://www.mfe.govt.nz/publications/fresh-water/national-policy-statement-freshwater-management-2014
[2] New Zealand Government (2017). National Policy Statement for Freshwater Management 2014 (amended 2017). New Zealand Government, Wellington.

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., Sorrell et al. 2006, Verburg et al. 2010, Larned et al. 2015). In the current project, the lake monitoring database used for the preceding national-scale report (Larned et al. 2015) was updated with data that had been collected between 2013 and December 2017. 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, 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
Chemical	Ammoniacal nitrogen	NH4N	mg m^{-3}
	Oxidised nitrogen	NO3N	mg m^{-3}
	Total nitrogen (unfiltered)	TN	mg m^{-3}
	Dissolved reactive phosphorus	DRP	mg m^{-3}
	Total phosphorus (unfiltered)	TP	mg m^{-3}
Biological	Chlorophyll *a*	CHLA	mg L^{-1}
	Escherichia coli	ECOLI	cfu 100 ml^{-1}
	Trophic Level Index	TLI	unitless

Secchi depth (referred to as SECCHI) 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. SECCHI 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 macrophtyes. 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 SECCHI data in addition to CHLA, TN and TP concentrations. However, SECCHI data were not available for all lakes in the current study. Moreover, SECCHI data are strongly influenced by factors that are independent of trophic state, such as fine glacial sediment and tannins. We used the TLI3 to maximize the number of sites in the TLI dataset. To ensure consistent calculations, we calculated TLI 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.

In addition to the nine variables listed in Table 2-1, we assessed the availability of data needed to analyse state and trends in planktonic cyanobacteria and bottom-water dissolved oxygen concentration. The data for bottom-water dissolved oxygen come from dissolved oxygen concentrations measured at the deepest point in each lake where oxygen-depth profiles are included in monitoring programmes. After assessing the number and geographic distribution of oxygen-depth profiles and planktonic cyanobacteria measurements in regional council datasets, and consultation with MfE, these variables were omitted from further analysis. Several regional councils had no corresponding data and most of the remaining council datasets comprised few sites or did not meet the sampling frequency and duration criteria we applied.

As noted in Section 1, we used attributes for lakes that have been defined by the NPS-FM to provide context to the water quality state analyses. Nationally applicable attribute states or bands are provided in the NPS-FM for five of the nine variables used in the current report: 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, TP refer to annual medians alone. The bands for CHLA and NH4N specify annual median and annual maximum values. The bands for ECOLI were updated in the 2017 amendments to the NPS-FM to include five bands that are defined using four statistics: median, 95th percentile, percent of samples exceeding 260 cfu 100 ml^{-1}, and percent of samples exceeding 540 cfu 100 ml^{-1} (New Zealand Government 2017).

2.2 Data acquisition

Lake water-quality monitoring data have been acquired periodically from regional councils for recent national scale analyses for MfE (Sorrell 2006, Verburg et al. 2010, Larned et al. 2015). 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: interrogation of data servers operated by individual regional councils and LAWA, requests to LAWA data managers for the most recent (2017) data, and direct requests to councils for data that were unavailable through data servers or LAWA. Regional council data servers (e.g., Hilltop and KiWIS servers) were interrogated using purpose-written R scripts to download water quality data for all available site × variable combinations. We used the data acquired through these three procedures to update the dataset used for the previous national-scale analysis (Larned et al. 2015). 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.

<u>Step 1. Reporting conventions</u>. The water-quality data received from councils and LAWA varied widely 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).

<u>Step 2. Monitoring site spatial information</u>. The following spatial data were associated with each lake monitoring site: site name, location and regional council identifier (if available), NZMS260 grid reference (converted from NZTM as necessary).

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. 2015).

Step 3. Reducing multiple measurements into single values for each lake x sampling date x variable combination. This step was used to process data from multiple sampling sites within lakes, 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/SECCHI data.

At the site level, we used the available metadata for each lake to flag non-representative sites. The sites excluded were generally located at or near a shoreline (rather than in the main body of the lake) or in secondary arms and outlets of larger lakes. Most lakes had at least one suitable monitoring site, but twelve (including all eight lakes in the Otago region: Lakes Dunstan, Hawea, Johnson, Onslow, Tuakitoto, Waihola, Wakatipu, and Wanaka) were left with no suitable sites and dropped out of our working dataset. The final dataset represented 213 monitoring sites on 155 lakes, of which 105 lakes were represented by a single site; 43 by two sites; 6 by three sites and 1 by four sites. For the 50 lakes with multiple monitoring sites, we averaged across sites within dates. In view of the monitoring-site averaging, the remainder of this report refers to 'lakes', and not to 'monitoring sites'.

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 in Larned et al. (2016). 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 (Patton et al. 2003, Horowitz 2013). Therefore, only TN and TP measured by the persulfate digestion method with unfiltered samples were retained for analysis. SECCHI 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 155 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
SECCHI	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 and TP	Procedures retained for CHLA, TN and TP

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 SECCHI 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 Time period for state analyses

The statistical robustness of the determinations of water quality state depends on the variability in the measurements between sampling occasions and the number of observations. This is particularly important for lakes that are close to a threshold associated with a water quality guideline or attribute state because the confidence that the assessment of state is ‘correct’ (e.g., that the lake has been correctly classified as either passing or failing a guideline) increases as the number of observations increase. As a general rule, the rate of increase in the confidence with which estimates of population statistics may be determined decreases as sample sizes increase above 30 (i.e., there are diminishing returns on increasing sample size with respect to confidence above this sample number; McBride 2005).

In this study, a period of five years represented a reasonable trade-off for most of the targets because it yielded a sample size of 30 or more for many lake × variable combinations. The five-year period for the state analyses is consistent with the 2009-2013 period used in the previous national water-quality state analyses (Larned et al. 2015). Because water quality data tends to be seasonal, 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 lake × variable combinations in the state analyses to those with measurements for at least 80% of the years (four out of five years) and at least 80% of the seasons in the period (either 48 of 60 months, or 16 of 20 quarters). Lake × variable combinations that did not comply with this rule were excluded from the state analysis. Note that the inclusion rule used in this study was slightly more stringent than the inclusion rule used in the previous study (Larned et al. 2015). In the 2015 study, each lake × variable combination was required to have at least one measurement in four out of five years, and at least 16 samples, with no restriction on the way the 16 samples were distributed across the five-year period.

3.1.2 Censored values in state analyses

Censored values were replaced by imputation for the purposes of calculating the state 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 it can accommodate multiple censoring limits. 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 where fewer than 24 total observations existed, censored values above the detection limit were replaced with 1.1 times the detection limit.

3.1.3 Calculation of percentiles

For each lake × variable combination, we characterised the current state using percentiles (5th, 20th, 25th, 50th, 75th, 80th, 95th) of the distribution of measured values for the period 2013 to 2017 (inclusive). All percentiles were calculated using the Hazen method.[3] For each water quality variable, the distribution of median values across multiple lakes were evaluated by aggregating lakes into the elevation × depth classes.

3.2 Water quality trend analyses

3.2.1 Sampling dates, seasons and time periods for analysis

Separate trend analyses were carried out for each water quality variable × lake combination that met the inclusion rules set out below, for three different time periods: 10, 20 and 28 years. Each time-period ended in late December 2017.

The processed lake dataset had variable starting and ending dates, variable sampling frequencies, and variable numbers of missing values. Inclusion rules (i.e., filtering rules) were used to ensure that for each variable, the data for each lake would provide a robust assessment of the trend. We used the filtering rules suggested by Helsel and Hirsch (1992), which restricted lake × variable combinations for trends in a given time period such that there were measurements for at least 90% of the years and at least 90% of seasons.

We used seasons defined by months preferentially in the trend analyses, and quarters when there were insufficient monthly observations. The trend analysis procedure accounted for seasonal variability in these monthly and quarterly data. We note that when there is more than one sample in a month or quarter, all samples can be used in a trend analysis resulting in increased statistical power and potentially different results. However, because our analyses are used to make regional comparisons and to contribute to spatial models, we elected to ensure that the lake-specific analyses had consistent statistical power.

3.2.2 Analyses of lake-specific trends

<u>Trend magnitude and confidence in trend direction</u>

The statistical analyses of water quality trends were performed using the LWP-Trends library, which comprises functions coded in the R statistical programming language. The statistical analyses of trends involves the evaluation of (1) the magnitude of the trend and (2) the confidence in the trend direction.

Trend magnitude was characterised by the Sen slope estimator (SSE; Hirsch *et al.,* 1982). The SSE is the slope parameter of a non-parametric regression, which 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 3-1).

[3] (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).

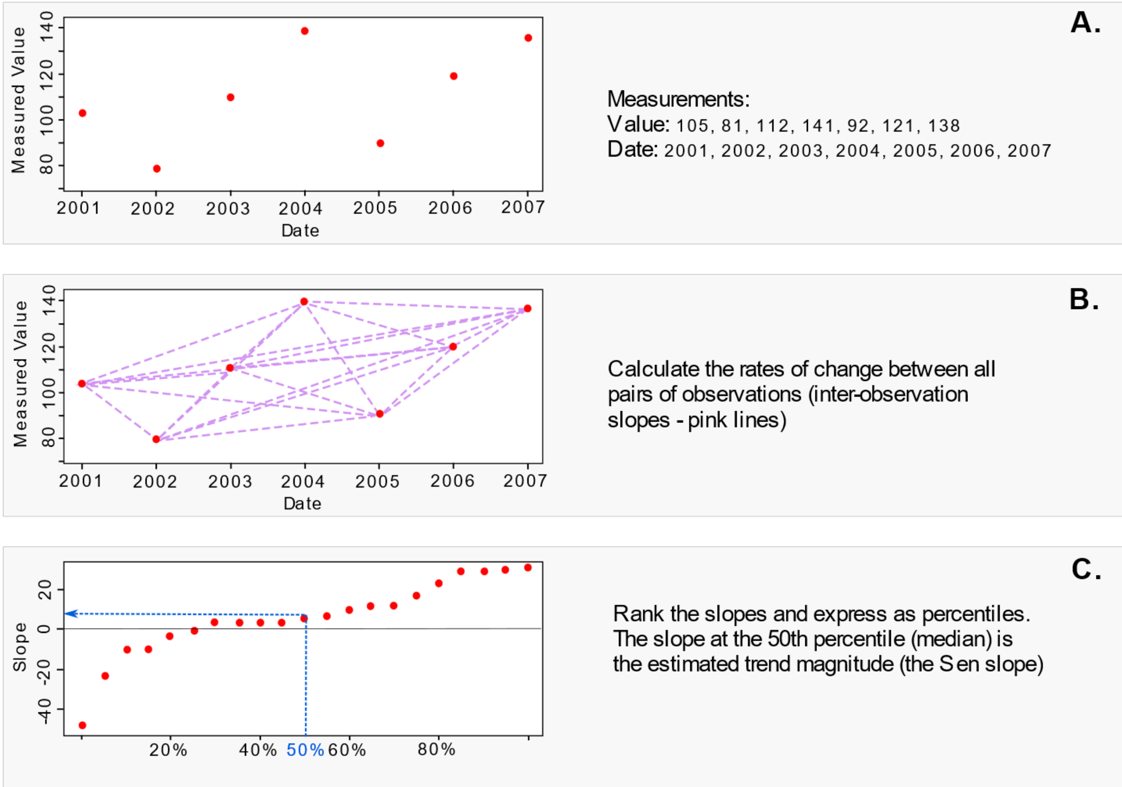

Figure 3-1: Pictogram of the steps taken in the trend analysis to calculate the Sen slope, which is used to characterise trend magnitude in the time-series of data for each lake × variable combination.

The seasonal version of the SSE is used in situations where there are significant ($p \leq 0.05$, as evaluated using a Kruskall Wallis test) differences in water quality measurements between 'seasons'. As noted above, seasons are defined primarily by sampling intervals, which were monthly or quarterly for all variables. The seasonal Sen slope estimator (SSSE) is the median of all inter-observation slopes within each season. Trend magnitudes for the variables measure at monthly or quarterly intervals that demonstrated significant seasonality were estimated with SSSE, and trend magnitudes in all other lake x variable combinations were analysed with SSE.

The Kendall test S and *p*-values are used by the LWP-Trends library to establish confidence in the trend direction (rather than using the Sen slope and its confidence intervals as used by Larned et al. 2015; the reasons for which are related to treatment of censored values and discussed in the following section). The Kendall test measures the rank correlation, which is a nonparametric correlation coefficient measuring the monotonic association between two variables, x and y. In water quality trend analysis, y is a sample of water quality measurements and x is the corresponding sample dates. Traditionally, the Kendall test is used to determine whether trends are statistically "significant" or "insignificant" (see Figure 3-2).

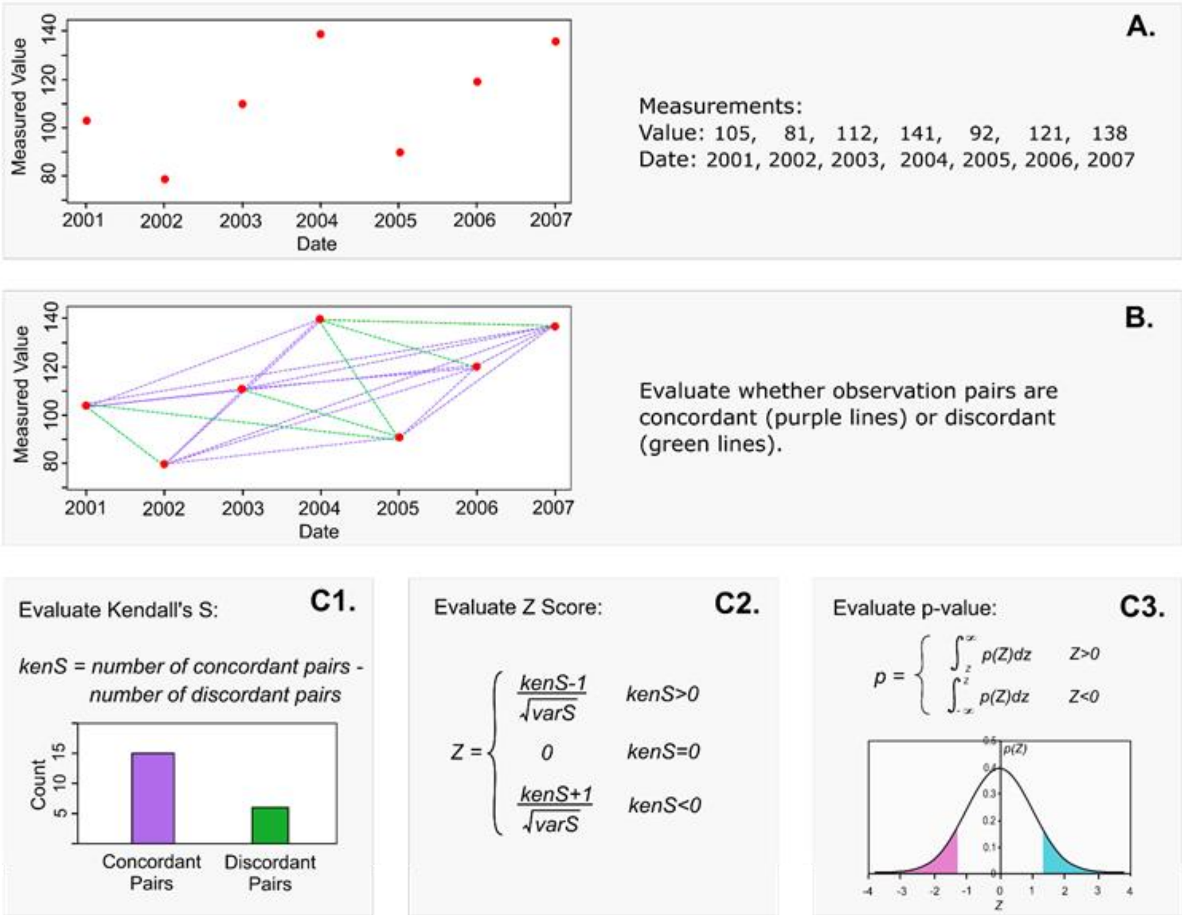

Figure 3-2: Pictogram of the steps taken in the trend analysis to calculate the Kendal S statistic and its p-value, which is used to characterise the confidence in trend direction.

In the LWP-Trends library and in the current report, confidence in the direction of each trend was evaluated by interpreting the Kendall *p*-value as a probability that the trend was decreasing as follows:

$$P(S < 0) = 1 - 0.5 \times pvalue$$

$$P(S > 0) = 0.5 \times pvalue,$$

where $pvalue$ is the *p*-value returned by Kendall test (either seasonal or non-seasonal), *S* is the S statistic returned by Kendall test (either seasonal or non-seasonal) and *P* is the probability that the trend was decreasing. The trend direction is interpreted as decreasing when $P > 0.5$ and increasing when $P < 0.5$. Note that if data are seasonal (i.e., Kruskall Wallis test $p \leq 0.05$), a seasonal version of the Kendall test is used to evaluate the $pvalue$ and P.

The trend direction is established with a 95% level of confidence if the probability associated with $S < 0$ (i.e., a decreasing trend) is ≥ 95%, or the probability associated with $S > 0$ (i.e., an increasing trend) s ≤ 5%. In both, these cases the trend is categorised as 'established with confidence' and when the probability the trend is decreasing is between the 90% confidence limits (i.e., is ≥5% and ≤95%), the trend is categorised as 'indeterminant'.

3.2.3 Handling censored values

Censored values in the data used to calculate Kendall's S and its *p*-value were robustly handled in the manner recommended by Hesel (2005, 2012). Briefly, for left-censored data (i.e., those data reported as less than a limit of detection), increases and decreases in a water quality variable were identified whenever possible. Thus, a change from a censored data entry of <1 to a measured value of 10 was considered an increase. A change from a censored data entry of <1 to a measured value 0.5 was considered a tie, as was a change from <1 to a <5, because neither can definitively be called an increase or decrease. Similar logic applied to right censored values. The information about ties was used in the calculation of the Kendall S statistic and its variance following Helsel (2012) and this provided for a robust calculation of the *p*-value associated with the Kendall test.

Note that as the proportion of censored values increases, the proportion of ties increases and confidence in the trend direction decreases. Therefore, the trend associated with lake × variable combinations with high proportions of censored observations tend to be categorised as indeterminant.

The inter-observation slope cannot be definitively calculated between any combination of observations in which either one or both are censored. Therefore, when SSE and SSSE (i.e., Sen slopes) are calculated by the LWP-Trends library, the censored data entries are replaced by their corresponding raw values (i.e., the numeric component of a censored data entry) multiplied by a factor (0.5 for left-censored and 1.1 for right-censored values). This ensures that any measured value that is equal to a raw value is treated as being larger than the censored value if it is left-censored value and smaller than the censored value if it is right-censored. 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 unaffected 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.

Helsel (2005) estimated that the impact of censored values on the Sen slope is negligible when fewer than 15% of the values are censored. However, this is a rule of thumb and is not always true. Depending on the arrangement of the data, a small proportion of censored values (e.g., 15% or less) could affect the computation of a Sen slope (Helsel 2012). To provide information about the robustness of the SSE and SSSE values, the supplementary output for every trend analysis includes the proportion of observations that were censored and whether the Sen slope (i.e., the median of all inter-observation slopes) was calculated from data containing censored observations. The estimate of the magnitudes decreases in reliability (i.e., the SSE and SSSE values), and confidence intervals of individual lake trends increase in width as the proportion of censored values increases. In addition, when there are censored values, greater confidence should be placed in the statistics returned by the Kendall tests (including the trend direction and the probability the trend was decreasing).

3.2.4 Differences in trend analysis methods to previous reports

The general approach to trend analyses in this study is consistent with the approach used in the most recent national-scale water-quality trend analyses (Larned et al. 2015, 2016). The current and previous studies all assessed monotonic changes in the central tendencies of water quality values through time and used the Sen slope estimator to characterise the magnitude of these changes. The current and previous studies also used evaluations of the confidence in the trend direction in lieu of statistical significance tests; this advancement distinguishes the studies by Larned et al. (2015, 2016) and the current study from earlier national-scale trend analyses (e.g., Ballantine et al. 2010). However, some steps in the trend analysis procedures used in the current study differ from all of the previous studies; most of these differences arise from improved methods for handling censored values.

In the studies by Larned et al. (2015, 2016), confidence in trend directions were evaluated using the Sen slope confidence intervals. If the symmetric confidence intervals around a Sen slope did not contain zero, the trend direction was considered to be established with confidence and the trend was classified as positive or negative. If the symmetric confidence intervals did contain zero, it was concluded that there were insufficient data to determine the trend direction at the nominated level of confidence, and the trend direction was classified as 'indeterminant'. Note that if two symmetric, one-sided 90% confidence intervals do not contain zero, the trend direction is established with 95% confidence, as explained in Larned et al*.* (2015) and McBride (in press). For the same reason, the analysis used in the current study categorises a trend as 'established with confidence' at 95% confidence when the probability that the trend is decreasing or increasing is ≤5% or ≥95% respectively, and as 'indeterminant' when the probability lies between these thresholds that define 90% (not 95%) confidence limits.

We recently identified a problem with the use of Sen slopes and their confidence intervals to make inferences about trend directions. and specifically, the treatment of censored values in confidence intervals. The problem concerns the effects of censored values on the accuracy of Sen slope estimates (as discussed above) and confidence intervals. Analytically the difference between a pair of censored values is not measurable and must be treated as zero, which is referred to as a 'tie'. Similarly, the difference between a measured value that is less than the numeric component of a censored value and that censored value is not measurable[4], and is also considered a tie. Replacement of censored values with imputed values can affect the identification of tied values, which reduces the robustness of the calculations of the confidence interval. While the imputation of censored values by Larned et al. (2015) was not strictly correct, the rule in that study that restricted lake × variable combinations to those with < 15% censored values ensured that imputation *per se* had minimal effects on estimates of trend magnitude or confidence intervals.

The approach used with censored values in the current study has two advantages compared with the previous studies. First, evaluations of confidence in trend directions for individual lakes are more reliable, irrespective of the proportion of censored observations. In turn, the methods used to aggregate lake trends are robust, because these procedures are based on levels of the confidence in the trend directions at individual lakes (discussed in detail in Section 3.2.5). Second, censored values can represent a large proportion of observations for some variables (e.g., DRP, NH4N). The procedures used in the current study reduced the need to exclude lake × variable combinations based on the proportion of censored observations (i.e., lakes with >15% censored values were not excluded as in previous studies). This had the advantage of preserving a larger number of lakes in each analysis and maximising spatial coverage. We did exclude some lake × variable combinations that had < 5 non-censored values and/or < 3 unique non-censored values, because these cases included so many ties that there was insufficient information to calculate Sen slopes and confidence intervals.

[4] An example of the numeric component of a censored value is the figure 0.05 in the data entry "< 0.05 mg L^{-1}".

3.2.5 Aggregation of individual lake trends

The aggregated results of analysis of water-quality trends are intended to provide an overview of recent water quality changes over a spatial domain of interest (e.g., environmental classes, regions, national). In the present study, we aggregated both trend magnitudes and trend directions across lake monitoring lakes. The distributions of trend magnitude across those lakes were characterised using box and whisker plots of the relative seasonal Sen slope estimates (RSSSE). Sen slopes were relativised by dividing the SSSE values by the trend periods to give estimates of temporal change in % yr^{-1}.

We used three different approaches for aggregating trend directions. For each approach, ‘improving trends’ corresponded to decreasing trends in nutrient and ECOLI concentrations and TLI scores, and increasing trends in SECCHI. Conversely, ‘degrading trends’ corresponded to increasing trends in nutrient and ECOLI concentrations and TLI scores, and decreasing trends in SECCHI.

In the previous national-scale water-quality trend analyses, lake-specific trends were aggregated by tabulating the numbers of lakes in three trend-direction categories (i.e., improving, degrading, and indeterminant) for each variable and each domain (Larned et al. 2015). In the current study, we retained the previous approach for continuity, and added two new approaches. The methods for the new approaches are set out below. Detailed descriptions of these approaches and comparisons with the previous approach are provided by Snelder and Fraser (2018).

The first new approach utilises the probability that the true trend was decreasing, which is derived from the Kendall test statistics (see Section 3.2.2). This probability facilitates a more nuanced inference rather than the ‘yes/no’ output corresponding to the trend-direction categories (i.e., increasing, decreasing, and indeterminant (McBride, in press). Confidence categories can be used to express the probability that a trend is improving (or its complement - degrading). Note that the conversion of the probability that a trend is decreasing to the probability it is improving (and its complement, degrading) depends on whether decreasing values represent improvement or degradation.

The confidence categories used in this study were adopted from those recommended by the Intergovernmental Panel on Climate Change (IPCC; Stocker et al. 2014). The categories and corresponding probability ranges are in Table 3-1. Note that confidence categories for degrading trends are the complement of the confidence categories for improving trends shown in Table 3-1, i.e., an "exceptionally unlikely" degrading trend is the same as a "virtually certain" improving trend.

Table 3-1: Level of confidence categories used to convey the probability that water quality was improving.

The same confidence categories are used by the Intergovernmental Panel on Climate Change (Stocker, 2014).

Categorical level of confidence	Probability (%)
Virtually certain	99–100
Extremely likely	95–99
Very likely	90–95
Likely	67–90
About as likely as not	33–67
Unlikely	10–33
Very unlikely	5–10

Extremely unlikely	1–5
Exceptionally unlikely	0–1

The confidence categories in Table 3-1 were used to aggregate the lake-specific trends in each water quality variable. Each lake trend was assigned a confidence category according to its evaluated probability and the categories shown in Table 3-1. The same confidence categories were used to map trends in each water quality variable, and to aggregate the lake-specific trends in each water quality variable at the national scale. We then calculated the proportion of lakes in each confidence category for each variable and summarised the proportions in a colour coded bar chart. Similar graphs were not used to summarise results across REC land-cover classes because the PIT statistics described below is a simpler way to represent aggregated trends across multiple domains.

The second approach also utilises the probability that the true trend was decreasing to provide a probabilistic estimate of the PIT within a domain of interest. For a given water quality variable, the trends at several monitoring lakes distributed across a domain of interest can be assumed to represent independent samples of the population of trends, at all lakes within that domain. Let the sampled lakes 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 and the value 0 with probability $q = 1 - p$. Therefore, $I_s = 1$ denotes an improving trend at lake $s \in \{1, \dots, S\}$ when the estimated $p_s \geq 0.5$ and a degrading trend as 0 when $p_s < 0.5$. Then, the estimated proportion of lakes with improving trends in the domain is:

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

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

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

PIT and its variance represent an estimate of the population proportion of improving trends and the uncertainty of that estimate. It is noted that the proportion of degrading trends is the complement of the result (i.e., 1 - PIT). The estimated variance of PIT can be used to construct 95% confidence intervals[5] around the PIT statistics as follows:

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

We calculated PIT and its confidence interval for all lake water quality variables and for domains of interest defined by the entire country, and by the lake elevation × depth classes defined in Section 2.3. We applied the PIT methodology to all trend periods (10-, 20- and 28-years). However, we note that the number of qualifying lakes for each water-quality variable decreases strongly as trend periods lengthen, and PIT statistics based on small numbers of lakes may be biased (i.e., the lakes used for calculating PIT statistics are not representative of the population within the domain of interest).

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

4 Results – lake state

Between 7 and 63 lakes met the filtering rules for the state analysis of nutrients, SECCHI, 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 generally lower than in the dataset used in the previous analysis of lake state (Larned et al. 2015). The primary reason is that the inclusion rules were slightly more stringent in the current study, to ensure that seasonal variation was accounted for (i.e., by requiring that data were available for 16 of the 20 quarters in the 5-year period). A second reason is that some councils stopped monitoring some variables during the 2013-2017 period used in the current study.

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 and DRP data were particularly sparse, with 7, 15 and 22 qualifying lakes, respectively.

The distributions of lake-medians for the lake water quality variables for the 2013-2017 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 SECCHI and the highest median CHLA, NH4N, TN, TP and TLI levels. Median SECCHI 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. The complete set of state analysis results is provided in the supplementary file "LakeStateResults2013-2017Inclusive_31_Oct18.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 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.

	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	19	19	10	4	NS	NS	7	4	63
DRP	6	11	3	1	NS	NS	1	NS	22
ECOLI	7	NS	4	NS	NS	NS	3	1	15
NH4N	18	19	10	4	NS	NS	7	4	62
NO3N	1	1	3	1	NS	NS	1	NS	7
SECCHI	14	19	8	3	NS	NS	5	3	52
TLI3	18	18	10	4	NS	NS	5	3	58
TN	19	19	10	4	NS	NS	7	4	63
TP	19	19	10	4	NS	NS	7	4	63

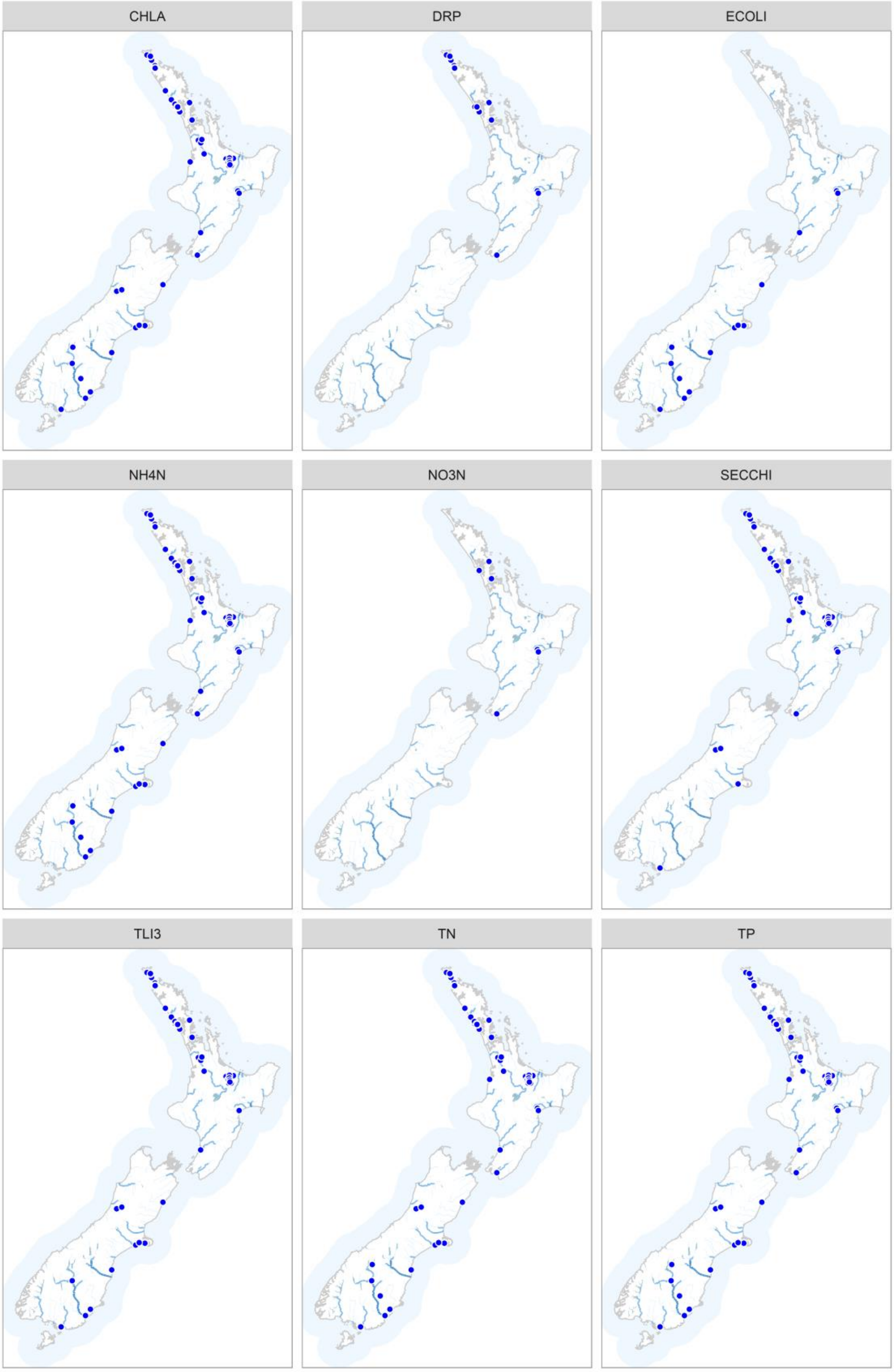

Figure 4-1: Locations of lakes used for state analyses of lake water quality.

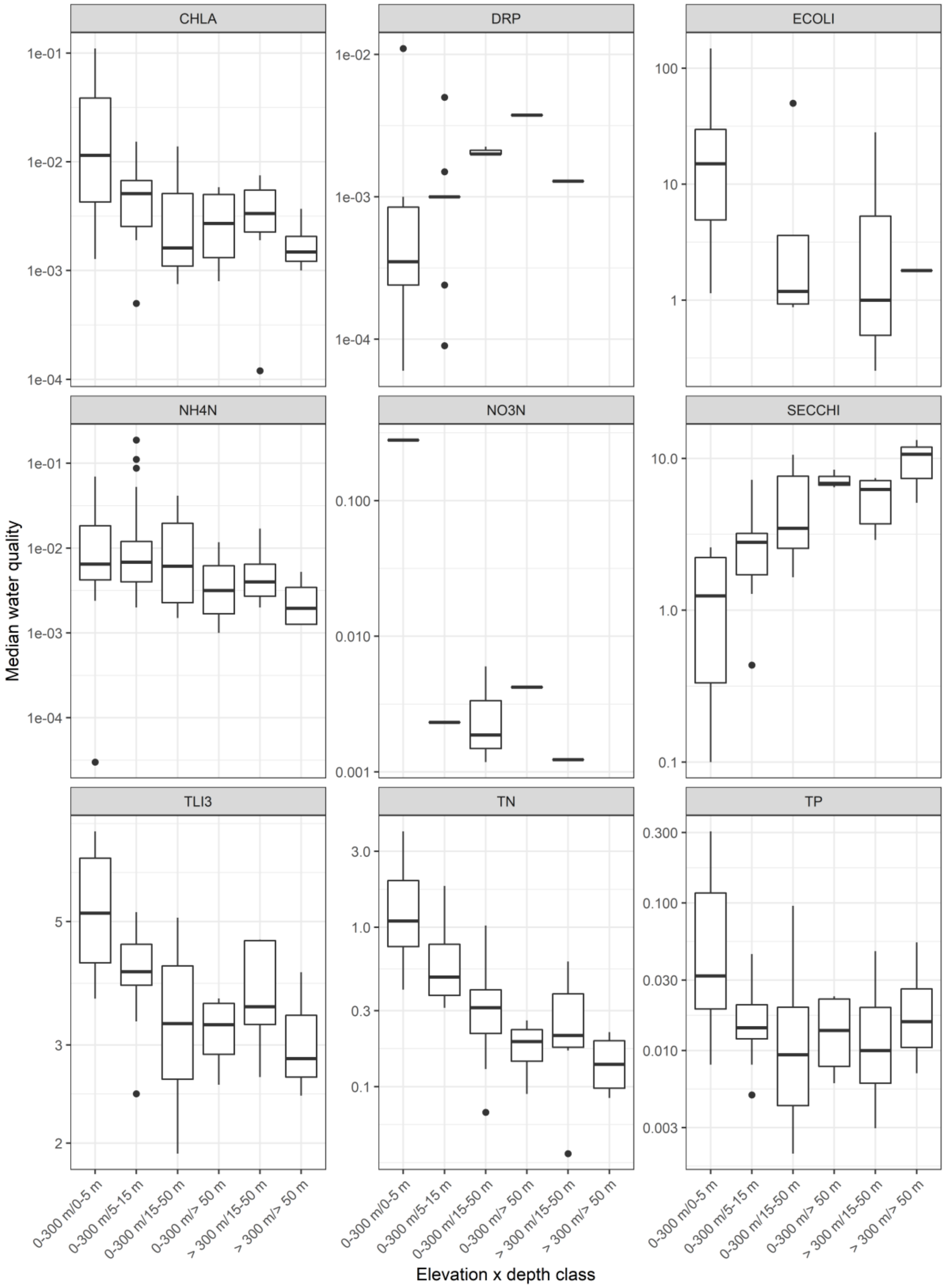

Figure 4-2: Lake water quality state in elevation × depth classes. Box-and-whisker plots show the distributions of lake medians within each class. The black horizontal line in each box indicates the median of lake medians, the box indicates the inter-quartile range and the whiskers indicate the 5th and 95th percentiles. Outliers are indicated by black circles. Note log-scale on Y-axes.

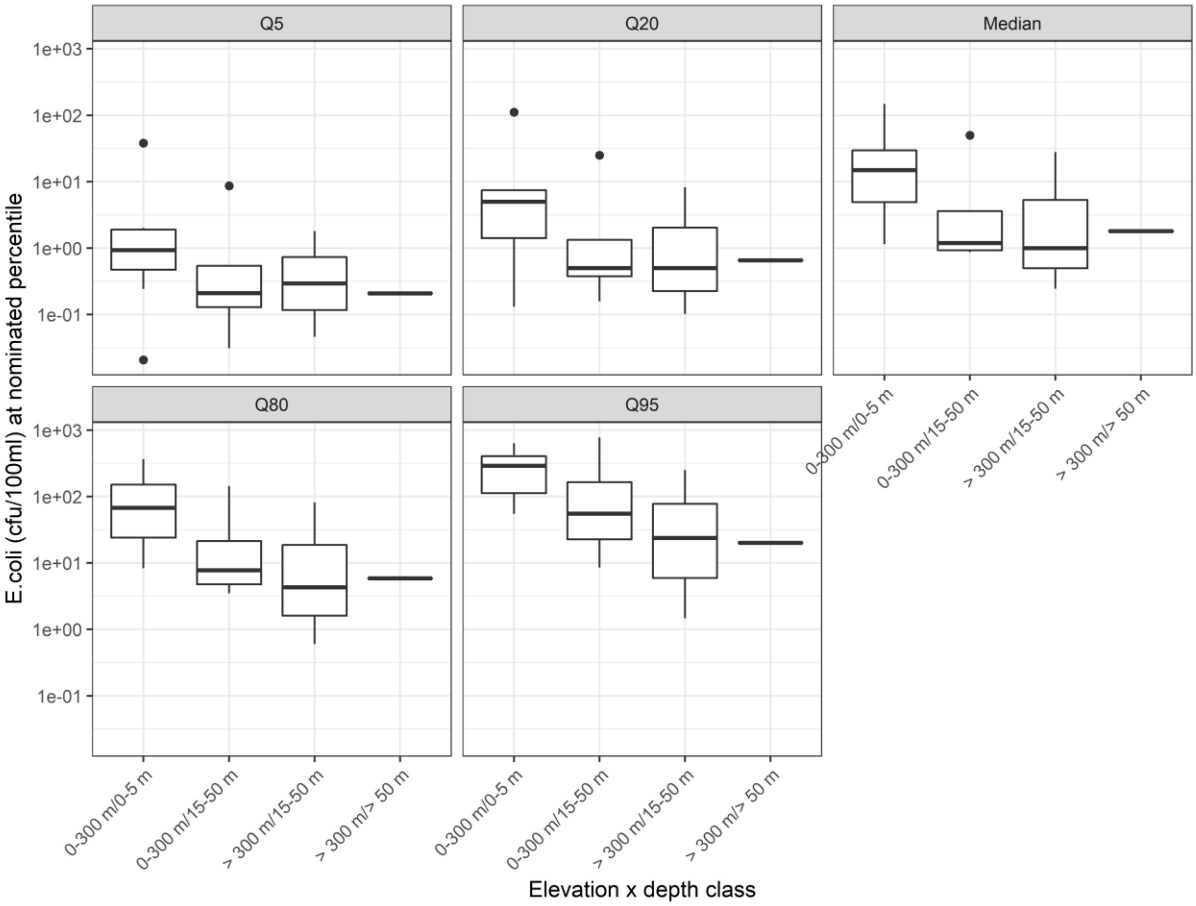

Figure 4-3: ECOLI concentrations in lake elevation × depth classes. Box-and-whisker plots show the distributions of lake percentiles within each class. The black horizontal line in each box indicates the median of lake percentiles, the box indicates the inter-quartile ranges, whiskers indicate the 5th and 95th percentiles, and open circles indicate outliers. Note log-scale on Y-axes.

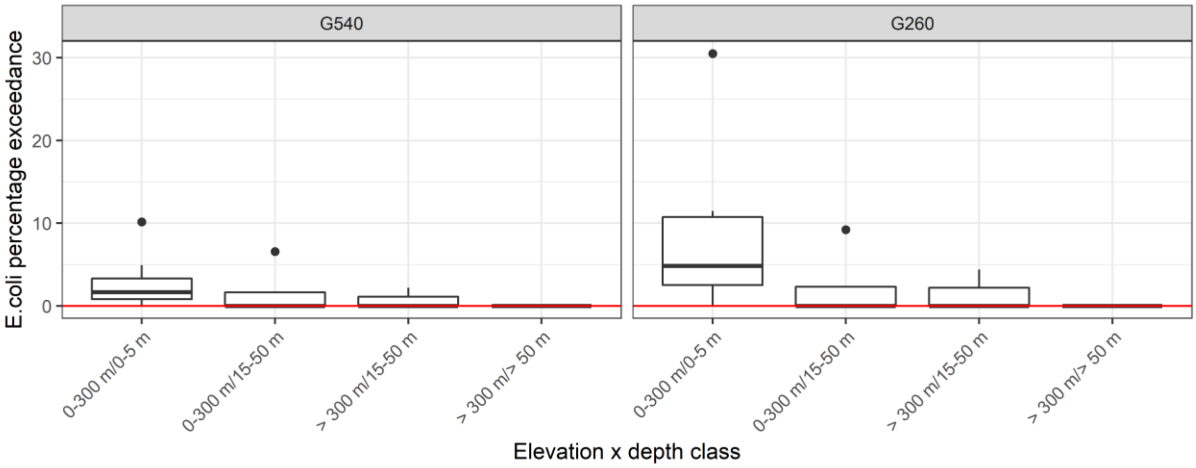

Figure 4-4: ECOLI percent exceedance in lake elevation × depth classes. Box-and-whisker plots show the distributions of percentage exceedance over 540 cfu 100 ml^{-1} (G540) and 260 cfu 100 ml^{-1} (G260) at lakes within elevation × depth classes. Black horizontal lines indicate the median of lake percentiles, boxes indicate the inter-quartile ranges, whiskers indicate the 5th and 95th percentiles, and open circles indicate outliers.

As noted in Section 2.1, the NPS-FM includes attribute bands for five variables included in this report: CHLA, TN, TP, NH4N and ECOLI. For each variable, the C/D band boundary corresponds to the national bottom-line. The CHLA attribute has separate bands for median and maximum concentrations. The TN attribute has separate bands for polymictic lakes and for seasonally stratified and brackish lakes. As noted in Section 2.3, lakes in the 15 – 50 m and > 50 m depth classes are likely to be seasonally stratified and shallower lakes are likely to be polymictic. Maximum depth data were available for most lakes in the dataset, and 15-m maximum depth was used to separate polymictic and stratifying lakes. This approach is consistent with the recommendations of the lake experts panel that informed the NPS-FM. The NH4N attribute concerns ammonia toxicity and specifies pH-adjusted NH4N concentrations. There were too few lakes with synoptic pH and NH4N data to evaluate the proportion of lakes that exceeded the national bottom line for pH-adjusted NH4N. The number and proportion of lakes that exceeded the national bottom lines for CHLA, TN, TP are in Table 4-2. The proportions of lakes that exceed the national bottom lines were highest in the low-elevation shallow class; over half of the low-elevation, shallow lakes in the dataset exceeded the national bottom lines for maximum CHLA and median TN concentrations.

Table 4-2: Number and proportions of lakes that exceeded the national bottom-lines for the NPS-FM CHLA, TN and TP attributes. National bottom line concentrations are in mg m^{-3}.

	Median CHLA	Maximum CHLA	Median TN		Median TP
			Seasonally stratified/brackish	Polymictic	
National bottom line	12	60	750	800	50
Exceedances (all lakes)	12 (19%)	21 (33%)	2 (8%)	17 (45%)	11 (17%)
Exceedances (0-300 m X 0-5 m lakes)	9 (47%)	12 (63%)	0	12 (63%)	9 (47%)

The attribute states or 'bands' for ECOLI have been updated in the 2017 amendments to the NPS-FM to include five states (designated A, B, C, D and E) that are based on four statistics: median, 95th percentile, percent of samples exceeding 260 cfu 100 ml^{-1}, and percent of samples exceeding 540 cfu 100 ml^{-1}. Each statistic has a numeric attribute state that corresponds to each band. The ECOLI attribute states were only determined at lakes with ≥ 60 samples in the 2013-2017 period, as per Footnote 1 in the *Escherichia coli* attribute table in the 2017 NPS-FM. We assigned each of these lakes to an ECOLI attribute state according to the rule in Footnote 2 of the attribute table, "Attribute state must be determined by satisfying all numeric attributes states". Of the 15 lakes where ECOLI was monitored, only six met the sample size criteria, and of these, four lakes were in band "A" and two lakes in band "B".

5 Results – lake trends

5.1 Ten-year trends

Between 8 and 62 lakes 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 were no qualifying lakes in two high-elevation, shallow lake classes, and no qualifying lake variable combinations in several other classes. The qualifying lakes are sparsely and unevenly distributed on both the North and South Islands (Figure 5-1). All lake locations, lake classes and numbers of sampling dates are included in the supplementary file "Trends10YearLake_31_Oct18.csv".

Table 5-1: Number of lakes in the elevation × depth lake classes used in the 10-year trend analyses. The lake numbers shown correspond to lakes that met the inclusion requirements in Section 3.2.1 (measurements are available for at least 90% of the years and at least 90% of seasons). NS: no qualifying lakes.

	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	18	18	11	4	NS	NS	7	4	62
DRP	6	11	6	1	NS	NS	1	NS	25
ECOLI	6	NS	4	NS	NS	NS	3	1	14
NH4N	17	18	11	4	NS	NS	7	4	61
NO3N	1	1	4	1	NS	NS	1	NS	8
SECCHI	14	17	9	4	NS	NS	5	3	52
TLI3	17	16	8	3	NS	NS	6	3	53
TN	18	18	11	3	NS	NS	7	4	61
TP	18	18	11	4	NS	NS	7	4	62

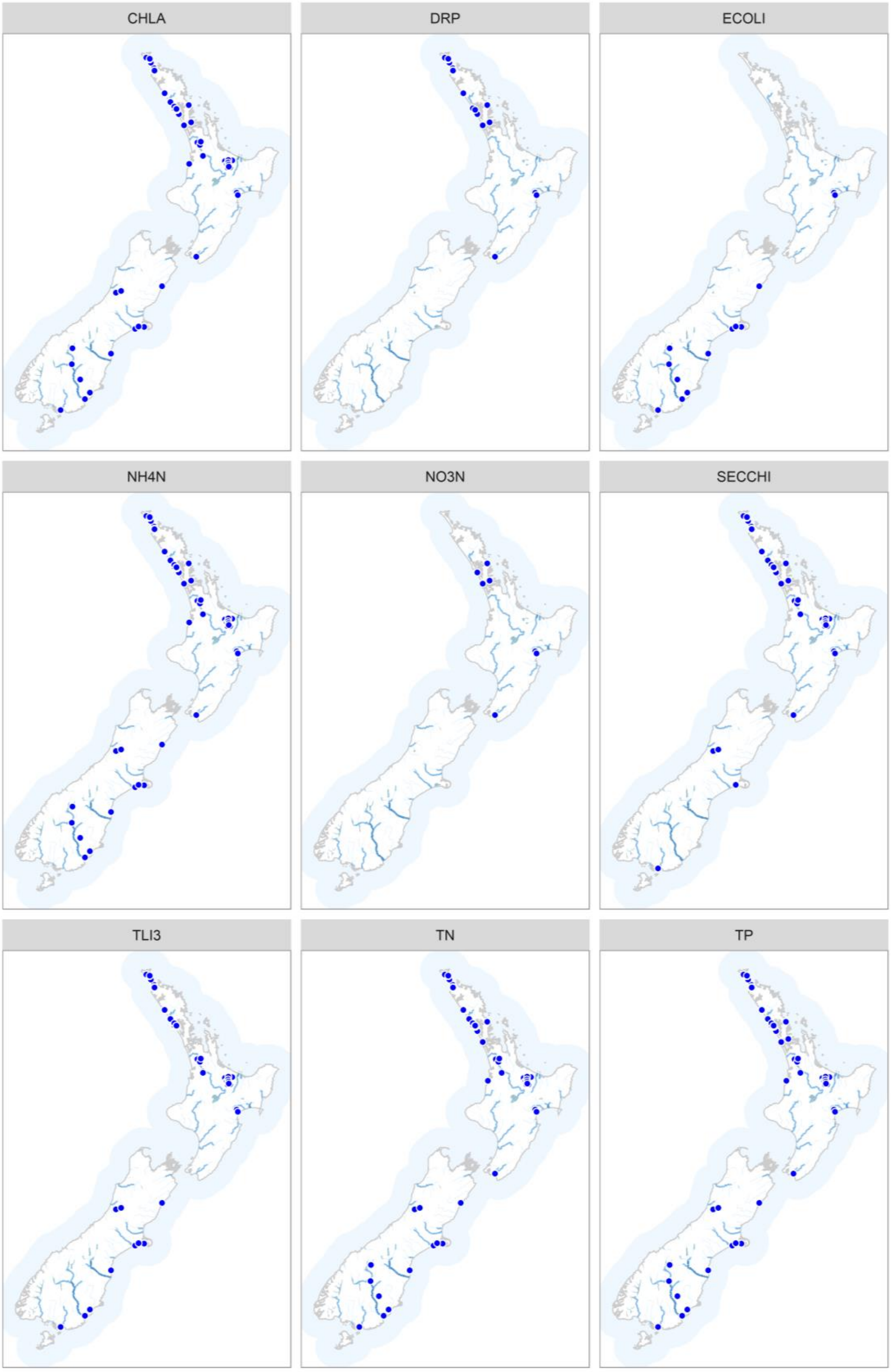

Figure 5-1: Locations of lakes used for 10-year trend analyses of water quality variables.

5.1.1 Trend Magnitude

Box-and-whisker plots were used to summarise the estimated trends for each of the lake water quality variables in each lake elevation × depth class for the 10-year period from 2008 – 2017 (Figure 5-2). All estimated trends are included in these plots, irrespective of the confidence in direction (as defined in Section 3.2.2). 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 (Figures 4-2, 4-3 and 4-4). Median trend magnitudes were largest for DRP, ECOLI, NH4N and NO3N in multiple lake classes (e.g., DRP and NO3N in the 0-300 m elevation, 15-50 m depth class), but the small number of lakes in these classes (1-6 lakes) makes the estimates of class medians unreliable. For classes with more lakes (6-18 lakes in Figure 5-2), trend magnitudes were generally less than 2% per year.

5.1.2 Trend Classification

The numbers and proportions of 10-year trends in four categories are summarised in Table 5-2. A large proportion of the trends for each of the nine variables (32 to 65% of the trends) were classified as “indeterminant”. Degrading, improving and indeterminant categories were used in the previous national-scale trend analysis, and the large proportions of indeterminant trends in the current study is consistent with the previous study (Larned et al. 2015). These results reflect the conservative approach used to infer trend directions; all cases where the 95% confidence intervals around the Sen slope include zero were categorised as indeterminant (Section 3.2.4). For the remaining trends, the proportions of lakes with improving trends in CHLA, and NH4N were substantially larger than the proportion with degrading trends, and the remaining variables all had at least as many improving trends as degrading trends. Only eight lakes qualified for analyses of trends in NO3N, so there is little evidence of a pattern in improving and degrading trends. A large proportion of the DRP lake trends (13 of 25 lakes) were not analysed due to the scarcity of non-censored values.

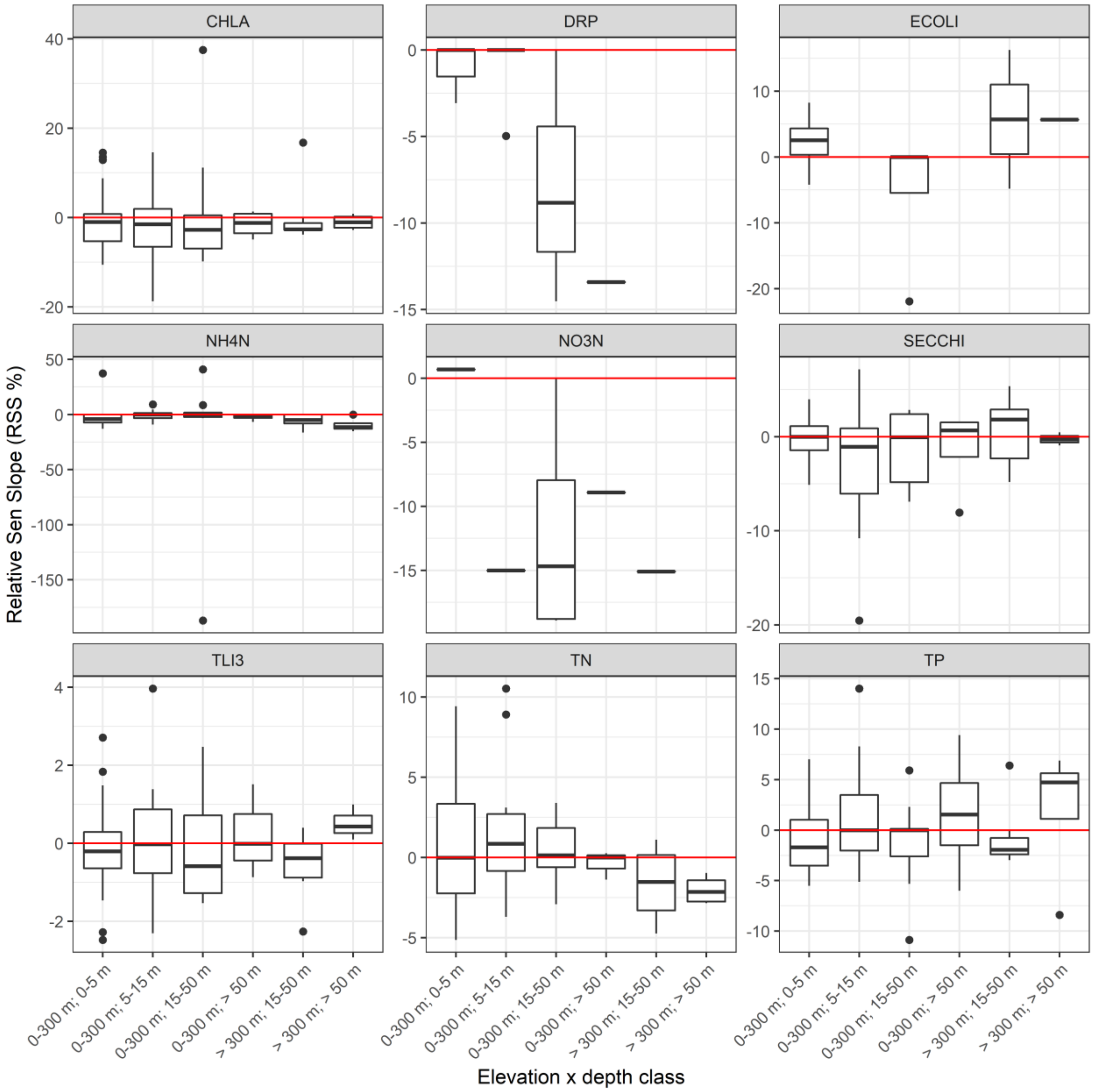

Figure 5-2: Summary of 10-year trend magnitudes in lake water quality variables, in lake elevation × depth classes. Box-and-whisker plots show the distributions of trends in each class. The black horizontal line in each box indicates the median of lake-specific trends, the box indicates the inter-quartile range, whiskers indicate the 5th and 95th percentiles, and open circles indicate outliers.

Table 5-2: Numbers and proportions of lakes in four trend categories for 10-year trends. The “not analysed” category corresponds to lake × variable combinations that met the inclusion requirements in Section 3.2.1, but did not meet the censored data requirements in Section 3.2.4. The classification of the remaining trends into degrading, improving and indeterminant categories follows the approach used in the previous national-scale river water quality trend analysis (Larned et al. 2015).

	Trend category			
Variable	**Degrading**	**Improving**	**Indeterminant (insufficient data)**	**Not analysed**
CHLA	8 (13%)	14 (23%)	40 (65%)	0 (0%)
DRP	0 (0%)	4 (16%)	8 (32%)	13 (52%)
ECOLI	2 (14%)	2 (14%)	9 (64%)	1 (7%)
NH4N	4 (7%)	19 (31%)	36 (59%)	2 (3%)
NO3N	0 (0%)	5 (62%)	3 (38%)	0 (0%)
SECCHI	14 (27%)	10 (19%)	28 (54%)	0 (0%)
TLI3	9 (17%)	13 (25%)	31 (58%)	0 (0%)
TN	14 (23%)	18 (30%)	29 (48%)	0 (0%)
TP	14 (23%)	16 (26%)	32 (52%)	0 (0%)

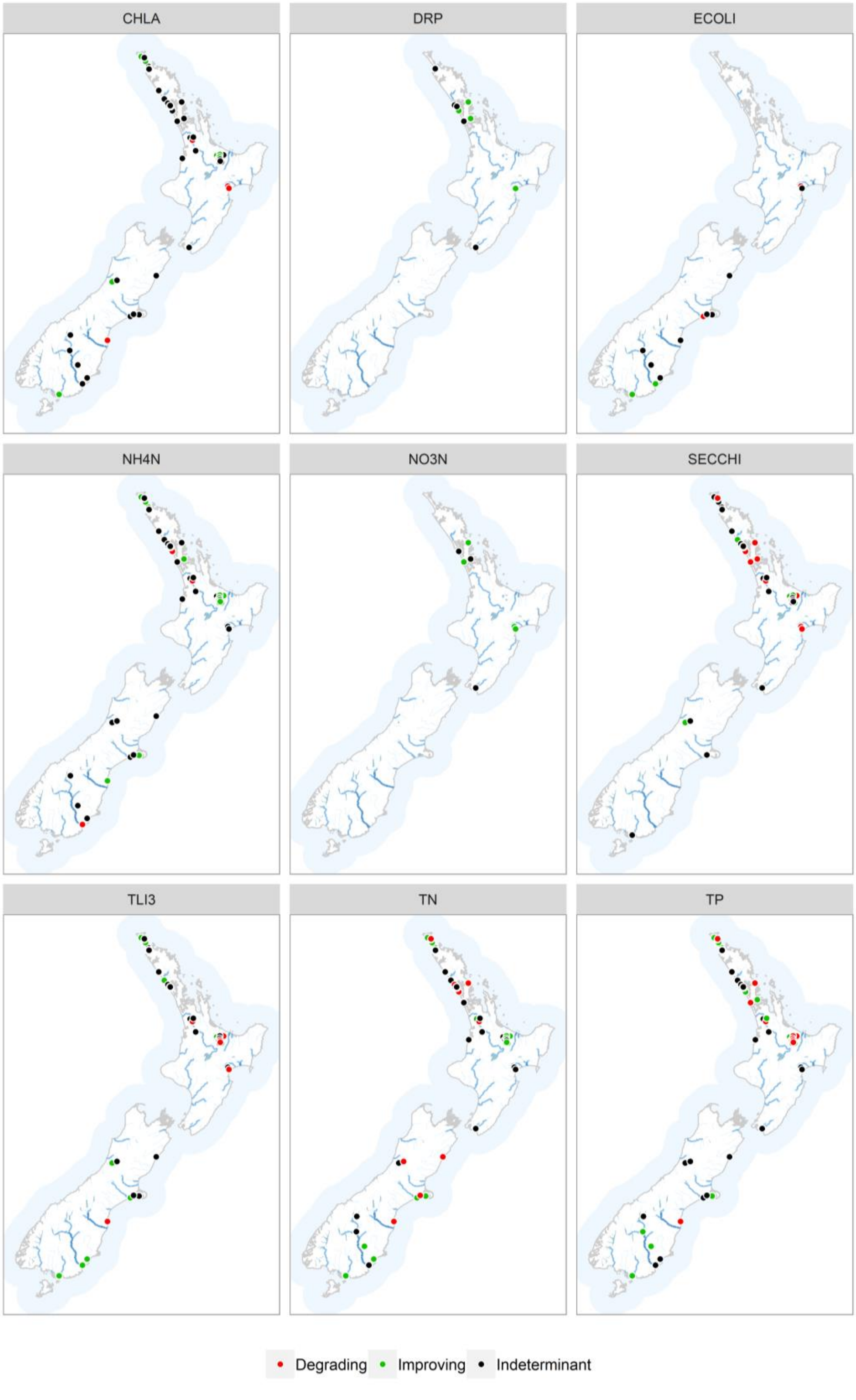

Figure 5-3: Lakes classified by 10-year trend categories (degrading, improving, indeterminant).

5.1.3 Probability of improvement

The levels of confidence listed in Table 3-1 were used to categorise the probability of an improving 10-year trend in each lake × variable combination. The mapped results for individual lakes are shown in Figure 5-4. Because probability of improvement is the complement of the probability of degradation, “unlikely” improvement could also be categorised as “likely” degradation. The maps indicate that for most water quality variables, those lakes previously categorised as indeterminant (shown in Figure 5-3) were about equally divided into likely-to-improve and unlikely-to-improve confidence categories (Figure 5-4). However, in the cases of CHLA, TLI3 and TP, most lakes that were previously categorised as indeterminant were subsequently classed as likely to be improving (i.e., the lakes were subsequently placed in the likely and very likely confidence categories).

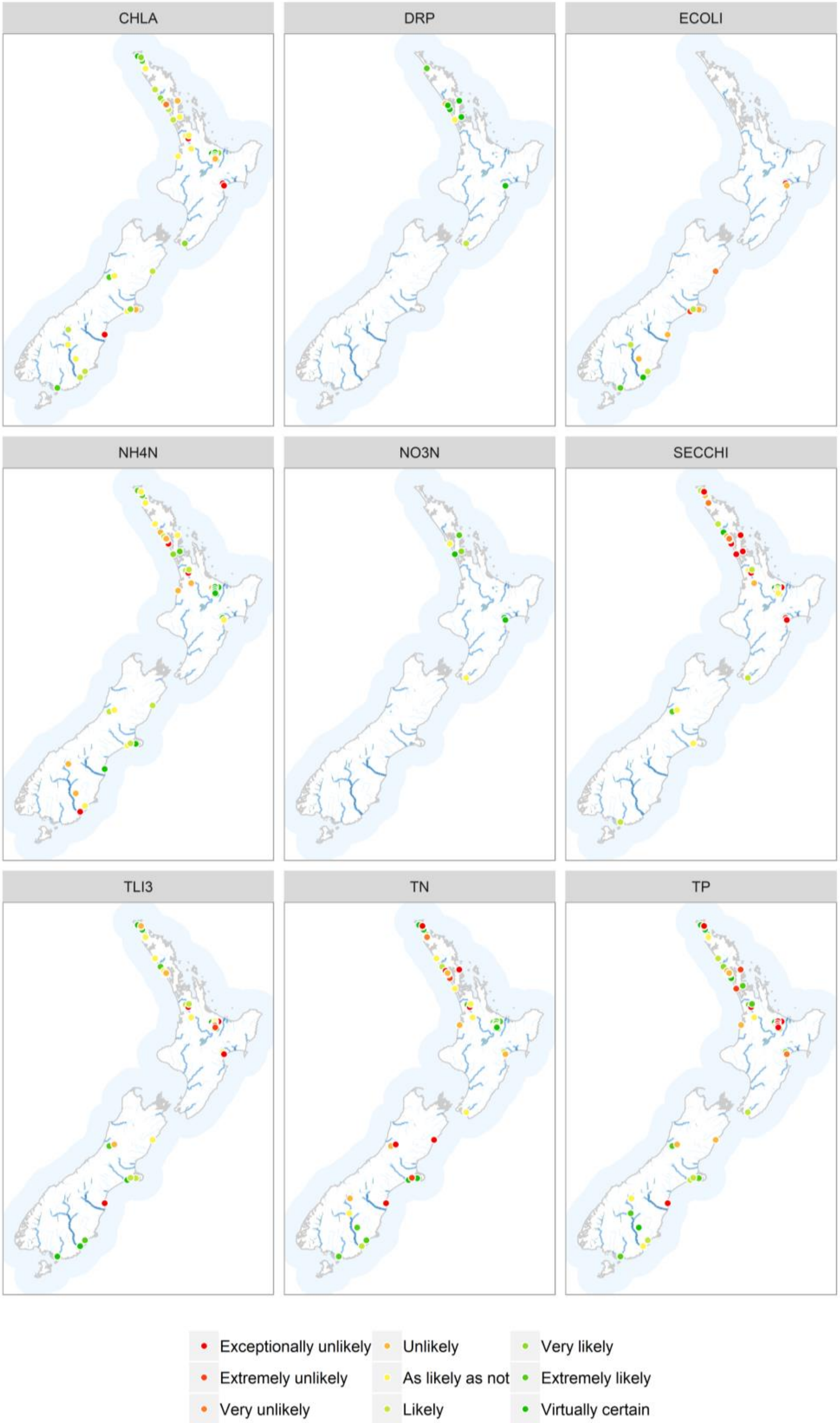

Figure 5-4: Lakes categorised by the probability that the 10-year trend in each variable is improving. Probability of improvement is expressed using the confidence categories in Table 3-1. Only lakes that met the sampling requirements outlined in Sections 3.2.1 and 3.2.5 are shown in the figure.

5.1.4 Aggregate Trends

Figure 5-5 shows the proportions of lakes for which 10-year trends indicated improvement at the nine confidence categories defined in Table 3-1. These plots provide national-scale assessments of the relative proportions of improving versus degrading lakes, based on the relative amounts of green and red in each bar.

The national-scale PIT and their confidence intervals are summarised in Table 5-3. The 10-year PIT statistics ranged from 39-88%. Five of the variables (CHLA, DRP, NH4N, NO3N and TP) had a majority of improving (i.e., >50%) trends at the 95% confidence level. The remaining four variables had 95% confidence intervals for the PIT that included 50% (ECOLI, SECCHI, TLI3, TN), and we cannot infer widespread degradation or improvement for these variables.

The 10-year PIT statistics and 95% confidence intervals for each water-quality variable and land-cover class are shown in Figure 5-6. Due to the small sample sizes within each lake class, the PIT statistics have large uncertainty bounds and clear patterns associated with lake classes could not be identified.

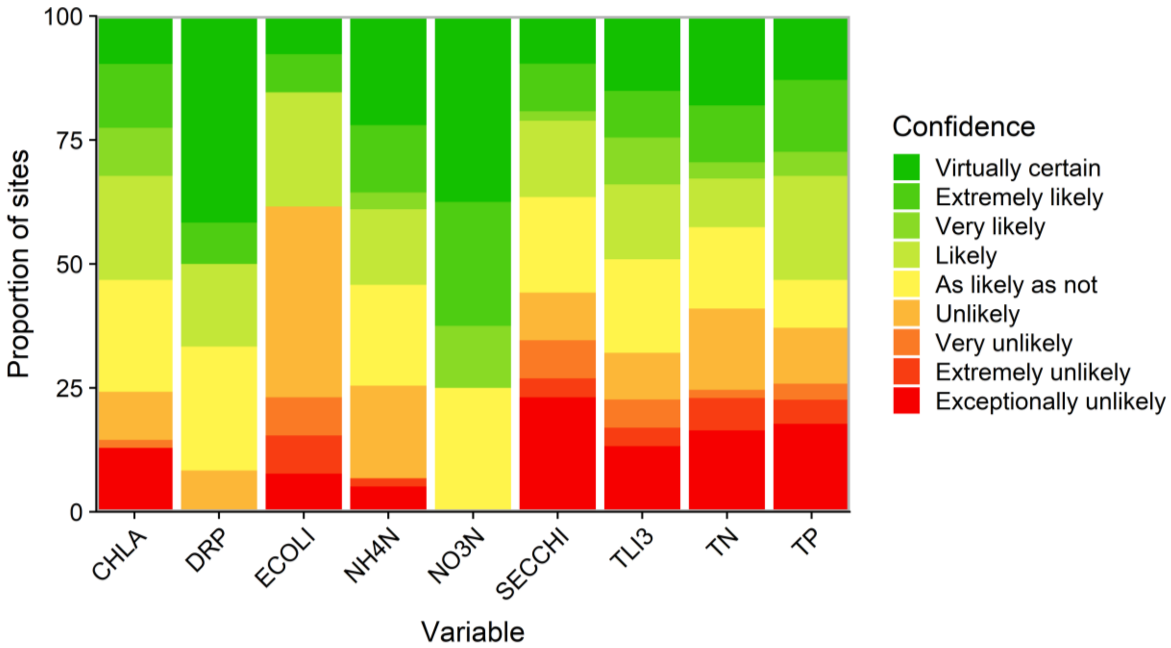

Figure 5-5: Summary plot representing the proportion of lakes with improving 10-year trends at each categorical level of confidence. The plot shows the proportion of lakes with improving trends at levels of confidence defined in Table 3-1.

Table 5-3: Proportions of improving trends (PIT) among 10-year trends. The lake × variable combinations that did not meet the sampling requirements outlined in Sections 3.2.1 and 3.2.5 were not included in the calculation of PIT statistic calculations.

Variable	Number of lakes	PIT (%)	95% confidence interval for PIT
CHLA	62	65.3	56.9 - 73.7
DRP	12	70.8	52.8 - 88.8
ECOLI	13	38.5	19.7 - 57.3
NH4N	59	67.8	59.2 - 76.4
NO3N	8	87.5	68.5 -100.0
SECCHI	52	45.2	36.8 - 53.6
TLI3	53	57.5	49.1 - 65.9
TN	61	50.8	43.4 - 58.2
TP	62	60.5	53.2 - 67.8

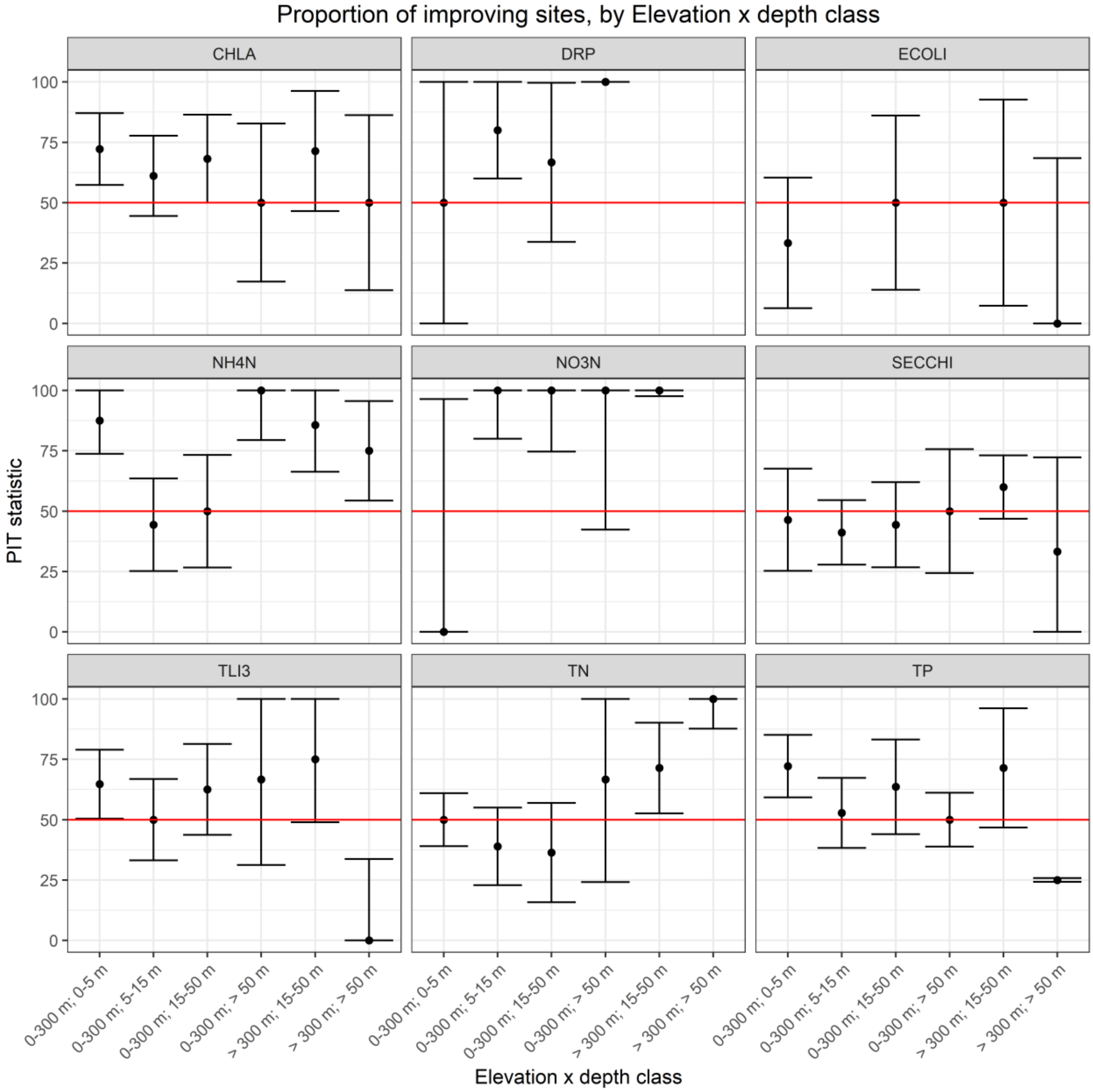

Figure 5-6: Proportions of improving trends (PIT) within lake elevation × depth classes for 10-year trends. Error bars are 95% confidence intervals.

5.2 Twenty-year trends

Between six and 21 lakes met the inclusion rules for the 20-year trend analysis of eight of the nine water quality variables (Table 5-4). No lakes met the inclusion rules for ECOLI. The number of lakes in each variable × lake class combination was very small (0-4 lakes). The majority of qualifying lakes were in the Bay of Plenty, Waikato and Auckland Regions, plus one lake in each of the Southland and West Coast Regions (Figure 5-7). All lake locations, lake classes and numbers of sampling dates are included in the supplementary file "Trends20YearLakes_Oct18.csv".

Table 5-4: Number of lakes in the elevation × depth lake classes used in the 20-year trend analyses. The lake numbers shown correspond to lakes that met the inclusion requirements in Section 3.2.1 (measurements available for at least 90% of the years and at least 90% of seasons). NS: no qualifying lakes.

	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	3	1	1	3	NS	NS	4	4	16
DRP	2	3	2	1	NS	NS	1	1	10
ECOLI	NS	NS	NS	NS	NS	NS	NS	NS	NS
NH4N	4	3	3	3	NS	NS	4	4	21
NO3N	NS	1	2	1	NS	NS	1	1	6
SECCHI	2	2	3	3	NS	NS	4	4	18
TLI3	NS	NS	1	3	NS	NS	4	3	11
TN	NS	NS	1	3	NS	NS	4	3	11
TP	4	2	3	3	NS	NS	4	4	20

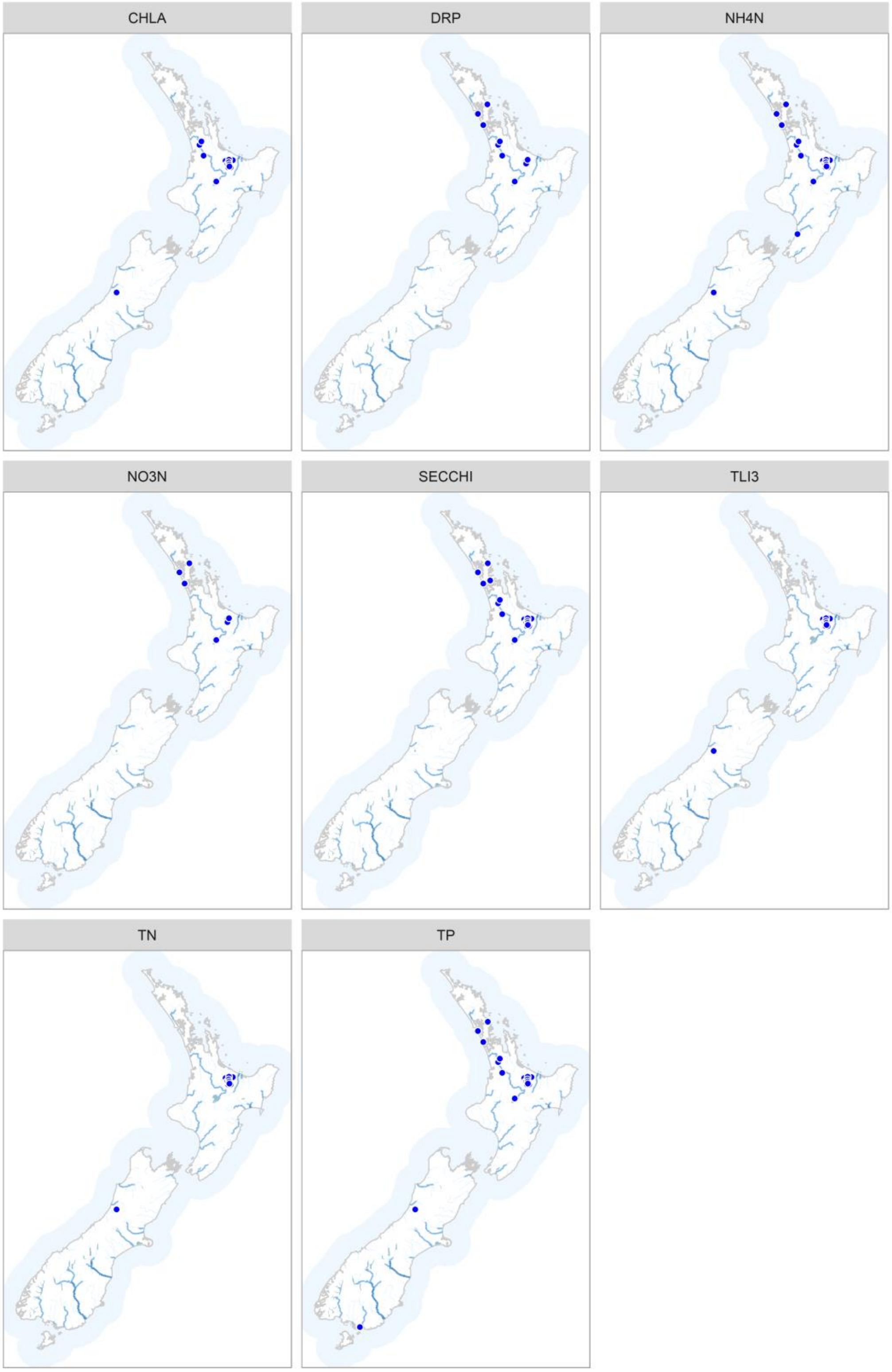

Figure 5-7: Locations of lakes used for 20-year trend analyses of water quality variables.

5.2.1 Trend Magnitude

Box-and-whisker plots were used to summarise the estimated trends for each of the lake water quality variables for the 20-year period from 1998 – 2017 (Figure 5-8). All estimated trends are included in these plots, irrespective of their significance (as defined in Section 3.2.2). The plots indicate that, with the exception on NH4N, lake elevation × depth classes did not account for a substantial amount of the variation in trends. Most of the classes with large median trend magnitudes in Figure 5-8 have too few lakes for reliable estimation. NH4N was the single variable for which there were a moderate number of lakes in some classes (four lakes) and relative high-magnitude trends. The median magnitudes of NH4N trends in the high elevation lake classes indicated improving trends of 3-6%.

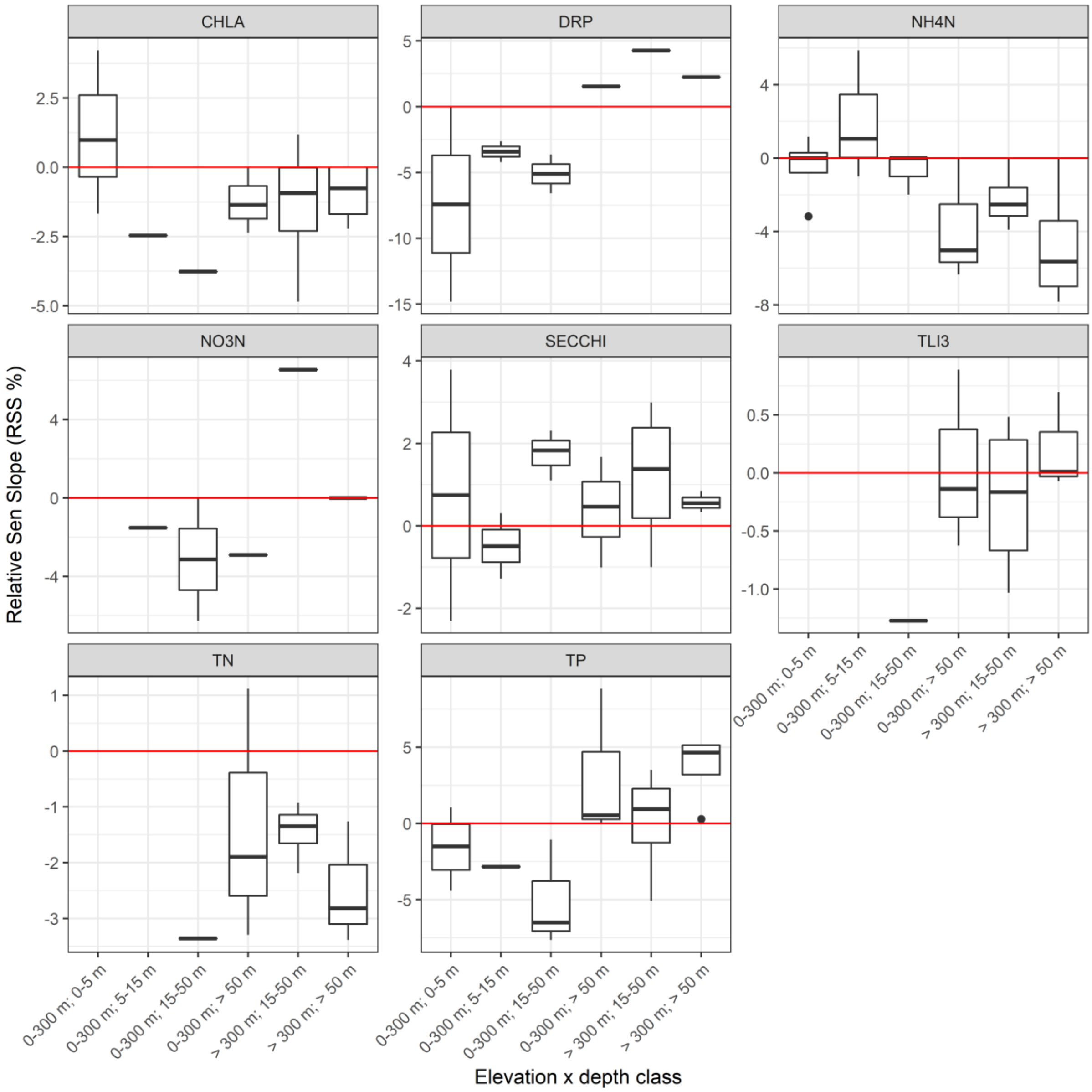

Figure 5-8: Summary of 20-year trend magnitudes in lake water quality variables, in lake elevation × depth classes. Box-and-whisker plots show the distributions of trends in each class. The black horizontal line in each box indicates the median of lake-specific trends, the box indicates the inter-quartile range, whiskers indicate the 5th and 95th percentiles, and open circles indicate outliers.

5.2.2 Trend classification

The numbers and proportions of 20-year trends in four categories are summarised in Table 5-5. A moderate-to-large proportion of the trends for the nine variables (10 to 50% of the trends) were classified as indeterminant. In addition, trends in DRP were not analysed for one lake due to high proportions of censored values. For the remaining trends, the proportions of lakes with improving trends in CHLA, DRP, NH4N, SECCHI and TN were substantially larger than the proportion with degrading trends. Only six lakes qualified for analyses of trends in NO3N, so there is little evidence of a pattern in improving and degrading trends.

Table 5-5: Numbers and proportions of lakes in four trend categories for 20-year trends. The "not analysed" category corresponds to lake × variable combinations that met the inclusion requirements in Section 3.2.1, but did not meet the censored data requirements in Section 3.2.4. The classification of the remaining trends into degrading, improving and indeterminant categories follows the approach used in the previous national-scale river water quality trend analysis (Larned et al. 2015). NS: no qualifying lakes.

	Trend category			
Variable	**Degrading**	**Improving**	**Indeterminant (insufficient data)**	**Not analysed**
CHLA	2 (12%)	8 (50%)	6 (38%)	0 (0%)
DRP	3 (30%)	5 (50%)	1 (10%)	1 (10%)
ECOLI	NS	NS	NS	NS
NH4N	1 (5%)	10 (48%)	10 (48%)	0 (0%)
NO3N	1 (17%)	2 (33%)	3 (50%)	0 (0%)
SECCHI	3 (17%)	9 (50%)	6 (33%)	0 (0%)
TLI3	4 (36%)	4 (36%)	3 (27%)	0 (0%)
TN	1 (9%)	10 (91%)	0 (0%)	0 (0%)
TP	6 (30%)	8 (40%)	6 (30%)	0 (0%)

Figure 5-9: Lakes classified by 20-year trend categories (degrading, improving, indeterminant).

5.2.3 Probability of improvement

The mapped probabilities that 20-year trends for each lake x variable combination were improving are shown in Figure 5-10. The maps indicate that for many variables, there are approximately equal numbers of increasing and decreasing trend directions for those lakes previously categorised as indeterminant. The maps also indicate that for most water quality variables, those lakes previously categorised as indeterminant (Figure 5-9) were about equally divided into likely and unlikely to improve confidence categories. However, in the case of SECCHI, most lakes that were previously categorised as indeterminant were subsequently classed as likely to be improving.

Figure 5-10: Lakes categorised by the probability that the 20-year trend in each variable is improving. Probability of improvement is expressed using the confidence categories in Table 3-1. Only lakes that met the sampling requirements outlined in Sections 3.2.1 and 3.2.5 are shown in the figure.

5.2.4 Aggregate trends

Figure 5-11 shows the proportions of lakes for which 20-year trends indicated improvement at the nine categorical levels of confidence defined in Table 3-1. The national-scale 20-year PIT statistics and their confidence intervals are summarised in Table 5-6. The small numbers of qualifying lakes for all variables (6-21 lakes) raises the possibility that the lakes used to calculate PIT statistics are not representative of all lakes nationally, as discussed in Section 3.2.5. The 20-year PIT statistics ranged from 55-91%. Five variables had a majority of improving trends at the 95% confidence level (CHLA, DRP, NH4N, SECCHI and TN). The remaining variables had 95% confidence intervals for PIT that included 50%, and we cannot infer widespread degradation or improvement for these variables.

The 20-year PIT statistics and 95% confidence intervals for each water-quality variable and lake class are shown in Figure 5-12. As noted above, the small numbers of qualifying lakes for all variable × class combination (1-4 lakes) raises the possibility that the PIT statistics are not representative of all lakes in the classes shown. For seven water quality variables (CHLA, DRP, NH4N, SECCHI, TLI3, TN and TP), there were a majority of improving trends in one or more lake classes at the 95% confidence interval. In contrast, TP appeared to be declining in the high-elevation deep-lake class (> 300 m elevation, > 50 m depth). For about 40% of the lake class × variable combinations for which PIT statistics were calculated (17 of the 43 combinations), the 95% confidence intervals included 50% and we cannot infer widespread degradation or improvement within those class × variable combinations.

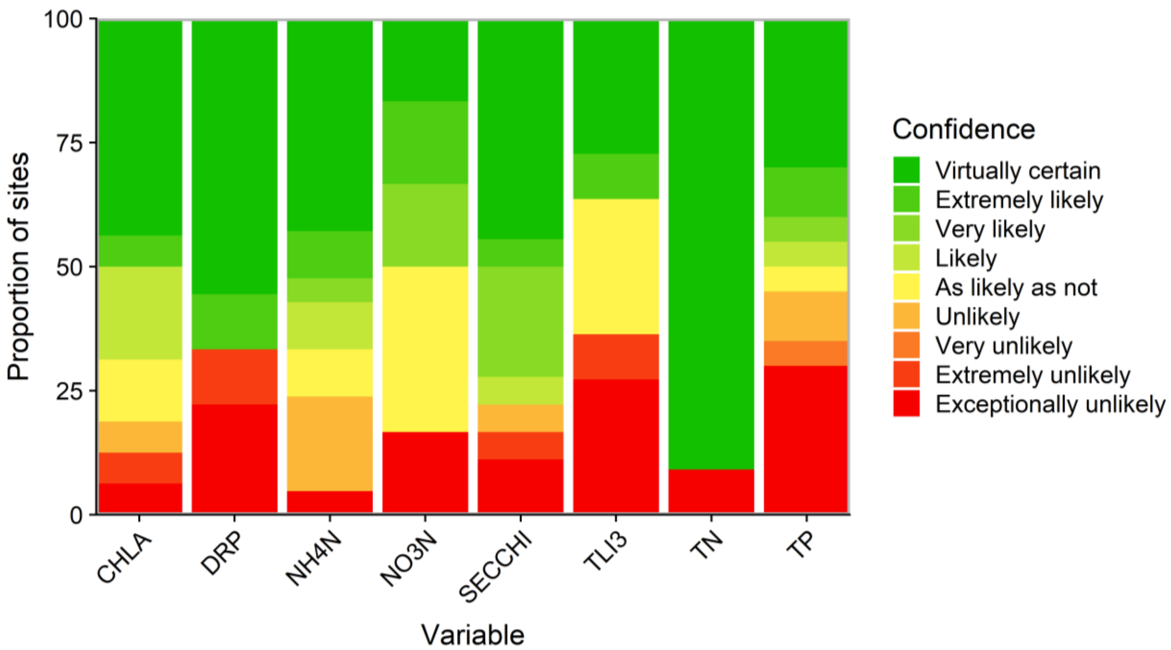

Figure 5-11: Summary plot representing the proportion of lakes with improving 20-year trends at each categorical level of confidence. The plot shows the proportion of lakes with improving trends at levels of confidence defined in Table 3-1.

Table 5-6: Proportions of improving trends (PIT) among 20-year trends. The lake × variable combinations that did not meet the sampling requirements outlined in Sections 3.2.1 and 3.2.5 were not included in the calculation of PIT statistic calculations.

Variable	Number of lakes	PIT (%)	95% confidence interval for PIT
CHLA	16	68.8	55.3 - 82.3
DRP	9	66.7	61.6 - 71.8
ECOLI	**0**	**NA**	**NA**
NH4N	21	71.4	59.6 - 83.2
NO3N	6	66.7	41.2 - 92.2
SECCHI	18	77.8	69.2 - 86.4
TLI3	11	54.5	39.2 - 69.8
TN	11	90.9	90.1 - 91.7
TP	20	55	45.6 - 64.4

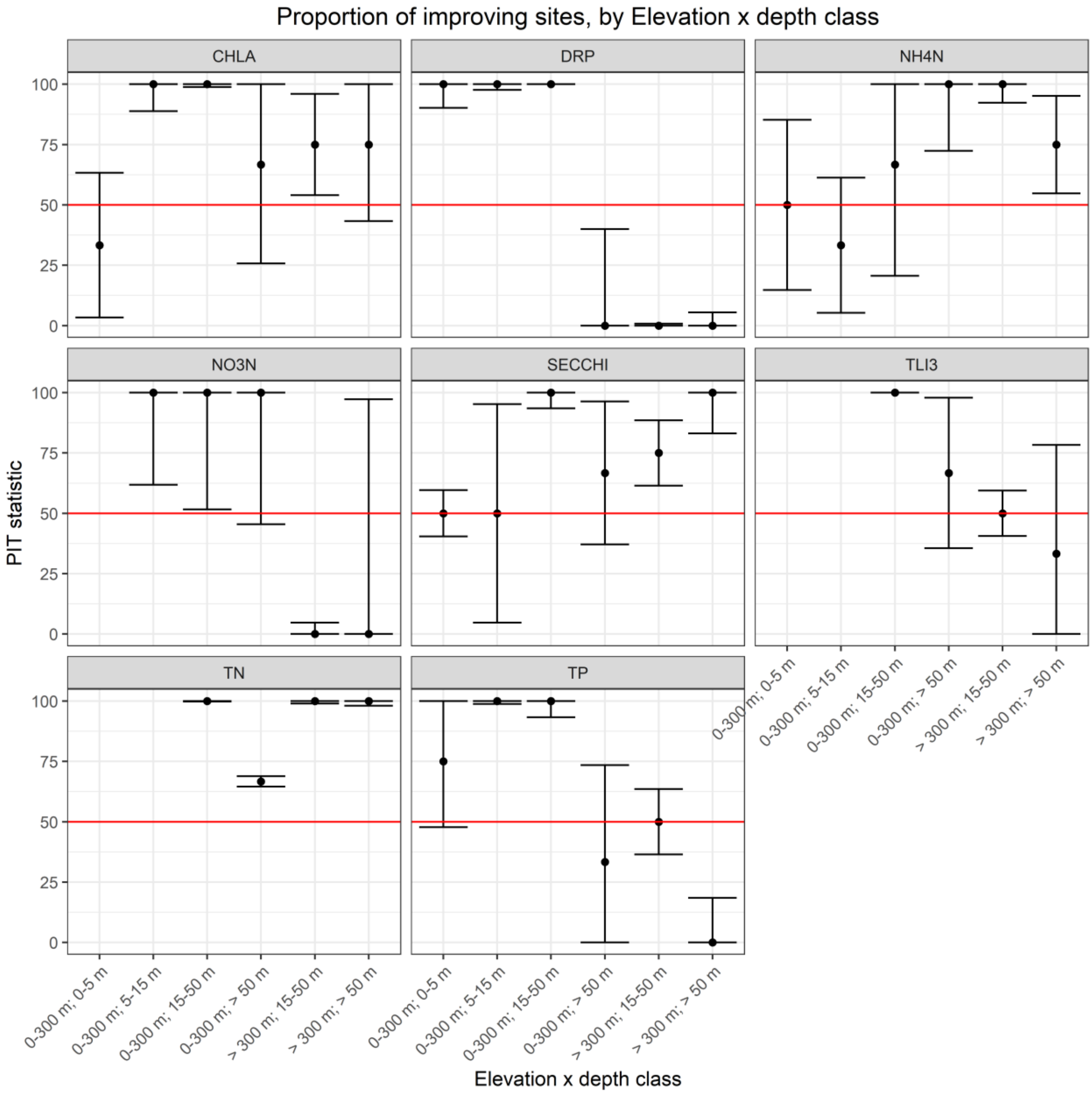

Figure 5-12: Proportions of improving trends (PIT) within lake classes for 20-year trends. Error bars are 95% confidence intervals.

5.3 Twenty-eight-year trends

Between one and 12 lakes met the inclusion rules for the 28-year trend analysis of eight of the nine water quality variables. No lakes met the inclusion rules for ECOLI. The number of lakes varied widely by lake class (Table 5-7). The qualifying lakes were limited to three regions: Bay of Plenty (maximum of eight lakes), Waikato (one lake) and Auckland (four lakes) (Figure 5-13). All lake locations, lake classes and numbers of sampling dates are included in the supplementary file "Trends28YearLake_31_Oct18.csv".

Table 5-7: Number of lakes in the elevation × depth lake classes used the 28-year trend analysis. The lake numbers shown refer to lakes that met the inclusion requirements in Section 3.2.1. NS: no qualifying lakes.

	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	NS	1	1	1	NS	NS	4	2	9
DRP	NS	2	2	NS	NS	NS	1	NS	5
ECOLI	NS	NS	NS	NS	NS	NS	NS	NS	NS
NH4N	NS	3	2	1	NS	NS	3	1	10
NO3N	NS	NS	NS	NS	NS	NS	1	NS	1
SECCHI	NS	1	3	2	NS	NS	4	2	12
TLI3	NS	NS	NS	1	NS	NS	3	1	5
TN	NS	NS	NS	1	NS	NS	3	1	5
TP	NS	3	2	1	NS	NS	3	1	10

Figure 5-13: Locations of lakes used for 28-year trend analyses of water quality variables.

5.3.1 Trend magnitude

Distributions of trend magnitudes for each of the water-quality variables for the 28-year period from 1990 – 2017 by lake class are shown in Figure 5-14. Most of the classes with large median trend magnitudes in Figure 5-8 have too few lakes for reliable estimation. The largest number of lakes in any variable × class combination was four, for CHLA and SECCHI in the > 300 m elevation, 15-50 m depth class. In both of these cases, the median trend magnitude for the class was small (< 0.5% y^{-1}).

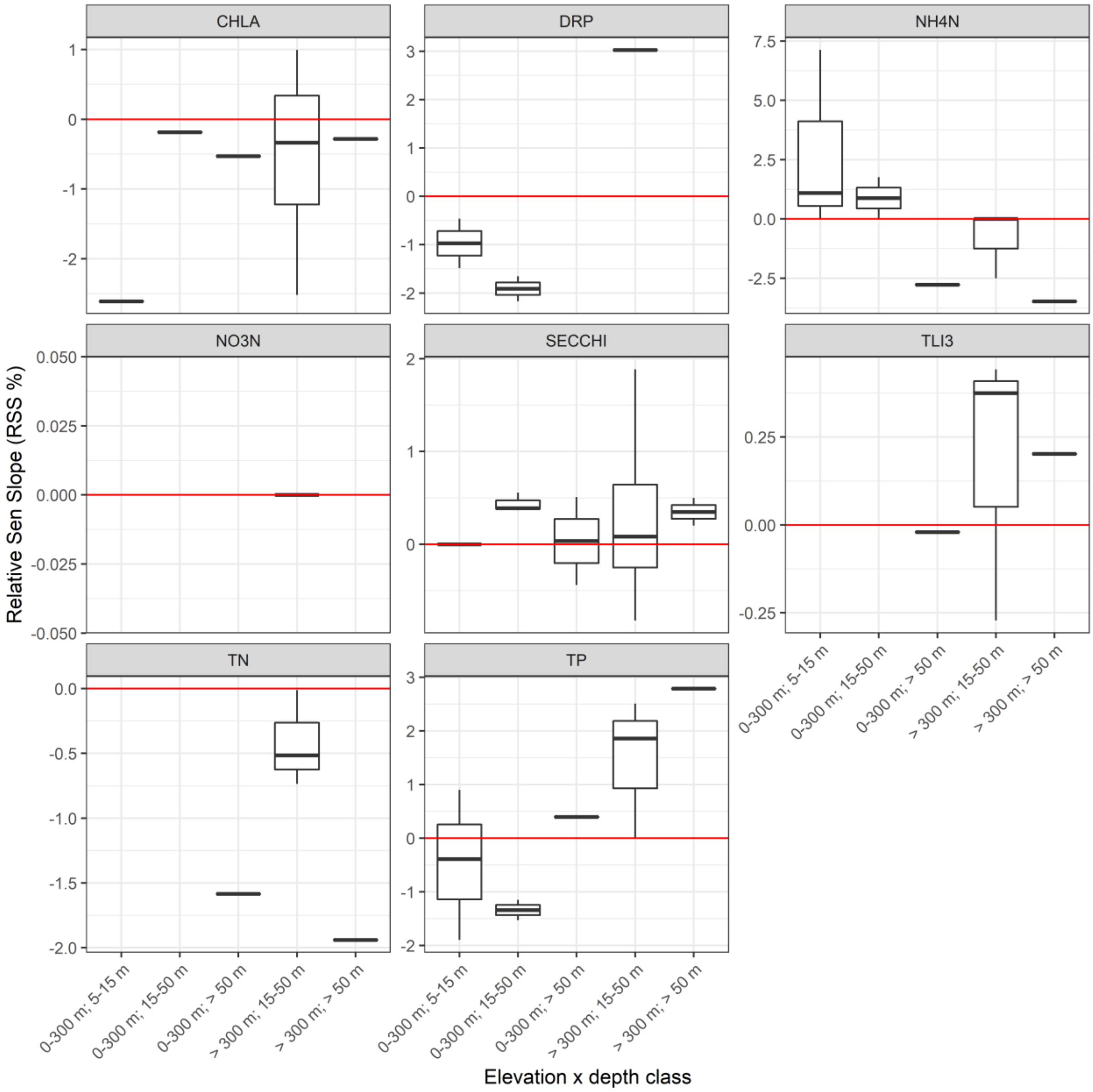

Figure 5-14: Summary of 28-year trend magnitudes in lake water quality variables, in lake elevation × depth classes. Box-and-whisker plots show the distributions of trends in each class. Black horizontal line in each box indicates the median of lake trends, box indicates the inter-quartile range, whiskers indicate the 5th and 95th percentiles, and open circles indicate outliers.

5.3.2 Trend classification

The numbers and proportions of 28-year trend trends in each lake category are summarised in Table 5-8. A moderate-to-large proportion of the trends for the eight variables (20 to 100% of the lake trends) were classified as indeterminant. For the remaining trends, the proportions of lakes with improving trends in DRP and TN were larger than the proportions with degrading trends, and the proportion of lakes with degrading trends in TLI3 were larger than the proportion with improving trends. Note that the small lake numbers for all variables make inferences about proportions of improving and degrading trends unreliable.

Table 5-8: Numbers and proportions of lakes in four trend categories for 28-year trends. The "not analysed" category corresponds to lake × variable combinations that met the inclusion requirements in Section 3.2.1, but did not meet the censored data requirements in Section 3.2.4. The classification of the remaining trends into degrading, improving and indeterminant categories follows the approach used in the previous national-scale river water quality trend analysis (Larned et al. 2015). NS: no qualifying lakes.

	Trend category			
Variable	**Degrading**	**Improving**	**Indeterminant (insufficient data)**	**Not analysed**
CHLA	1 (11%)	2 (22%)	6 (67%)	0 (0%)
DRP	1 (20%)	4 (80%)	0 (0%)	0 (0%)
ECOLI	NS	NS	NS	NS
NH4N	2 (20%)	3 (30%)	5 (50%)	0 (0%)
NO3N	0 (0%)	0 (0%)	1 (100%)	0 (0%)
SECCHI	1 (8%)	3 (25%)	8 (67%)	0 (0%)
TLI3	2 (40%)	0 (0%)	3 (60%)	0 (0%)
TN	0 (0%)	4 (80%)	1 (20%)	0 (0%)
TP	4 (40%)	3 (30%)	3 (30%)	0 (0%)

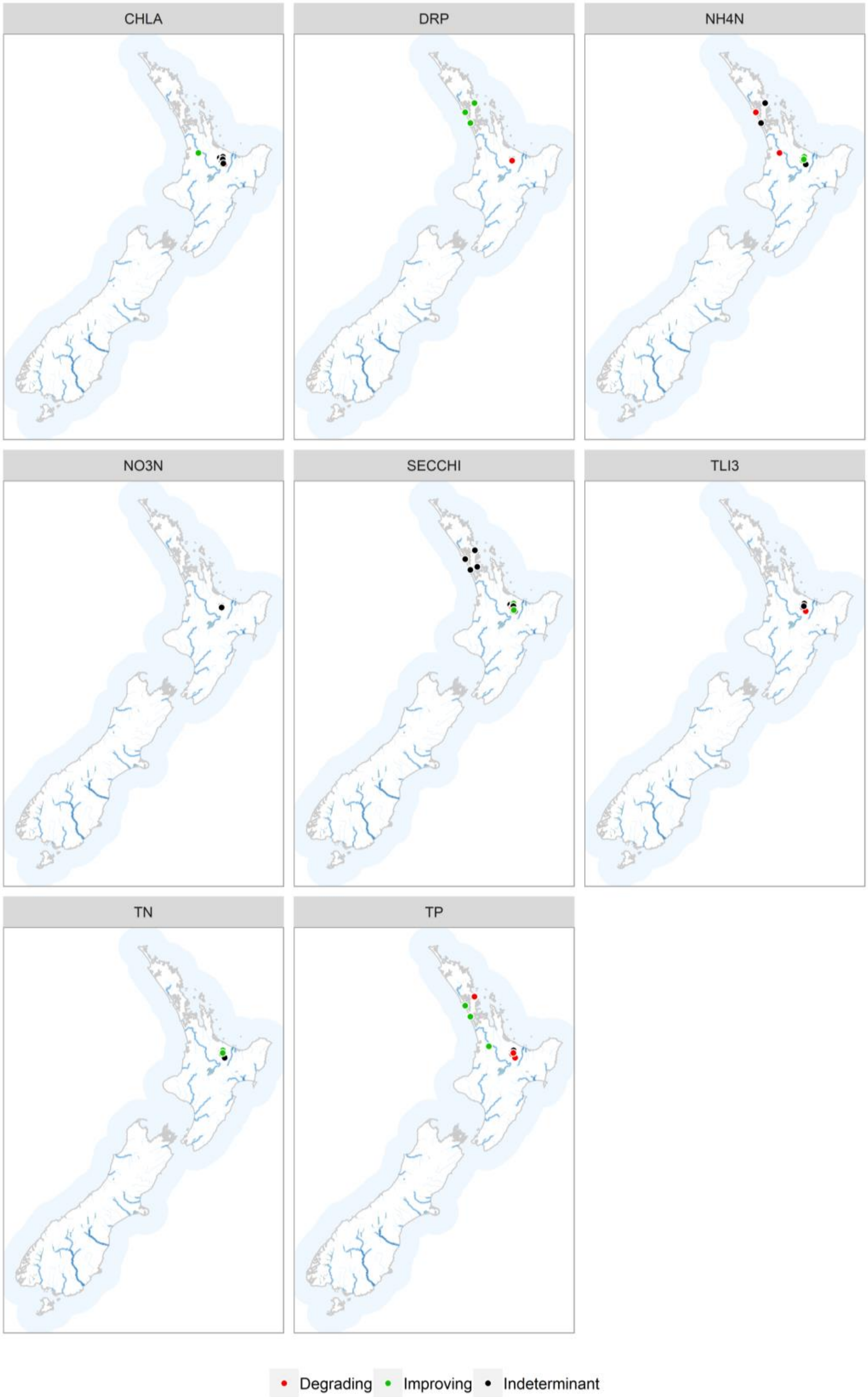

Figure 5-15: Lakes classified by 28-year trend categories (degrading, improving, indeterminant).

5.3.3 Probability of improvement

The mapped probabilities that 28-year trends for each variable were improving are shown in Figure 5-16. A comparison of Figures 5-15 and 5-16 indicate that most of the trends in CHLA, SECCHI and NH4N that were classified as indeterminant in Figure 5-15 were classified as improving in Figure 5-16, based the confidence categories in Table 3-1.

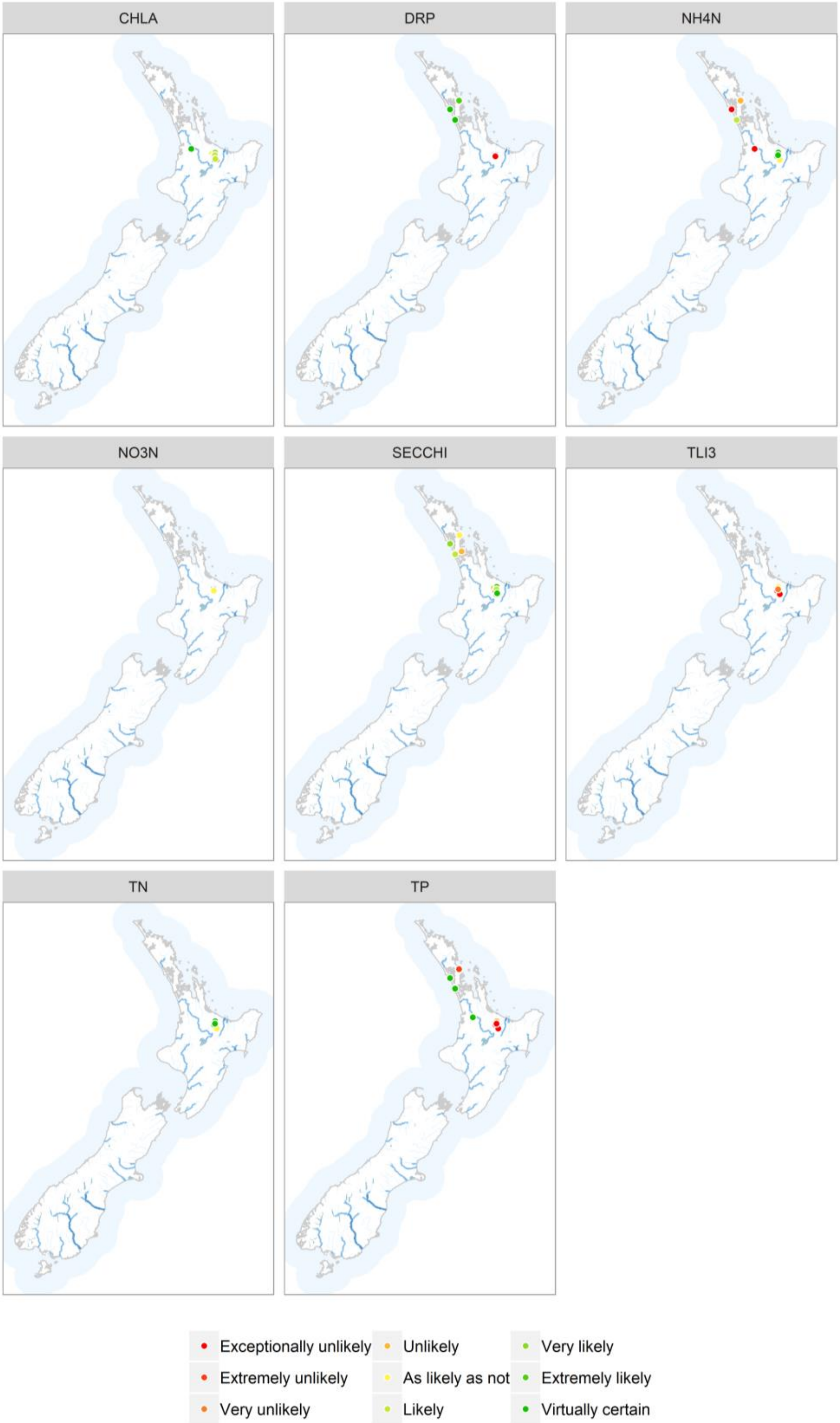

Figure 5-16: Lakes categorised by the probability that the 28-year trend in each variable is improving. Probability of improvement is expressed using the confidence categories in Table 3-1. Only lakes that met the sampling requirements outlined in Sections 3.2.1 and 3.2.5 are shown in the figure.

5.3.4 Aggregate trends

Figure 5-17 shows the proportions of lakes for which 28-year trends indicated improvement at the nine categorical levels of confidence defined in Table 3-1. The national-scale 28-year PIT statistics and their confidence intervals are summarised in Table 5-9. The 28-year PIT statistics ranged from 33-100%. Four variables had a majority of improving trends at the 95% confidence level (CHLA, DRP, NH4N and TN). Three variables had 95% confidence intervals for PIT that included 50% (SECCHI, TLI3, TP), and we cannot infer widespread degradation or improvement for these variables. There was a single qualifying lake for NO3N, and no qualifying lakes for ECOLI. The small numbers of qualifying lakes for all variables (1-12 lakes nationally) raises the possibility that the lakes used to calculate PIT statistics are not representative of all lakes nationally, as discussed in Section 3.2.4.

The 28-year PIT statistics and 95% confidence intervals for each water-quality variable and lake class are shown in Figure 5-18. As noted above, the small numbers of qualifying lakes for all variable × class combination (1-4 lakes) raises the possibility that the PIT statistics are not representative of all lakes in the classes shown. In half of the cases (15 out of 30 lake class × variable combination), the 95% confidence intervals included 50% and we cannot infer widespread degradation or improvement within those class × variable combinations. In one lake class (> 300 m elevation, 15-50 m depth), NH4N and TN appeared to have a majority of improving trends and TLI3 a majority of degrading trends. For the remaining class × variable combinations, there were only one or two qualifying lakes, and no inferences were made.

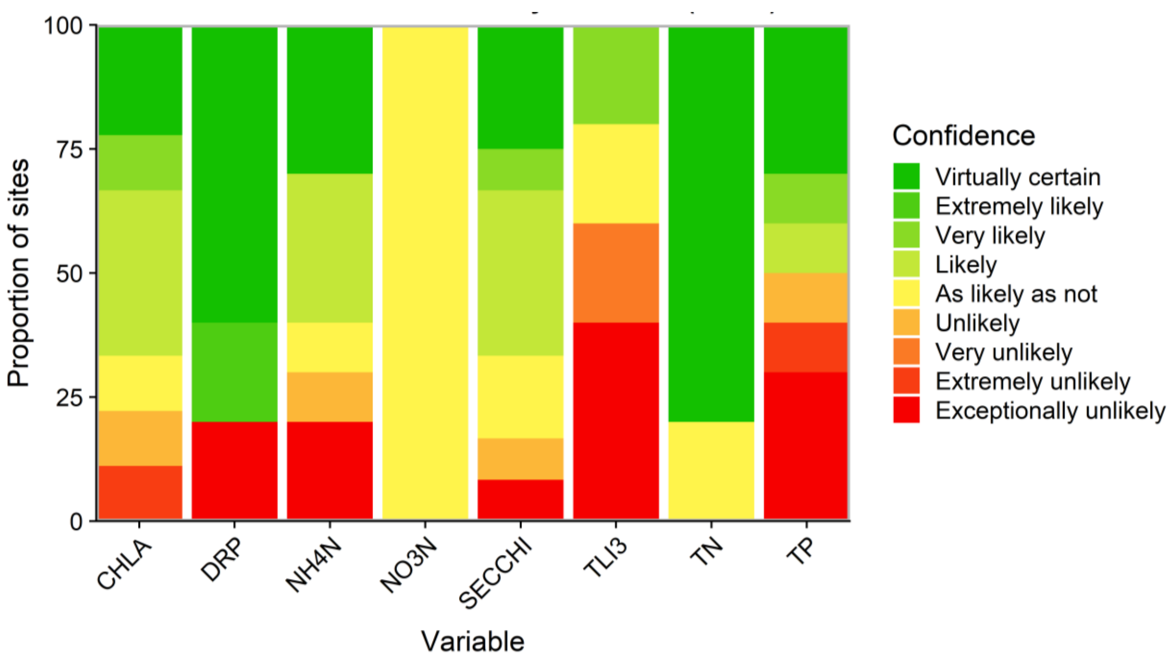

Figure 5-17: Summary plot representing the proportion of lakes with improving 28-year time-period trends at each categorical level of confidence. The plot shows the proportion of lakes with improving trends at levels of confidence defined in Table 3-1.

Table 5-9: Proportions of improving trends (PIT) among 28-year trends. The lake × variable combinations that did not meet the sampling requirements outlined in Sections 3.2.1 and 3.2.5 were not included in the calculation of PIT statistic calculations. NS: no qualifying lakes.

Variable	Number of lakes	PIT (%)	95% confidence interval for PIT
CHLA	9	77.8	55.1 - 100.0
DRP	5	80	72.0 - 88.0
ECOLI	NS	NS	NS
NH4N	10	70	52.0 - 88.0
NO3N	1	100	4.2 - 100.0
SECCHI	12	66.7	47.7 - 85.7
TLI3	5	40	16.1 - 63.9
TN	5	100	80.4 - 100.0
TP	10	50	37.1 - 62.9

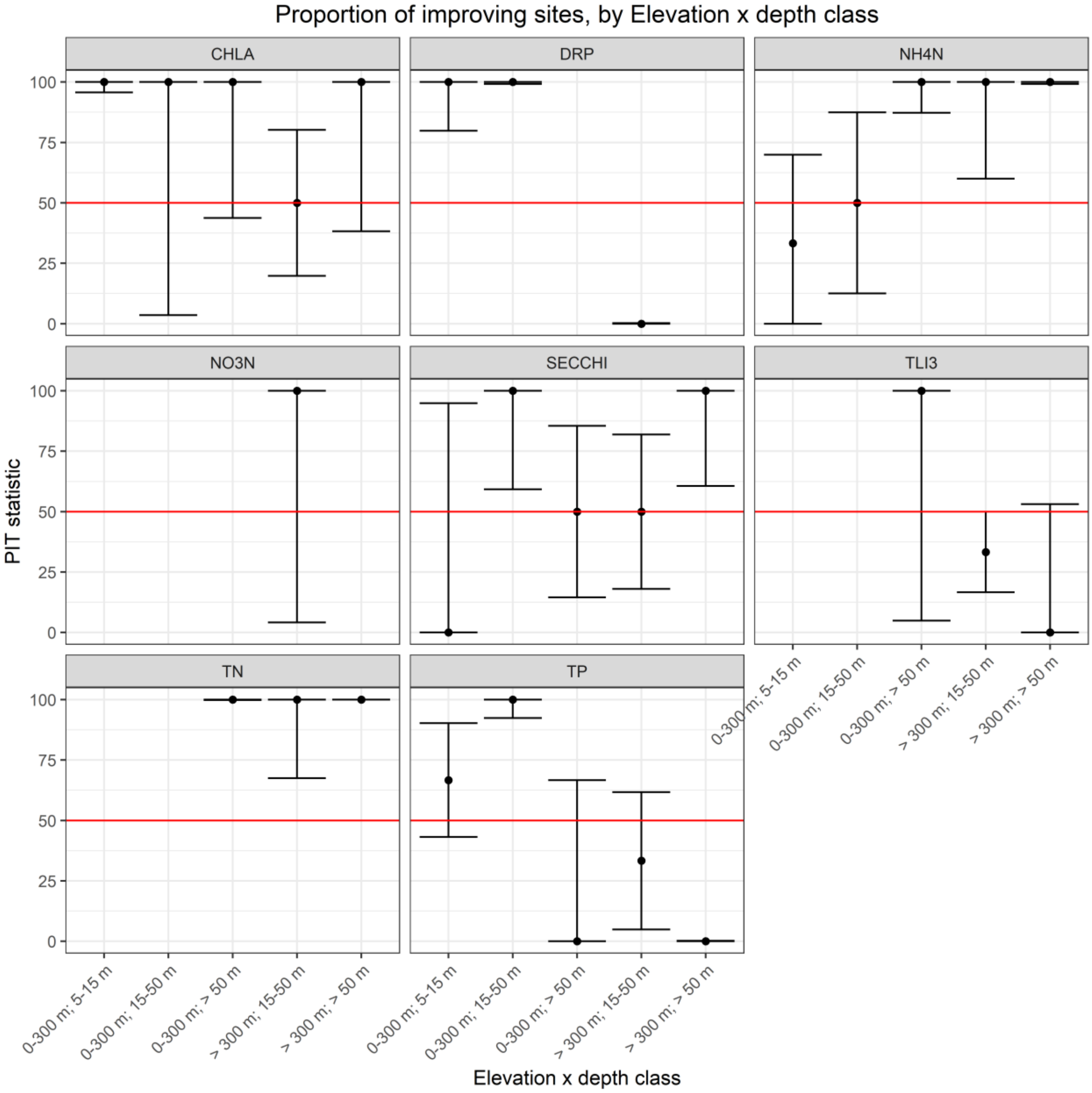

Figure 5-18: Proportions of improving trends (PIT) within lake classes for 28-year trends. Error bars are 95% confidence intervals.

6 Discussion

The primary purpose of the lake state and trend analyses reported here is to provide MfE with information required for reporting on the freshwater domain and for policy development. The detailed information for each lake used in the analyses is contained in the supplementary files that accompany this report. The lakes and their water quality conditions can be aggregated in many ways to meet different information requirements (e.g., grouped by region or environmental class, or 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.

As with the previous national-scale water quality state analyses (Larned et al. 2015), we used a five-year period ending in the immediate past year to represent recent water quality state (for this report the period is 2013 to December 2017). This period represents a trade-off between ensuring sufficient sampling dates to provide robust summary statistics, and minimising the influence of long-term temporal trends on estimates of current state. Longer periods would have also reduced the number of lakes in the dataset and therefore, spatial coverage. The five-year period ensured that there were at least 30 samples for each lake, for variables that are measured at monthly intervals, as recommended by McBride (2005).

In the previous national-scale lake water quality analyses, we used the same procedures for managing censored data in both the state and trend datasets. In the current study, different procedures were used for censored data in the state dataset and in the trends datasets. For the state dataset, we used the same methods used by Larned et al. (2015): sites with more that 50% imputed data were excluded, and for qualifying sites, censored data were replaced with imputed values using procedures based on regression-on-order-statistics and survival analysis.

In contrast to the state dataset, we did not replace censored values in the trends datasets used for the current study; the inclusion of imputed values in trend datasets in previous reports was not strictly correct because the imputation process cannot account for the time order of samples (Snelder 2018). In addition, the approach adopted in the current study only excluded sites based on censored values in extreme situations: where there were so many ties caused by censored values that Sen slopes and confidence intervals could not be calculated (Section 3.2.4). This approach also differs from that of the previous study, where all site × variable combinations with > 15% censored values were excluded from the trend analyses. Retaining all but the most extreme cases in the current study maximised the spatial coverage of sites. The assessments of trend directions in this report were carried out using both the methods set out in the previous national-scale water quality trend analyses (Larned et al. 2015), and new methods. By showing the results derived from both approaches, the effects of the new methods are apparent. In both the previous and the current reports, we replaced traditional significance tests about trend directions (which posit that the trend slope is exactly zero) with inferential information about trend direction, including confidence intervals. As noted in the previous report and McBride (in press), true trend slopes cannot be zero, and the traditional hypothesis is *a priori* false. While the replacement of significance tests represented an advancement in trend analyses, there was room for further improvement. The use of the 'indeterminant' trend category in the previous report to indicate cases in which there are insufficient data for inferring trend directions with stated levels of confidence was one area with potential for improvement. For example, in the 10-year trend analyses reported using the previous method, between 38 and 65% of the within-lake trends were categorised as indeterminant at the 95% confidence level (Table 5-2). Unfortunately, the indeterminant trend category has been misconstrued as indicating 'stable trends', i.e., water quality that does not change over the observation period. More generally, categorising large proportions of site trends as indeterminant can be viewed as a substantial loss of information, including the numbers of trends that are very likely to have a particular direction, but at a confidence level below 95%. The two new approaches introduced in this report reduce such information loss:

- The first approach involves subdividing the group of trends formerly categorised as indeterminant into eight subgroups with progressively decreasing probabilities of improving trends, as described in Section 3.2.6.
- The second approach involves the use of the PIT statistic and its confidence interval to estimate the proportion of improving a population of sites in a given domain, including all of New Zealand.

We recommend adopting the new approaches set out in this report to increase the information yield from trend analyses, and ultimately, from regional council and national monitoring programmes. We recognise that progressive changes in data analysis methods can impede comparisons between consecutive reports. To alleviate that problem, we provided results of trend analyses using both the methods of Larned et al. (2015) and the new methods, and we recommend presenting the results in parallel as we have in the current report. We note that the current report does not represent the last word in water-quality data analysis; further advancements are inevitable and beneficial.

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 155 lakes. This number was reduced to a maximum of 63 lakes (for state analysis) and 61 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 156 (Larned et al. 2015), 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 in 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 no qualifying lakes in two of the eight elevation × depth classes that we used (Tables 4-1 and 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 is indicated by the uneven distribution of lake × variable combinations in Tables 4-1 and 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 Acknowledgments

We thank Dan Elder, Annabelle Ellis and Sheree De Malmanche (Ministry for the Environment) for project support and feedback. We thank Abi Loughnan and the LAWA team for data updates. We thank the many regional council staff who provided additional data and information about monitoring programmes. We thank Clive Howard-Williams and Graham McBride for reviewing the report, and Janine Pineda-Verhoogh for formatting.

8 References

Burns N, Bryers G, Bowman E. (2000) *Protocol for monitoring lake trophic levels and assessing trends in trophic state.* Ministry for the Environment, Wellington, New Zealand.

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., Hirsch, R.M. (1992) Statistical methods in water resources. Elsevier Publishers.

Hesel, D.R. (2005) Nondetects and data analysis. Statistics for censored environmental data. Wiley-Interscience.

Hesel, D.R. (2012) Statistics for censored environmental data using MINITAB and R, second edition. Wiley.

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 ST, Snelder T. (2012) National Environmental Monitoring and Reporting: Network design principles for national freshwater monitoring. New Zealand Ministry for the Environment. *NIWA Report* CHC2012-027.

Larned, S., Snelder, T., Unwin, M., McBride, G., Verburg, P., McMillan, H. (2015) Analysis of water quality in New Zealand lakes and rivers: Data sources, data sets, assumptions, limitations, methods and results. New Zealand Ministry for the Environment. *NIWA Client Report* CHC2015-033.

McBride, G.B. (2005) Using statistical methods for water quality management: issues, problems and solutions (Vol. 19). John Wiley & Sons.

McBride, G.B. (2019) Has water quality improved or been maintained? A quantitative assessment procedure. *Journal of Environmental Quality*. doi:10.2134/jeq2018.03.0101.

New Zealand Government. (2014) *National Policy Statement for Freshwater Management 2014*, New Zealand Government. http://www.mfe.govt.nz/publications/fresh-water/national-policy-statement-freshwater-management-2014.

Ministry for the Environment (2017) *Swimming categories for E. coli in the Clean Water Package: a summary of the categories and their relationship to human health risk from swimming*. Ministry for the Environment (MfE), Wellington, New Zealand.

New Zealand Government. (2017). *National Policy Statement for Freshwater Management 2014* (amended 2017). New Zealand Government, Wellington.

Patton CJ, Kryskalla JR. (2003) Evaluation of alkaline persulfate digestion as an alternative to Kjeldahl digestion for determination of total and dissolved nitrogen and phosphorus in water. *US Geological Survey Water Resources Investigations Report,* 3-4174. http://nwql.usgs.gov/pubs/WRIR/WRIR-03-4174.pdf.

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

Snelder T., Wood, S, Atalah J. (2016) Strategic assessment of New Zealand's freshwaters for recreational use: a human health perspective. LWP Client Report 2016-011.Sorrell, B. (2006). Lake water quality in New Zealand. Status in 2006 and recent trends 1990–2006. *Ministry for the Environment Publication*, 832.

Sorrell, B., M. Unwin, K. Dey, and H. Hurren. (2006) A snapshot of Lake Water Quality. *NIWA Client Report* CHC2006–145.

Verburg, P., Hamill, K., Unwin, M., J. Abell, J. (2010) Lake water quality in New Zealand 2010: status and trends. New ZealandMinistry for the Environment. *NIWA Client Report* HAM2010-107.