Water quality state and trends in New Zealand rivers

Analyses of national data ending in 2017

Prepared for Ministry for the Environment

December 2018

Prepared by:
Scott Larned
Amy Whitehead
Caroline Fraser LWP
Ton Snelder LWP
Jing Yang

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:	2018347CH
Report date:	December 2018
NIWA Project:	MFE18502

Quality Assurance Statement		
	Reviewed by:	Neale Hudson
	Reviewed by:	Graham McBride
	Formatting checked by:	Fenella Falconer
	Approved for release by:	Clive Howard-Williams

Contents

Tables

Figures

Executive summary

Background and brief

The New Zealand Ministry for the Environment (MfE) and Statistics New Zealand use the results from analyses of river water quality state and trends to inform policy development and meet their requirements for environmental reporting on the freshwater domain under the Environmental Reporting Act 2015. The data used for these analyses come from regional council state-of-the-environment (SoE) monitoring programmes and NIWA's National River Water Quality Network (NRWQN). MfE have commissioned national-scale analyses of river water quality data periodically since 2003. The current study was commissioned to analyse river water quality state and trend for the period ending in late 2017.

The two outcomes required from this analysis of river water quality data are accurate estimates of current state and temporal trends at individual monitoring sites. In this study, we use several approaches to aggregate results, including River Environment Classification (REC) land-cover classes, and continuous land-cover data. The principal output of the study are site-specific results that have been provided to MfE as supplementary files. These site-specific results may then be aggregated and summarised in different ways (e.g., by environmental class, region, entire nation) to meet other environmental reporting requirements and to better inform policy-makers.

The brief for this work consisted of seven major steps:

1. Compile river water quality data from regional councils, Land and Water Aotearoa (LAWA) and NIWA.
2. Organise and process the data, including error correction, application of reporting conventions and links to spatial data for each site.
3. Assess the suitability of data for 13 physical, chemical, microbial and ecological variables for statistical analyses and apply site inclusion rules.
4. Carry out analyses of water quality state, including comparisons of state at monitoring sites aggregated by River Environment Classification (REC) land-cover classes, and relationships between water quality state and high-intensity agricultural land cover.
5. Estimate river flows for each site and sampling date, to adjust trend analyses for the extraneous effects of flow variation.
6. Carry out trend analyses using 10-, 20- and 28-year periods ending in late 2017, including comparisons of trends at sites aggregated by REC land-cover classes. The 28-year trend period corresponds to the period of record for NRWQN monitoring sites and a smaller number of long-term council sites.
7. Assess water quality trends at the national scale using two approaches: categorical levels of confidence and a statistical analysis of the proportions of improving trends.

Methods

Data acquisition and processing

We used three procedures to acquire updated data: interrogation of data servers operated by individual regional councils, Land Air Water Aotearoa (LAWA), and NIWA (for NRWQN data); 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. These data were organised into a consistent format and stored in a single RData file.

Data processing was carried out in four steps: 1) application of consistent conventions for variable names, site identifiers, date and time formats, units of measurement, and other data structure elements; 2) correction of errors identified using time-series plots and quantile plots (e.g., transcription errors and scale problems caused by inconsistent units (e.g., concentrations in mg L^{-1} and $\mu g\ L^{-1}$); 3) exclusion of data generated using non-comparable methods (e.g., total nitrogen and total phosphorus concentration data derived from filtered water samples); 4) attachment of spatial information to the data for each monitoring site, including NZMS260 grid reference (converted from NZTM as necessary), NZReach number, REC classes and catchment land cover data.

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 monitoring variables (of the 13 variables assessed) were selected for use in the state and trends analyses: visual clarity (CLAR), turbidity (TURB), concentrations of nitrate-nitrogen (NO3N), ammoniacal nitrogen (NH4N), total nitrogen (TN), dissolved reactive phosphorus (DRP), total phosphorus (TP), the faecal bacterium *Escherichia coli* (ECOLI), and macroinvertebrate community index scores (MCI).

Four other candidate variables (total suspended sediment concentration, areal cover of deposited fine sediment, periphyton biomass, and areal cover of the cyanobacterium *Phormidium*), were omitted for use for several reasons, including: several regional councils had no data, most of the remaining council datasets comprised few sites or did not meet the sampling frequency and duration criteria.

State analyses

The state dataset consisted of data for the nine variables listed above, for the 2013-2017 period, at sites for which measurements were available in at least 90% of the sampling intervals in that period (i.e., at least 54 of 60 months or 18 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 site × variable combination, concentration or measured value percentiles were calculated and the site medians used in two subsequent steps of the state analysis. First, site medians were grouped by REC land-cover classes for inter-class comparison. Second, linear regressions were used to relate median water-quality state to proportions of high-intensity agricultural land cover in the catchments upstream of the monitoring sites. In addition, the state dataset was used to assess river monitoring sites against attribute states that are set out in the National Policy Statement for Freshwater Management (NPS-FM). We determined the ECOLI attribute state for monitoring sites and determined the number of river monitoring sites at which the NPS-FM bottom-lines for NO3N and NH4N toxicity were exceeded.

Trend analyses

The trend assessment utilised data for the nine variables listed above, for the 10-, 20-, and 28-year periods ending in December 2017. For all variables except MCI, the site inclusion rule required that measurements be available for at least 90% of each year in the trend period, and for at least 90% of the seasons. MCI is generally calculated from macroinvertebrate samples that are collected annually, so the site inclusion rule was limited to 90% of the years in the trend period.

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 site × variable combinations for which more than 15% of the data consisted of censored entries were excluded. In the current study, we modified these approaches to improve estimations of Sen slopes and confidence intervals, and to reduce the number of site exclusions (thereby increasing spatial coverage).

For each site and sampling date, the corresponding daily average river flow was estimated, using measured flow (for sites near flow recorders), or estimates derived from the TopNet hydrological model, corrected using flow-duration curves. In a second shift from the previous analyses, we discontinued automated flow adjustment of data for all sites used in trend analyses. Instead, flow adjustments were applied only to site × variable combinations for which reliable water-quality-flow relationships existed. Where the water quality-flow relationship was poor, trend analyses were carried out without flow adjustment.

Trend assessment utilised estimates of trend magnitude made with the Sen slope estimator, and estimates of the confidence in the trend direction, made using the P-values from Kendall tests. The seasonal version of the Sen slope estimator was used for variables measured seasonally (i.e., monthly or quarterly), and for which variability in the water quality variable was significantly explained by season.

The trends for all site × variable combinations were classified in two ways. The first approach used four trend direction categories: improving, degrading, indeterminant and not analysed (the approach 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 second, new approach classified trends into nine confidence categories on basis of a 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. These approaches involved aggregating multiple sites into environmental (REC land-cover classes), or a spatial domain covering the entire country. Other environmental or spatial domains may also be used. The first approach used the nine confidence categories described above, following which the proportion of sites in each category was tallied.

The second approach used the same probabilities of improving trends from individual sites to estimate the proportion of improving trends (PIT) for all sites in the domain. The PIT statistic and its 95% confidence intervals were calculated for each water quality variable within each REC land-cover class, and nationally.

Results

Water quality state

The summaries of river state indicated that variation in median nutrient and ECOLI concentrations and CLAR was partly explained by REC land-cover classes. Median concentrations of all nutrients and ECOLI were lowest and CLAR and MCI highest in the natural class. Nutrient and ECOLI concentrations were highest in the urban class, closely followed by the pasture class.

Approximately 1% of sites (6 of 567 sites) in the pastoral land cover class exceeded the NOF bottom-line for NO3N annual median concentrations and four of those sites also exceeded the bottom-line for 95th percentile concentrations. No sites in the other land use classes exceeded NO3N bottom lines. Less than 1% of sites (2 of 532 sites) in the pastoral land cover class exceeded the pH-adjusted NH4N median bottom line, and 1.7% of pastoral sites (9 of 532 sites) and 1.7% of urban sites (1 of 58 sites) exceeded the pH-adjusted NH4N maximum bottom line. No sites in the other land use classes exceeded NH4N bottom lines. The classification of monitoring sites by ECOLI attribute states indicated that most sites in the natural and exotic forest land-cover classes were in the A attribute state, most sites in the pastoral land-cover class were in the D and E attribute states, and most sites in the urban land-cover class were in the E attribute state.

Regressions of site medians for the nine variables on high-intensity agricultural land cover in the upstream catchment of each site indicated that the concentrations of each nutrient and ECOLI increased, and MCI scores and visual clarity decreased, with increasing proportions of high-intensity agricultural land cover.

Water quality trends

In this summary, we first set out results of the 10-, 20-, and 28-year trend analyses in terms of trend magnitude (percent change in a water quality variable per year). We then summarise the trend analysis results in terms of trend direction (improving or degrading). As noted above, the analyses of trend directions included the method used in the previous national-scale trend analysis, a new approach in which all trends are classified into nine categorical confidence categories, and a new approach to estimate the proportion of improving trends (the PIT statistic). For brevity, the following summary is based on the PIT statistics for each water quality variable at the national level and within land-cover classes.

The magnitudes of 10-, 20- and 28-year trends did not vary strongly or consistently between land cover classes. However, the following patterns were evident:

- Median 10-year trend magnitudes were largest for CLAR, DRP, TP and TURB in the urban land-cover class; in each case the trend direction indicated improving conditions.
- The median 20-year trend magnitudes were largest for NH4N and TP in the urban land-cover class (declining by over 2% per year), and for TN and TURB in the exotic forest class (increasing by approximately 2% per year).
- The median 28-year trend magnitude was largest for NH4N in the natural land-cover class indicating improving conditions (declining by over 2% per year).

The national scale PIT statistics for each water quality variable are shown in the following table. All values in the table are estimates of the proportion of improving sites with respect to the corresponding water quality variable.

Table i: Trends in river water quality variables according to proportion of improving trends (PIT). Blue font = improvement in water quality at most sites. Red = degradation in water quality at most sites. Green = = inferences regarding improvement or degradation in water quality at most sites cannot be made at the specified confidence level.

Variable	Proportion of sites across New Zealand indicating improving trends (%)		
	10-year trend (2008-2017)	20-year trend (1998-2017)	28-year trend (1990-2017)
CLAR	65.1	49.1	79.5
DRP	55.0	64.3	54.9
ECOLI	52.1	67.4	68.8
MCI	44.7	35.4	47.1
NH4N	72.2	78.2	85.8
NO3N	56.5	41.3	49.1
TN	49.5	45.1	32.5
TP	71.4	81.3	64.5
TURB	50.1	35.4	35.1

A comparison of the 10-, 20- and 20-year trends in this table reveal several changes in the balance of improving and degrading trends: 1) a predominance of degrading 20-year trends in NO3N shifted to a predominance of improving 10-year trends; 2) a predominance of degrading 20- and 28-year trends in TN shifted to roughly equal proportions of degrading and improving 10-year trends; and 3) a predominance of improving 20- and 28-year trends in ECOLI shifted to roughly equal proportions of degrading and improving 10-year trends. In contrast to these changes between trend periods, the predominance of improving trends in NH4N and TP has persisted between all trend periods, and the predominance of degrading trends in MCI scores has persisted from the 20- to 10-year period.

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.

1 Introduction

The New Zealand Ministry for the Environment (MfE) and Statistics New Zealand (Stats NZ) use analyses of river water quality state and trends to inform policy development, and to meet their requirements for environmental reporting on the freshwater domain under the Environmental Reporting Act 2015. In this report, we use "river water quality" as a general term to refer to the physical, chemical and biological variables that are included in river state-of-environment (SoE) monitoring programmes. In a previous report for MfE, we provided water quality state and trends based on monitoring data from 365-577 river monitoring sites (depending on the variable); the time-series for each site × 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 end dates for monitoring sites in the new compilation are in December 2017.

The brief for this work consisted of seven major steps:

1. Compile river water quality data from regional councils, Land and Water Aotearoa (LAWA) and NIWA.
2. Organise and process the data, including error correction, application of reporting conventions and links to spatial data for each site.
3. Assess the suitability of data for 13 physical, chemical, microbial and ecological variables for statistical analyses and apply site inclusion rules.
4. Carry out analyses of water-quality state, including comparisons of state at monitoring sites aggregated by River Environment Classification (REC) land-cover classes, and relationships between water quality state and high-intensity agricultural land cover.
5. Estimate river flows for each site and sampling date, to adjust trend analyses for the extraneous effects of flow variation.
6. Carry out trend analyses using 10-, 20- and 28-year periods ending in late 2017, including comparisons of trends at sites aggregated by REC land-cover classes. The 28-year trend period corresponds to the period of record for NRWQN monitoring sites and a smaller number of long-term council sites.
7. Assess water quality trends at the national scale using two approaches: categorical levels of confidence and a statistical analysis of the proportions of improving trends.

As an additional step, we used the water-quality state dataset to assess river monitoring sites 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). We determined the ECOLI attribute state for monitoring sites and determined the number of river monitoring sites at which the NPS-FM bottom-lines for NO3N and NH4N toxicity were exceeded.

The main components of the current report are detailed methods for data processing and analysis, summaries of water-quality state and trends at the national scale and within four contrasting land-cover classes, and supplementary files with site-specific results and spatial data for each site. The detailed methods and tabulated, site-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 methodology.

The methods used in the current study include several advances on our previous national-scale water-quality trend analyses: 1) a modified statistical procedure was used to determine the directions of trends and the magnitudes of trends (and associated confidence); 2) the previous approach of automatically flow-adjusting water-quality data for all sites was replaced with a site-based approach where water quality-flow relationships were assessed, and flow-adjustments were applied only where strong water quality-flow relationships existed; 3) a new approach was used for aggregating trend directions from multiple sites within a given environmental or spatial domain (e.g., an environmental class), based on the likelihood that water quality was improving for each variable; and 4) a second new procedure was used to make probabilistic estimates of the proportions of improving trends (PIT) for each variable within a domain.

2 Data acquisition, organisation and processing

New Zealand regional and district councils carry out SoE monitoring at > 1000 river 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 ("variables") are measured at these sites. In addition, water quality and biological monitoring had been carried out by NIWA since 1989 at the river sites that make up the National River Water Quality Network (NRWQN).

Council and NRWQN river monitoring data are periodically acquired and federated into databases for preparation of national-scale SoE reports and to investigate monitoring performance. In the current project, the river monitoring database used for the preceding national-scale report (Larned et al. 2015) was updated with data collected between 2013 and December 2017. In this section we describe the water quality variables, data sources and organisation of the river database, and explain the data processing procedures used to derive datasets suitable for state and trend analyses.

2.1 Water quality variables

We assessed river water quality using nine variables that characterise physical, chemical and microbiological conditions, and macroinvertebrate community composition (Table 2-1). Unless otherwise stated, we made no distinction between data collected at regional council sites and NRWQN sites, and we refer to the sites collectively as the "river monitoring network". Data for physical, chemical and microbiological variables were derived from monthly or quarterly samples; macroinvertebrate data came from annual samples.

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

Variable type	Variable	Abbreviation	Units
Physical	Visual clarity	CLAR	m
	Turbidity	TURB	NTU
Chemical	Ammoniacal nitrogen	NH4N	mg/m^3
	Nitrate 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
Microbiological	*Escherichia coli*	ECOLI	cfu/100 mL
Macroinvertebrate	Macroinvertebrate Community Index	MCI	unitless

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

Turbidity (TURB) refers to light scattering by suspended particles. Turbidity is generally measured *in situ* with hand-held nephelometers or with a bench-top nephelometer in a laboratory, using grab samples of water from the monitoring site. Both types of nephelometers are calibrated with standard light-scattering solutions (e.g., formazin), and the sensor reading is not absolute light scattering, but light-scattering relative to the standard solution, in 'nephelometric turbidity units' (NTU). Nephelometric turbidity is generally inversely correlated with visual water clarity (Davies-Colley and Smith 2001), but unlike visual clarity, turbidity measurements do not account for the optical effects (i.e., absorption) of dissolved materials.

The five nutrient species (NO3N, NH4N, DRP, TN and TP) were included because they influence the growth of benthic river algae (periphyton) and vascular plants (macrophytes), and because nitrate and ammonia can be toxic to aquatic organisms at elevated concentrations. Nutrient enrichment from point and non-point source discharges is strongly associated with intensive land use in New Zealand (Larned et al. 2016, Snelder et al. 2018). Nutrient enrichment can promote excessive growth of 'nuisance' periphyton and macrophytes that can, in turn, degrade river habitat, increase daily fluctuations in dissolved oxygen and pH, impede flows, block water intakes, and cause water colour and odour problems. At elevated concentrations, nitrate and ammonia can be toxic to river fish and invertebrates (Hickey 2013, 2014). Mechanisms of nitrate and ammonia toxicity include reduced oxygen transport by haemoglobin, carcinogenic nitrosamine formation, and disruption of ion transport across cell membranes (Camargo et al. 2005).

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

In addition to the physical, chemical and microbiological variables described above, we used the New Zealand Macroinvertebrate Community index (MCI) as a biotic indicator of general river health. MCI scores are calculated using tolerance values for the macroinvertebrate taxa present in benthic samples. Tolerance values are weighting factors that correspond to the relative abundance of taxa along stressor gradients. We used the non-quantitative MCI rather than the quantitative (qMCI) or semi-quantitative (sqMCI) forms of MCI because some council datasets do not include invertebrate abundance data (Stark and Maxted 2007). Non-quantitative MCI scores are based on presence/absence data which are widely available. Physical and chemical variables and ECOLI are measured monthly or quarterly, whereas the invertebrate samples used to calculate MCI scores are generally collected once each summer. Due to the difference in sampling frequency, trend analyses of MCI scores were carried out using a different procedure to that used for the other variables (see Section 3.2.1).

Four additional river water quality variables were considered for analysis: total suspended sediment concentration, areal cover of deposited fine sediment, periphyton biomass, and areal cover of the cyanobacterium *Phormidium*. After assessing the number and geographic distribution of measurements for these variables, and following 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 rivers that have been incorporated in the NPS-FM to provide context to the water quality state analyses. Attribute states or bands are identified for three of the nine variables used in the current report: NO3N, NH4N and ECOLI. The NPS-FM attributes for rivers include two forms of nitrogen, NO3N and NH4N, but these attributes are based on nitrate and ammonia toxicity, rather than their potential to stimulate periphyton and macrophyte growth. However, the NPS-FM does have an attribute based on periphyton, which indirectly requires nitrogen and phosphorus management to prevent excessive growth. NO3N and NH4N concentrations associated with toxic effects are generally much higher than concentrations associated with proliferations of periphyton and macrophytes.

The attribute bands for NO3N and NH4N are defined in terms of annual median and annual maximum values. The bands for ECOLI were updated in the 2017 amendments to the NPS-FM to include five bands for each of four statistics: median, 95th percentile, and the proportion of samples exceeding 260 cfu 100 ml^{-1} or 540 cfu 100 ml^{-1} respectively, expressed as a percent (Ministry for the Environment 2017).

2.2 Data acquisition

River water-quality monitoring data have been acquired periodically from regional councils and NIWA for recent national scale analyses for MfE (Ballantine et al. 2010, Unwin and Larned 2013, Larned et al. 2015). For each successive analysis, data were used to update a database comprising site information, sampling dates and measurements of a wide range of monitoring variables. The database also contains metadata (e.g., methods, alternative variable labels, analytical detection limits). Until the current project, these 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, Land Air Water Aotearoa (LAWA), and NIWA (for NRWQN data); 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 datasets, as described in the next section.

2.3 Data processing

River 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., MCI).

Step 2. 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.

Step 3. Monitoring site spatial information. The following spatial data were associated with each river monitoring site: site name, location and regional council identifier (if available), NZMS260 grid reference (converted from NZTM as necessary), and NZReach number. NZReaches are unique river network section identifiers stored in the River Environment Classification (REC) geodatabase (Snelder et al. 2010). Sites were mapped to reveal and correct georeferencing errors. To provide a measure of national consistency, we used the NRWQN data when an NRWQN site coincided with a regional council river monitoring site.

In addition to the site-specific spatial data listed above, the catchment upstream of each monitoring site was delineated using the digital network in the REC. Each catchment is linked to a wide range of spatial data in the REC. For the current report, the following spatial data were extracted for each site: land cover data from the Land Cover Database Version 4.1 (LCDB4)[1] and the categorical REC classes. The LCDB4 comprises proportional cover of 33 land-cover classes, generated from satellite imagery collected in summer 2012-13). The REC classes are composed of multiple hierarchical levels, each corresponding to a factor that influences river environmental conditions (Snelder and Biggs 2002). In the current study, we grouped river monitoring sites into REC land-cover classes and pooled across the three higher hierarchical levels (climate, topography and geology). This approach results in substantial variation in water quality within land-cover classes, while ensuring that classes with relatively few monitoring sites have sufficient data for statistical analyses. In previous studies of New Zealand river water quality, REC land-cover classes were shown to account for a substantial level of variability in some water-quality variables (Larned et al. 2004, 2016). As in the previous studies, four land-cover classes were used: pastoral (P), exotic forest (EF), urban (U) and a natural (N) category that incorporates the indigenous forest, tussock, scrub and bare-land categories. Following the classification rules in Snelder and Biggs (2002), river sites were classified as exotic forest or natural if those categories accounted for the largest proportion of the upstream catchment area, unless pastoral land exceeded 25% of the catchment, in which case the segment was classified as pastoral, or where urban land exceeded 15% of the catchment, in which case the segment was classified as urban.

Step 4. Comparable field and laboratory methods. The next data processing step was to assess methodological differences between data sources in the measurement of water quality variables. For most 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.

[1] lris.scinfo.org.nz

The data produced by multiple procedures used to measure ECOLI, NO3N, CLAR, TURB and MCI were pooled, assuming that the different procedures gave comparable results. In contrast, some procedures used to measure TN and TP are unlikely to give comparable results. Most councils and the NRWQN 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) from which TN is calculated as TKN + NO3N. At least one council uses filtered samples for the data labelled TN and TP, although the results derived from filtered samples are more correctly labelled total dissolved nitrogen and total dissolved 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.

At the completion of the data processing steps, our dataset comprised 1304 river monitoring sites, with values for some or all of the variables listed in Table 2-1.

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 analytical "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 rivers often include DRP, TP and NH4N measurements that are below detection limits, and ECOLI and CLAR 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, and the procedure used for trend analyses is set out in Section 3.2.3.

Table 2-2: Measurement procedures for water quality variables. MCI procedures are from Stark et al. (2001). Where multiple measurement procedures existed, "Procedures retained" refers to data generated by a preferred procedures that were retained for analysis in this study.

Variable	Measurement procedure(s)	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, filtered samples	Molybdenum blue colourimetry, unfiltered samples
CLAR	Black-disk Horizontal clarity tube	Both procedures (presumed to give comparable results)
TURB	Field or laboratory nephelometer	Both procedures (presumed to give comparable results)
MCI	Collection procedures C1, C2, C3, C4 Processing procedures P1, P2, P3	All procedures (presumed to give comparable presence/absence data for calculating non-quantitative MCI scores

3 Analysis methods

3.1 Water quality state analyses

3.1.1 Time period for state analyses

The statistical robustness with which water quality state may be determined depends on the variability in the measurements between sampling occasions and the number of observations. This is particularly important for sites 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 site has been correctly classified as either passing or failing a guideline) increases as the number of observations increase. As a general rule, the increase in the confidence with which estimates of population statistics may be determined slow for sample sizes greater than 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 that was 30 or more for many sites and variable combinations (i.e., five years of monthly observations, where observations that are counted for some variables are for flows below the 50th percentile). 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 site × variable combinations in the state analyses to those with measurements for at least 90% of the sampling intervals in that period (at least 63 of 70 months or 18 of 20 quarters). Site × variable combinations that did not comply with these rules were excluded from the state analysis.

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 river site and variable, we characterised the current state using percentiles (5th, 20th, 25th, 50th, 75th, 80th, 95th) derived from the distribution of measured values for the period 2013 to 2017 (inclusive), with the exception of MCI, where we used the time period 1 July 2012 – 30 June 2017 (to prevent splitting summer samples into two calendar years). All percentiles were calculated using the Hazen method.[2]

3.1.4 Relationships between water quality state and catchment land cover

We used linear regressions to relate water-quality state to proportions of high-intensity agricultural land cover in the catchments upstream of the monitoring sites. The proportion of high-intensity agricultural land cover was defined as the sum of proportional land cover in three LCDB4 classes (high-producing exotic grassland, short-rotation crops, and orchards and vineyards). The same composite classification for high-intensity agricultural land cover was used in previous national-scale water-quality analyses (McDowell et al. 2013, Larned et al. 2016). In addition to high-intensity agricultural land cover, we considered urban and natural land cover as predictor variables. However, examination of land cover data indicated that the range of urban land cover represented by the sites in our dataset was inadequate (> 90% of sites had < 10% urban cover), and natural land cover was strongly negatively correlated with high-intensity agricultural land cover (r = -0.68, n = 1304). All variable values were log-transformed to improve the normality of residuals.

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 × site combination that met the inclusion rules set out below, for three different time periods:10, 20 and 28 years. With the exception of MCI, each of the time periods ended in late December 2017. For MCI, we used the time period 1 July 2012 – 30 June 2017, in order to capture complete summer sampling seasons.

The processed dataset had a range of start and end dates, a range of sampling frequencies, and different numbers of missing values. Site inclusion rules (i.e., filtering rules) were used to ensure that for each variable, data for each site would enable robust trend assessment. We used the filtering rules suggested by Helsel and Hirsch (1992), which restricted trend assessment in a given time period only to those site and variable combinations where measurements were available for at least 90% of the years and at least 90% of seasons.

For assessments of trends in water quality variables other than MCI, we used seasons defined by months preferentially, and quarters when there were insufficient monthly observations. The trend analysis procedure accounted for seasonal variability in these monthly and quarterly data. Macroinvertebrates are generally sampled annually at SoE monitoring sites, so these data do not represent seasonal variability. For some sites and variables, more than one sample within some seasons exist, and for some sites, MCI scores were available for more than one invertebrate sample within some years. In these cases, we used the median of the values for the season (or the year for the invertebrate samples) to ensure consistent statistical power across sites. We note that when more than one sample in a season exist, all samples can be used in a trend analysis, increasing statistical power and potentially providing different results. However, because our analyses are used to make regional comparisons and to contribute to spatial models, we elected to ensure that the site-specific analyses had consistent statistical power.

[2] (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).

3.2.2 Analyses of site-specific trends

Trend magnitude and confidence in trend direction

The statistical analyses of water quality trends were performed using the LWP-Trends library, which comprises functions coded in the R statistical programming language. Statistical trend analysis 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 nonparametric 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).

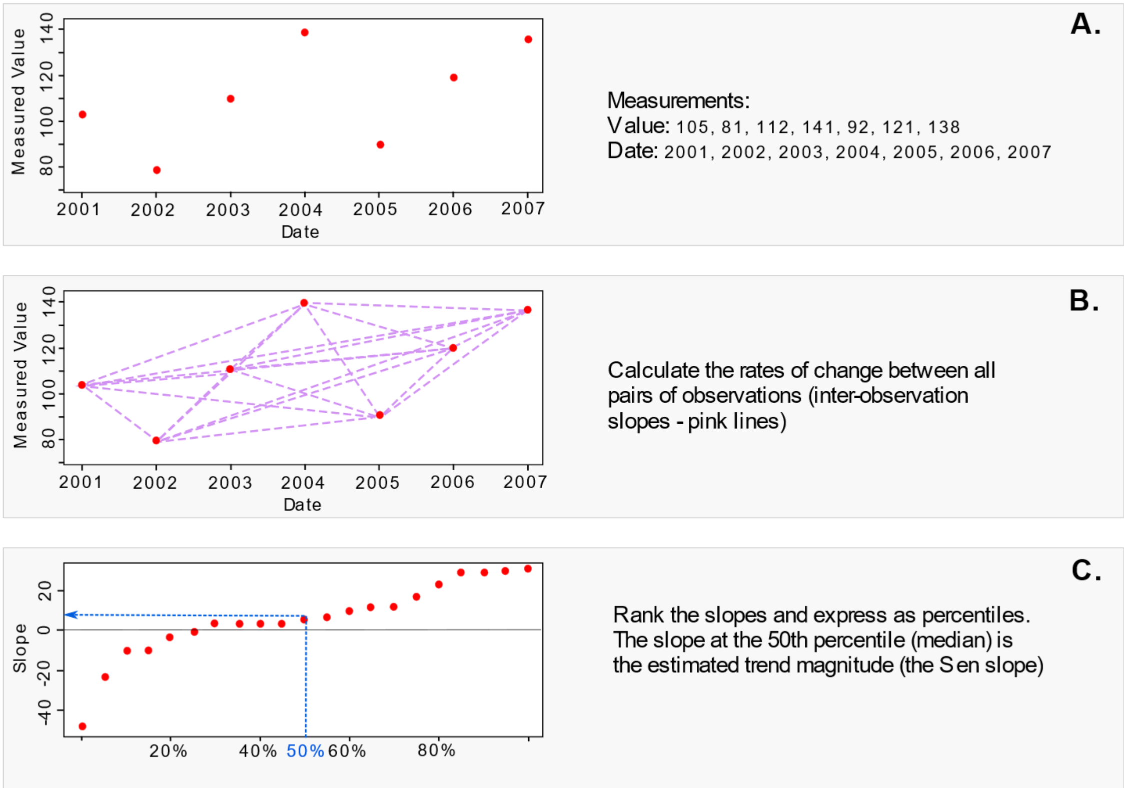

Figure 3-1: Pictogram of the steps taken in the trend analysis to calculate the Sen slope. which is used to characterise trend magnitude in the time-series of data for each site × variable combination.

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 except MCI. 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 annual MCI scores and all other site × 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 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).

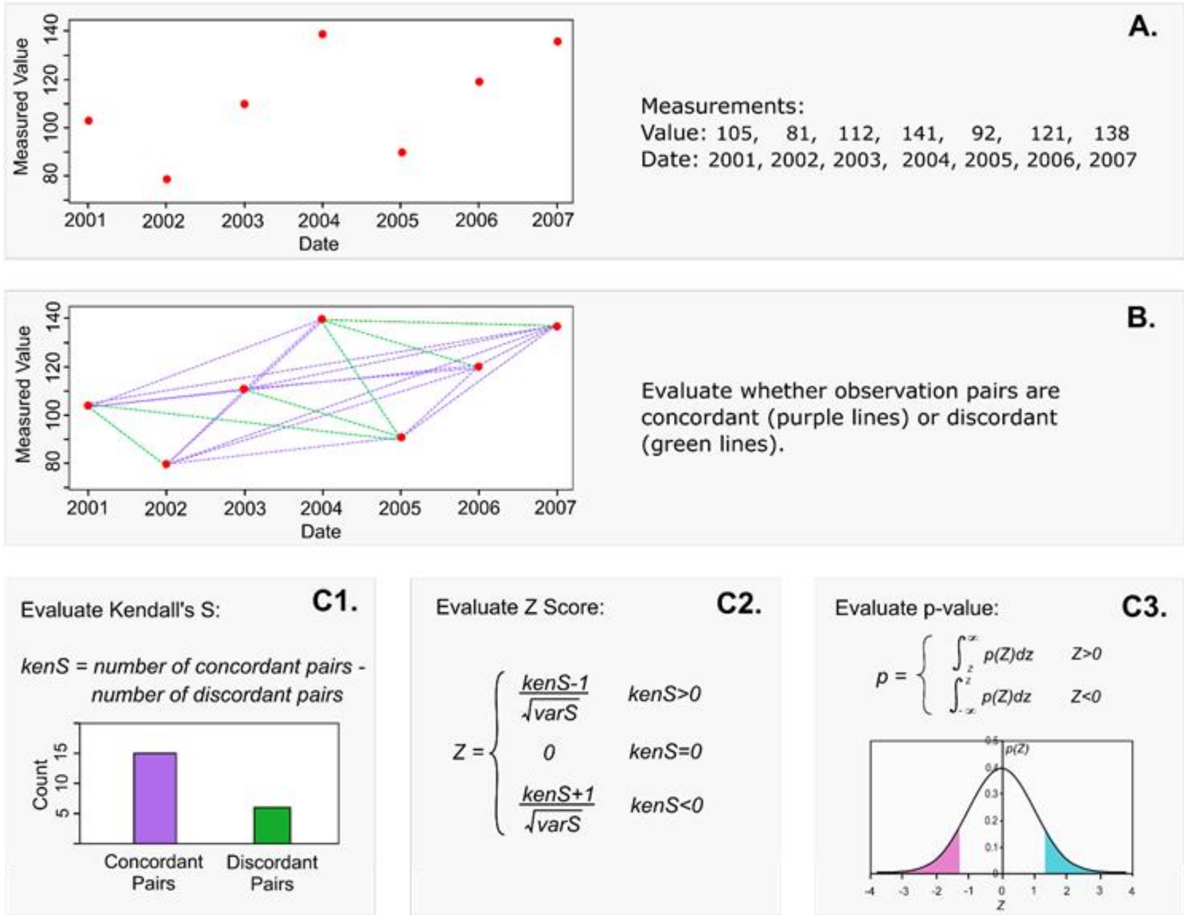

Figure 3-2: Pictogram of the steps taken in the trend analysis to calculate the Kendal S statistic and its p-value, which are used to characterise confidence in trend directions.

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 ≤ 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 site × 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 site 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 from 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 review). 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[3], 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 site × 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 sites are more reliable, irrespective of the proportion of censored observations. In turn, the methods used to aggregate site trends are robust, because these procedures are based on levels of the confidence in the trend directions at individual sites (discussed in detail in Section 3.2.6). 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 site × variable combinations based on the proportion of censored observations (i.e., sites with >15% censored values were not excluded

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

as in previous studies). This had the advantage of preserving a larger number of sites in each analysis and maximising spatial coverage. We did exclude some site × 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.

3.2.5 Flow-adjustment procedures

Flow rate at the time that a river water quality measurement is made can affect the observed values for some water quality variables because values may decrease systematically with increasing flow (e.g., dilution effects on contaminant concentrations), or increase with increasing flow (e.g., wash-off effects on contaminant concentrations) (Smith et al. 1996). Different mechanisms may dominate at different sites so that the same water quality variable can exhibit positive or negative relationships with flow (Snelder et al. 2018).

Adjusting water quality observations to account for the effect of flow (hereafter flow adjustment) or any other covariate decreases variability and increases statistical power (i.e., increases the likelihood of detecting a trend with certainty; Helsel and Hirsch, 1992). In addition, a trend in a water-quality variable may arise because there is a relationship between time and flow on the sample occasions (i.e., a trend in the flow on sample occasions such as increasing or decreasing flow with time). Flow adjustment may change the direction and/or magnitude of a trend in a water-quality variable. Previous studies have provided trend analyses based on both flow adjusted and raw data (e.g., Ballantine et al. 2010; Larned et al. 2015).

Flow adjustment requires that water quality observations are associated with the flow at the time of sampling. In this study, flow estimates for each monitoring site and date were based on measured or modelled daily average flow. For monitoring sites with flow recorders on the same reach, daily average flows were calculated from measured flow. However, most river monitoring sites iar not on a reach with a flow recorder, and daily average flows for these sites were estimated by hydrological modelling. We used predicted flows from the TopNet hydrological model, corrected using flow-duration curves, which were in turn estimated with random forest models (Booker and Snelder 2012; Booker and Woods 2014). TopNet is a spatially distributed time-stepping model that combines water-balance models with a kinematic wave channel-routing algorithm (McMillan et al. 2013).

In this study we followed the recommendations of Snelder (2018) concerning flow adjustment of water quality variables. In particular, we did not rely on the automated flow adjustment procedure used by Larned et al. (2015), because unsupervised fitting of regression models to relationships between water quality observation relationships and flows can result in the selection of unreliable models. Instead, we inspected the models and used expert judgement to choose the most suitable model based on the homoscedasticity (constant variance) of the regression residuals and plausibility of the shape of the fitted model. We considered LOESS, GAM and log-log models. In this study, log-log models were found to be the most appropriate for all site variable combinations for which there were detectable relationship between water column measures and flow; the LOESS and GAM methods generally produced implausible relationships due to their flexibility. When the relationship between flow and a water quality variable was poor, no flow adjustment was performed. Given the large number of site-variable combinations, we applied a general rule to define whether flow adjustment would be performed. Where the log-log relationship yielded an R^2 value greater than 20%, we flow adjusted the data. For poorer fits, we used the raw data (i.e., did not flow adjust). The R^2 threshold was determined from visual examination of all flow-water-quality relationships and was selected as a threshold that provided a balance between reducing concentration variance due to the covariate relationship, and the risk of selecting implausible models of the relationship between water column measures and flow.

3.2.6 Aggregation of site trends

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

The distributions of trend magnitude across sites were characterised using box and whisker plots of the relative Sen slope estimates (RSSE) and relative seasonal Sen slope estimates (RSSSE). Sen slopes were relativised by dividing the SSE and SSSE values by the duration of the trend period to give estimates of temporal change in % yr^{-1}.

We used three different approaches for aggregating trend directions. For each approach, 'improving trends' corresponded to decreasing trends in nutrient and ECOLI concentrations and turbidity, and increasing trends in CLAR and MCI. Conversely, 'degrading trends' corresponded to increasing trends in nutrient and ECOLI concentrations and turbidity, and decreasing trends in CLAR and MCI.

In the previous national-scale water-quality trend analyses, site-specific trends were aggregated by tabulating the numbers of sites 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 review). 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 Figure 3-1. Note that confidence categories for degrading trends are the complement of the confidence categories for improving trends shown in Figure 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 categorical levels of confidence in Figure 3-1 were used to aggregate the site-specific trends in each water quality variable. Each site trend was assigned a categorical level of confidence that the trend was improving according to its evaluated probability and the categories shown in Figure 3-1

The categorical levels of confidence in Figure 3-1 were also used to map trends in each water quality variable at each qualifying site, and to aggregate the site-specific trends in each water quality variable at the national scale. We then calculated the proportion of sites 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 proportion of improving trends (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 proportion of improving site-specific trends (PIT) within a spatial or environmental domain. For a given water-quality variable, the trends at multiple monitoring sites distributed across a domain of interest can be assumed to represent independent samples of the population of trends, for all sites within that domain.

Let the sampled sites within this domain be indexed by s, so that $s \in \{1, \dots, S\}$ and let I be a random Bernoulli distributed variable which takes the value 1 with probability p and the value 0 with probability $q = 1 - p$. Therefore, $I_s = 1$ denotes an improving trend at site $s \in \{1, \dots, S\}$ when the estimated $p_s \geq 0.5$ and a degrading trend as 0 when $p_s < 0.5$. Then, the estimated proportion of sites 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 site 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, within a spatial or environmental domain, 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[4] 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 water quality variables and for domains of interest defined by the entire country, and by the four REC land-cover classes defined in Section 2.3, exotic forest, natural, pastoral, and urban.

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

4 Results – river state

Between 587 and 887 river monitoring sites met the filtering rules for the state analysis of nutrients, ECOLI, CLAR and MCI; the number of qualifying sites varied by water quality variable and by REC land cover class (Table 4-1). The geographic distribution of sites is shown in Figure 4-1. The sites are reasonably well-distributed, although there are gaps in the central North and central South Islands. The complete set of state analysis results is provided in the supplementary file "StateResultsWaterQuality_2013-2017_INCLUSIVE_27_Nov18.csv".

The distributions of site-median values of the nine water quality variables for the 2013-2017 period are summarized as box-and-whisker plots, with sites grouped by REC for the land cover (Figure 4-2). The plots in Figure 4-2 indicate that water quality state (i.e., site medians for nutrients, ECOLI, MCI and CLAR) was highly variable, with some of the variation explained by the land cover classes. Sites in the different land cover classes had different water quality characteristics, both in terms of their central tendencies (indicated by the median of the median site values) and their variation (indicated by the boxes and whiskers in Figure 4-2). For example, median TN was highest and least variable in the urban class. The lowest land-cover class median for TN and with large variability occurred in the natural class. Median concentrations of all nutrients and ECOLI were lowest and CLAR and MCI highest in the natural class. In contrast, nutrient and ECOLI concentrations were highest in the urban class, closely followed by the pasture class.

The distribution of ECOLI concentration percentiles (5th, 20th, 50th, 80th and 95th) are shown in Figure 4-3, and the distribution of the 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.The site median of each ECOLI concentration percentile varied across REC land cover classes in the same order (from highest to lowest): urban, pastoral, exotic forest, natural (Fig. 4-3). The medians of site G260 and G540 values also varied across land-cover classes in the same order (from highest to lowest): urban, pastoral, exotic forest, natural (Fig. 4-4).

Table 4-1: Number of river monitoring sites by REC land cover class and water quality variable that were included in the state analyses of nutrients, ECOLI, CLAR, TURB and MCI. The site numbers shown refer to sites where less than 50% of the values for a variable were censored, and ≥ 30 values were available, distributed over at least four of the five years from 2013 to 2017.

Variable	Total	Exotic Forest	Natural	Pasture	Urban
CLAR	587	23	167	380	17
DRP	877	31	206	578	62
ECOLI	866	31	205	568	62
MCI	832	29	266	494	43
NH4N	882	31	207	582	62
NO3N	855	30	204	560	61
TN	764	19	180	521	44
TP	740	19	166	511	44
TURB	878	31	207	580	60

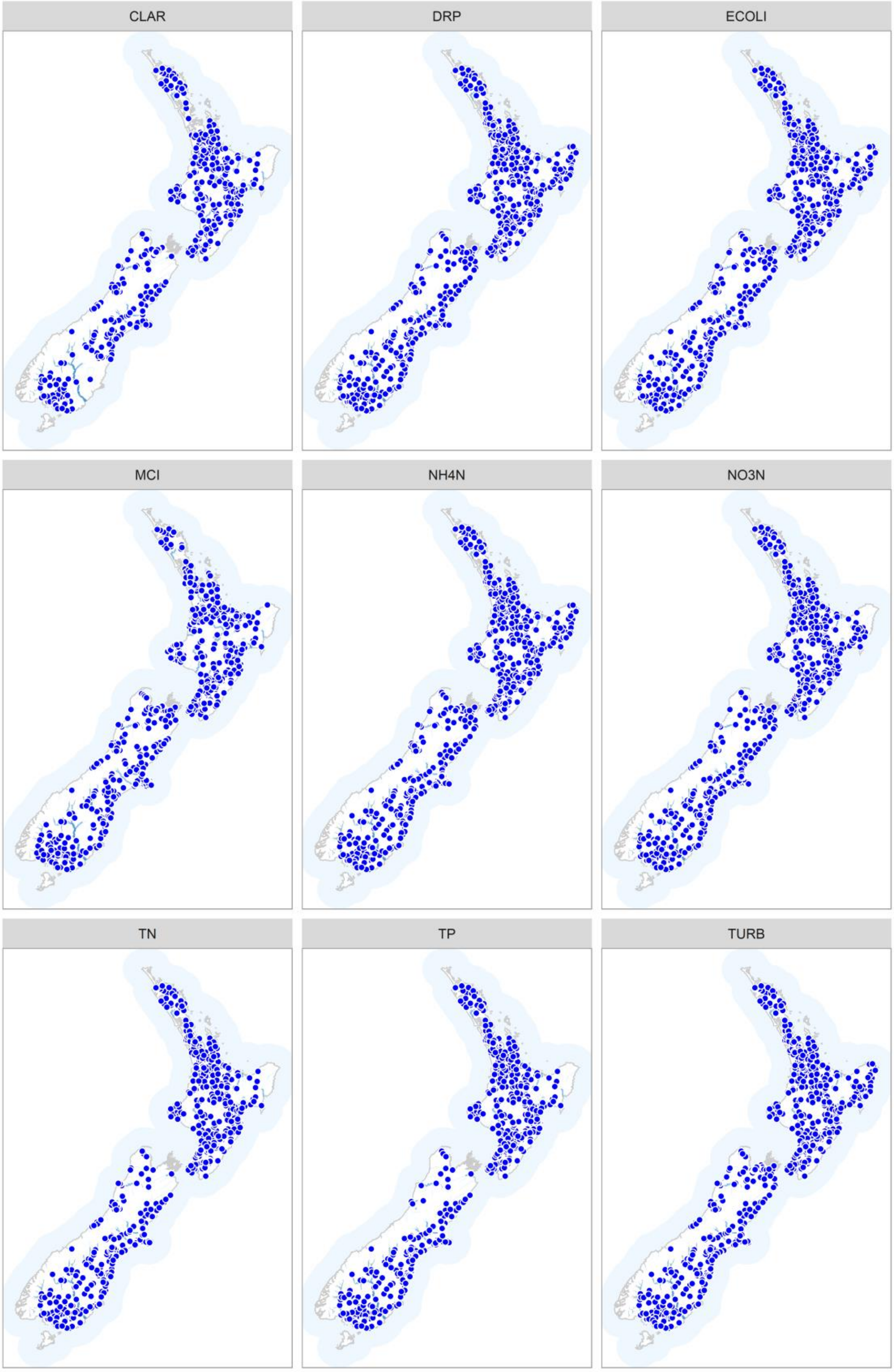

Figure 4-1: River water quality monitoring sites used for state analyses of nutrients, ECOLI, CLAR, TURB and MCI.

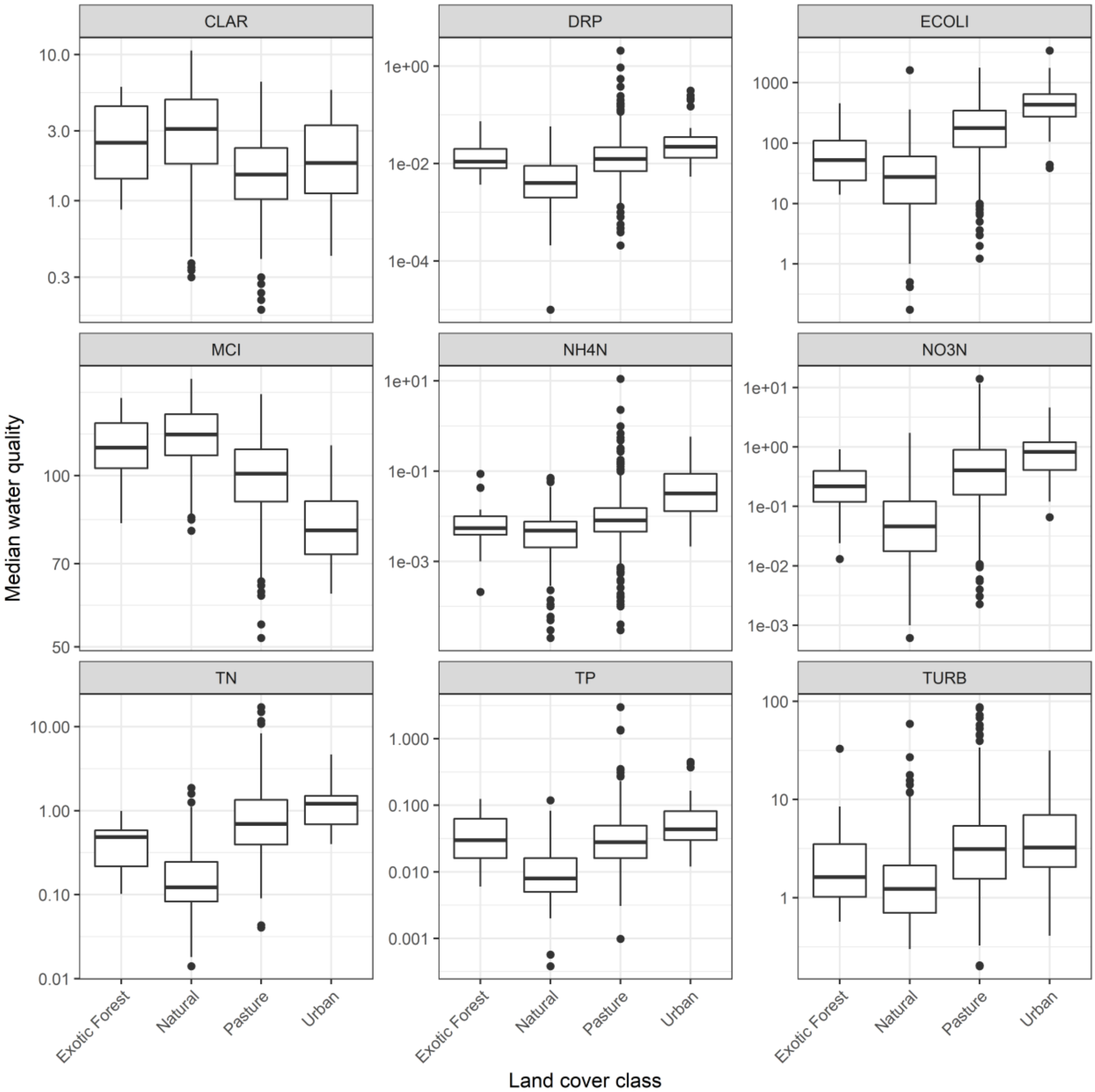

Figure 4-2: River water quality state in REC land cover classes. Box-and-whisker plots show the distributions of monitoring site medians within land cover classes. For y-axes units of measure refer to Table 2-1. Black horizontal line in each box indicates the median of site medians, box indicates the inter-quartile range, whiskers indicate the 5th and 95th percentiles, and closed circles indicate outliers. Note log-scale on Y-axes.

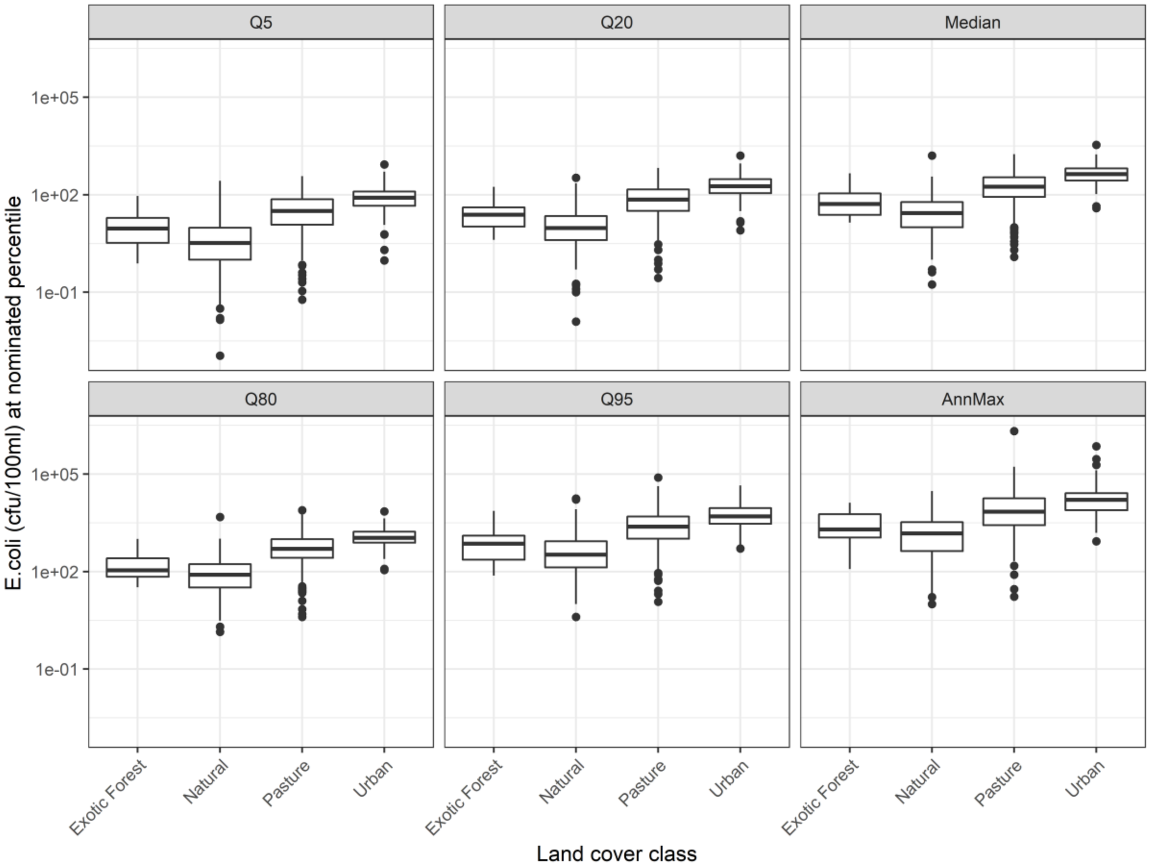

Figure 4-3: ECOLI concentrations in REC land cover classes. Box-and-whisker plots show the distributions of monitoring site percentiles within land cover classes. Black horizontal line in each box indicates the median of site percentiles, box indicates the inter-quartile range, whiskers indicate the 5th and 95th percentiles, and open circles indicate outliers. Note log-scale on Y-axes.

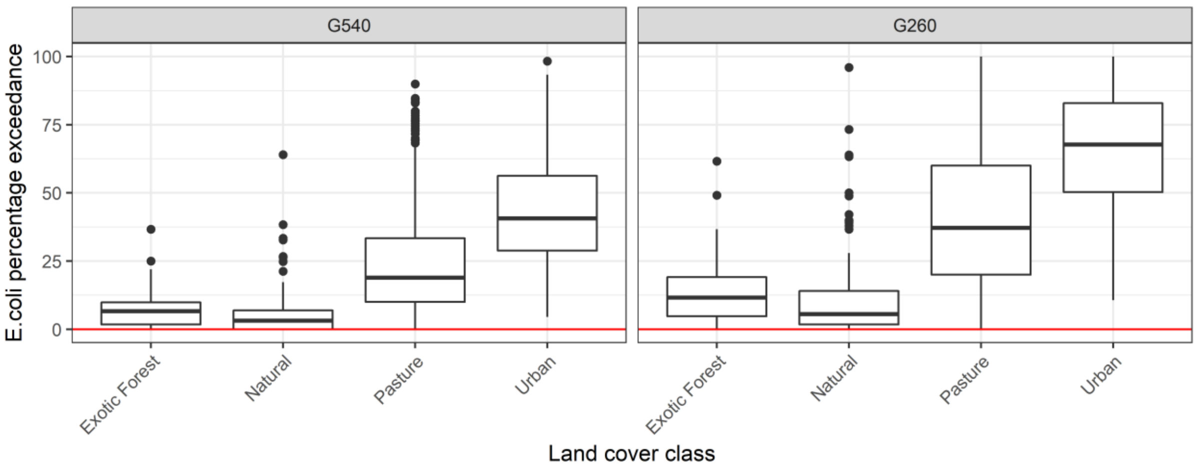

Figure 4-4: ECOLI percent exceedance in REC land cover classes. Box-and-whisker plots show the distributions of percentage exceedance over 540 cfu 100 ml^{-1} (G540) and 260 cfu 100 ml^{-1} (G260) at river monitoring sites within land cover classes. Black horizontal line in each box indicates the median of percent exceedances, box indicates the inter-quartile range, 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 NO3N, NH4N and ECOLI. The NO3N attribute has four numeric attribute states (or bands A-D) for annual median and annual 95th percentile values. The C/D band boundaries correspond to national bottom-lines for the annual median (6.9 mg N L^{-1}) and annual 95th percentile (9.8 mg N L^{-1}).

The NH4N attribute has numeric attribute states for annual median and annual maximum values, which apply to NH4N concentrations adjusted to pH 8. The national bottom-lines for pH-adjusted NH4N are 1.30 mg N L^{-1} for the median and 2.20 mg N L^{-1} for the maximum. To compare NH4N concentrations at monitoring sites to the NPS-FM bottom-line, we adjusted the measured NH4N concentrations to pH 8 using the conversion ratios in the Draft Guide to Attributes in the NPS-FM[5]. The adjustments were limited to monitoring site–date combinations where both pH and NH4N were measured.

Approximately 1% of sites in the pastoral land cover class exceeded the NO3N annual median bottom line for toxicity and four of those sites also exceeded the 95th percentile bottom line (Table 4-2). No sites in the other land use classes exceeded NO3N bottom lines. Less than 1% of sites in the pastoral land cover class exceeded the pH-adjusted NH4N median bottom line, and 1.7% of pastoral sites and 1.7% of urban sites exceeded the pH-adjusted NH4N maximum bottom line (Table 4-2). No sites in the other land use classes exceeded NH4N bottom lines.

Table 4-2: Number and proportions of river monitoring sites that exceeded the national bottom-lines for the NPS-FM NO3N and NH4N toxicity attributes.

		NO3N	
Land-cover class	Sites	Sites exceeding median (% sites)	Sites exceeding 95th percentile (% sites)
Exotic forest	30	0	0
Natural	206	0	0
Pastoral	567	6 (1%)	4 (0.7%)
Urban	61	0	0
		NH4N	
Land-cover class		**Sites exceeding median (% sites)**	**Sites exceeding maximum (% sites)**
Exotic forest	28	0	0
Natural	160	0	0
Pastoral	532	2 (0.4%)	9 (1.7%)
Urban	58	0	1 (1.7%)

[5] (http://www.mfe.govt.nz/publications/fresh-water/draft-guide-attributes-appendix-2-national-policystatement-freshwater).

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 sites 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 sites 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".

The distribution of river monitoring sites across the ECOLI attribute states within each land cover class is shown in Table 4-3. A total of 375 monitoring sites had ≥ 60 sampling dates in the 2013-2017 period. Of those sites, the majority in the natural and exotic forest land-cover classes were in the A attribute state, although the exotic forest class was limited to 14 sites. Most of the sites in the pastoral land-cover class were in the D and E attribute states, and most of the sites in the urban land-cover class were in the E attribute state.

Table 4-3: River monitoring sites in the ECOLI attribute states specified in the NPS-FM. Values are numbers of river monitoring sites, and proportions of sites within each attribute state in parentheses).

Attribute State	Land-cover class				Total
	Exotic forest	Natural	Pastoral	Urban	
A	7 (8.9%)	54 (68.4%)	18 (22.8%)	0 (0%)	79
B	3 (7.3%)	12 (29.3%)	25 (61%)	1 (2.4%)	41
C	0 (0%)	1 (16.7%)	5 (83.3%)	0 (0%)	6
D	3 (2.6%)	14 (12.3%)	92 (80.7%)	5 (4.4%)	114
E	1 (0.7%)	3 (2.2%)	94 (69.6%)	37 (27.4%)	135

The regression results indicated that the concentrations of each nutrient and ECOLI increased, and MCI scores and visual clarity decreased, with increasing proportions of high-intensity agricultural land cover in the upstream catchment (Figure 4-5). Agricultural land cover explained 8%–47% of the variation in log-transformed water-quality variables; these relationships were strongest for median TN, NO3N, TP and ECOLI concentrations and MCI scores.

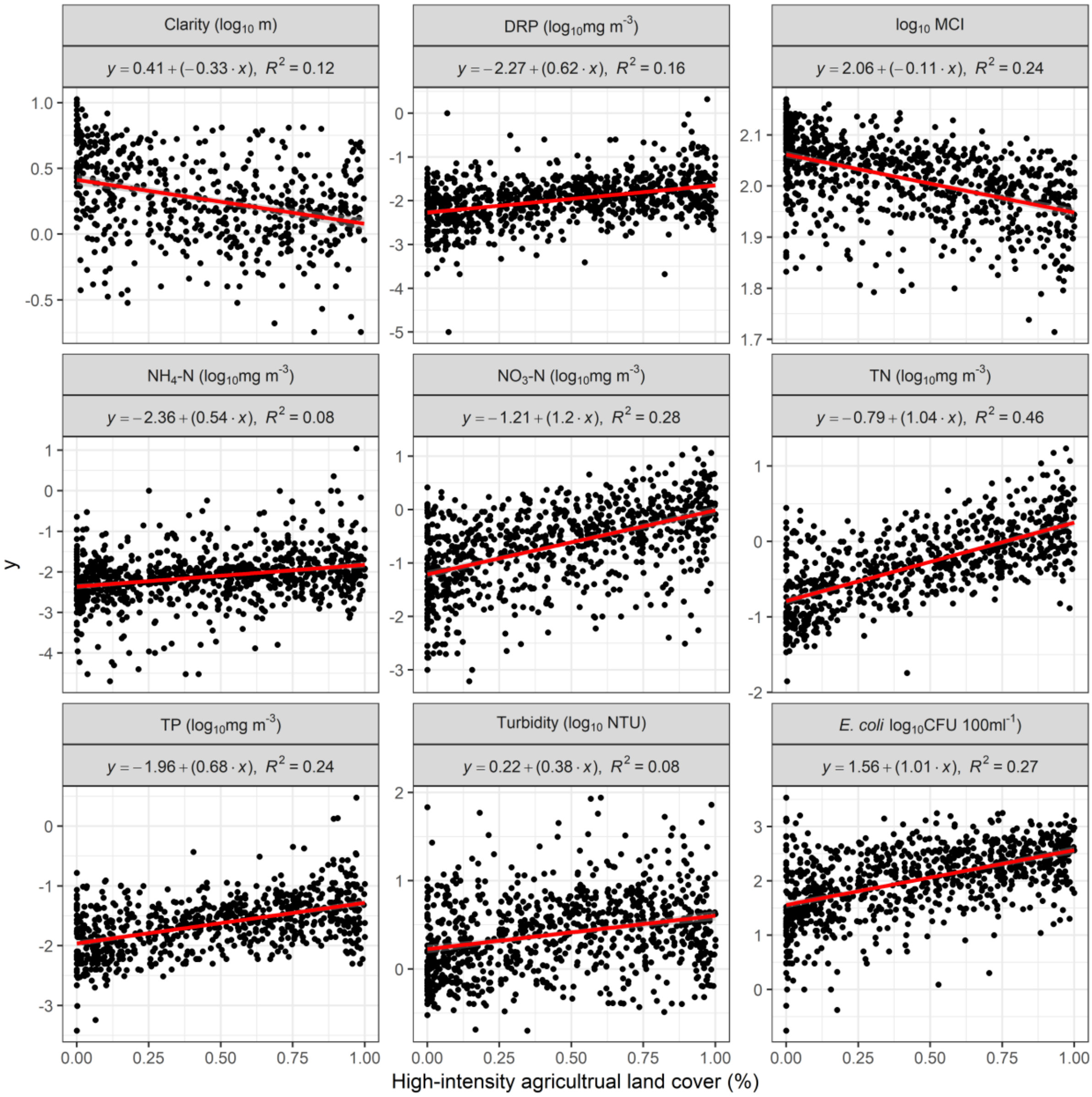

Figure 4-5: Relationships between median water-quality state and proportion of a catchment under high-intensity agricultural land cover in the catchments above monitoring sites in the state data set. Solid lines indicate least squares linear regression models.

5 Results – river trends

5.1 Ten-year trends (2008-2017)

Between 457 and 791 river monitoring sites met the filtering rules for the 10-year trend analysis of nutrients, ECOLI, TURB, MCI and CLAR (Table 5-1). The qualifying sites were reasonably well-distributed geographically (Figure 5-1), with gaps in the central North and South islands and the West Coast. All site locations, land cover classes and numbers of sampling dates are included in the supplementary file "RiverTrends_AllPeriods_FlowAdjusted_27_Nov18.csv".

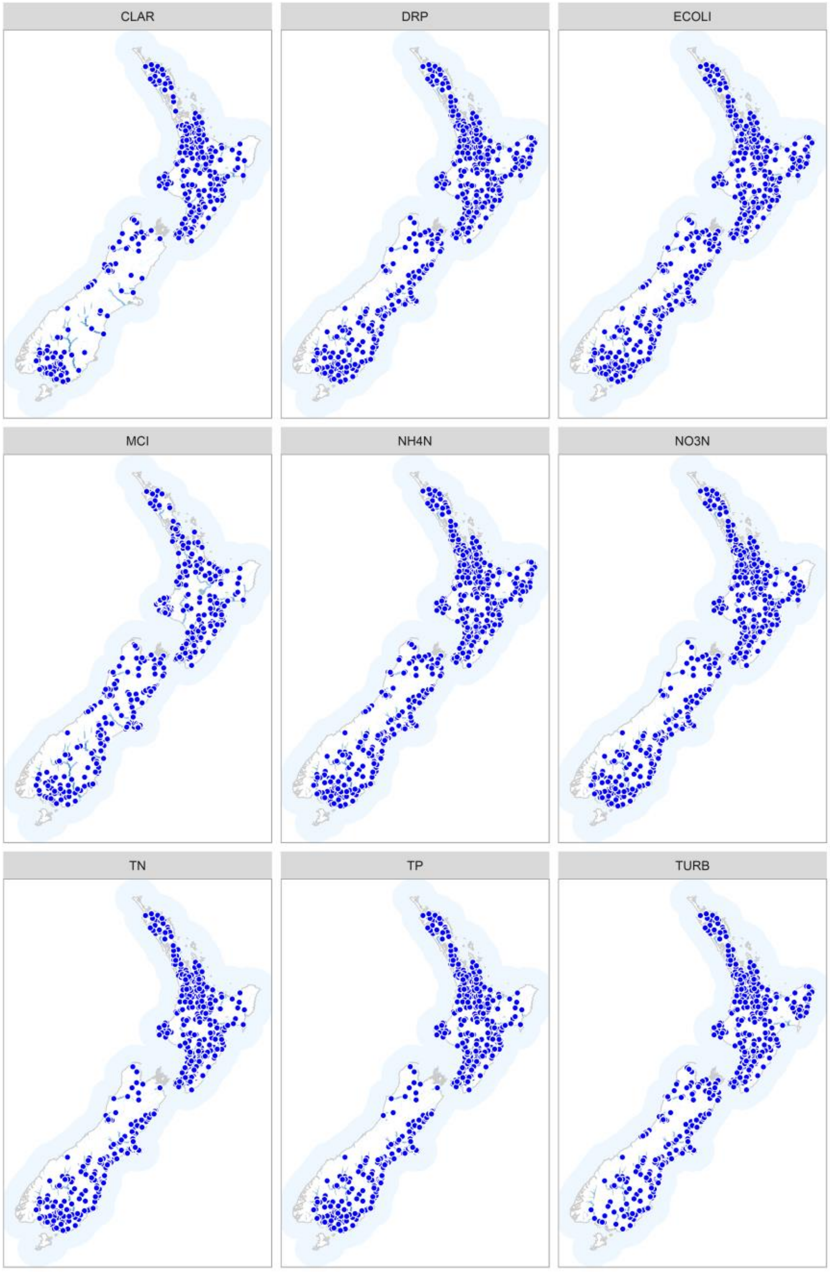

Figure 5-1. River water quality monitoring sites used for 10 year trend analyses of nutrients, ECOLI, CLAR, TURB and MCI.

Table 5-1: Number of river monitoring sites by REC land cover class and water quality variable included in the 10-year trend analyses of nutrients, ECOLI, CLAR, TURB and MCI. The site numbers shown refer to sites that met the site inclusion requirements in Section 3.2.1 (measurements were available for at least 90% of the years and at least 90% of seasons).

Variable	Number of sites				
	Total	Exotic Forest	Natural	Pasture	Urban
CLAR	457	24	131	291	11
DRP	771	30	181	504	56
ECOLI	753	31	186	480	56
MCI	573	19	196	329	29
NH4N	791	30	186	519	56
NO3N	752	30	179	488	55
TN	663	19	154	452	38
TP	664	19	151	455	39
TURB	718	28	172	467	51

5.1.1 Trend magnitude

Box and whisker plots were used to summarise the estimated trends for each of the water quality variables for the 10-year period from 2008 – 2017 across the four land cover classes (Figure 5-2). All estimated trends are included in these plots, irrespective of the confidence in direction (as defined in Section 3.2.2). These plots indicate that land cover classes did not account for a substantial amount of the variation in trends for any variable. This contrasts with the state analyses of river variables, where water-quality state clearly varied between land cover classes (Figures 4-2, 4-3 and 4-4). Median trend magnitudes were largest for CLAR, DRP, TP and TURB in the urban land-cover class; in each case the trend direction indicated improving conditions.

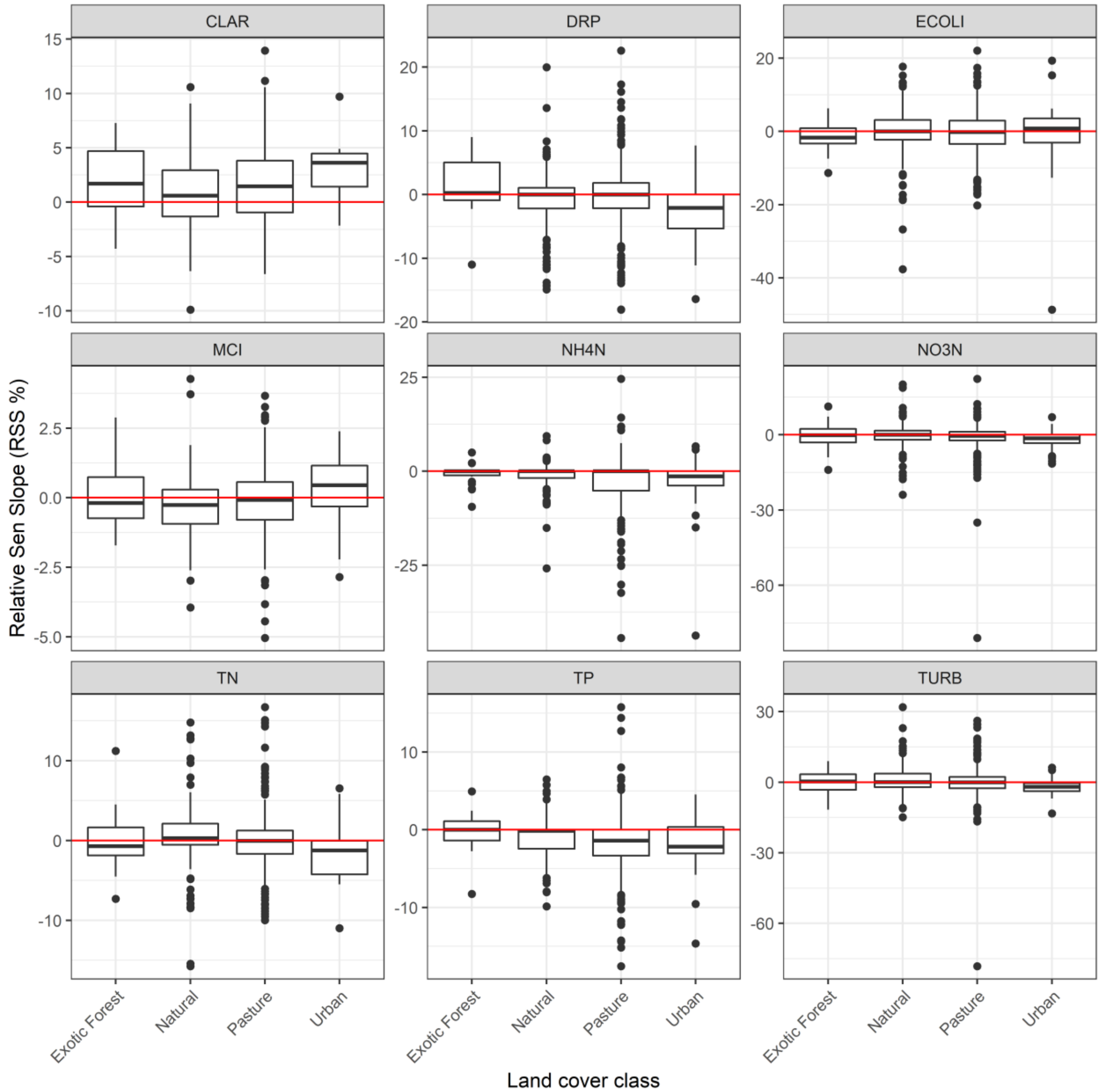

Figure 5-2. Summary of 10-year flow adjusted trends. Box-and-whisker plots show the distributions of site trends within REC land cover classes. Black horizontal line in each box indicates the median of site trends, box indicates the inter-quartile range, whiskers indicate the 5th and 95th percentiles, and open circles indicate outliers.

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 (54 to 88% of the site 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). In addition, a small number of site × variable combinations (0-7%) were not analysed in the current study due to very high proportions of censored values. For the remaining trends, the proportions of sites with improving trends in CLAR, NH4N, NO3N and TP were substantially larger than the proportion with degrading trends, and the proportions of sites with degrading trends in MCI was larger than the proportion with improving trends. However, the great majority of MCI trend (88%) were indeterminant.

Some geographic patterns in 10-year trends are apparent in Figure 5-3. For example, the majority of degrading DRP trends and the majority of improving CLAR trends are in the North Island (although this pattern is influenced by the poor sampling distribution in the South Island). NO3N trends in Northland and Hawke's Bay are dominated by improving trends. NH4N trends in the lower North Island and in Canterbury are dominated by improving trends.

Table 5-2: Numbers and proportions of sites in four trend categories for 10-year, flow-adjusted trends. The "Not analysed" category corresponds to site × variable combinations that met the site inclusion requirements in Section 3.2.1 (measurements are available for at least 90% of the years and at least 90% of seasons), but did not meet the censored data requirements in Section 3.2.4 (i.e., there were < 5 non-censored values and/or < 3 unique non-censored values). The classification of the remaining site 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).

Variable	Trend category			Not analysed
	Degrading	Improving	Indeterminant	
CLAR	29 (6%)	141 (31%)	287 (63%)	0 (0%)
DRP	156 (20%)	180 (23%)	414 (54%)	21 (3%)
ECOLI	104 (14%)	104 (14%)	540 (72%)	5 (1%)
MCI	49 (9%)	21 (4%)	503 (88%)	0 (0%)
NH4N	47 (6%)	218 (28%)	467 (59%)	59 (7%)
NO3N	111 (15%)	190 (25%)	448 (60%)	3 (0%)
TN	144 (22%)	129 (19%)	387 (58%)	3 (0%)
TP	48 (7%)	201 (30%)	414 (62%)	1 (0%)
TURB	122 (17%)	111 (15%)	485 (68%)	0 (0%)

Figure 5-3: Water quality monitoring sites classified by 10-year trends. Derived using the three categories defined by Larned et al. (2015): degrading, improving, indeterminant.

5.1.3 Probability of improvement

The levels of confidence listed in Table 3-3 were used to categorise the probability of an improving 10-year, flow-adjusted trend in each site-variable combination. The spatial distributions of categorised individual sites 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 sites previously categorised as indeterminant (shown in Figure 5-3) were about equally divided into likely and unlikely to improve level-of-confidence categories (Figure 5-4). However, in the cases of NH4N and TP, most sites that were previously categorised as indeterminant were classed as likely to be improving (i.e., the sites were subsequently placed in the likely and very likely level-of-confidence categories.

Figure 5-4: Water quality monitoring sites 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 sites 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 sites for which 10-year, flow-adjusted trends indicated improvement at the nine categorical levels of confidence defined in Table 3-3. These plots provide national-scale assessments of the relative proportions of improving versus degrading sites, based on the relative amounts of green and red in each bar.

The national-scale proportions of improving trends (PIT) and their confidence intervals are summarised in Table 5-3. The 10-year PIT statistics ranged from 45-72%. MCI had a majority (i.e., <50%) of degrading trends, at the 95% confidence level. Five of the variables had a majority of improving (i.e., >50%) trends, at the 95% confidence level (CLAR, DRP, NH4N, NO3N and TP). The remaining three variables had 95% confidence intervals for the PIT that included 50% (ECOLI, TN, TURB), 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. For five of the nine water quality variables (CLAR, DRP, NO3N, TN, TURB), the PIT statistic was highest (i.e., the greatest proportion of improving trends) in the urban land-cover class. In contrast, the PIT statistic was < 50% for ECOLI in the urban land-cover class. The PIT statistics also indicated that there were a majority of degrading trends in ECOLI, MCI, TN and TURB at sites in the natural land cover class.

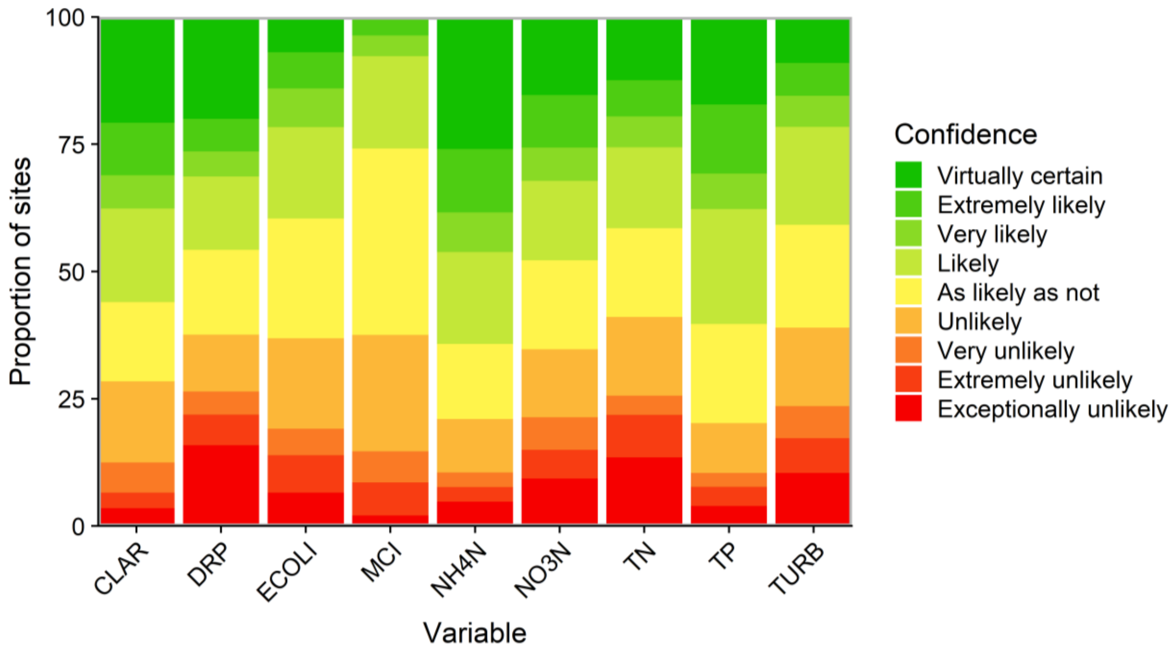

Figure 5-5: Summary plot representing the proportion of sites with improving 10-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with improving trends at levels of confidence defined in Table 3.1.

Table 5-3: **Proportions of improving trends (PIT) for 10-year time period.**

Variable	Number of sites	PIT (%)	95% confidence interval for PIT (%)
CLAR	457	65.1	62.2 - 68.0
DRP	750	55	52.8 - 57.2
ECOLI	748	52.1	49.6 - 54.6
MCI	573	44.7	41.4 - 48.0
NH4N	731	72.2	70.0 - 74.4
NO3N	749	56.5	54.1 - 58.9
TN	660	49.5	47.0 - 52.0
TP	663	71.4	68.9 - 73.9
TURB	718	50.1	47.6 - 52.6

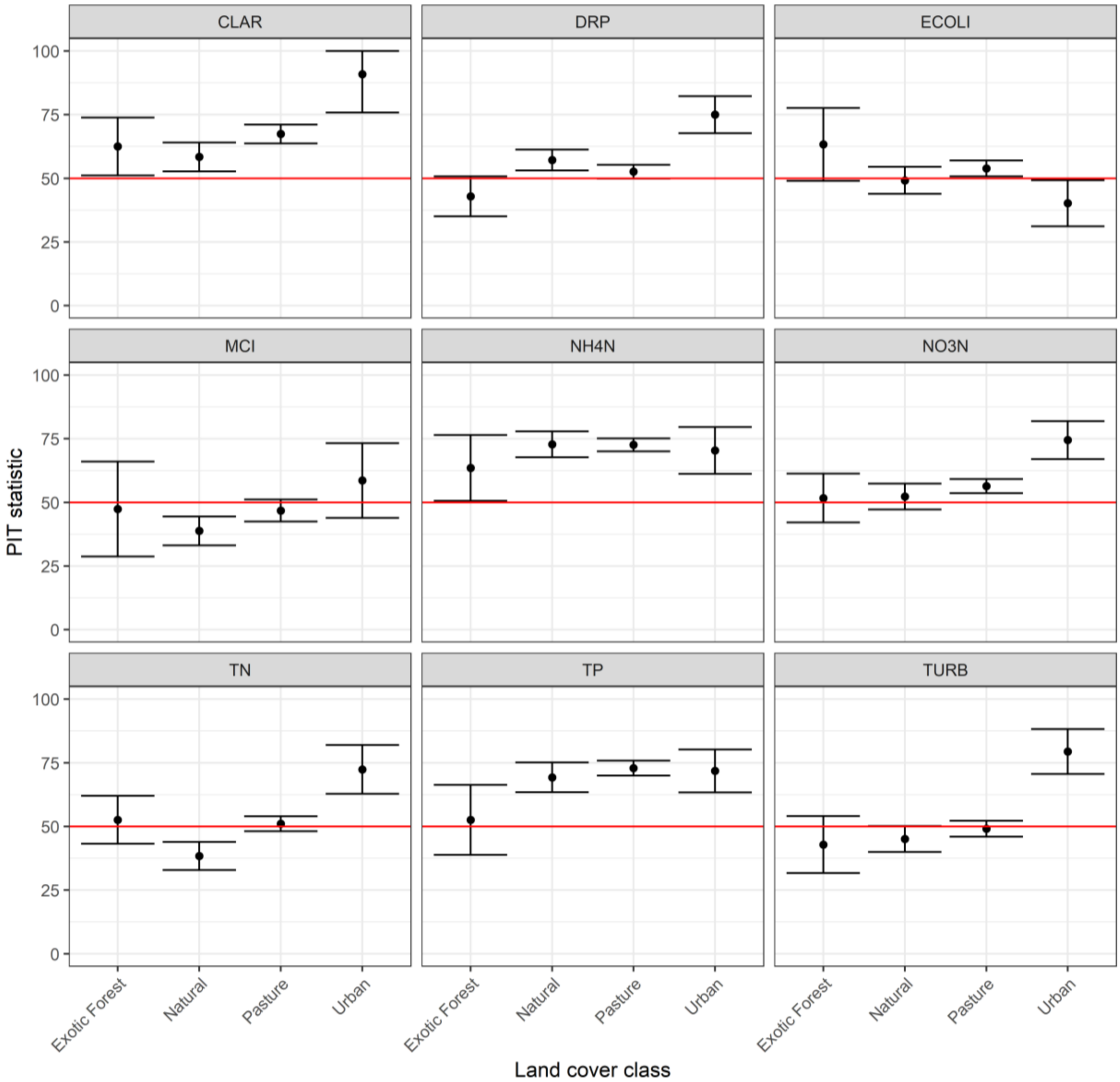

Figure 5-6: Proportions of improving trends (PIT) within REC land-cover classes for 10-yeartrends. Error bars are 95% confidence intervals.

5.2 Twenty-year trends (1998-2017)

Between 79 and 332 river monitoring sites met the filtering rules for the 20-year trend analyses of nutrient and ECOLI concentrations, and for TURB, MCI and CLAR (Table 5-4). The qualifying sites were reasonably well-distributed geographically (Figure 5-7), with gaps in the central North and South Islands and the West Coast. Turbidity was very poorly samples (only 10 sites) and ECOLI was limited to only a small number of regions. Site locations, land cover classes and numbers of sampling dates are included in the supplementary file "RiverTrends_AllPeriods_FlowAdjusted_27_Nov18.csv ".

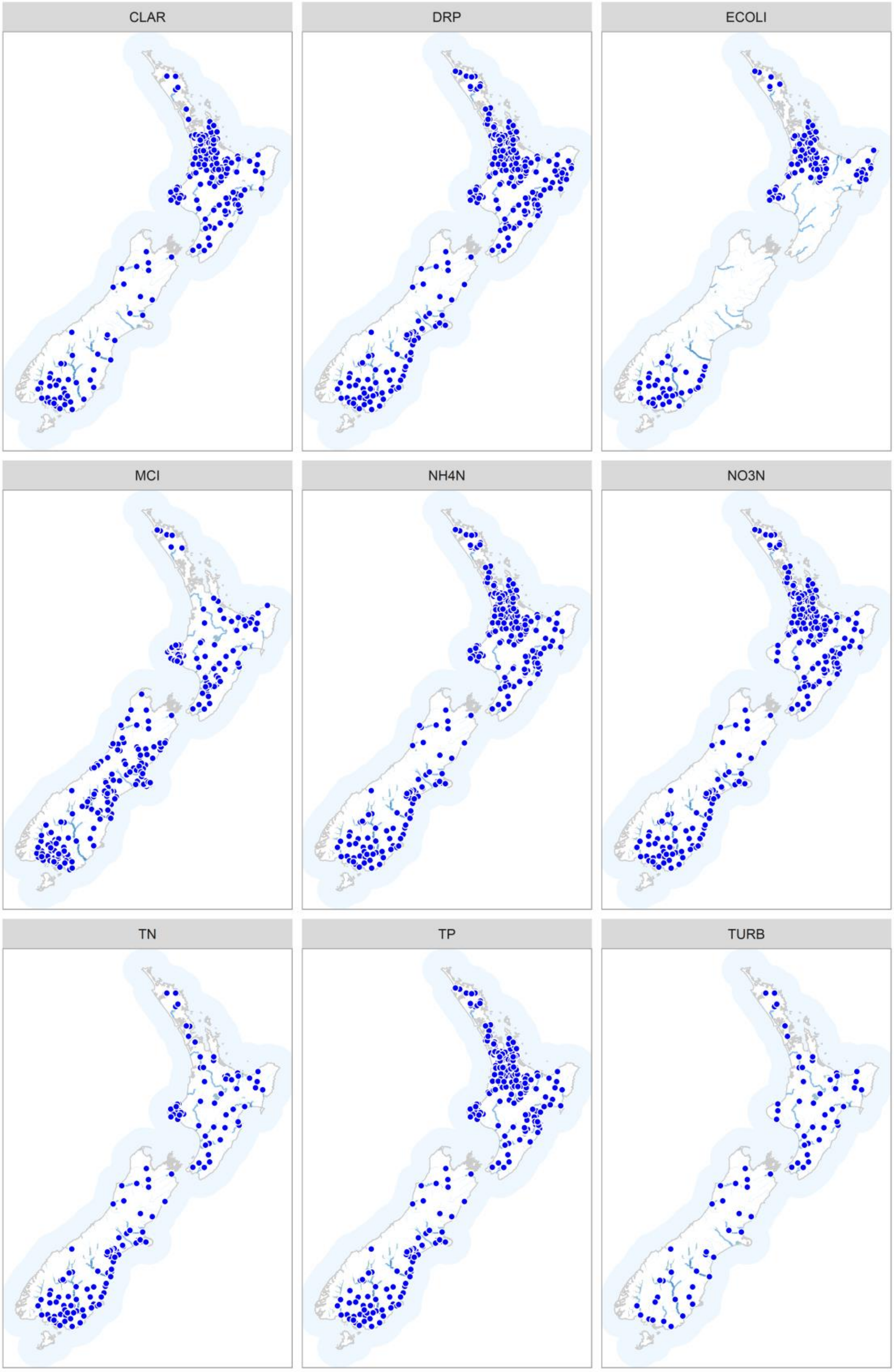

Figure 5-7: River water quality monitoring sites used for 20 year trend analyses of nutrients, ECOLI, CLAR, TURB and MCI.

Table 5-4: Number of river monitoring sites by land cover class and water quality variable that were included in the 20-year trend analyses of nutrients, ECOLI, CLAR and MCI. The site numbers shown refer to sites that met the site inclusion requirements in Section 3.2.1 (measurements are available for at least 90% of the years and at least 90% of seasons).

Variable	Total	Exotic Forest	Natural	Pasture	Urban
CLAR	230	11	60	155	4
DRP	331	14	71	231	15
ECOLI	152	10	27	108	7
MCI	332	4	119	200	9
NH4N	316	13	71	218	14
NO3N	309	13	68	214	14
TN	162	4	51	100	7
TP	307	13	70	210	14
TURB	79	3	34	42	0

5.2.1 Trend magnitude

Distributions of trend magnitudes for each of the water-quality variables in four land-cover classes for the 20-year period from 1998 – 2017 are shown in 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 some exceptions, land-cover classes did not account for a substantial amount of the variation in trend. This contrasts with the state analyses of river variables, where water-quality state clearly varied between land cover classes (Figures 4-2, 4-3 and 4-4). Exceptions included NH4N and TP in the urban land-cover class, for which the median magnitudes indicated reductions of over 2% per year, and TN and TURB in the exotic forest class, for which the median magnitudes indicated increases of approximately 2% per year.

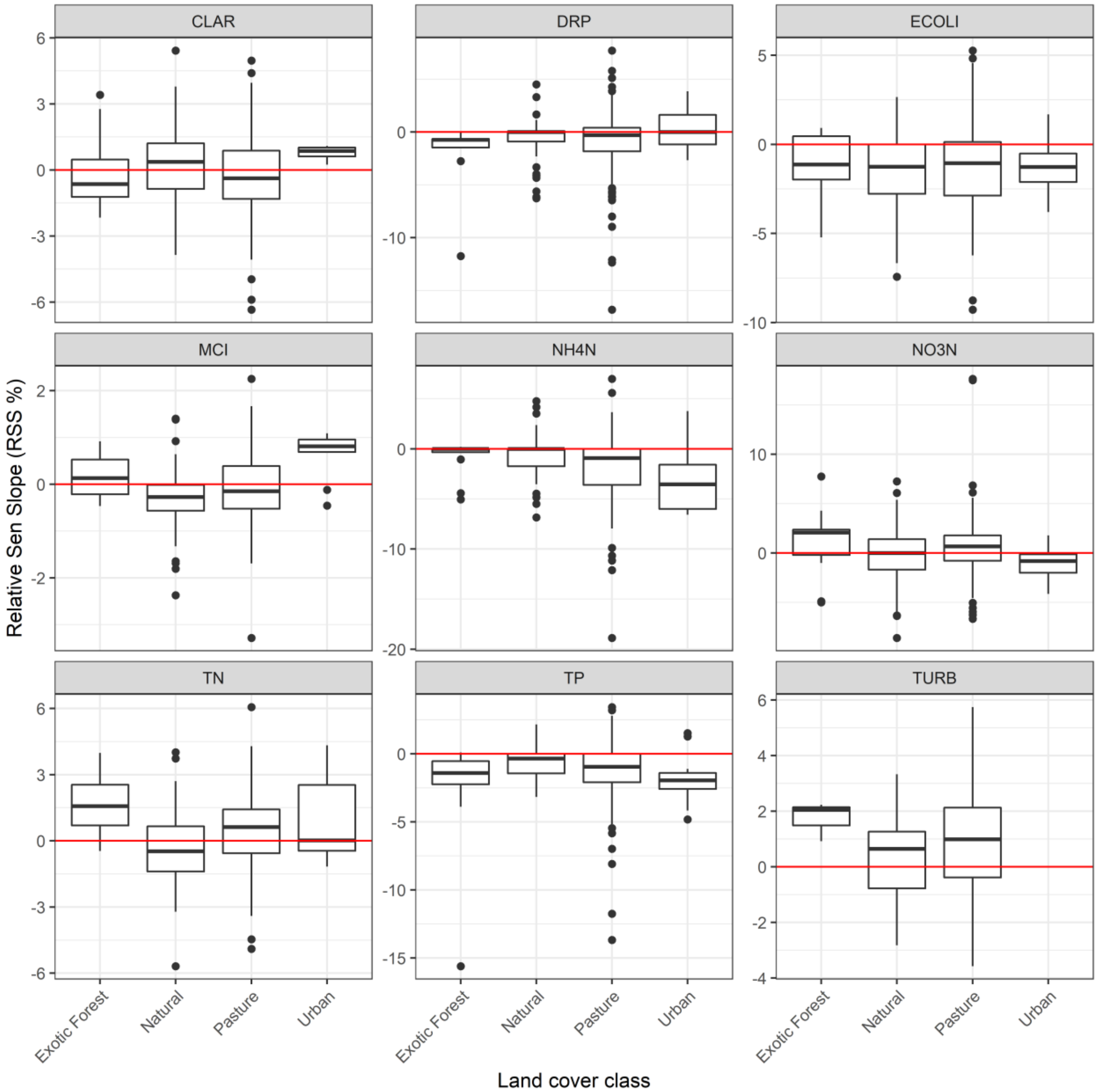

Figure 5-8: Summary of 20-year trends. Box-and-whisker plots show the distributions of site trends within land cover classes**.** Black horizontal line in each box indicates the median of site trends, box indicates the inter-quartile range, whiskers indicate the 5th and 95th percentiles, and open circles 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 (20 to 65% of the site trends) were classified as indeterminant. In addition, trends in DRP were not analysed for three sites due to very high proportions of censored values. For the remaining trends, the proportions of sites with improving trends in DRP, ECOLI, NH4N, and TP were substantially larger than the proportion with degrading trends, and the proportions of sites with degrading trends in MCI, NO3N, TN and TURB were larger than the proportions with improving trends.

Some geographic patterns in 20-year trends are apparent in Figure 5-9. For example, the majority of degrading DRP trends and the majority of improving CLAR trends were in the central North Island (although this pattern is influenced by the poor sampling distribution in the South Island). NO3N trends were primarily improving in Northland and Hawke's Bay, and degrading in the central North Island, Canterbury and Southland.

Table 5-5: Numbers and proportions of sites in four trend categories for 20-year trends. The classification of site 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).

Variable	Trend category			Not analysed
	Degrading	Improving	Indeterminant	
CLAR	78 (34%)	61 (27%)	91 (40%)	0 (0%)
DRP	67 (20%)	135 (41%)	126 (38%)	3 (1%)
ECOLI	14 (9%)	53 (35%)	85 (56%)	0 (0%)
MCI	70 (21%)	46 (14%)	216 (65%)	0 (0%)
NH4N	27 (9%)	138 (44%)	133 (42%)	18 (6%)
NO3N	141 (46%)	91 (29%)	77 (25%)	0 (0%)
TN	66 (41%)	44 (27%)	52 (32%)	0 (0%)
TP	26 (8%)	170 (55%)	111 (36%)	0 (0%)
TURB	36 (46%)	13 (16%)	30 (38%)	0 (0%)

Figure 5-9: Water quality monitoring sites classified by 20-year trends. The three categories defined according to Larned et al. (2015): degrading, improving, indeterminant.

5.2.3 Probability of improvement

The distributions of probabilities that 20-year trends for each site × 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 sites previously categorised as having insufficient data. The maps also indicate that for most water quality variables, those sites previously categorised as indeterminant (Figure 5-9) were about equally divided into likely and unlikely to improve confidence categories. However, in the cases of NH4N and TP, most sites that were previously categorised as indeterminant were dominated by improving trends. Conversely, MCI trends that were previously categorised as indeterminant were more likely to indicate degrading trends.

Figure 5-10: Water quality monitoring sites categorised by the probability that the 20-year trend in each variable is improving.

5.2.4 Aggregate trends

Figure 5-11 shows the proportions of sites for which 20-year, flow-adjusted trends indicated improvement at the nine categorical levels of confidence defined in Table 3-3. The national-scale 20-year PIT statistics and their confidence intervals are summarised in Table 5-6. The 20-year PIT statistics ranged from 35-81%. Four variables had a majority of degrading trends at the 95% confidence level (MCI, TN, NO_3N, and TURB). Four other variables had a majority of improving trends at the 95% confidence level (DRP ECOLI, NH4N, and TP). CLAR had 95% confidence intervals for PIT that included 50%, and we cannot infer widespread degradation or improvement in CLAR.

The 20-year PIT statistics and 95% confidence intervals for each water-quality variable and land-cover class are shown in Figure 5-12. For four of the nine water-quality variables (CLAR, ECOLI, MCI, NO3N), the PIT statistic was highest in the urban land-cover class. In contrast, the PIT statistic was < 50% for DRP in the urban land-cover class. The PIT statistics also indicated a majority of degrading trends in CLAR, NO3N, TN and TURB in the exotic forest, in CLAR, MCI, NO3N and TN in the pastoral class, and in MCI and TURB in the natural class.

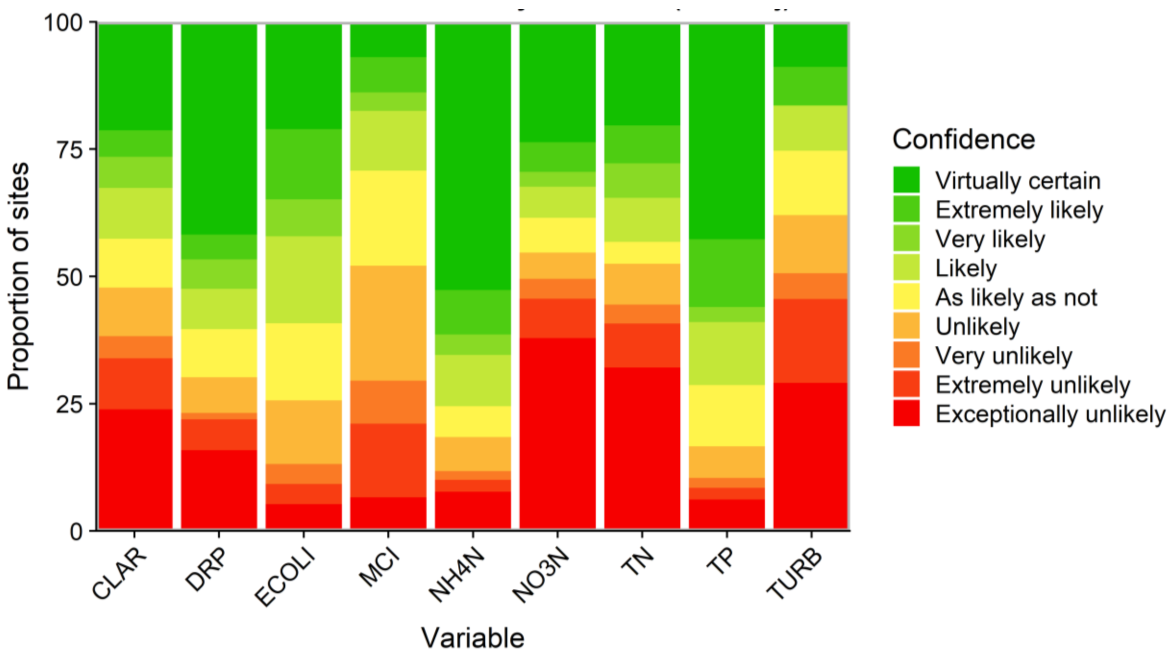

Figure 5-11: Summary plot representing the proportion of sites with improving 20-year trends at each categorical level of confidence. The plot shows the proportion of sites with improving trends at levels of confidence defined in Table 3.1.

Table 5-6: **Proportions of improving trends (PIT) among 20-year trends.**

Variable	No. of sites	PIT (%)	95% confidence interval for PIT (%)
CLAR	230	49.1	45.8 - 52.4
DRP	328	64.3	61.8 - 66.8
ECOLI	152	67.4	62.5 - 72.3
MCI	332	35.4	31.9 - 38.9
NH4N	298	78.2	75.7 - 80.7
NO3N	309	41.3	38.9 - 43.7
TN	162	45.1	41.8 - 48.4
TP	307	81.3	78.4 - 84.2
TURB	79	35.4	29.3 - 41.5

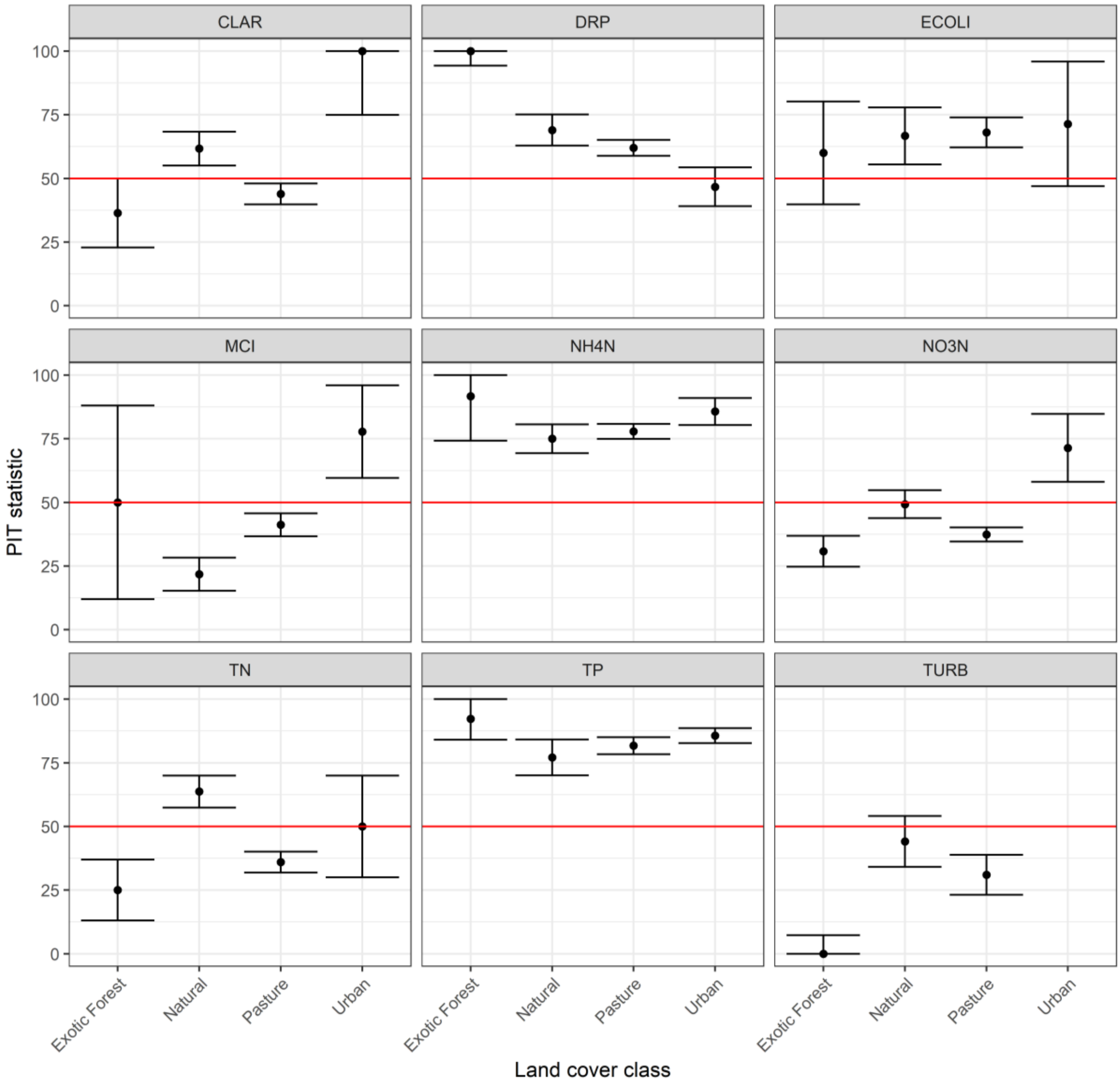

Figure 5-12: Proportions of improving trends by REC land-cover class for 20-year-trends. Error bars are 95% confidence intervals.

5.3 Twenty-eight year trends (1990-2017)

Between 16 and 122 river monitoring sites met the filtering rules for the 28-year trend analysis of nutrient and ECOLI concentrations, TURB, MCI and CLAR (Table 5-7). The qualifying sites were reasonably well-distributed geographically (Figure 5-13), with gaps in the central North and South islands and the West Coast. ECOLI was poorly represented, with only 16 sites. Sites in the exotic forest and urban land-cover classes were scarce for each variable (0-6 sites). Site locations, land cover classes and numbers of sampling dates are included in the supplementary file "RiverTrends_AllPeriods_FlowAdjusted_27_Nov18.csv ".

Figure 5-13: River water quality monitoring sites used for 28-year trend analyses of nutrients, ECOLI, CLAR, TURB and MCI.

Table 5-7: Number of river monitoring sites by land cover class and water quality variable that were included in the 28-year trend analyses of nutrients, ECOLI, CLAR and MCI. The site numbers shown refer to sites that met the site inclusion requirements in Section 3.2.1.

Variable	Total	Exotic Forest	Natural	Pasture	Urban
CLAR	78	3	34	41	0
DRP	122	5	39	72	6
ECOLI	16	2	0	13	1
MCI	70	1	36	33	0
NH4N	109	4	39	60	6
NO3N	112	4	39	63	6
TN	83	3	36	42	2
TP	110	4	39	61	6
TURB	77	3	34	40	0

5.3.1 Trend magnitude

Distributions of trend magnitudes for each of the water-quality variables for the 28-year period from 1990 – 2017 in four land-cover classes are shown in Figure 5-14. The plots indicate that land cover classes did not account for a substantial amount of the variation in trends for most variables. The two land-cover classes represented by numerous sites for most variables, natural and pasture, had similar medians and quartile ranges, with the exception of NH4N, for which the median trend magnitude in the natural class indicated a reduction of over 2% per year. The median trend magnitudes for NH4N, NO3N and TP in the urban class also indicated reductions, but there were only six monitoring sites in the urban class.

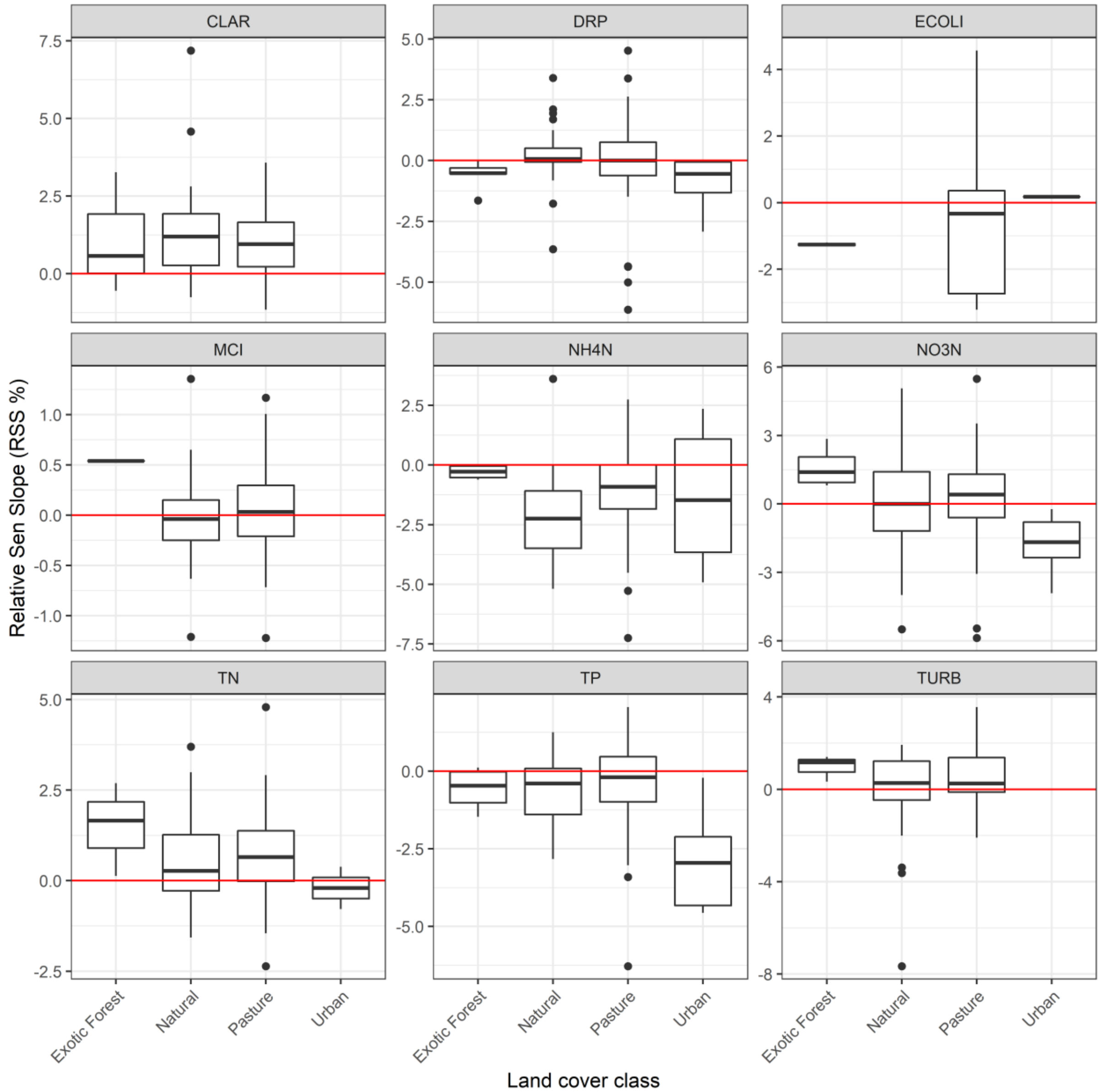

Figure 5-14: Summary of 28-year trends. Box-and-whisker plots show the distributions of site trends in REC land cover classes. Black horizontal line in each box indicates the median of site trends, box indicates the inter-quartile range, whiskers indicate the 5th and 95th percentiles, and open circles indicate outliers.

5.3.2 Trend classification

The numbers and proportions of 28-year trend trends in four categories are summarised in Table 5-8. A moderate-to-large proportion of the trends for the nine variables (25 to 64% of the site trends) were classified as indeterminant. In addition, trends in NH4N were not analysed for three sites due to very high proportions of censored values. For the remaining trends, the proportions of sites with improving trends in CLAR, NH4N, and TP were substantially larger than the proportions with degrading trends, and the proportion of sites with degrading trends in TN and TURB were larger than the proportion with improving trends. The proportion of sites with improving trends in ECOLI was larger than the proportions with degrading trends, but the numbers of sites were very small (9 sites total).

Some geographic patterns in 28-year trends are apparent in Figure 5-15. For example, trends in CLAR and NH4N are predominately improving across the South Island. Qualifying sites for analyses of ECOLI trends were limited to the Gisborne District and two sites in the Bay of Plenty Region.

Table 5-8: Numbers and proportions of sites in four trend categories for 28-year trends. The classification of site 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).

Variable	Trend category			Not Analysed
	Degrading	Improving	Indeterminant	
CLAR	5 (6%)	52 (67%)	21 (27%)	0 (0%)
DRP	33 (27%)	37 (30%)	52 (43%)	0 (0%)
ECOLI	2 (12%)	7 (44%)	7 (44%)	0 (0%)
MCI	13 (19%)	12 (17%)	45 (64%)	0 (0%)
NH4N	6 (6%)	70 (64%)	30 (28%)	3 (3%)
NO3N	49 (44%)	36 (32%)	27 (24%)	0 (0%)
TN	44 (53%)	18 (22%)	21 (25%)	0 (0%)
TP	21 (19%)	48 (44%)	41 (37%)	0 (0%)
TURB	33 (43%)	11 (14%)	33 (43%)	0 (0%)

Figure 5-15: Water quality monitoring sites classified by 28-year trends. The three categories used are according to Larned et al. (2015): degrading, improving, indeterminant.

5.3.3 Probability of improvement

The probabilities that 28-year trends for each variable were improving are shown spatially in Figure 5-16. The maps indicate that most of the trends in TN at South Island sites that were previously categorised as indeterminant (Figure 5-15), were categorised as degrading in Figure 5-16. Conversely, most of the trends in CLAR and NH4N on the North Island that were previously categorised as indeterminant (Figure 5-15), were classified as improving in Figure 5-16. Most of the trends in ECOLI (limited to sites the Bay of Plenty Region and Gisborne District) were also classified as improving in Figure 5-16.

Figure 5-16: Water-quality monitoring sites categorised by the probability that the 28-year trend in each variable is improving.

5.3.4 Aggregate trends

Figure 5-17 shows the proportions of sites for which 28-year trends indicated improvement at the nine categorical levels of confidence defined in Table 3-3. The national-scale 28-year PIT statistics and their confidence intervals are summarised in Table 5-9. The 28-year PIT statistics ranged from 32-86%. Two variables had a majority of degrading trends at the 95% confidence level (TN and TURB). Four other variables had a majority of improving trends at the 95% confidence level (CLAR, ECOLI, NH4N, TP). Three variables had 95% confidence intervals for PIT that included 50% (DRP, MCI, NO3N), and we cannot infer widespread degradation or improvement for these variables.

Table 5-9: Proportions of improving trends (PIT) among 28-year trends.

Variable	No. of sites	PIT (%)	95% confidence interval for PIT
CLAR	78	79.5	74.4 - 84.6
DRP	122	54.9	50.6 - 59.2
ECOLI	16	68.8	53.7 - 83.9
MCI	70	47.1	39.3 - 54.9
NH4N	106	85.8	82.1 - 89.5
NO3N	112	49.1	45.4 - 52.8
TN	83	32.5	27.8 - 37.2
TP	110	64.5	59.8 - 69.2
TURB	77	35.1	28.8 - 41.4

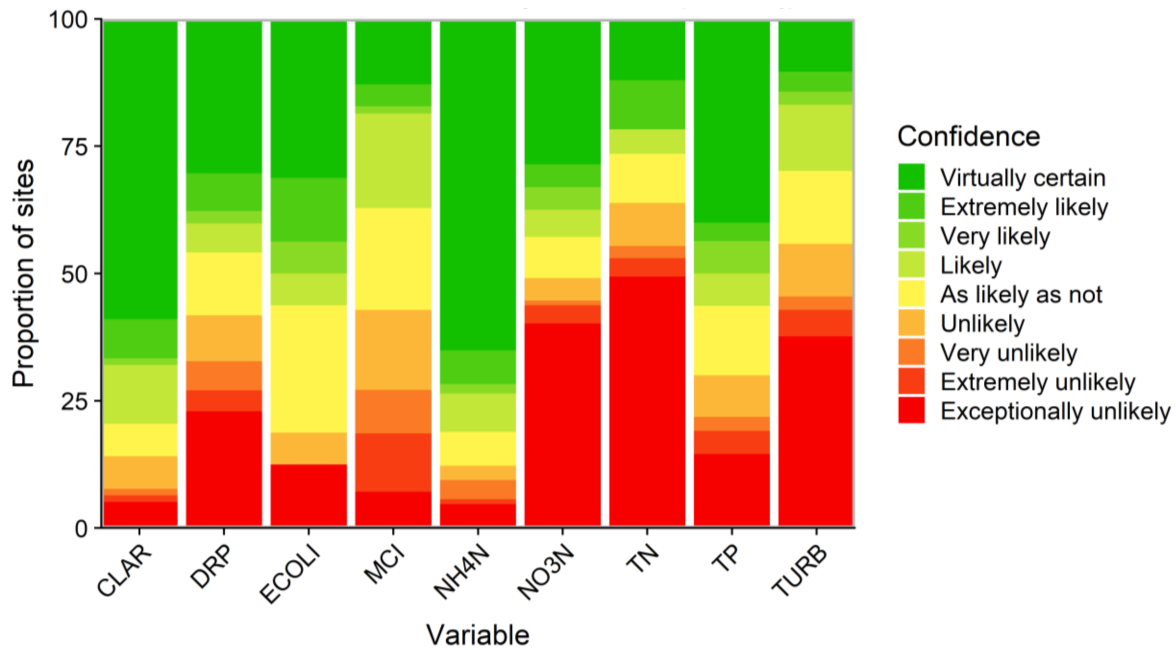

Figure 5-17: Summary plot representing the proportion of sites with improving 28-year trends at each categorical level of confidence. The plot shows the proportion of sites with improving trends at levels of confidence defined in Table 3.1.

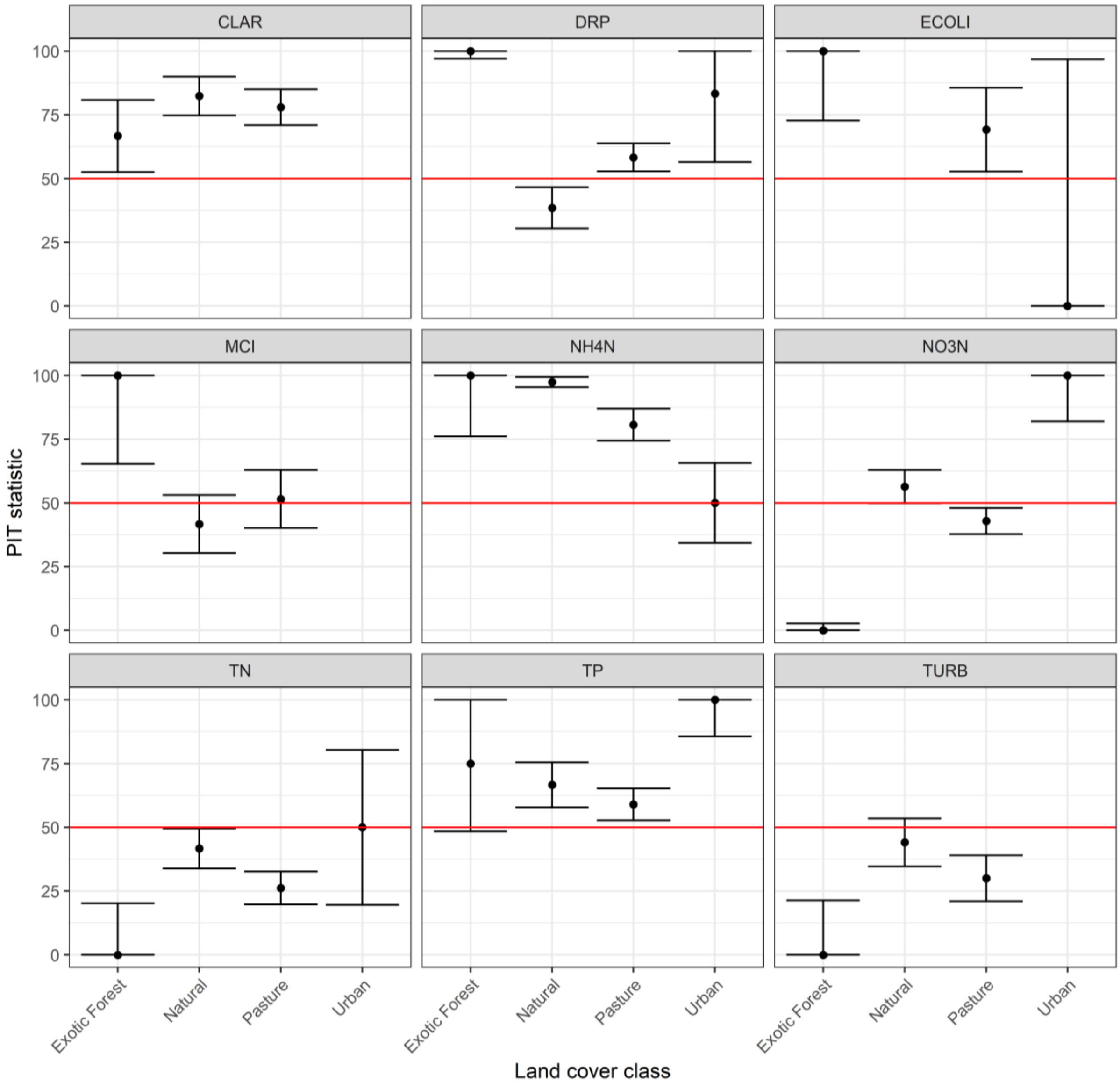

Figure 5-18: Proportions of improving trends by REC land-cover class for 28-year trends. Error bars are 95% confidence intervals.

5.4 Comparisons of trend directions between 10-, 20- and 28-year periods

The national scale PIT statistics for each water quality variable are shown in Table 5-10, which combines the results in Tables 5-3, 5-6 and 5-9. A comparison of the 10-, 20- and 20-year trends in this table reveal several changes in the balance of improving and degrading trends:

1. a predominance of degrading 20-year trends in NO3N shifted to a predominance of improving 10-year trends;
2. a predominance of degrading 20- and 28-year trends in TN shifted to roughly equal proportions of degrading and improving 10-year trends; and
3. a predominance of improving 20- and 28-year trends in ECOLI shifted to roughly equal proportions of degrading and improving 10-year trends.

In contrast to these changes between trend periods, the predominance of improving trends in NH4N and TP has persisted between all trend periods, and the predominance of degrading trends in MCI scores has persisted from the 20- to 10-year period.

Table 5-10: National-scale PIT statistics. Values are estimated percentages of river sites with improving trends across New Zealand. Blue font: majority of sites improving. Red font: majority of sites degrading. Green font: cannot infer improvement or degradation at most sites because the 95% confidence intervals for the PIT statistic included 50%.

Variable	10-year trend (2008-2017)	20-year trend (1998-2017)	28-year trend (1990-2017)
CLAR	65.1	49.1	79.5
DRP	55.0	64.3	54.9
ECOLI	52.1	67.4	68.8
MCI	44.7	35.4	47.1
NH4N	72.2	78.2	85.8
NO3N	56.5	41.3	49.1
TN	49.5	45.1	32.5
TP	71.4	81.3	64.5
TURB	50.1	35.4	35.1

6 Discussion

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

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

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

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 sites and therefore, spatial coverage. The five-year period ensured that there were at least 30 samples for each site, for variables that are measured at monthly intervals, as recommended by McBride (2005).

In the previous national-scale river 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 dataset. 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 dataset in the current study; the inclusion of imputed values in trend datasets is 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 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 review), 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 great potential for improvement. For example, in the 10-year trend analyses reported using the previous method, between 54 and 88% of the site-specific 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. Finally, we note that the current report does not represent the last word in water-quality data analysis; further advancements are inevitable and beneficial.

7 Acknowledgements

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 Graham Bryers (NIWA) for assistance with NRWQN data. We thank Fenella Falconer for formatting the reporting, and Neale Hudson, Graham McBride and Clive Howard-Williams for reviewing the draft report.

8 References

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

Booker, D.J., Snelder, T.H. (2012) Comparing methods for estimating flow duration curves at ungauged sites. *Journal of Hydrology* 434: 78-94.

Booker, D.J., Woods, R.A. (2014) Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments. *Journal of Hydrology* 508: 227239.

Camargo, J.A., Alonsom, A., Salamanca, A. (2005) Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates. *Chemosphere* 58: 1255-1267.

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.

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.

Hickey, C.W. (2013) Updating nitrate toxicity effects on freshwater aquatic species. New Zealand Ministry for Business, Innovation and Employment Envirolink. *NIWA Client Report* HAM2013-009.

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

Horowitz, A.J. (2013) A review of selected inorganic surface water quality-monitoring practices: are we really measuring what we think, and if so, are we doing it right? *Environmental Science & Technology* 47:2471-2486.

Larned, S.T., Scarsbrook, M., Snelder, T., Norton, N., Biggs, B. (2004) Water quality in low-elevation streams and rivers of New Zealand: recent state and trends in contrasting land-cover classes. *New Zealand Journal of Marine and Freshwater Research* 38: 347-366.

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.

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

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

McBride, G.B. In review. Has water quality improved or been maintained? A quantitative assessment procedure. *Journal of Environmental Quality*.

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

McMillan, H.K., Hreinsson, E.O., Clark, M.P., Singh, S.K., Zammit, C., Uddstrom, M.J. (2013) Operationalhydrological data assimilation with the recursive ensemble Kalman filter. *Hydrology and Earth System Sciences* 17: 21–38.

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

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. http://www.mfe.govt.nz/publications/fresh-water/swimming-categories-e-coli-clean-water-package-summary-of-categories-and.

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

New Zealand Government. (2017) National Policy Statement for Freshwater Management 2014 (amended 2017). https://www.mfe.govt.nz/publications/fresh-water/national-policy-statement-freshwater-management-2014-amended-2017.

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.

Smith, D.G., McBride, G.B., Bryers, G.G., Wisse, J., Mink, D.F. (1996) Trends in New Zealand's national river water quality network. *New Zealand Journal of Marine and Freshwater Research* 30: 485-500.

Snelder, T. (2018) Assessment of recent reductions in *E. coli* and sediment in rivers of the Manawatū-Whanganui Region. LWP Report 2017-06.

Snelder, T., Biggs, B.J.F. (2002) Multiscale river environment classification for water resources management. *Journal of the American Water Resources Association* 38: 1225-1239.

Snelder, T., Biggs, B., Weatherhead, M. (2010) New Zealand River Environment Classification user guide. New Zealand Ministry for the Environment.

Snelder, T., Larned, S.T., McDowell, R. (2018) Anthropogenic increases of catchment nitrogen and phosphorus loads in New Zealand. *New Zealand Journal of Marine and Freshwater Research 52*: 336-361.

Snelder, T., Fraser, C. (2018) Aggregating trend data for environmental Reporting. LWP Client Report 2018-01.

Stark, J.D., Maxted, J.R. (2007) A user guide for the Macroinvertebrate Community Index. Cawthron Report 1166.

Unwin, M., Larned, S.T. (2013) Statistical models, indicators and trend analyses for reporting national-scale river water quality. New Zealand Ministry for the Environment. *NIWA Client Report* CHC2013-033.