Water quality state and trends in New Zealand Rivers

Analyses of national data ending in 2020

Prepared for Ministry for the Environment

February 2022

Prepared by:
Amy Whitehead
Caroline Fraser (LWP)
Ton Snelder (LWP)
Kathy Walter
Simon Woodward
Christian Zammit

For any information regarding this report please contact:

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

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

Phone +64 3 348 8987

NIWA CLIENT REPORT No: 2021296CH
Report date: February 2022
NIWA Project: MFE21502

Revision	Description	Date
Version 1.0	Final Report	8 October 2021
Version 1.1	Updated Final Report	18 November 2021
Version 1.2	Updated Final Report	25 February 2022

Quality Assurance Statement		
	Reviewed by:	Scott Larned
	Formatting checked by:	Rachel Wright
	Approved for release by:	Helen Rouse

Contents

Tables

Figures

Executive summary

Background and brief

The New Zealand Ministry for the Environment (MfE) and Stats NZ Tatauranga Aotearoa 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 trends for the period ending in December 2020.

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. The principal outputs 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. In this study, we used several approaches to aggregate results, including River Environment Classification (REC) land-cover classes, and continuous land-cover data.

The brief for this work consisted of eight 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 nine 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 30-year periods ending in December 2020, including comparisons of trends at sites aggregated by REC land-cover classes. The 30-year trend period corresponds to the period of record for NRWQN monitoring sites and a smaller number of long-term council sites.
7. Evaluate trends for each of the water quality variables at each monitoring site for rolling windows of 10-years duration starting in 1990 and incrementing by one year (ending 31 December) to a final window ending in 2020 (i.e., a total time period of 30 years).
8. Assess water quality trends at the national scale using two approaches: categorical levels of confidence and a statistical analysis of the proportions of decreasing trends.

In addition, we used the water-quality state dataset from Step 4 to grade river monitoring sites for water quality variables that are used to define numeric attribute states in the National Policy Statement for Freshwater Management 2020.

Methods

Data acquisition and processing

We used three procedures to acquire updated data for the current report: requests to Land Air Water Aotearoa (LAWA) data managers for available data from regional councils acquired for the annual LAWA refresh, interrogation of data servers operated by individual regional councils and NIWA (for NRWQN 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); 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 spatial coordinates, nzsegment 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 MfE), nine monitoring variables were selected for use in the state and trends analyses: visual clarity (CLAR), turbidity (TURB), concentrations of nitrate-nitrite-nitrogen (NNN), 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).

State analyses

The state dataset consisted of data for the nine variables listed above, for the 2016-2020 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 numeric attribute states that are set out in the National Policy Statement for Freshwater Management (NPS-FM).

Trend analyses

The trend assessment utilised data for the nine variables listed above, for the 10-, 20-, and 30-year periods ending in December 2020. 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.

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. 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 analyses produced estimates of trend rate made with the Sen slope estimator, and estimates of the confidence in the trend direction, from Kendall tests. The seasonal version of the Sen slope estimator was used for variables measured seasonally (i.e., monthly, bi-monthly or quarterly), and for which seasons accounted for a significant amount of the variability in a site × variable combination.

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

Two approaches were used to evaluate patterns of trends at the national scale and within environmental classes. Both approaches involved aggregating multiple sites into two domains, an environmental domain (REC land-cover classes) and a spatial domain covering the entire country. 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 confidence of decreasing trends from individual sites to estimate the proportion of decreasing trends (P_d) for all sites in the domain. The Pd 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.

The majority of sites (66%) were graded below the national bottom line for the NPS-FM *E.coli* combined numeric attribute state (i.e., most were graded D or E). Over 95% of sites in the urban land cover class, and 75% of sites in the pastoral land cover class were below the bottom line. Very few sites (1-10%) were below the bottom line for the ammonia (toxicity) or nitrate (toxicity) attributes. For the suspended fine sediment attribute, 38% of sites were below the bottom line, including 25% of sites in the "Natural" land cover class. For the macroinvertebrate attribute, 26% of sites were below the bottom line, including over 90% of sites in the "Urban" land cover class.

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 30-year trend analyses in terms of trend rate (percent change in a water quality variable per year). We then summarise the trend analysis results in terms of trend direction (increasing or decreasing). As noted above, the analyses of trend directions included the classification of all trends into nine categorical confidence categories, and an evaluation of the probabilistic proportion of decreasing trends (the P_d statistic). For brevity, the following summary is based on the P_d statistics for each water quality variable at the national level and within land-cover classes.

The magnitudes of 10-, 20- and 30-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 30-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 1: Trends in river water quality variables according to proportion of decreasing trends (P_d). Magenta cells: majority of sites decreasing. Blue cells: majority of sites increasing. Yellow cells: cannot infer increases or decreases at most sites because the 95% confidence intervals for the P_d statistic included 50%.

Variable	10-year trend (2011-2020)	20-year trend (2001-2020)	30-year trend (1991-2020)
CLAR	39.9	40.5	34.8
DRP	57.7	71.8	70.7
ECOLI	39.4	47.9	73.7
MCI	67.8	66.8	66.7
NH4N	65.6	63.5	**66.7**
NNN	46.8	42.7	37.8
TN	51.4	46.4	42.6
TP	49.9	82.1	75.6
TURB	63.9	57.4	29.3

A comparison of the 10-, 20- and 30-year trends in this table reveal several changes in the balance of increasing and decreasing trends: 1) a predominance of decreasing 30-year trends in ECOLI shifted to a predominance of increasing 10-year trends; 2) a predominance of increasing 30-year trends in TURB shifted to a predominance of decreasing 10-year trends; and 3) a predominance of decreasing 20- and 30-year trends in TP shifted to roughly equal proportions of degrading and improving 10-year trends. In contrast to these changes between trend periods, the predominance of increasing trends in CLAR and NNN and decreasing trends in DRP, MCI and NH4N have persisted between all trend periods.

The analysis for the rolling 10-year trend windows provided some insights into changes in aggregate trends over time. The national-scale proportions of decreasing trends (P_d) did not exhibit monotonic changes in the P_d score for any of the variables. There were quasi-periodic fluctuations in P_d that varied between variables. The magnitude of the fluctuations was greatest for CLAR (ranging from 19% to 68%) and smallest for MCI (ranging from 54% to 69%). MCI and NH4N consistently had a majority of sites (i.e., >50%) that had decreasing trends.

1 Introduction

The New Zealand Ministry for the Environment (MfE) and Stats NZ Tatauranga Aotearoa 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 587-882 river monitoring sites (depending on the variable); the time-series for each site × variable combination had an ending date in December 2017 (Larned et al. 2018). In the current report, we have undertaken a new data compilation in order to report updated states and trends; the end dates for monitoring sites in the new compilation are in December 2020.

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 calculations of statistical 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 30-year periods ending in late 2020, including comparisons of trends at sites aggregated by REC land-cover classes. The 30-year trend period corresponds to the period of record for NRWQN monitoring sites and a smaller number of long-term council sites.
7. Carry out 10-year rolling trends, including comparisons of trends at sites aggregated by REC land-cover classes.
8. Assess water quality trends at the national scale using two approaches: categorical levels of confidence and a statistical analysis of the proportions of declining trends.

As an additional step, we used the water-quality state dataset to grade river monitoring sites for water quality variables that are used to define numeric attribute states in the National Policy Statement for Freshwater Management (NPS-FM; New Zealand Government 2020).

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.

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. 2018) was updated with data collected between 2018 and December 2020. 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". Where NIWA and regional councils both monitor at the same site, we treated the data as two separate datasets (using the same practice as LAWA). 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 l^{-1}
	Nitrate + nitrite nitrogen	NNN	mg l^{-1}
	Total nitrogen (unfiltered)	TN	mg l^{-1}
	Dissolved reactive phosphorus	DRP	mg l^{-1}
	Total phosphorus (unfiltered)	TP	mg l^{-1}
Microbiological	*Escherichia coli*	ECOLI	cfu 100 ml^{-1}
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 (NNN, 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. All MCI data were supplied by the collecting agency as calculated scores rather than raw invertebrate data. 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).

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; Larned et al. 2015, 2018; Unwin et al. 2010; Unwin and Larned 2013). 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). Since 2018, these data have been stored in an RData file.

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

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

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: lawaid, site name, location and regional council identifier (if available), latitude and longitude (WGS84), and nzsegment number. Nzsegments 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.

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 5.0 (LCDB5)[1] and the categorical REC classes. The LCDB5 comprises proportional cover of 33 land-cover classes, generated from satellite imagery collected in summer 2018-19). 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

[1] https://lris.scinfo.org.nz/layer/104400-lcdb-v50-land-cover-database-version-50-mainland-new-zealand/

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.

The data produced by multiple procedures used to measure ECOLI, NNN, 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 (Horowitz 2013; Patton and Kryskalla 2003). 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 1155 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 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)
NNN	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 Grading of monitoring sites

Water quality state for river monitoring sites was graded based on attributes and associated attribute state bands defined by the National Objectives Framework (NOF) of the NPS-FM (New Zealand Government 2020) (Table 3-1).

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

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

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

A site can be **graded** for each attribute by assigning it to attribute bands (e.g., a site can be assigned to the A band for the nitrate toxicity attribute). The grades are referred to as 'NOF grades' in the results below. Site grading is done by using the numeric attribute state (e.g., annual median nitrate-nitrogen) as a **compliance statistic**. The value of the compliance statistic for a site is calculated from a record of the relevant water quality variable (e.g., the median value is calculated from the observed monthly nitrate-nitrogen concentrations). The site's compliance statistic is then compared against the numeric ranges associated with each attribute band and a grade assigned for the site (e.g., an annual median nitrate-nitrogen concentration of 1.3 mg/l would be graded as "B-band", because it lies in the range >1.0 to ≤2.4 mg/l). Note that for attributes with more than one numeric attribute state, we have provided a grade for each numeric attribute state (e.g., for the nitrate (toxicity) attribute, grades are defined for both the median and 95th percentile concentrations).

Table 3-1 provides a summary of the NOF attributes and numeric attribute states calculated as part of this study. In addition to the NOF attributes in Table 3-1, we also report on water quality state for

Total Nitrogen (TN), Total Phosphorous (TP), raw (not pH adjusted) Ammoniacal Nitrogen (NH4N), and Turbidity (TURB). For these variables, we report the median of the observations.

We used median site NNN concentrations rather than NO3N concentrations to grade sites in terms of the NOF nitrate toxicity attribute for the following reasons. The biological mechanisms of nitrite toxicity in freshwater animals are relatively well characterised (Camargo et al. 2005). The primary mechanism is methaemoglobinemia, which occurs when nitrite converts haemoglobin to methaemoglobin. Methaemoglobin cannot bind oxygen, resulting in decreased oxygen transport and tissue hypoxia. Additional mechanisms of nitrite toxicity include electrolyte imbalance and, possibly, the formation of carcinogenic N-nitroso compounds. Nitrite is either ingested directly or generated internally through the bacterial reduction of ingested nitrate. The direct toxic effects of ionic nitrate on freshwater animals have not been identified with certainty. Instead nitrate toxicity is predominantly indirect, mediated by the conversion of nitrate to nitrite by gut bacteria, and the subsequent direct effects of nitrite listed above. Therefore 'nitrate toxicity' is more accurately described as 'nitrate/nitrite toxicity' (Gehl 2009).

The most common laboratory method for analysing nitrate and nitrite in water samples from NZ freshwater monitoring programmes involves the reduction of all nitrate to nitrite, followed by the colorimetric measurement of nitrite-N concentration based on a standard curve (Table 2-2). The results represent the sum of nitrate-N and nitrite-N (abbreviated NNN) in the water samples. Therefore, because 'nitrate toxicity' results from the combined ingestion of nitrate and nitrite, and most regional councils monitor NNN concentrations in their freshwater SOE programmes, we used NNN for site grading in lieu of NO3N. This approach differs from that used in the previous national reporting (Larned et al. 2018), but maximises available nitrate toxicity data and aligns with the approach taken by LAWA.

3.1.2 Handling censored values

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

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

3.1.3 Time period for assessments and minimum data requirements

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

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

NPS-FM Reference – NOF Attribute	Calculation guidance	Numeric attribute state description	Units	Abbreviated name
A2A; Table 5 – Ammonia (toxicity)	Based on temperature and pH adjusted Ammoniacal-N	Median concentration of Ammoniacal-N	mg l^{-1}	NOF.NH4N.Med
		Maximum concentration of Ammoniacal-N	mg l^{-1}	NOF.NH4N.Max
A2A; Table 6 – Nitrate (toxicity)		Median concentration of NNN	mg l^{-1}	NOF.NNN.Med
		95th percentile concentration of NNN	mg l^{-1}	NOF.NNN.p95
A2A.; Table 8 - Suspended fine sediment	Median of 5 years of at least monthly samples (at least 60 samples)	Median visual clarity	m	NOF.CLAR.Med
A2A; Table 9 - *Escherichia coli*	minimum of 60 samples over a maximum of 5 years,	% exceedances over 260 cfu 100 mL^{-1}	%	NOF.ECOLI.260
		% exceedances over 540 cfu 100 mL^{-1}	%	NOF.ECOLI.540
		Median concentration of *E. coli*	cfu 100 ml^{-1}	NOF.ECOLI.Med
		95th percentile concentration of *E. coli*	cfu 100 ml^{-1}	NOF.ECOLI.p95
A2B; Table 14 - Macroinvertebrates	State calculated as 5 year median based on observations between Dec-Mar	Median MCI score	-	NOF.MCI.Med
A2B; Table 20 - DRP		Median concentration of DRP	mg l^{-1}	NOF.DRP.Med
		95th percentile concentration of DRP	mg l^{-1}	NOF.DRP.p95

The grading assessments were based on a compliance statistic, (e.g., the median value of the observations), made for the 5-year time period to end of December 2020. For MCI, this time period was shifted by 6 months (the 5-year period to end of June 2020), to align with water years, in order to prevent splitting summer samples into two calendar years. The start and end dates for this period were determined by the availability of quality assured data (see Section 2), MfE reporting time periods and consideration of statistical precision of the compliance statistics used in the grading of state. The statistical precision of the compliance statistics depends on the variability in the water quality observations and the number of observations. For a given level of variability, the precision of a compliance statistic increases with the number of observations. This is particularly important for sites that are close to a threshold defined by an attribute band because the confidence that the assessment of state is 'correct' (i.e., that the site has been correctly graded) increases with the precision of the compliance statistics (and therefore with the number of observations). As a general rule, the rate of increase in the precision of compliance statistics slows for sample sizes greater than 30 (i.e., there are diminishing returns on increasing sample size with respect to precision (and therefore confidence in the assigned grade) above this number of observations; McBride 2005).

In this study, a period of five years represented a reasonable trade-off for grading assessments because it yielded a sample size of 30 or more for many sites and variable combinations). The five-year period for the state analyses is consistent with the 2013-2017 period used in the previous national water-quality state analyses (Larned et al. 2018). Because water quality data tend to fluctuate seasonally, it is also important that each season is well-represented over the period of record. In New Zealand, it is common to sample either monthly or quarterly, and in these cases, seasons are defined by months or quarters. We therefore applied a rule that restricted site × variable combinations in the state analyses to those with measurements for at least 90% of the sampling intervals in that period (at least 56 of 60 months or 18 of 20 quarters). Site × variable combinations that did not comply with these rules were excluded from the state analysis. For annually sampled macroinvertebrate variables, which are generally less variable than physical or chemical water quality variables, the nominated minimum sample size requirement was 4 (with samples in at least 4 years).

For grading the suspended fine sediment and *E. coli* attributes, the NPS-FM requires 60 observations over 5 years. For monthly monitoring, this requires collection of all monthly observations (i.e., no missing data). For this study, we relaxed the rule to require observations for 90% of months over the 5-year period (54 observations). Both this relaxation and our default sample number are subjective choices. Therefore, within the supplementary files we provide state assessments for all sites regardless of whether they meet the filtering rules, as well as details about the number of observations and number of years with observations. This will allow MfE to apply tighter or more lenient filtering rules as required.

3.1.4 Calculation of percentiles and compliance statistics

For each 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 2016 to 2020 (inclusive), with the exception of MCI, where we used the time period 1 July 2015 – 30 June 2020 (to prevent splitting summer samples into two calendar years). All percentiles were calculated using the Hazen method.[2]

[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).

For compliance statistics specified as “Annual” (maximum, median, 95th percentile) in the NPS-FM, we calculated these compliance statistics over the entire 5-year state period.

MCI monitoring patterns were found to often be irregular, and although generally sampled one time per summer, occasionally councils took more samples over some summer periods. In order to reduce bias towards summers with a greater number of samples, the MCI median compliance statistic was calculated based on the median value of the median MCI over each water year. The NPS-FM MCI attribute requires that the compliance statistic is only calculated based on samples from December-March. NEMS guidance proposes the appropriate period to collect MCI data is November to April, as the NPS-FM period may be overly restrictive. The results for the MCI attribute presented within this report are based on a compliance statistic using the NPS-FM sample months. However, in the supplementary output files, we have also included the percentiles for data complying with the NEMS guidance, and for all observed data.

3.1.5 pH adjustment of ammonia

Ammonia is toxic to aquatic animals. When in solution, ammonia occurs in two forms: the ammonium cation (NH_4^+) and unionised ammonia (NH_3); the relative proportions of the forms are strongly dependent on pH (and temperature). Unionised ammonia is more toxic to fish than ammonium, hence the total ammonia toxicity increases with increasing pH (and/or temperature) (ANZECC and ARMCANZ 2000). The NPS-FM 2020 attributes related to ammoniacal-N concentrations in freshwater require a correction to account for pH and temperature. Despite this requirement, the results in the current report are not temperature-corrected due to insufficient temperature data. We applied a pH correction to NH4N to adjust values to equivalent pH 8 values, following the methodology outlined in Hickey (2014). For pH values outside the range of the correction relationship (pH 6-9), the maximum (pH<6) and minimum (pH>9) correction ratios were applied. pH adjustment of ammonia was performed after imputation of censored values (Section 3.1.2). In results tables and figures adjusted ammoniacal-N is abbreviated as “NH4N (Adj.)”.

3.1.6 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 LCDB5 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 (Larned et al. 2016, 2018; McDowell et al. 2013). 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 = -0.65, n = 1375). 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 analyses

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

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

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

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

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

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

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

The application of filtering rules for variables that are measured at quarterly intervals or more frequently requires two steps. First, retain sites for which observations are available for at least *X*% of the years in the period. Second, retain sites for which observations are available for at least *Y*% of the sample intervals. For variables that are measured annually such as MCI, the filtering rules are applied by retaining sites for which values are available for at least *X*% of the years in the trend period.

In this study, we used filtering rules applied by Larned et al. (2018), which set X and Y to 90%. Further, the definition of seasons was flexible in order to maximise the number of sites that were included. If the site failed to comply with filter rule (2) when seasons were set as months, a coarsening of the data to bi-monthly seasons was applied and the filter rule (2) was reassessed, and then repeated with seasons as quarterly if bi-monthly seasons failed to comply with filter rule (2). If the data then complied with filter rule (2), the trend results based on the coarser (i.e., bi-monthly or quarterly) seasons were retained for reporting. It is noted that this decision implies a tolerance of variable levels of statistical power and temporal representativeness across the sites that were included in the analysis. In this study, we also included bi-months as an intermediate coarseness between months and quarters, as this is a historically used sampling interval for some regional councils.

For MCI, we allowed both annual and bi-annual sampling intervals, as a number of regional councils have routinely monitored MCI bi-annually. If a site failed to comply with filter rule (2) when seasons were set as biannual, a coarsening of the data to annual sampling intervals was applied and only filter rule (1) was applicable.

3.2.2 Handling censored values

Censored values are managed in a special way by the non-parametric trend assessment methods described in Section 3.2.5. It is therefore important that censored values are correctly identified in the data. Detection limits or reporting limits that have changed through the trend period (often due to analytical changes) can induce trends that are associated with the changing precision of the measurements rather than actual changes in the variable. This possibility needs to be accounted for in the trend analysis and this is another reason that it is important that censored values are correctly identified in the data.

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

The water quality datasets included a small number of left censored values that were much larger than the apparent detection limit at any given time (outliers). Unsupervised application of the hi-censor filter in these circumstances can lead to the unnecessary loss of statistical power in the assessment. To avoid this problem, we employed the following approach. We expected that systematic changes in detection limit would be relatively consistent for a variable across a regional council. To explore patterns in detection level, we plotted left censored data over time by variable and regional council and used these plots to identify the occurrence of outliers. We identified a maximum realistic detection level for each variable and regional council and capped the hi-censor level at these values. We did not apply the hi-censor filter for CLAR, as left censored values varied in magnitude over time, and differences in censor level did not appear to be associated with systematic changes in detection levels. Similarly, left censored values for *E. coli* were highly variable, although there was a systematic change in the lowest censor values with time. To account for this, we applied a maximum hi-censor value for *E. coli* of 10 cfu/100ml across all regional councils.

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

3.2.3 Flow adjustment

Where water quality observations are made in a river and are associated with a solute or particulate matter (e.g., a concentration or an optical measure such as clarity or turbidity) some of the variation can be associated with the river flow (i.e., discharge) at the time the observation was made. The observed values can vary systematically with flow rate due to two kinds of physical processes. The water quality observations may decrease systematically with increasing flow due to the effect of dilution of the contaminant, or increase with increasing flow due to wash-off of the contaminant (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. Some water quality variables can be associated with a combination of dilution and wash off with increasing flow. For example, a portion of the *E. coli* load may come from point sources discharges such as sewage treatment plants (dilution effect), but another portion may be derived from surface wash-off. Increasing flow in this situation may result is an initial dilution at the low end of the discharge range, followed by an increase with discharge at higher values of discharge.

Trend analysis seeks to quantify the relationship between the water quality observations and time. In this context, flow can be considered as a “covariate”; a variable that is also related to the water quality observations but whose influence is confounding the water quality – time relationship that we are interested in. Statistical analysis can be used to remove the influence of the covariate on the water quality observations. For river data, this statistical analysis is called “flow adjustment”.

Flow adjustment has two purposes. First, it can increase the statistical power of the trend assessment (i.e., increase the confidence in the estimate of direction and rate of the trend) by removing some of the variability that is associated with flow. Second, it removes any component of the trend that can be attributed to a trend in the flow data (e.g., a trend in the flow on sample occasions such as increasing or decreasing flow with time).

Flow adjustment involves fitting a model that describes the relationship between the water quality observation and flow, and then using the residuals of the model instead of the original water quality observations in the subsequent trend assessment steps. 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 are 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).

We considered four alternative regression models to describe the relationship between the water quality observations and flow: log-log regression, locally estimated scatterplot smoothing (LOESS, with spans of 0.7 and 0.9) and generalised additive models (GAM). Censored values were represented during model fitting by raw values (i.e., the numeric component of the censored values) multiplied by a 0.5 for detection limit censoring and 1.1 for reporting limit censoring.

The next step was to select the best model from the alternatives. We used expert judgement to choose the most suitable model based at least three considerations: (1) the homoscedasticity (constant variance) of the regression residuals, (2) model goodness of fit measures and (3) plausibility of the shape of the fitted model. We note that model goodness of fit measure alone should not be relied on because they can indicate good model performance but describe unrealistic relationships. This is particularly likely when more flexible models are used such as LOESS and GAM models and therefore these models should be used with caution.

When the relationship between flow and the water quality variable was poor, we concluded that that there was not a systematic relationship between the observations and flow. In this case, no model was selected, no flow adjustment was performed and the trend assessment was performed on the raw data. Choosing not to flow adjust took into consideration the balance between the potential to reduce variance in the observations, and the risk of selecting an implausible/inappropriate model of the relationship between the observations and flow.

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. We did not perform flow adjustment on MCI data.

Flow adjusted trends are not reported in the results section but are provided in supplementary files.

3.2.4 Seasonality assessment

For many site/variable combinations, observations vary systematically by season (e.g., by month or quarter). In cases where seasons are a major source in variability, accounting for the systematic seasonal variation should increase the statistical power of the trend assessment (i.e., increase the confidence in the estimate of direction and rate of the trend). The purpose of a seasonality assessment was to identify whether seasons explain variation in the water quality variable. If this was true, then seasonal versions of the trend assessment procedures were used at the trend assessment step (Section 3.2.5).

We evaluated seasonality using the Kruskall-Wallis multi-sample test for identical populations. This is a non-parametric ANOVA that determines the extent to which season explains variation in the water quality observations. Following Hirsch *et al*. (1982), we identified site/variable combinations as being seasonal based on the *p*-value from the Kruskall-Wallis test with α=0.05. For these sites/variable combinations, subsequent trend assessments followed the "seasonal" variants, described in Section 3.2.5.

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

3.2.5 Analysis of trends

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

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

Assessments of trend directions

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

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

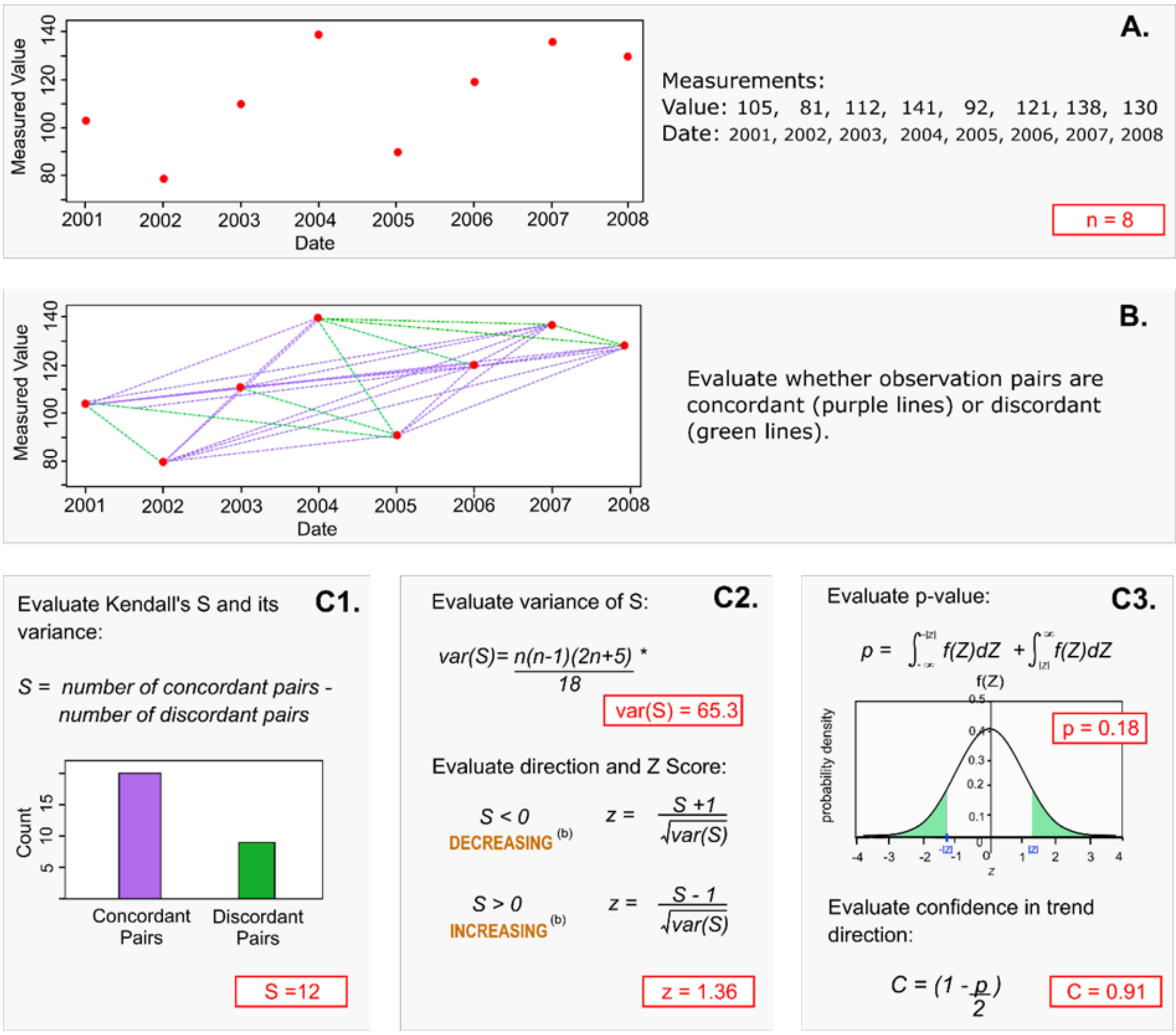

Figure 3-1: Schematic diagram demonstrating how the Kendall *S* statistic and confidence in trend direction (C) is calculated. See text for calculation details (from Snelder et al. 2021).

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

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

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

NHST produces a categorical measure of confidence in the trend assessment (i.e., significant or not significant) based on rejection of the null hypothesis when the *p*-value is smaller than an arbitrary significance level. In this study we define a continuous measure of confidence, which we call confidence in the trend direction (C). Confidence in the trend direction is calculated as:

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

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

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

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

Assessments of trend rates

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

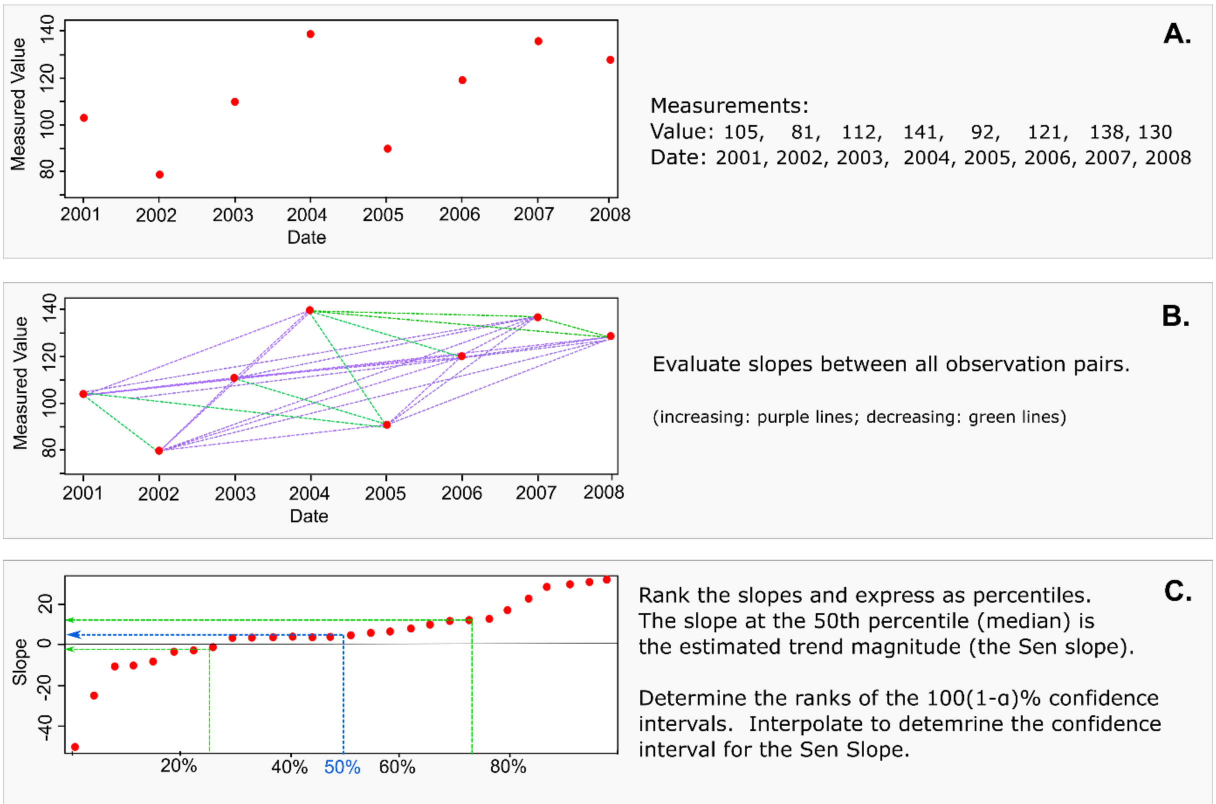

Figure 3-2: Schematic diagram of the calculation of the Sen slope, which is used to characterise trend rate (from Snelder et al. 2021).

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

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

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

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

3.2.6 Interpretation of trends

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

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

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

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

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

1. When a large proportion of the values were censored (data has <5 non-censored values and/or <3 unique non-censored values). This arises because trend analysis is based on examining differences in the value of the variable under consideration between all pairs of sample occasions. When a value is censored, it cannot be compared with any other value and the comparison is treated as a “tie” (i.e., there is no change in the variable between the two sample occasions). When there are many ties there is little information content in the data and a meaningful statistic cannot be calculated.
2. When there is no, or very little, variation in the data, which also results in ties. This can occur because laboratory analysis of some variables has low precision (i.e., values have few or no significant figures). In this case, many samples have the same value, and this then results in ties.

3.2.7 Aggregation of site trends

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

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

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

The second approach also utilises the confidence that the true trend was decreasing to provide a probabilistic estimate of the proportion of decreasing site-specific trends (P_d) within a geographic or environmental domain. Note that P_d is equivalent to the proportion of improving trends (PIT) statistic reported in Larned et al. (2018), without the additional subjective step of defining the directions that indicate improvement or degradations for each water quality variable. For a given water-quality variable, the trends at multiple monitoring sites distributed across a domain of interest can be assumed to represent independent samples of the population of trends, for all sites within that domain.

The statistic P_d is calculated by letting the sampled sites within this domain be indexed by s, so that $s \in \{1, \dots, S\}$ and let I be a random Bernoulli distributed variable which takes the value 1 with probability $p = C_d$ and the value 0 with probability $q = 1 - C_d$ (where C_d is the confidence that the trend was decreasing, as described in Section 3.2.5). Therefore, $I_s = 1$ denotes a decreasing trend at site $s \in \{1, \dots, S\}$ when the estimated $p_s \geq 0.5$ and an increasing trend as 0 when $p_s < 0.5$. Then, the estimated proportion of sites with decreasing trends in the domain is:

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

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

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

P_d and its variance represent an estimate of the population proportion of decreasing trends, within a spatial or environmental domain, and the uncertainty of that estimate. It is noted that the proportion of increasing trends is the complement of the result (i.e., 1 - P_d). The estimated variance of P_d can be used to construct 95% confidence intervals[3] around the P_d statistics as follows:

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

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

We calculated P_d and its confidence interval for all water quality variables and for domains of interest defined by the entire country, and by the four REC land-cover classes (defined in Section 2.2) exotic forest, natural, pastoral, and urban. We also calculated P_d for each of the rolling 10-year time windows.

4 Results – river state

Between 834 and 973 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 "RiverState_2016to2020_v210916.csv". The metadata for each water quality and invertebrate monitoring sites are provided in the supplementary files "RiverMetaData_WQ_v211008.csv" and "RiverMetaData_Inverts_v211008.csv", respectively

The distributions of site-median values of the nine water quality variables for the 2016-2020 period are summarized as box-and-whisker plots, with sites grouped by REC land cover classes (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, whereas the natural land-cover class had the lowest median for TN and with quite large variability. 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 (Figure 4-4). 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 (Figure 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 that met the data requirements outlined in Section 3.1.3. Note NH4N (Adj.) is pH-adjusted ammoniacal-N.

Variable	Total	Exotic Forest	Natural	Pasture	Urban
CLAR	715	24	173	480	38
DRP	973	30	207	652	84
ECOLI	967	30	207	646	84
MCI	857	54	245	510	48
NH4N	973	30	207	652	84
NH4N (Adj.)	885	30	193	618	44
NNN	946	30	191	644	81
TN	938	28	195	635	80
TP	901	28	195	630	48
TURB	834	26	172	558	78

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

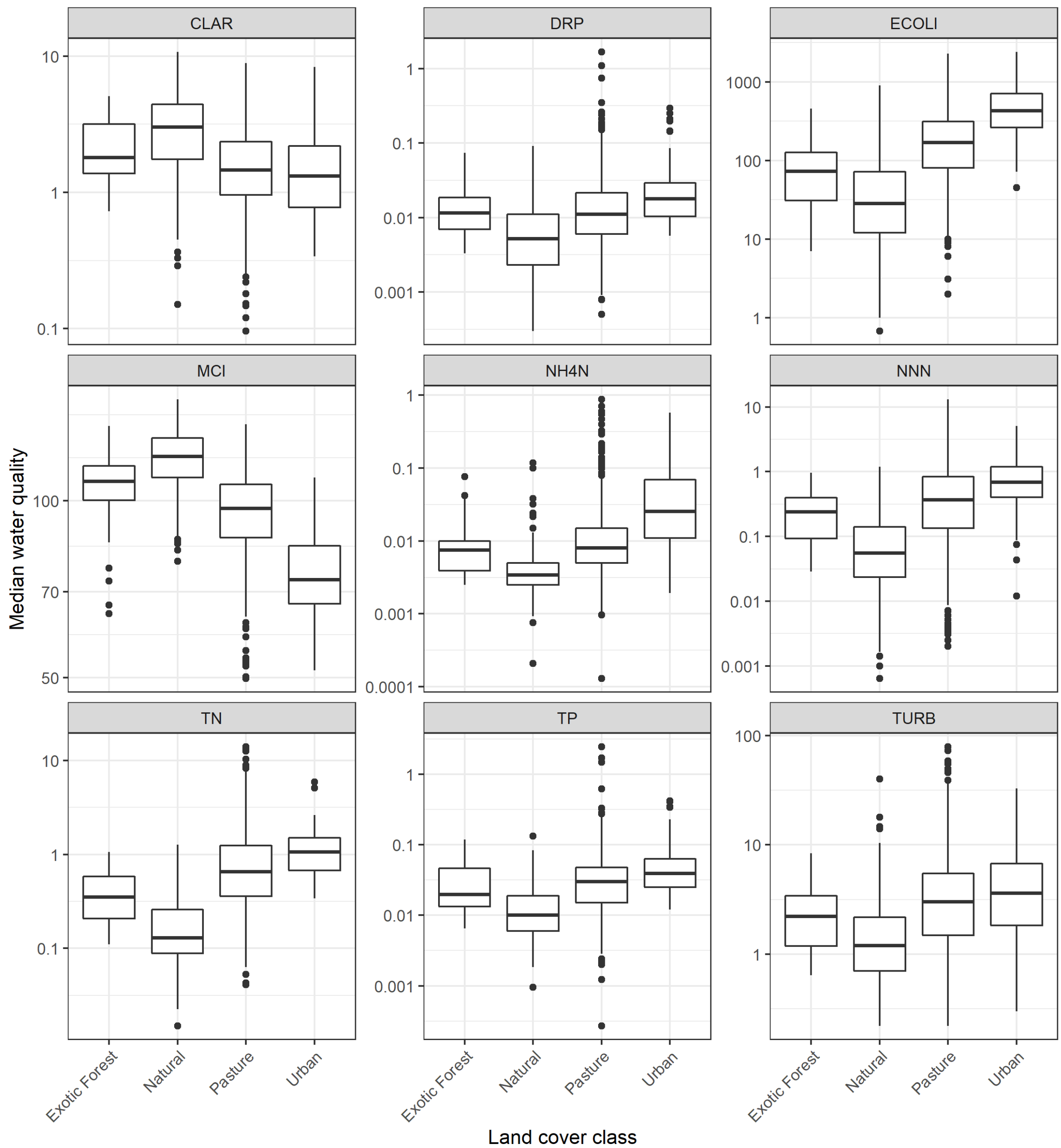

Figure 4-2: River water quality state in REC land cover classes. Box-and-whisker plots show the distributions of monitoring site medians within land cover classes. For y-axes units of measure refer to Table 2-1. Black horizontal line in each box indicates the median of site medians, and the box indicates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

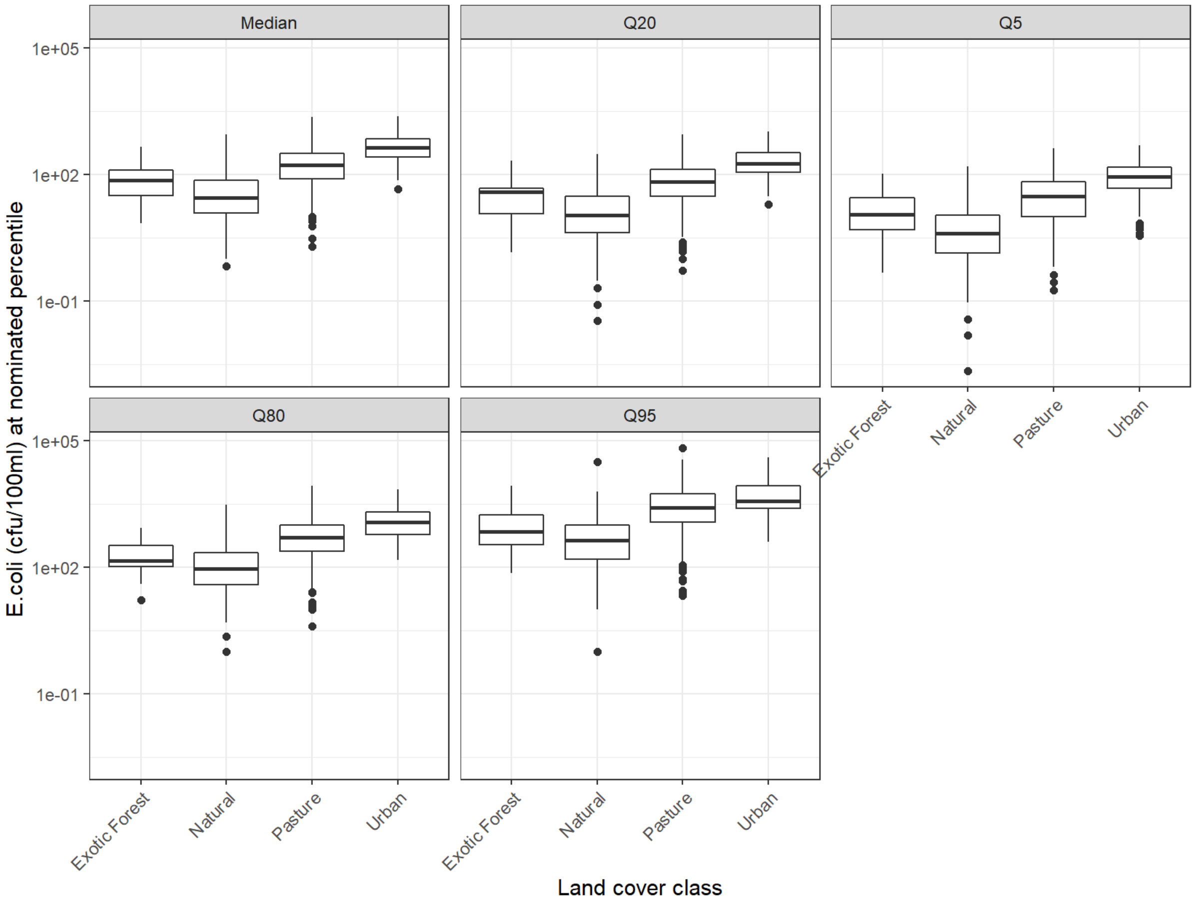

Figure 4-3: ECOLI concentrations in REC land cover classes. Box-and-whisker plots show the distributions of monitoring site percentiles within land cover classes. Black horizontal line in each box indicates the median of site percentiles, and the box indicates the inter-quartile range (IQR). Whiskers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

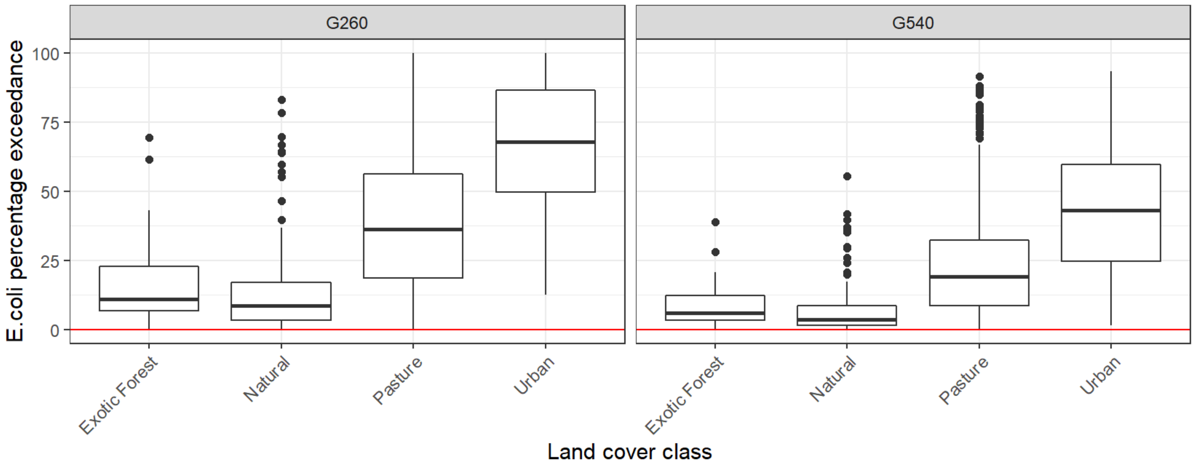

Figure 4-4: ECOLI percent exceedance in REC land cover classes. Box-and-whisker plots show the distributions of percentage exceedance over 540 cfu 100 ml-1 (G540) and 260 cfu 100 ml-1 (G260) at river monitoring sites within land cover classes. Black horizontal line in each box indicates the median of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

4.1 NOF grades

Table 4-2 provides a summary of water quality grades for each NPS-FM attribute, demonstrating the number and percentage of sites that are classified in each NOF grade. Figure 4-5 and Figure 4-6 provide maps for each attribute showing the sites coloured by their evaluated NOF grade. Figure 4-7 and Figure 4-8 shows the percentage of sites belonging to each grade, by land cover category and variable.

The majority of sites (66%) were graded below the national bottom line for the NPS-FM *E.coli* combined numeric attribute state (i.e., most were graded D or E). Over 95% of sites in the urban land cover class, and 75% of sites in the pastoral land cover class were below the bottom line. Very few sites (1-10%) were below the bottom line for the ammonia (toxicity) and nitrate (toxicity) attributes. For the suspended fine sediment attribute (NOF.CLAR.med), 32% of sites were below the bottom line, including 20% of sites belonging to the "Natural" land cover class. For the macroinvertebrate attribute (numeric attribute state of median MCI), 26% of sites were below the bottom line, including over 90% of sites in the "Urban" land cover class. There is no national bottom line for the DRP attribute, but 27% and 19% of sites received D grades for the median and 95th numeric attribute states, respectively. Many of the lowest DRP grades were located in Taranaki and Bay of Plenty, which may in part reflect local geological conditions.

Table 4-2: Summary of the number and percentage (in brackets) of sites assigned to NOF grades. Cells shown in grey are for grades that are below the NOF national bottom line.

Numeric attribute state	NOF Grade				
	A	B	C	D	E
NOF.CLAR.Med	338 (47%)	79 (11%)	67 (9%)	231 (32%)	0 (0%)
NOF.DRP.Med	303 (31%)	182 (19%)	221 (23%)	267 (27%)	0 (0%)
NOF.DRP.p95	468 (48%)	150 (15%)	167 (17%)	188 (19%)	0 (0%)
NOF.ECOLI.Combined	195 (20%)	109 (11%)	29 (3%)	329 (34%)	305 (32%)
NOF.ECOLI.G260	372 (38%)	122 (13%)	35 (4%)	159 (16%)	279 (29%)
NOF.ECOLI.G540	198 (20%)	179 (19%)	199 (21%)	149 (15%)	242 (25%)
NOF.ECOLI.Med	482 (50%)	0 (0%)	0 (0%)	201 (21%)	284 (29%)
NOF.ECOLI.p95	212 (22%)	114 (12%)	35 (4%)	606 (63%)	0 (0%)
NOF.MCI.Median	52 (6%)	240 (28%)	341 (40%)	224 (26%)	0 (0%)
NOF.NH4N.Max	445 (50%)	351 (40%)	77 (9%)	12 (1%)	0 (0%)
NOF.NH4N.Med	795 (90%)	78 (9%)	12 (1%)	0 (0%)	0 (0%)
NOF.NNN.Med	795 (84%)	107 (11%)	36 (4%)	8 (1%)	0 (0%)
NOF.NNN.p95	725 (77%)	164 (17%)	52 (5%)	5 (1%)	0 (0%)

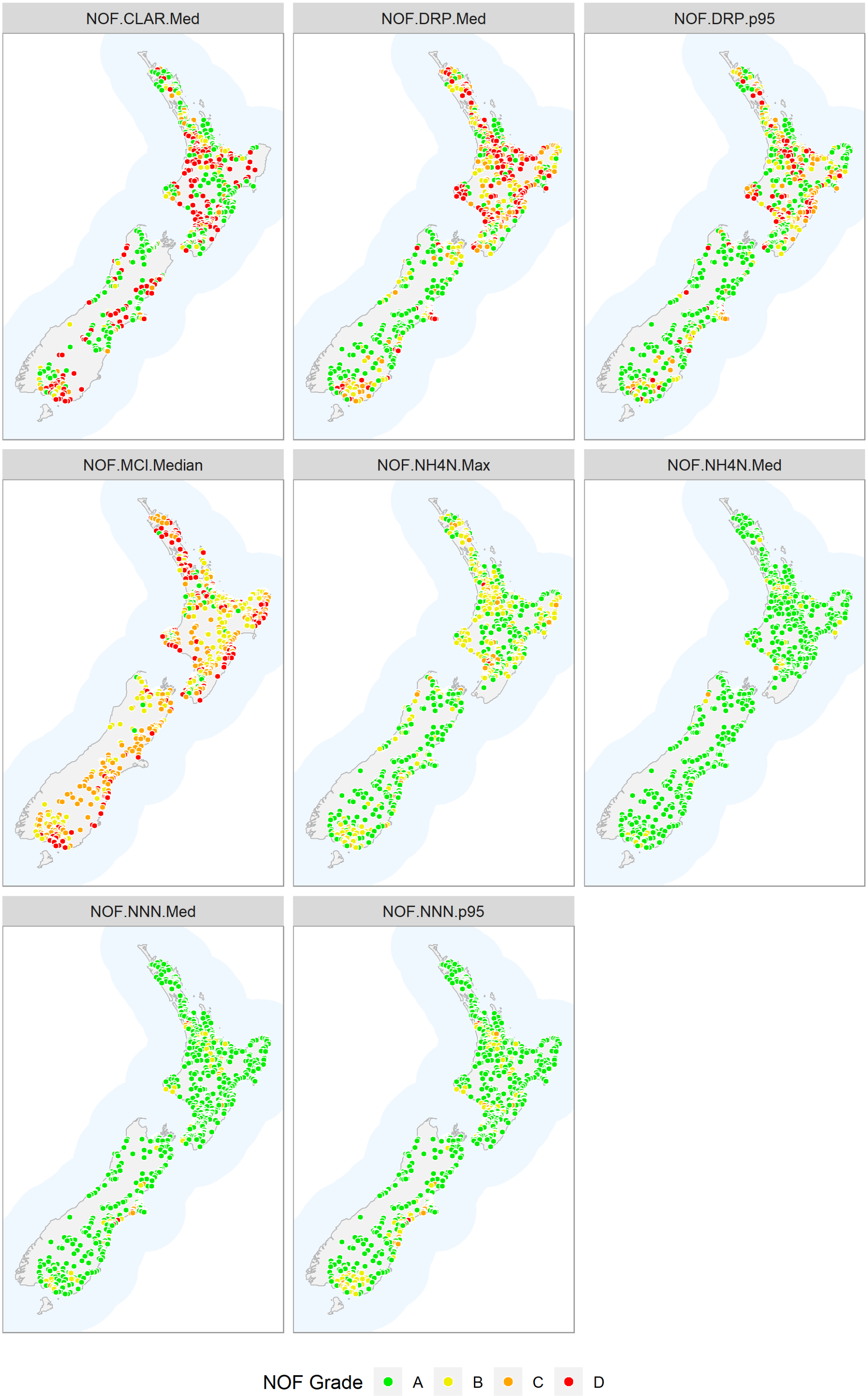

Figure 4-5: Maps showing NOF grades for physico-chemical attributes.

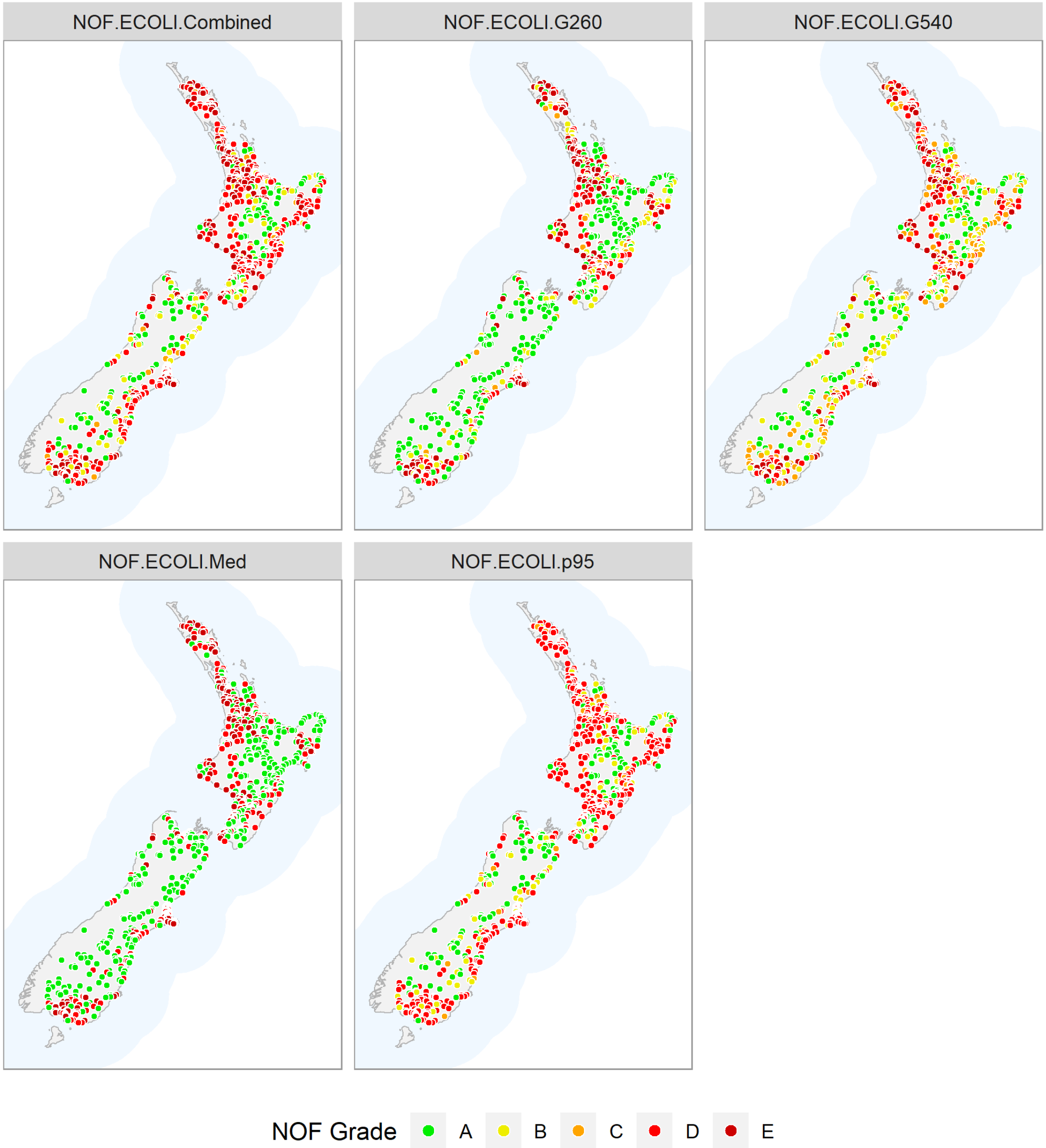

Figure 4-6: Maps showing NOF grades for the E. coli attribute.

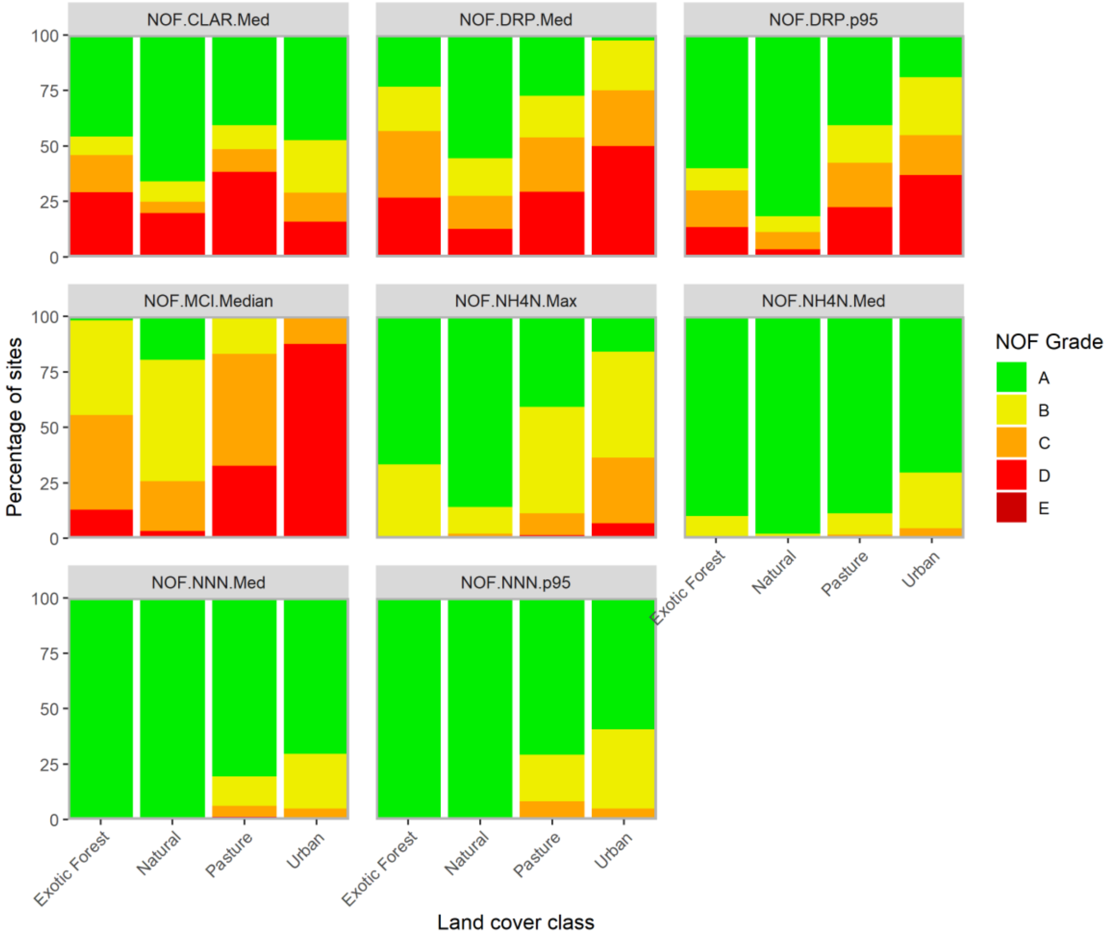

Figure 4-7: Stacked bar charts showing the percentage of physico-chemical sites assigned each NOF grade, by land cover class and NOF attribute.

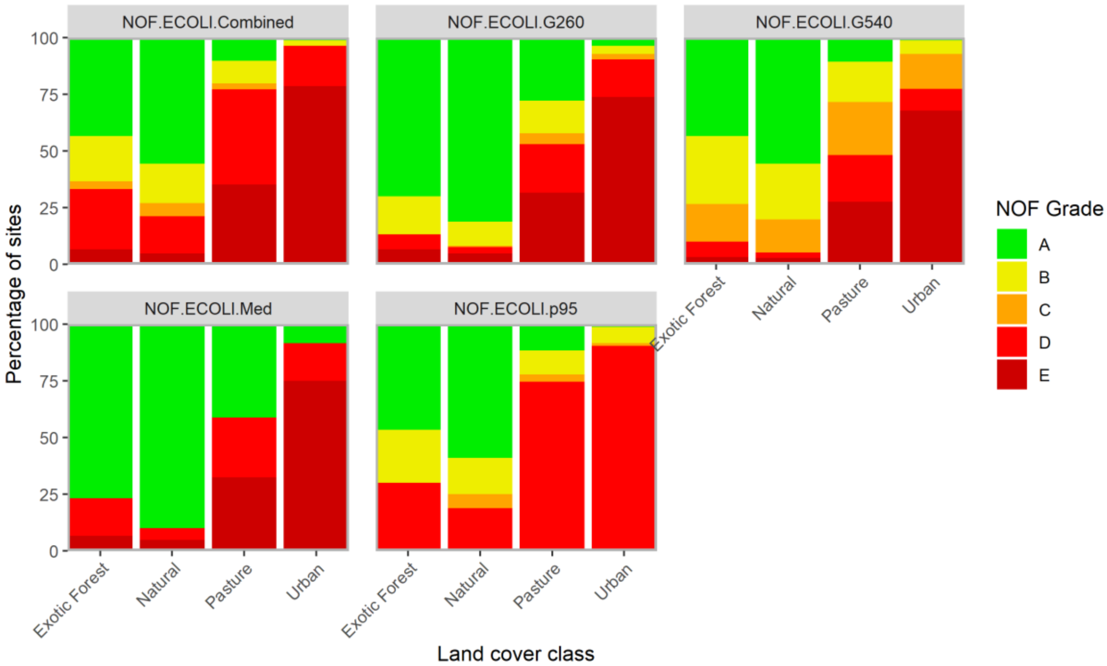

Figure 4-8: Stacked bar charts showing the percentage of E. coli sites assigned each NOF grade, by land cover class and NOF attribute.

4.2 Relationships between water quality state and catchment landcover

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 4-9). Agricultural land cover explained 9%–47% of the variation in log-transformed water-quality variables; these relationships were strongest for median TN, NNN, TP and ECOLI concentrations and MCI scores.

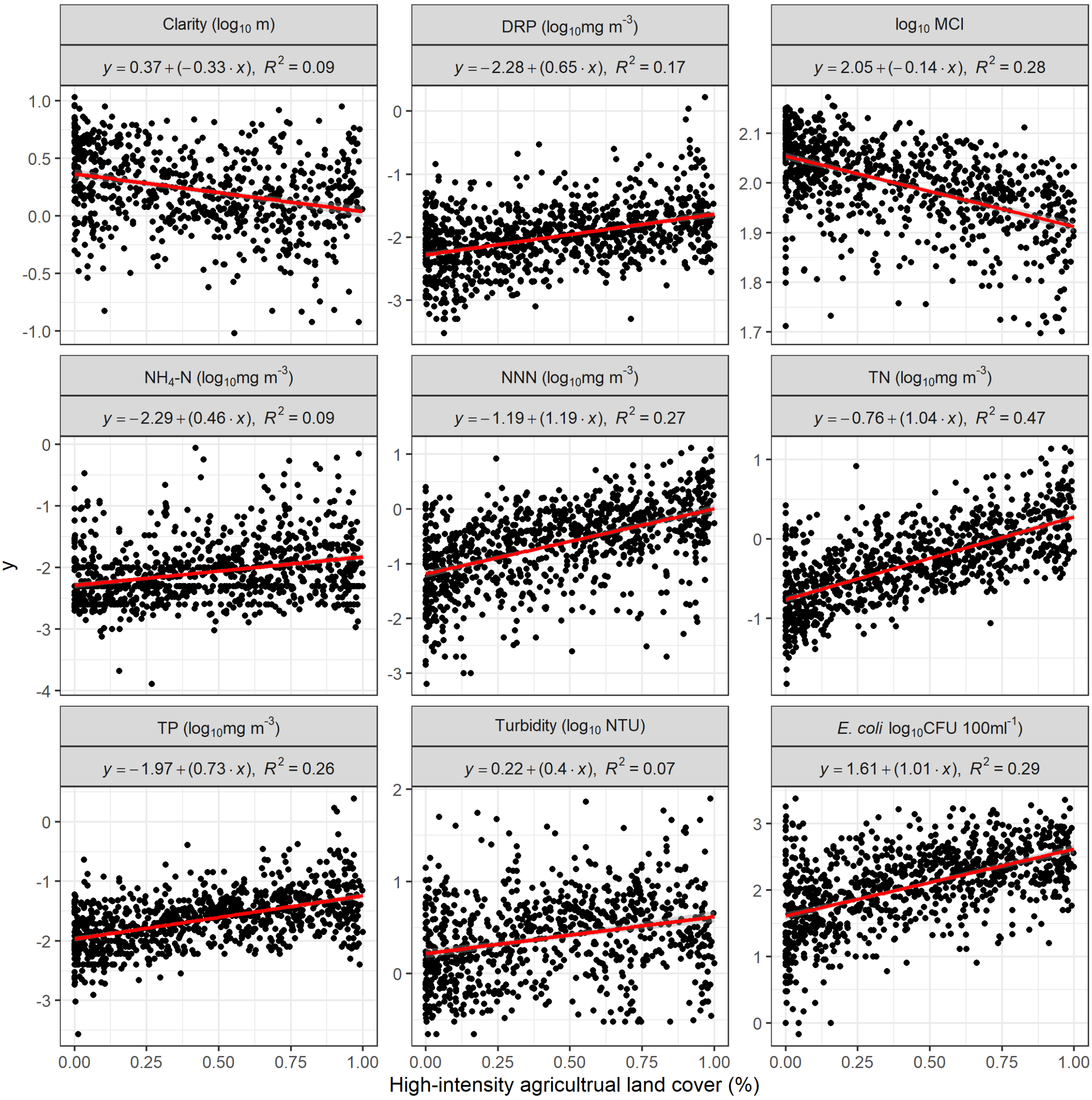

Figure 4-9: Relationships between median water-quality state and proportion of a catchment under high-intensity agricultural land cover in the catchments above monitoring sites in the dataset. Solid lines indicate least squares linear regression models.

5 Results – river trends

Results presented in this section all pertain to raw (i.e., not flow adjusted) trends. Raw and flow adjusted trends for all time periods are provided in the supplementary files "RiverTrends_to2020_v210916.csv" and "FlowAdj_RiverTrends_to2020_210922.csv" respectively.

5.1 Ten-year trends (2011-2020)

Between 512 and 828 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. There were large gaps in the South Island for CLAR. All site locations, land cover classes and numbers of sampling dates are included in the supplementary file "RiverTrends_to2020_v210916.csv".

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	512	21	134	329	28
DRP	819	26	164	555	74
ECOLI	818	28	177	538	75
MCI	681	20	184	432	45
NH4N	828	26	168	559	75
NNN	794	25	162	536	71
TN	720	15	143	510	52
TP	703	15	143	509	36
TURB	747	23	148	505	71

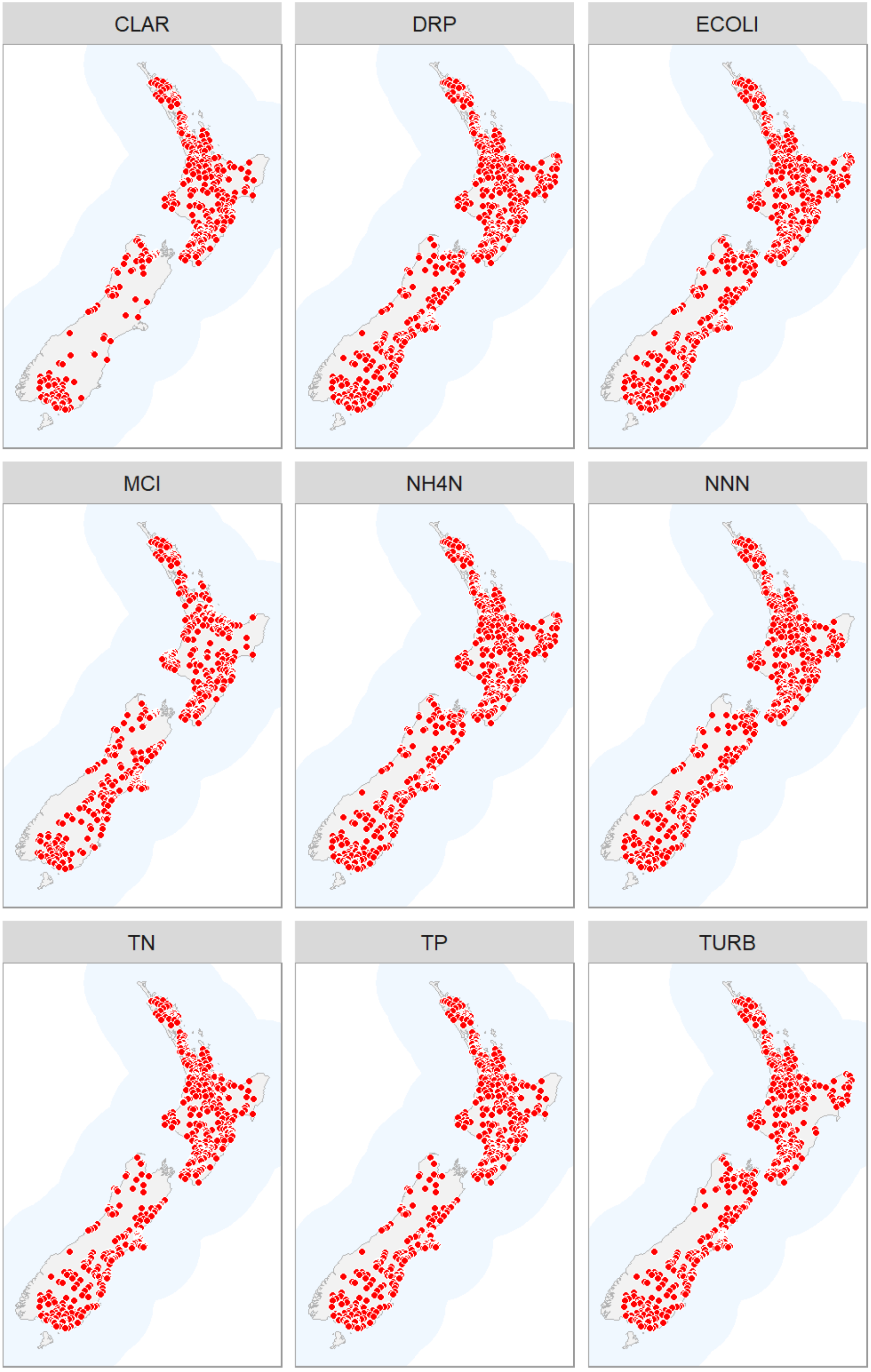

Figure 5-1: River water quality monitoring sites used for 10-year trend analyses of nutrients, ECOLI, CLAR, TURB and MCI.

5.1.1 Trend rate

Box and whisker plots were used to summarise the estimated trend rates for each water quality variable for the 10-year period from 2011 – 2020 across the four land cover classes (Figure 5-2). All estimated trend rates are included in these plots, irrespective of the level of confidence in the assessment (as defined in Section 3.2.5). These plots indicate that land cover classes did not account for a substantial amount of the variation in trend rates for any variable. This contrasts with the state analyses of river variables, where water-quality state clearly varied between land cover classes (Figure 4-2, Figure 4-3, Figure 4-4).

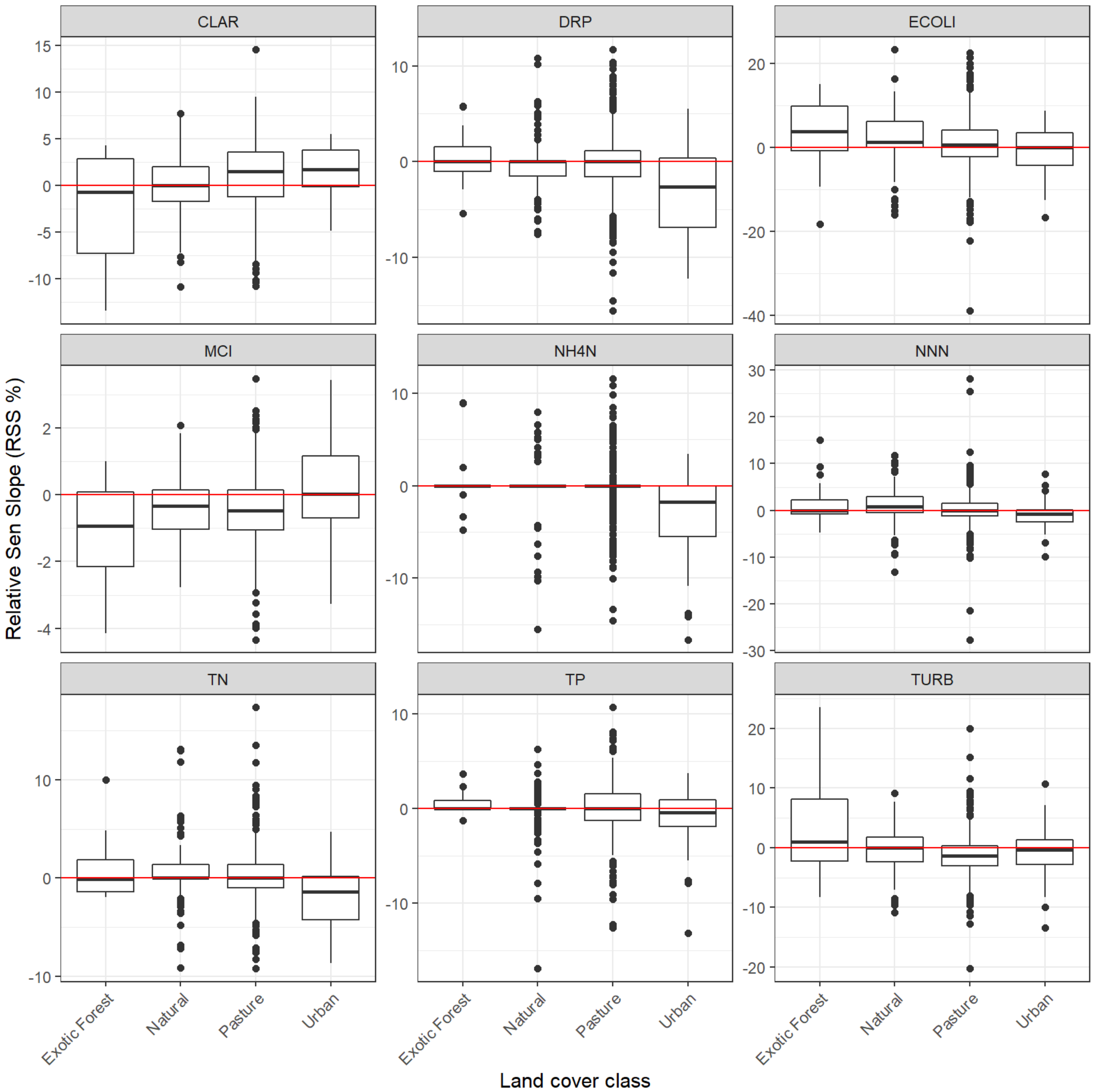

Figure 5-2: Summary of 10-year raw trend rates. Box-and-whisker plots show the distributions of site trend rates within REC land cover classes. Black horizontal line in each box indicates the median of site trend rates and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as and black circles.

5.1.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence of a decreasing 10-year, raw (i.e., not flow adjusted) trend for each site × variable combination. The spatial distributions of categorised individual sites are shown in Figure 5-3. Because confidence that a trend is decreasing is the complement of the confidence that a trend is increasing, "unlikely" decrease, can also be categorised as "likely" increase. Also note, that for MCI and CLAR, decreasing trends indicate degradation, whereas for all other variables decreasing trends indicate improvement.

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

5.1.3 Aggregate trends

Figure 5-4 shows the proportions of sites belonging to each of the nine categorical levels of confidence for C_d defined in Table 3-2 for the 10-year, raw trends. These plots provide a national scale summary of the assessed confidence in trend direction across sites.

The national-scale proportions of decreasing trends (P_d) and their confidence intervals are summarised in Table 5-2. The 10-year P_d statistics ranged from 40-68%. ECOLI and CLAR had a majority (i.e., P_d <50%) of increasing trends at the 95% confidence level. Four of the variables had a majority of decreasing (i.e., P_d>50%) trends at the 95% confidence level (DRP, MCI, NH4N and TURB). The remaining three variables had 95% confidence intervals for the P_d that included 50% (TP, NNN and TN) and we cannot infer widespread increases or decreases for these variables.

The 10-year P_d statistics and 95% confidence intervals for each water-quality variable and land-cover class are shown in Figure 5-5. For eight of the nine water quality variables (all but TURB) the greatest proportion of decreasing trends (P_d<50% for MCI and CLAR, or P_d>50% for all other variables) was in the urban land-cover class. The P_d statistics also indicated that there were a majority of degrading trends (at the 95% confidence level) in ECOLI, MCI and NNN at sites in the natural land-cover class.

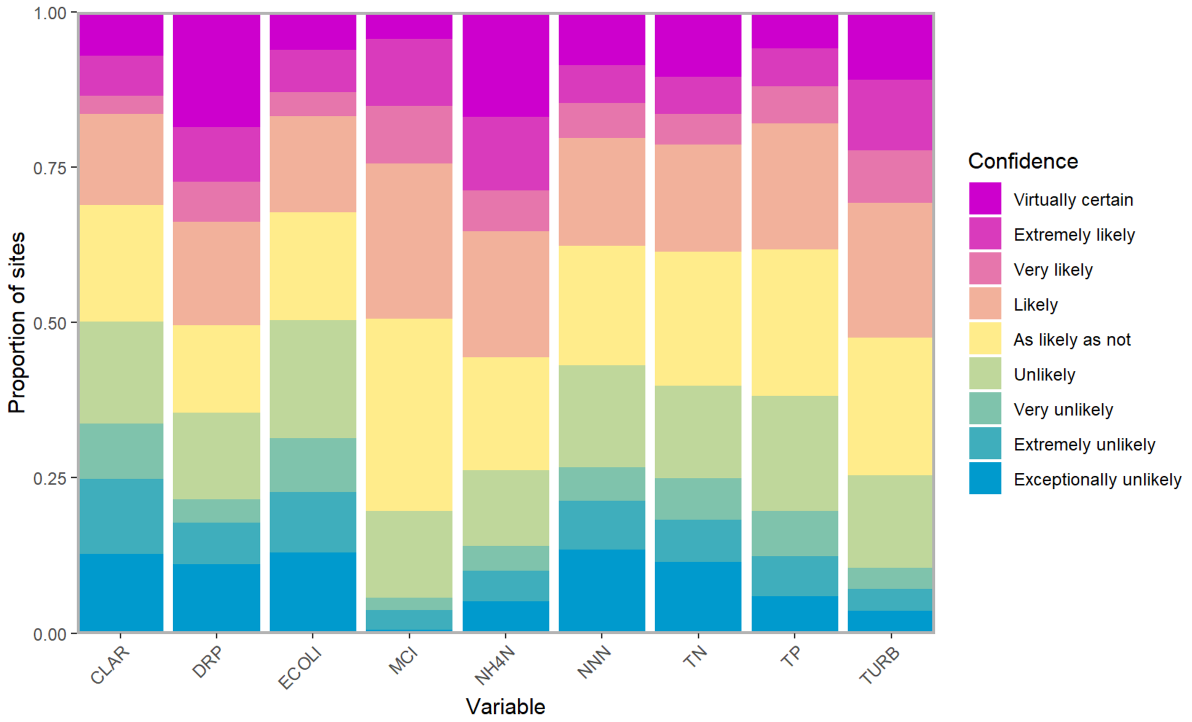

Figure 5-4: Summary plot representing the proportion of sites with decreasing 10-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with decreasing trends at levels of confidence defined in Table 3-2.

Table 5-2: Proportions of decreasing trends (P_d) for 10-year time period.

Variable	Number of sites	P_d (%)	95% confidence interval for P_d (%)
CLAR	512	39.9	37.0 - 42.8
DRP	778	57.7	55.5 - 59.9
ECOLI	809	39.4	37.0 - 41.8
MCI	681	68.4	65.5 - 71.3
NH4N	710	65.6	63.2 - 68.0
NNN	788	46.8	44.4 - 49.2
TN	713	51.4	48.9 - 53.9
TP	701	49.9	47.2 - 52.6
TURB	747	63.9	61.4 - 66.4

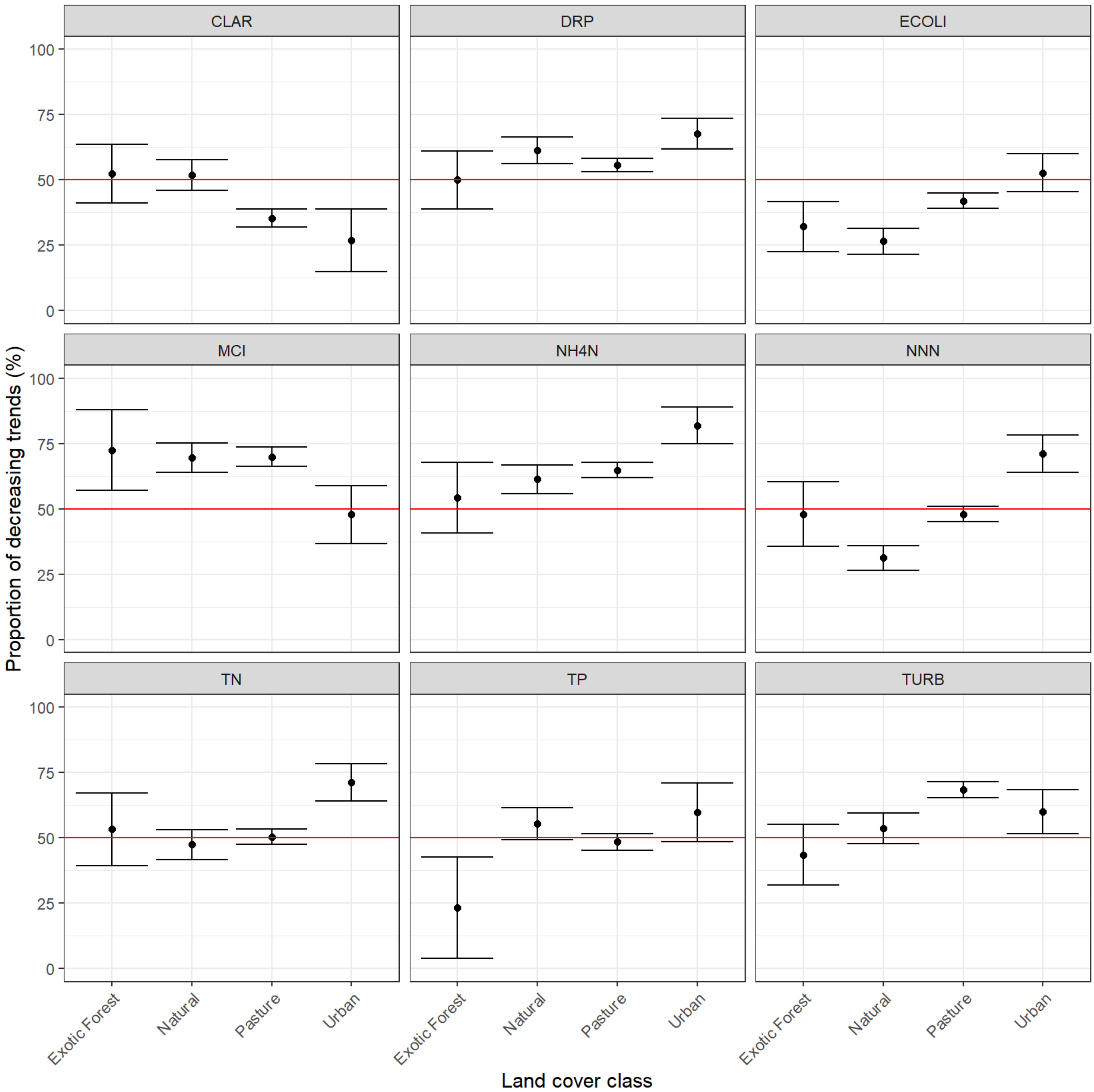

Figure 5-5: Proportions of decreasing trends (P_d) within REC land-cover classes for 10-year trends. Error bars are 95% confidence intervals.

5.2 Twenty-year trends (2001-2020)

Between 115 and 389 river monitoring sites met the filtering rules for the 20-year trend analysis of nutrients, ECOLI, TURB, MCI and CLAR (Table 5-3). The qualifying sites were reasonably well-distributed geographically (Figure 5-6), with gaps in the central North and South islands and the West Coast. The distributions of qualifying sites for ECOLI are concentrated in the north of the North Island and the south and east of the South Island. All site locations, land cover classes and numbers of sampling dates are included in the supplementary file "RiverTrends_to2020_v210916.csv".

Table 5-3: Number of river monitoring sites by REC land cover class and water quality variable included in the 20-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	264	17	70	169	8
DRP	389	20	83	261	25
ECOLI	268	16	57	177	18
MCI	345	12	75	235	23
NH4N	380	19	86	251	24
NNN	364	19	79	243	23
TN	194	5	48	126	15
TP	334	12	70	234	18
TURB	115	5	37	63	10

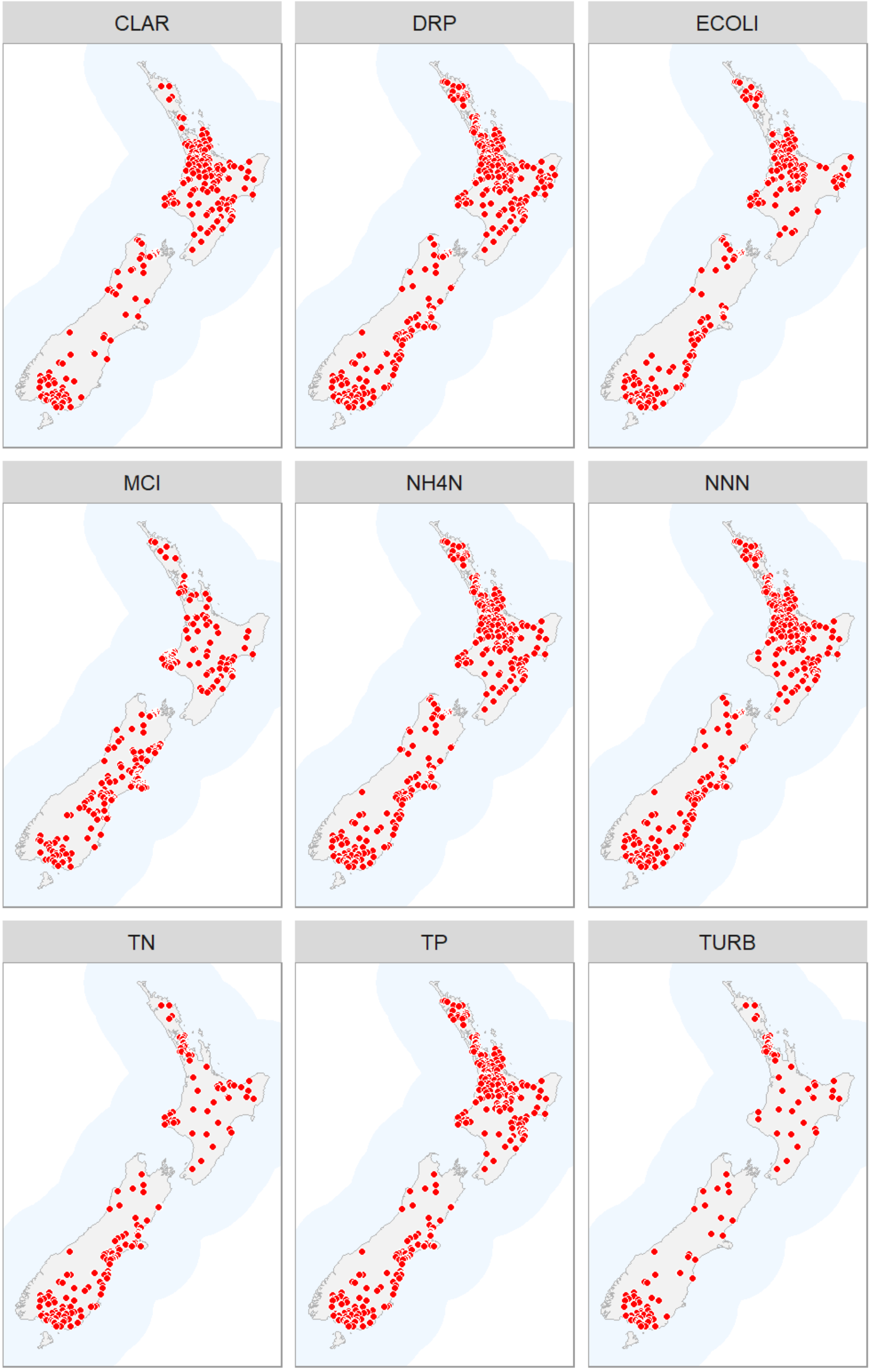

Figure 5-6: River water quality monitoring sites used for 20-year trend analyses of nutrients, ECOLI, CLAR, TURB and MCI.

5.2.1 Trend rate

Box and whisker plots were used to summarise the estimated trend rates for each of the water quality variables for the 20-year period from 2001 – 2020 across the four land cover classes (Figure 5-7). All estimated trend rates are included in these plots, irrespective of the level of confidence in the assessment (as defined in Section 3.2.5). These plots indicate that land cover classes did not account for a substantial amount of the variation in trend rates for any variable, with the exception TURB (although it is noted that TURB has a much lower sample size than other variables). This contrasts with the state analyses of river variables, where water-quality state clearly varied between land cover classes (Figure 4-2, Figure 4-3, Figure 4-4). Median absolute trend rates were largest (> 1.5%) for TP and NH4N in the urban land-cover class (in both cases the trend direction was decreasing) and NNN and TURB for the exotic forest land cover class (in both cases the trend direction was increasing).

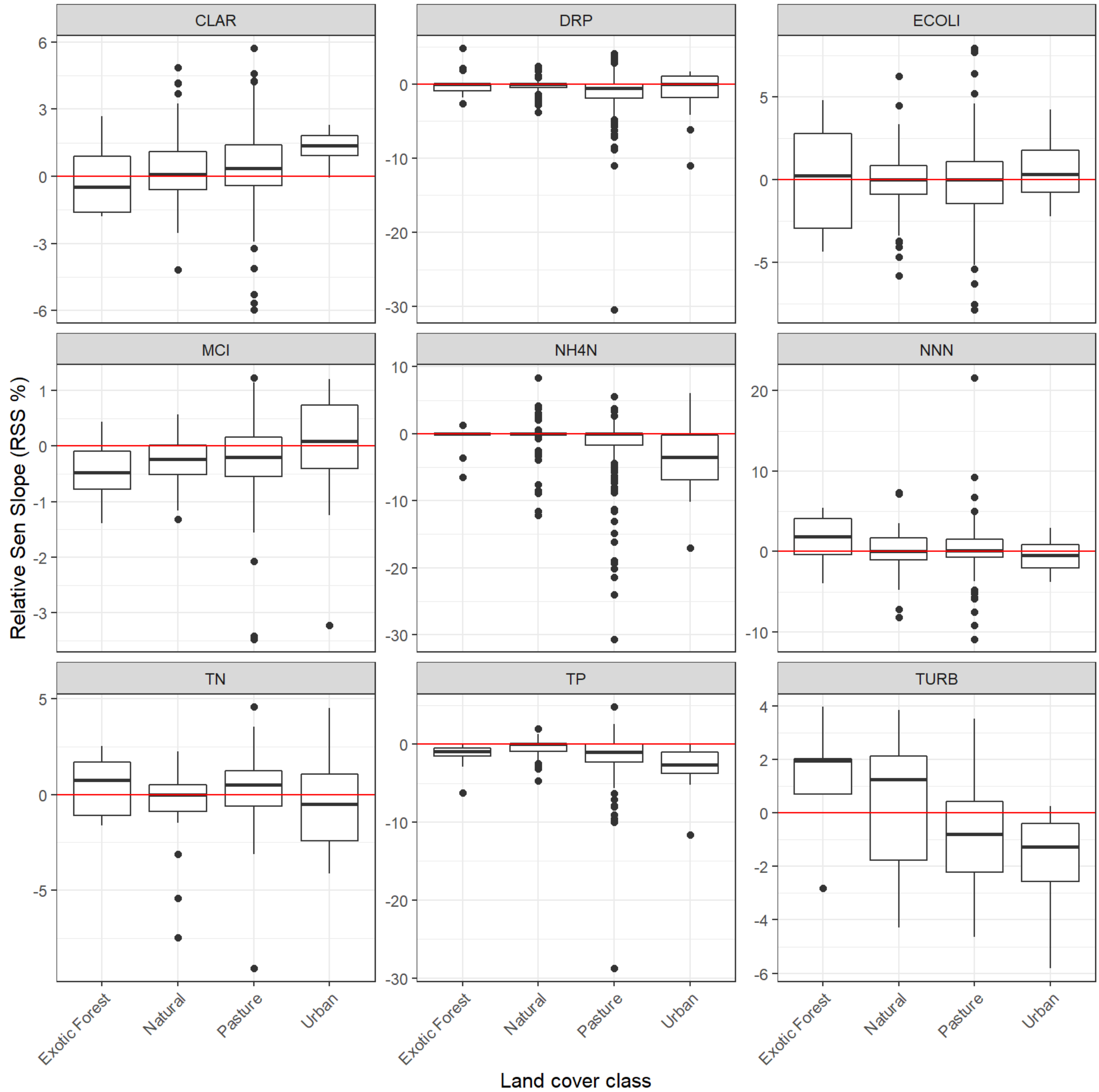

Figure 5-7: Summary of 20-year raw trend rates. Box-and-whisker plots show the distributions of site trend rates within REC land cover classes. Black horizontal line in each box indicates the median of site trend rates, and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as and black circles.

5.2.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence of a decreasing 20-year, raw trend for each site × variable combination. The spatial distributions of categorised individual sites are shown in Figure 5-8. Because confidence that a trend is decreasing is the complement of the confidence that a trend is increasing, "unlikely" decrease, could also be categorised as "likely" increase. Also note, that for MCI and CLAR, decreasing trends indicate degradation, whereas for all other variables decreasing trends indicate improvement.

Figure 5-8: Water quality monitoring sites categorised by the confidence that the 20-year trend is decreasing (C_d) for each variable. C_d is expressed using the confidence categories in Table 3-2. Only sites that met the sampling requirements outlined in Section 3.2.1 are shown.

5.2.3 Aggregate trends

Figure 5-9 shows the proportions of sites belonging to each of the nine categorical levels of confidence for C_d defined in Table 3-2 for the 20-year, raw trends. These plots provide a national-scale summary of the assessed confidence in trend direction across sites.

The national-scale proportions of decreasing trends (P_d) and their confidence intervals are summarised in Table 5-4. The 20-year P_d statistics ranged from 41-82%. CLAR had a majority (i.e., P_d<50%) of increasing trends, at the 95% confidence level. Four of the variables had a majority of decreasing (i.e., P_d>50%) trends, at the 95% confidence level (MCI, DRP, NH4N, and TP). The remaining three variables had 95% confidence intervals for the P_d that included 50% (NNN, TN and TURB), and we cannot infer widespread decreases or increases for these variables.

The 20-year P_d statistics and 95% confidence intervals for each water-quality variable and land-cover class are shown in Figure 5-10. For seven of the nine water quality variables (all but DRP and ECOLI) the greatest proportion of decreasing trends (P_d<50% for MCI and CLAR, or P_d>50% for all other variables) was in the urban land-cover class.

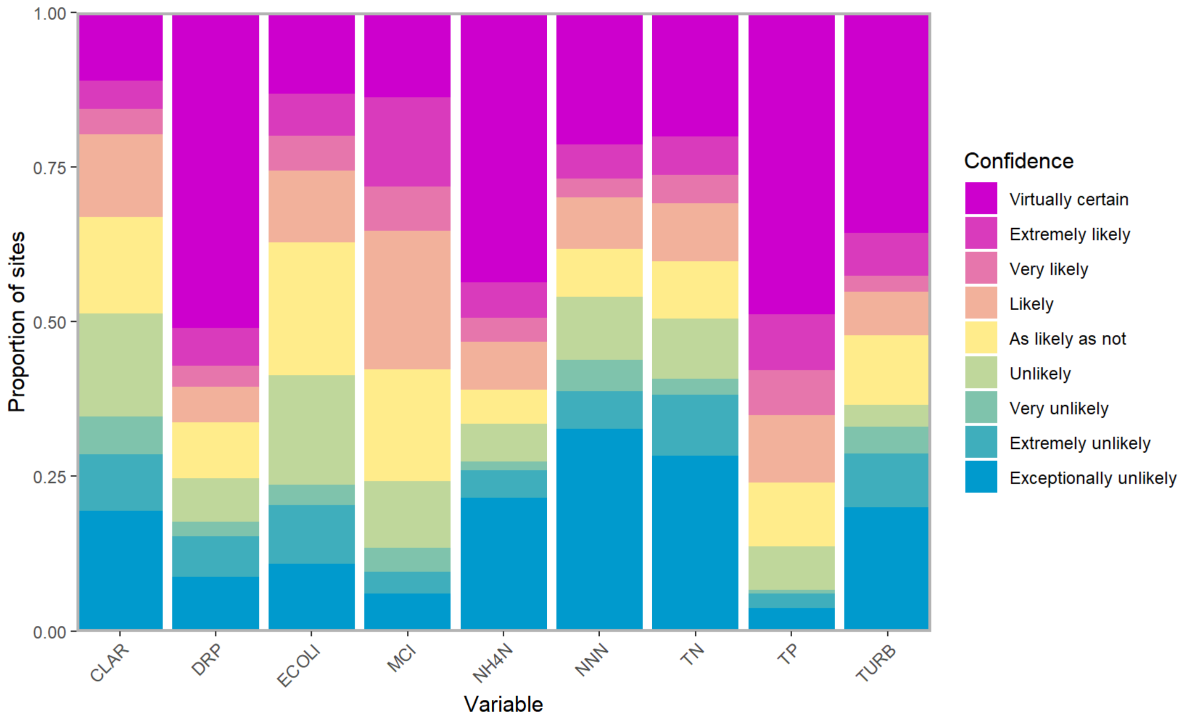

Figure 5-9: Summary plot representing the proportion of sites with decreasing 20-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with decreasing trends at levels of confidence defined in Table 3-2.

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

Variable	Number of sites	P_d (%)	95% confidence interval for P_d (%)
CLAR	263	40.5	36.8 - 44.2
DRP	380	71.8	69.4 - 74.2
ECOLI	266	47.9	44.0 - 51.8
MCI	345	65.5	62.0 - 69.0
NH4N	362	63.5	61.3 - 65.7
NNN	361	42.7	40.2 - 45.2
TN	194	46.4	42.9 - 49.9
TP	330	82.1	79.4 - 84.8
TURB	115	57.4	53.3 - 61.5

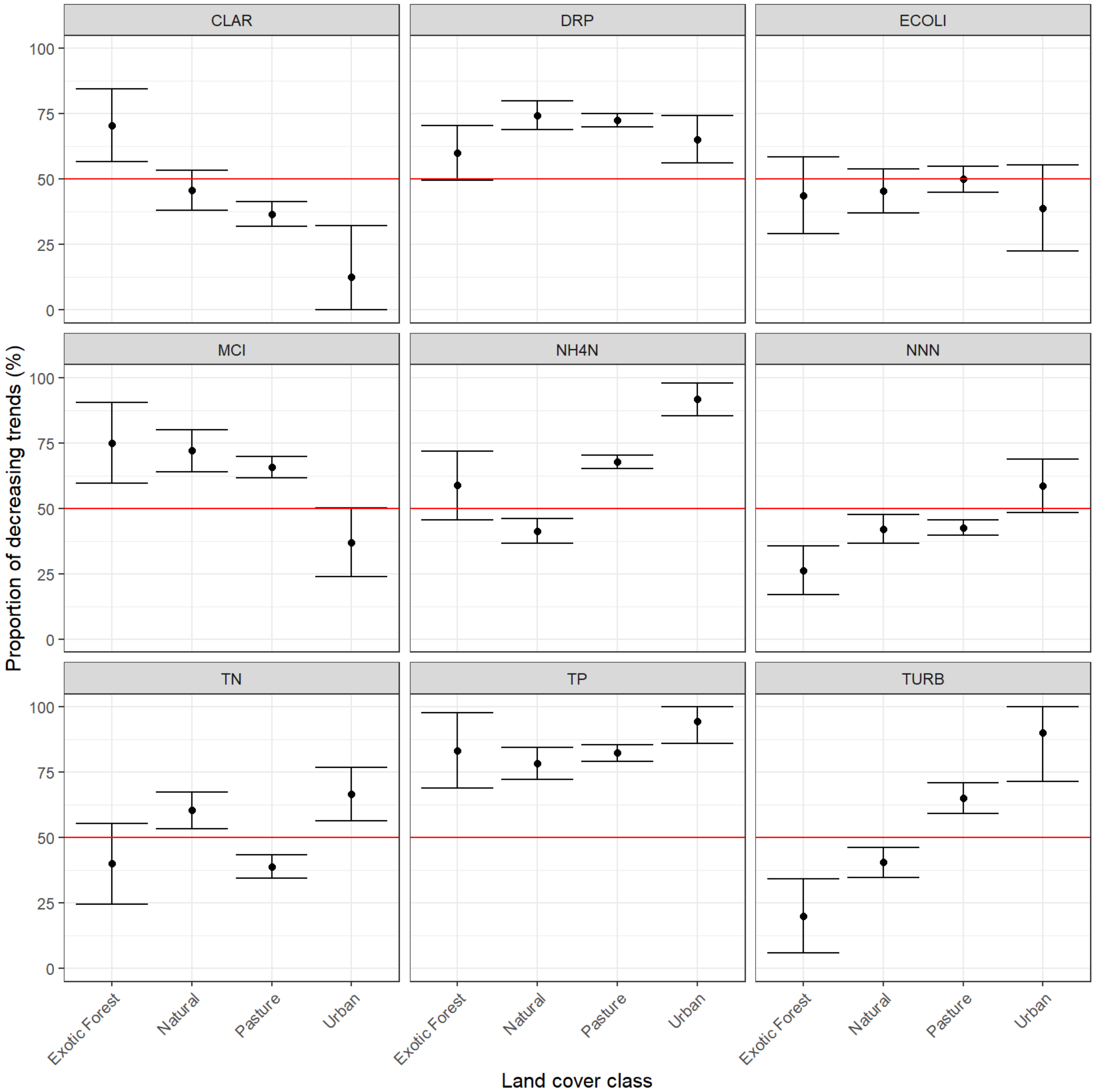

Figure 5-10: Proportions of decreasing trends (P_d) within REC land-cover classes for 20-year trends. Error bars are 95% confidence intervals.

5.3 Thirty-year trends (1991-2020)

Between 19 and 179 river monitoring sites met the filtering rules for the 30-year trend analysis of nutrients, ECOLI, TURB, MCI and CLAR (Table 5-5). The qualifying sites were reasonably well-distributed geographically (Figure 5-11), although ECOLI was limited to sites in the eastern North Island, and sites were overrepresented for the Waikato region for CLAR, DRP, NNN, NH4N and TP. The exotic forest and urban land-cover classes were poorly represented. All site locations, land cover classes and numbers of sampling dates are included in the supplementary file "RiverTrends_to2020_v210916.csv".

Table 5-5: Number of river monitoring sites by REC land cover class and water quality variable included in the 30-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	115	6	32	77	0
DRP	179	9	37	124	9
ECOLI	19	2	1	15	1
MCI	33	0	8	25	0
NH4N	164	8	36	113	7
NNN	164	8	37	113	6
TN	68	3	26	35	4
TP	160	8	37	108	7
TURB	58	3	26	29	0

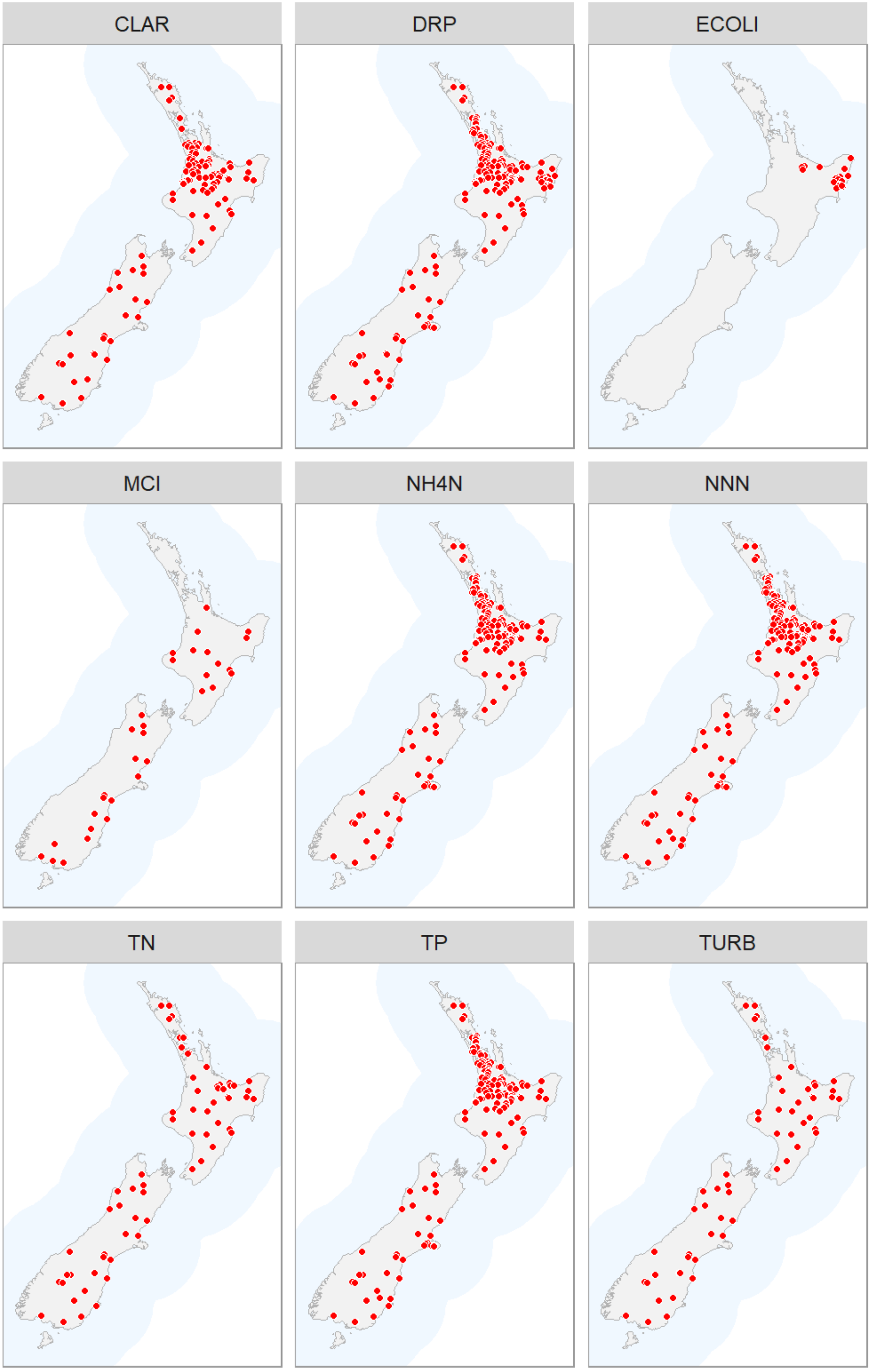

Figure 5-11: River water quality monitoring sites used for 30-year trend analyses of nutrients, ECOLI, CLAR, TURB and MCI.

5.3.1 Trend rate

Box and whisker plots were used to summarise the estimated trend rates for each of the water quality variables for the 30-year period from 1991 – 2020 across the four land cover classes (Figure 5-12). All estimated trend rates are included in these plots, irrespective of the level of confidence in these assessments (as defined in Section 3.2.5). The interpretation of Figure 5-12 should take into account the site numbers described in Table 5-5, which shows that numbers of sites by land-class for the 30-year period were generally very low or zero for the urban and exotic forest land-cover classes, and low across all land-cover classes for MCI and ECOLI.

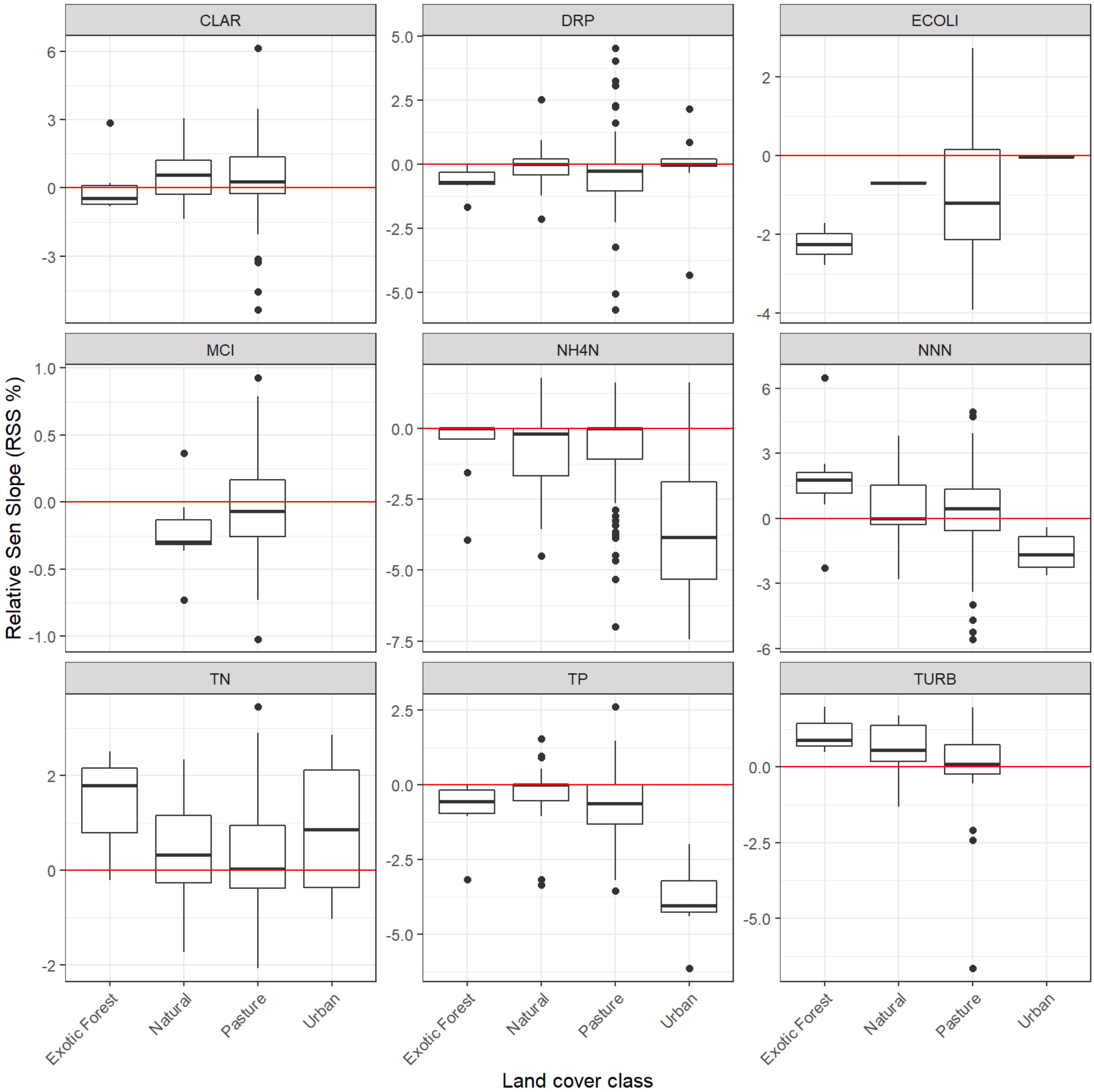

Figure 5-12: Summary of 30-year raw trend rates. Box-and-whisker plots show the distributions of site trend rates within REC land cover classes. Black horizontal line in each box indicates the median of site trend rates, and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as and black circles.

5.3.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence of a decreasing 30-year, raw trend for each site × variable combination. The spatial distributions of categorised individual sites are shown in Figure 5-13. Because confidence that a trend is decreasing is the complement of the confidence that a trend is increasing, “unlikely” decrease, could also be categorised as “likely” increase. Also note, that for MCI and CLAR, decreasing trends indicate degradation, whereas for all other variables decreasing trends indicate improvement.

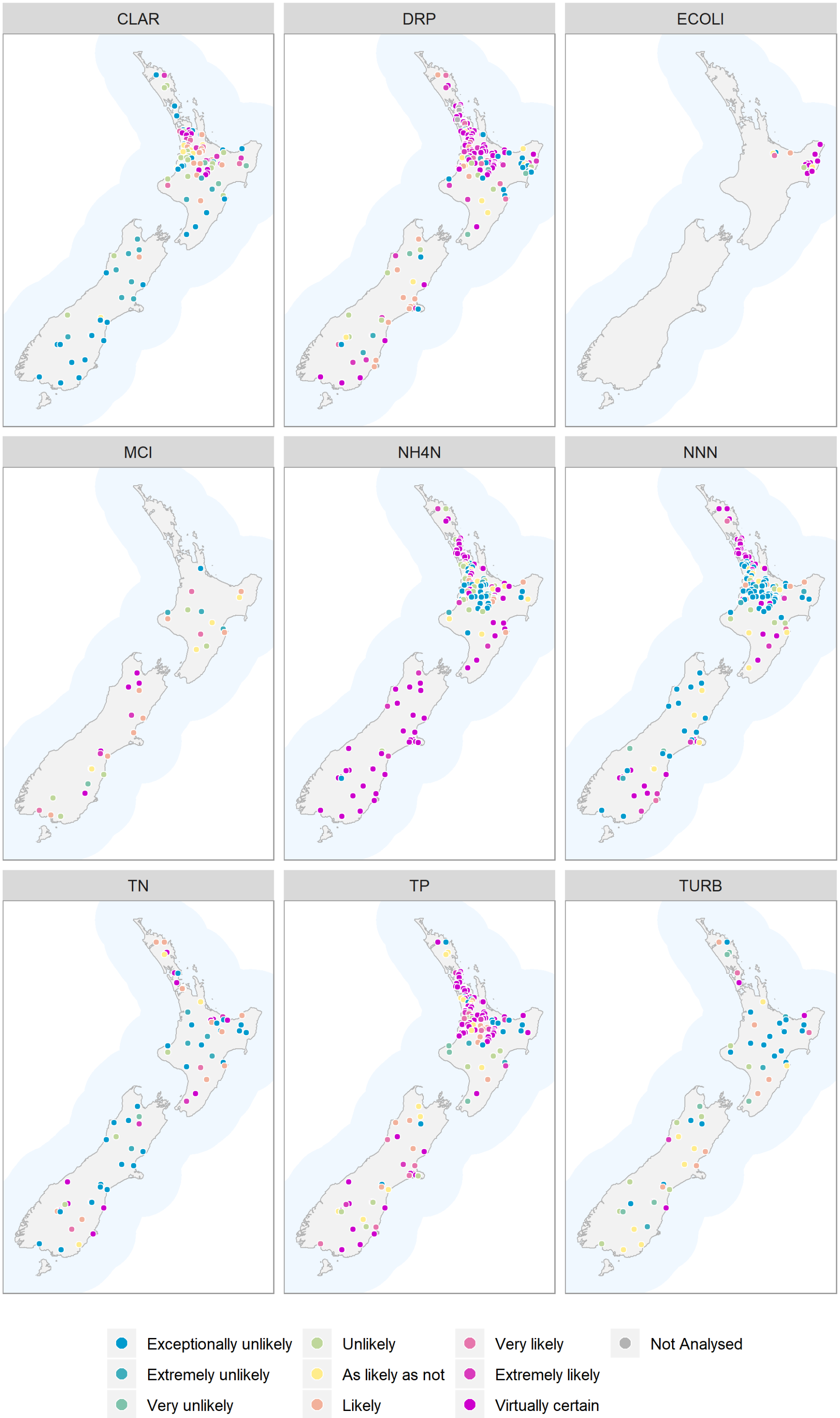

Figure 5-13: Water quality monitoring sites categorised by the confidence that the 30-year trend is decreasing (C_d) for each variable. C_d is expressed using the confidence categories Table 3-2. Only sites that met the sampling requirements outlined in Section 3.2.1 are shown in the figure.

5.3.3 Aggregate Trends

Figure 5-14 shows the proportions of sites belonging to each of the nine categorical levels of confidence for C_d defined in Table 3-2 for the 30-year, raw trends. These plots provide a national-scale summary of the assessed confidence in trend direction across sites.

The national-scale proportions of decreasing trends (P_d) and their confidence intervals are summarised in Table 5-6. The 30-year P_d statistics ranged from 29-76%. CLAR, NNN and TURB had a majority (i.e., P_d<50%) of increasing trends, at the 95% confidence level. Three of the variables had a majority of decreasing (i.e., P_d>50%) trends, at the 95% confidence level (DRP, TP, ECOLI). The remaining three variables had 95% confidence intervals for the P_d that included 50% (MCI, NH4N, and TN), and we cannot infer widespread increases or decreases for these variables.

The 30-year P_d statistics and 95% confidence intervals for each water-quality variable and land-cover class are shown in Figure 5-14. For water quality variable x land-cover class combinations that included more than 20 sites, the P_d statistics indicated that there were a majority (at the 95% confidence level) of increasing trends in NNN, and of decreasing trends in DRP and TP at sites in the pasture land-cover class.

The interpretation of Figure 5-14, Table 5-6 and Figure 5-15 should take into account the site numbers described in Table 5-5, which shows that numbers of sites by land-class for the 30-year period were generally very low or zero for the urban and exotic forest land-cover classes, and low across all land-cover classes for MCI and ECOLI.

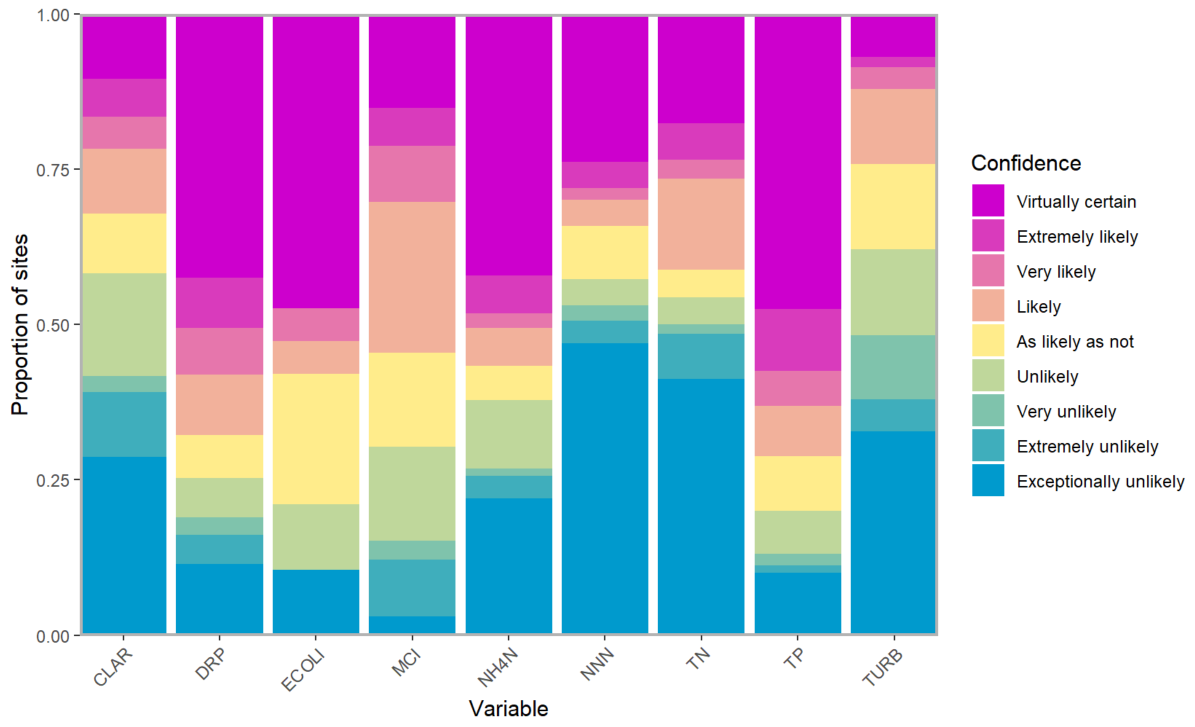

Figure 5-14: Summary plot representing the proportion of sites with decreasing 30-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with decreasing trends at levels of confidence defined in Table 3-2.

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

Variable	Number of sites	P_d (%)	95% confidence interval for P_d (%)
CLAR	115	34.8	29.7 - 39.9
DRP	174	70.7	67.4 - 74.0
ECOLI	19	73.7	60.8 - 86.6
MCI	33	66.7	55.3 - 78.1
NH4N	164	59.1	56.0 - 62.2
NNN	164	37.8	34.9 - 40.7
TN	68	42.6	37.3 - 47.9
TP	160	75.6	72.1 - 79.1
TURB	58	29.3	21.7 - 36.9

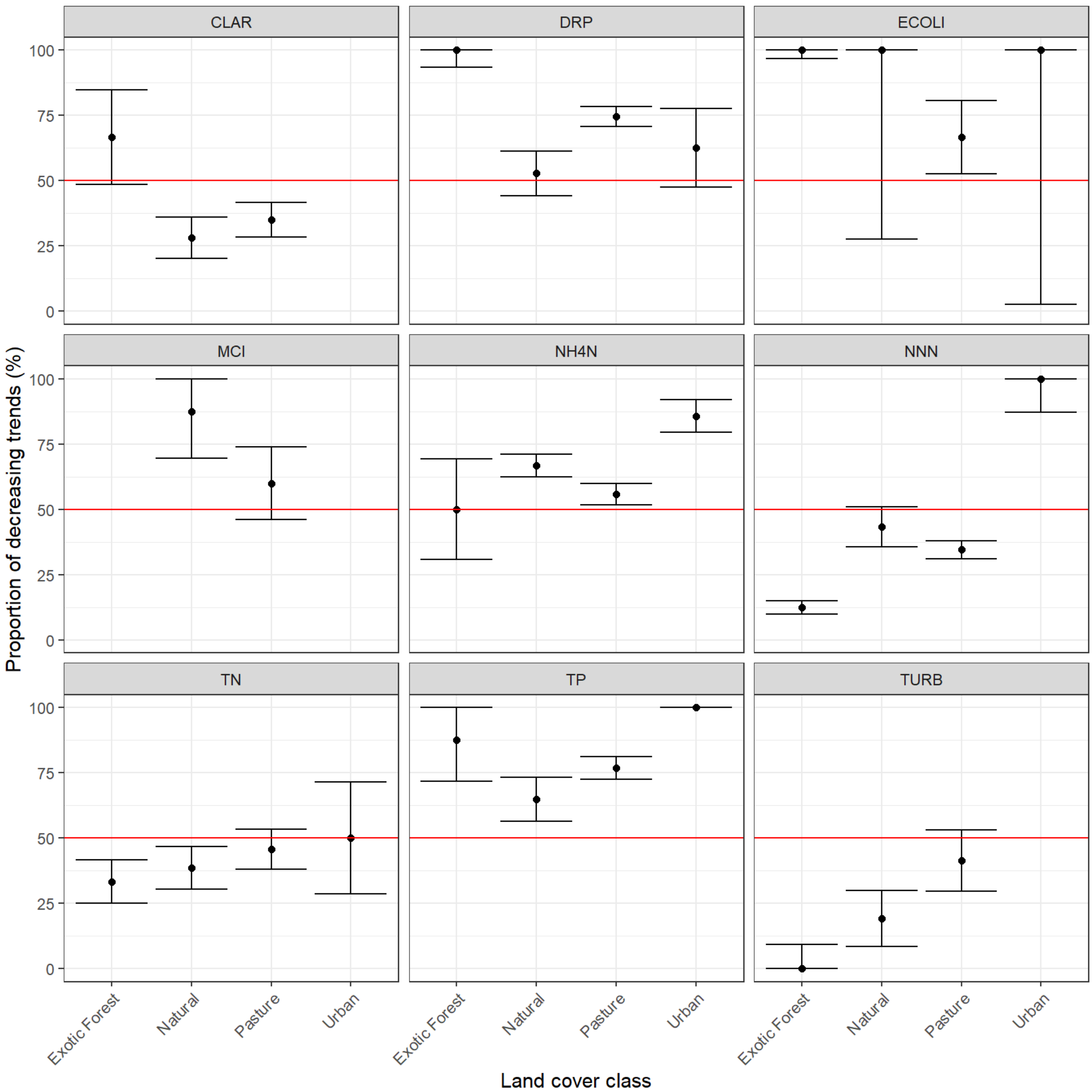

Figure 5-15: Proportions of decreasing trends (P_d) within REC land-cover classes for 30-year trends. Error bars are 95% confidence intervals.

5.4 Comparisons of trend directions between 10-, 20- and 30-year periods

The national scale P_d statistics for each water quality variable are shown in Table 5-7, which combines the results in Table 5-2, Table 5-4 and Table 5-6. A comparison of the 10-, 20- and 20-year trends in this table reveal several changes in the balance of decreasing and increasing trends:

1. a predominance of decreasing 30-year trends in ECOLI, shifted to roughly equal proportions of increasing and decreasing 20-year trends and then to a predominance of increasing 10-year trends;
2. a predominance of increasing 30-year trends in TURB, shifted to roughly equal proportions of increasing and decreasing 20-year trends and then to a predominance of decreasing 10-year trends;
3. a predominance of decreasing 20- and 30-year trends in TP shifted to roughly equal proportions of increasing and decreasing 10-year trends.

In contrast to these changes between trend periods, the predominance of increasing trends in CLAR and NNN and decreasing trends DRP, NH4N and MCI have persisted between all trend periods.

Table 5-7: National-scale P_d statistics. Values are estimated percentages of river sites with decreasing trends across New Zealand. Magenta cells: majority of sites decreasing. Blue cells: majority of sites increasing. Yellow cells: cannot infer increases or decreases at most sites because the 95% confidence intervals for the P_d statistic included 50%.

Variable	10-year trend (2011-2020)	20-year trend (2001-2020)	30-year trend (1991-2020)
CLAR	39.9	40.5	34.8
DRP	57.7	71.8	70.7
ECOLI	39.4	47.9	73.7
MCI	67.8	66.8	66.7
NH4N	65.6	63.5	**66.7**
NNN	46.8	42.7	37.8
TN	51.4	46.4	42.6
TP	49.9	82.1	75.6
TURB	63.9	57.4	29.3

5.5 Rolling ten-year trends

The number of river monitoring sites that met the filtering rules for inclusion in each of the ten-year windows in the rolling trends analyses ranged from 6 to 105 for the first time window, and from 512 to 828 for the most recent time window. The changes in the number of sites that were included in the rolling trend analysis over time and by variable are shown in Figure 5-16. The number of time windows for each site by variable combination that complied with the filtering rules are mapped in Figure 5-17. All site locations, land cover classes and numbers of sampling dates are included in the supplementary file “RiverTrends_to2020_v220225.csv”.

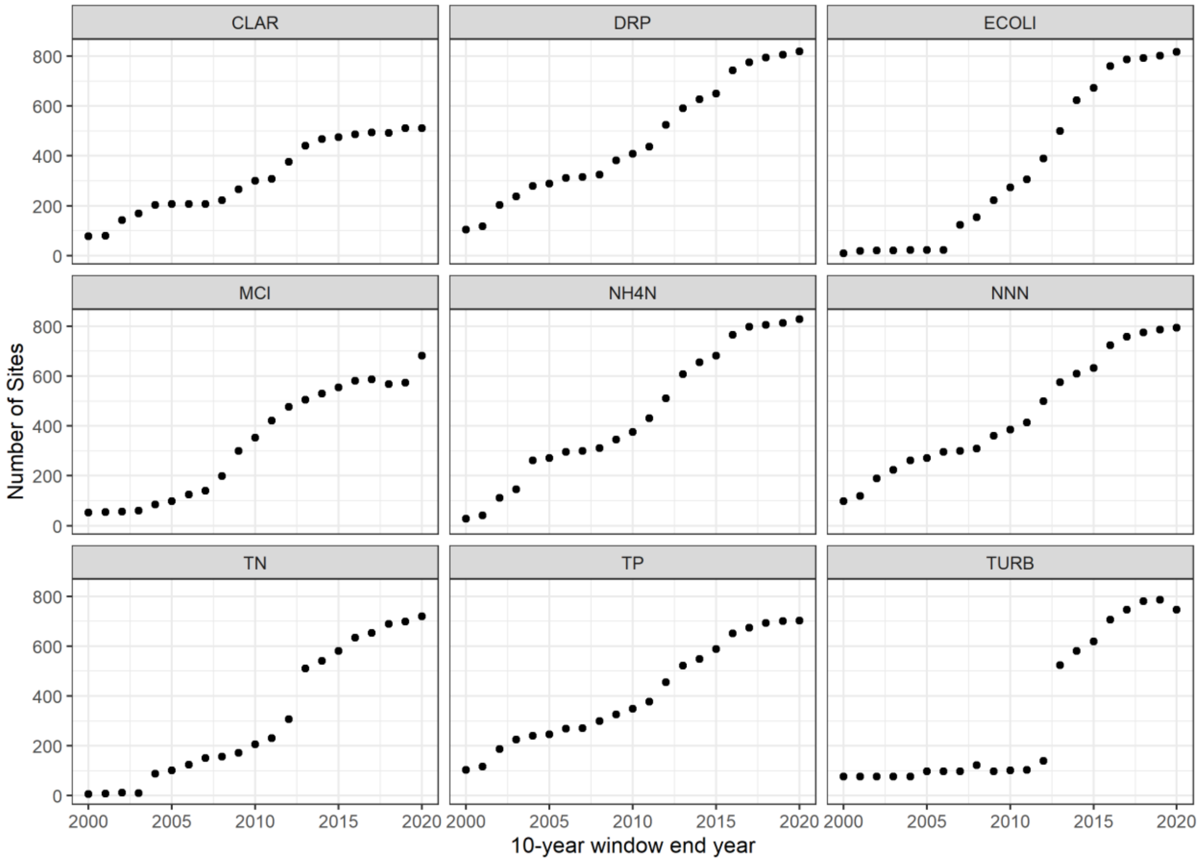

Figure 5-16: Number of sites that were included for each 10-year window included in the rolling trends analysis by variable.

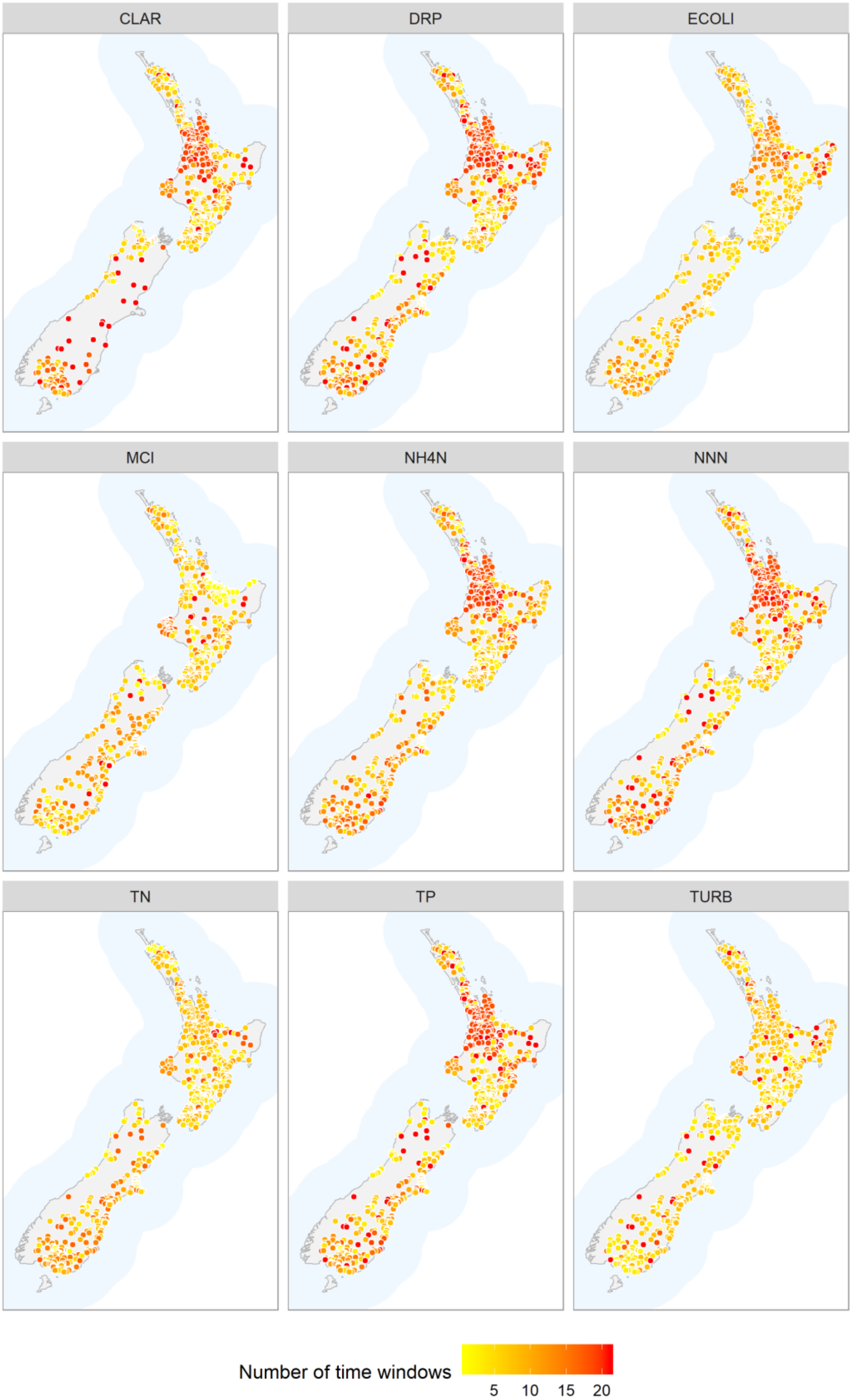

Figure 5-17: Number time windows that each river water quality monitoring site was included the 10-year rolling trend analyses of nutrients, ECOLI, CLAR, TURB and MCI.

5.5.1 Trend rate

Box and whisker plots summarise the trends assessed for each 10-year time window and each water quality variable (Figure 5-18). All assessed trends are included in these plots, irrespective of the level of confidence in the assessment (see Section 3.2.5). Time windows are only shown where the sample size was at least 200 sites. This arbitrary cut-off is intended to minimise bias that might be present in a small sample size but maximise the number of time windows that were reported. The plots show quasi-periodic fluctuations in median RSS for all variables.

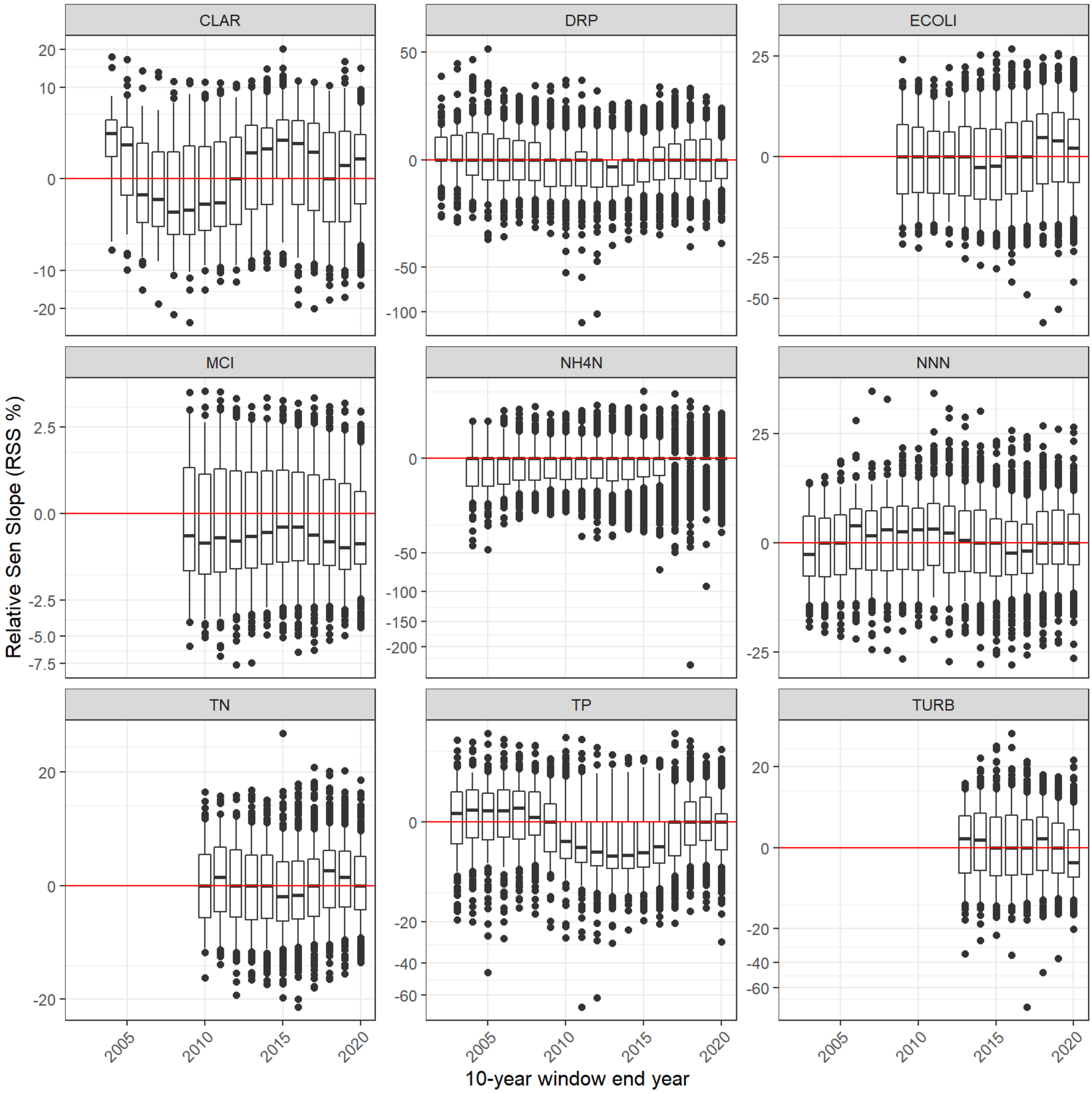

Figure 5-18: Summary of raw trend rates for rolling 10-year windows. Box-and-whisker plots show the distributions of relative trend rates (i.e., Sen slopes) within each ten-year window. Black horizontal line in each box indicates the median of site trend rates, and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5*IQR from the box. Data beyond the whiskers are shown as and black circles. The red line indicates a trend rate of zero. Note, y-axis has a signed square root transformation. Units for each variable are given in Table 2-1.

5.5.2 Aggregate trends

Figure 5-19 shows the proportions of raw site trends assigned to each of the nine categorical levels of confidence for C_d defined in Table 3-2 for each of the 10-year time window, These plots provide a national-scale summary of the assessed confidence in trend direction across sites and the rolling 10-year time windows. Time windows are only shown where the sample size was at least 200 sites.

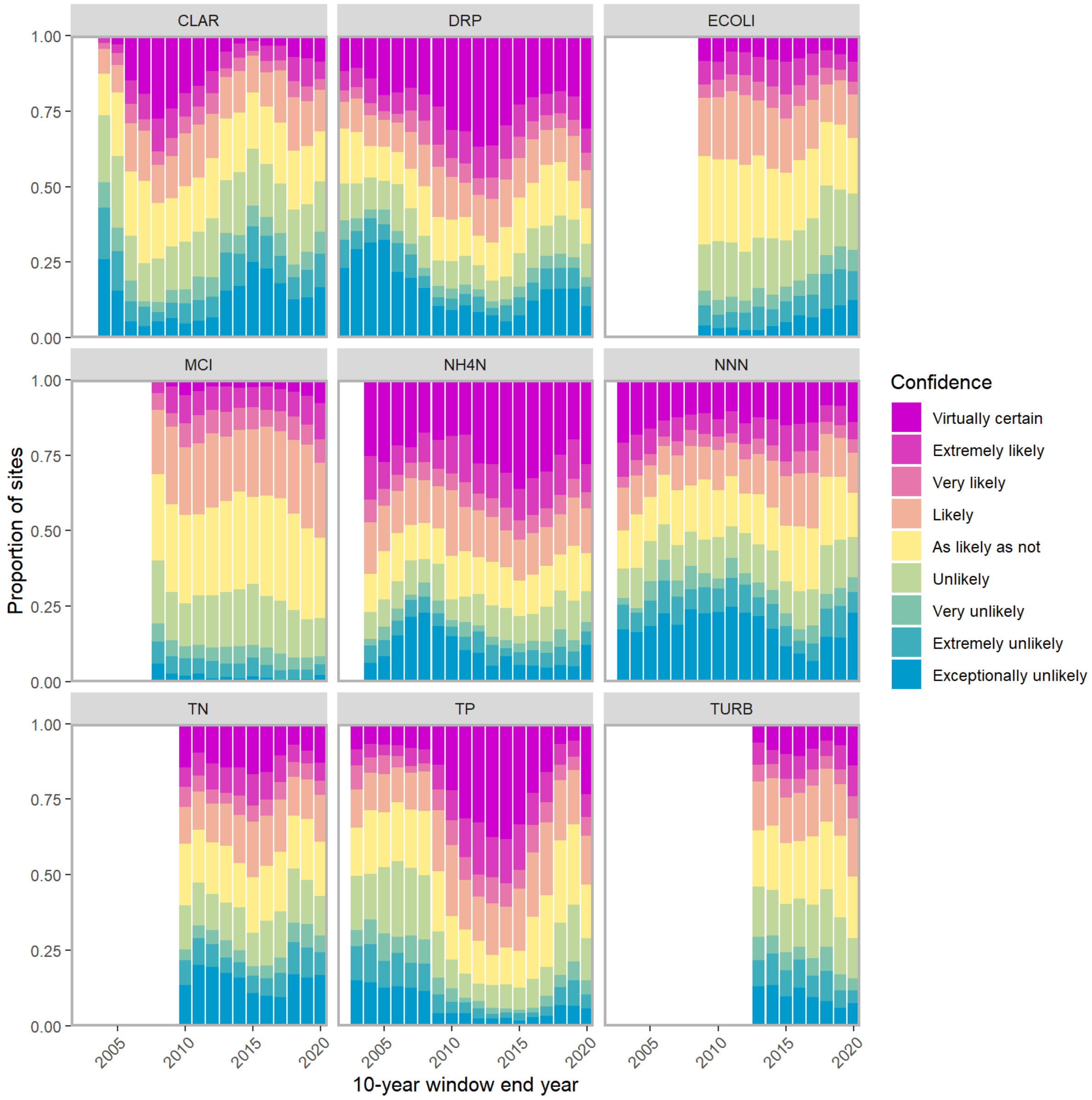

Figure 5-19: Summary plot representing the proportion of decreasing raw site trends at each categorical level of confidence for each of the rolling 10-year windows. The plot shows the proportion of sites with decreasing trends at levels of confidence defined in Table 3-2.

The national-scale proportions of decreasing trends (P_d) and their confidence intervals for each variable and time window with more than 200 sites are summarised in Figure 5-20. No variables exhibited monotonic changes in the P_d score. There were quasi-periodic fluctuations in P_d that varied between variables. The magnitude of the fluctuations was greatest for CLAR (ranging from 19% to 68%) and smallest for MCI (ranging from 54% to 69%). MCI and NH4N consistently had a majority of sites (i.e., >50%) that had decreasing trends. However, that the majority of sites were decreasing was not established at the 95% confidence level for two and three time windows for MCI and NH4N, respectively.

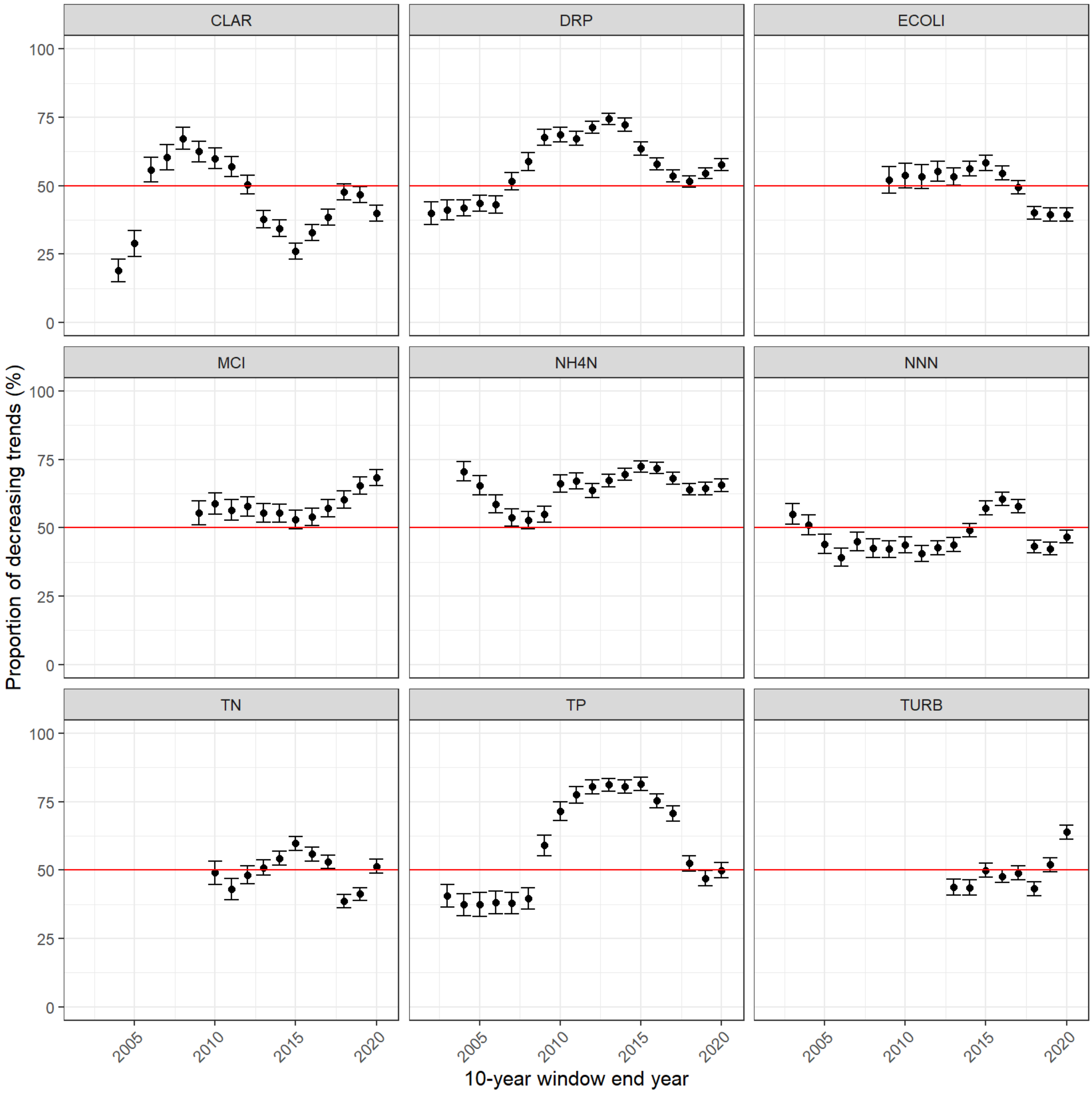

Figure 5-20: Summary plot representing the proportion of decreasing raw site trends (P_d) for each of the rolling 10-year windows. The error bars indicate the 95% confidence intervals for P_d.

6 Discussion

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

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

The detailed information for each 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.

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

Changes in method:

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

Changes in terminology and reporting:

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

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

7 Acknowledgements

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

8 Glossary of abbreviations and terms

CLAR	Visual Clarity
DRP	Dissolved reactive phosphorus
ECOLI	*Escherichia coli*
LAWA	Land Air Water Aotearoa
MCI	Macroinvertebrate community index
MfE	Ministry for the Environment
NH4N	Ammoniacal nitrogen
NNN	Nitrate + nitrite nitrogen
NOF	National Objectives Framework
NPS-FM	National Policy Statement for Freshwater Management
NRWQN	National River Water Quality Network
REC	River Environment Classification
SoE	State of Environment
TN	Total nitrogen
TP	Total phosphorus
TURB	Turbidity

9 References

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

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

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: 227–239.

Camargo, J.A., Alonso, 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.

Gehl, K. (2009) *Nitrate/nitrite toxicity*. Case studies in Environmental Medicine. Agency for Toxic Substances and Disease Registry. http://www.atsdr.cdc.gov/csem

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

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J. (2020) *Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, Book 4, Chapter A3*. 458 pp. https://doi.org/10.3133/tm4a3.

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

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

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

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

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

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

Larned, S.T., Whitehead, A.L., Snelder, T.H., Fraser, C., Yang, J. (2018) *Water quality state and trends in New Zealand rivers: Analyses of national data ending in 2017*. NIWA Client Report 2018347CH prepared for the Ministry for the Environment. NIWA, Christchurch.

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

McDowell, R.W., 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.

McMillan, H.K., Hreinsson, E.Ö., Clark, M.P., Singh, S.K., Zammit, C., Uddstrom, M.J. (2013) Operational hydrological 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, Wellington.

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

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

Smith, D.G., McBride, G.B., Bryers, G.G., Wisse, J., Mink, D.F.J. (1996) Trends in New Zealand's National River Water Quality Network. *New Zealand Journal of Marine and Freshwater Research*, 30: 485–500.

Snelder, T.H., 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.H., Fraser, C., Larned, S.T., Whitehead, A.L. (2021) *Guidance for the analysis of temporal trends in environmental data*. NIWA Client Report 2021017WN prepared for Envirolink (MBIE). NIWA, Christchurch.

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

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

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

Unwin, M.J., Larned, S.T. (2013) *Statistical models, indicators and trend analyses for reporting national-scale river water quality*. NIWA client CHC2013-033 report prepared for the Ministry for the Environment. NIWA, Christchurch.

Unwin, M.J., Snelder, T.H., Booker, D.J., Ballantine, D., Lessard, J. (2010) *Predicting water quality in New Zealand rivers from catchment-scale physical, hydrological and land cover descriptors using Random Forest models*. NIWA Client Report CHC2010-037. NIWA, Christchurch.