Earth Sciences New Zealand

Water quality state and trends in New Zealand rivers

Analysis of national data ending in 2024

Prepared for Ministry of the Environment

February 2026

Prepared by:
Doug Booker, Rachel Smith, David Wood, Caroline Fraser (LWP), Ton Snelder (LWP)

For any information regarding this report please contact:

Doug Booker
Water Quality Scientist
Freshwater Modelling Group
+64 3 343 8050
doug.booker@niwa.co.nz

New Zealand Institute for Earth Science Limited
PO Box 8602
Riccarton
Christchurch 8440

Phone +64 3 348 8987

Client Report No: 2025331CH
Report date: February 2026
Project No: MFE26501

Revision	Description	Date
Version 1.0	Final version sent to client	24 October 2025
Version 1.1	Updated MCI trends	26 February 2026

Quality Assurance Statement		
	Reviewed by:	Bruce Dudley
	Formatting checked by:	Nic McNeil
	Approved for release by:	Phil Jellyman

Contents

Tables

Figures

Executive summary

Purpose

The New Zealand Ministry for the Environment (MfE) and Stats NZ Tatauranga Aotearoa (Stats NZ) use results from national-scale analyses of river water quality originating from state-of-the-environment (SoE) monitoring to inform policy development and meet their requirements for environmental reporting on the freshwater domain under the Environmental Reporting Act 2015. The current study analysed river water quality variables representing physical, chemical, microbial, and ecological conditions observed across multiple SoE monitoring sites for the period ending in December 2024. Following conventions used in previous national environmental reporting and in consultation with MfE, nine monitoring variables were analysed for current state and temporal trend: visual clarity (clarity), turbidity (turbidity), concentrations of nitrate-nitrite-nitrogen (NNN), ammoniacal nitrogen (NH_4-N), total nitrogen (TN), dissolved reactive phosphorus (DRP), total phosphorus (TP), the faecal bacterium *Escherichia coli* (*E. coli*), and macroinvertebrate community index scores (MCI). Various organisations monitor these nine water quality variables across many sites for various purposes. A large volume of observation and associated metadata therefore exist, but are sometimes held in different formats, different units, and using different data conventions. The purpose of this report was collation, standardisation, and preliminary analysis of available water quality data for subsequent use in environmental reporting.

Methods

A data request was made directly to all regional councils and unitary authorities to obtain all water quality data collected routinely for SoE purposes from commencement of monitoring to 31 December 2024. We also obtained relevant water quality data from the National River Water Quality Network (NRWQN) collected by Earth Sciences New Zealand (ESNZ). Relevant censoring information, site metadata, and descriptions of measurement methods were also requested. Questions that arose after initial data collation about data quality, completeness, or purpose were resolved using subsequent queries to data providers.

Data were collated into a common format to ensure consistent conventions for variable names, site identifiers, date and time formats, units of measurement, and other data structure elements. Transcription errors, incomplete censoring information, or potentially erroneous values were identified and resolved following data inspection and additional queries with data providers. Predefined rules for data exclusion were applied based on information supplied by data providers such as data quality, non-SoE monitoring purposes, or non-comparable data collection methods where they were known. Site metadata were used to collate information describing the upstream catchment characteristics for each monitoring site. The upstream proportion of aggregated agricultural, urban, and exotic forest land cover was used to represent a gradient of human modified land cover.

Duration and frequency of observations for each site × variable combination were assessed against predetermined criteria for inclusion in statistical analysis. For several variables, some observations were provided as censored values because they were reported as a value less than an analytical detection limit or greater than a reporting limit. Predefined methods were used to consistently replace censored values by imputation prior to analysis of state and robust methods were used to handle censored values during analysis of trends.

We assessed state by calculating summary statistics for each site × variable combination within a 5-year period ending December 2024. Site medians were grouped by four evenly split classes of human modified land cover. Linear regressions were used to inspect the relationship between median water quality state and human modified land cover. Site × variable combinations were assigned to a predefined attribute state defined by bands labelled A to E.

The trend rate for 10-, 20- and 30-year periods ending in December 2024 was assessed using the Sen slope estimator. The seasonal version of the Sen slope estimator was applied for variables measured seasonally (i.e., monthly, bi-monthly or quarterly), and where within-site seasonal variability was detected. Kendall tests provided estimates of the confidence in the trend direction. Trends for all site × variable combinations were classified using predefined categories describing the confidence that a given trend was improving. Temporal transience of trends was evaluated for each variable at each site using rolling windows of 10-years duration starting in 1990 and incrementing by one year to a final window ending December 2024.

Two approaches were used to evaluate aggregate trends in each water quality variable at the national scale and within human modified land cover classes. The first approach tallied the proportion of sites in each category of trend direction and confidence. The second approach used the confidence in trend direction and confidence for individual sites to estimate the proportion of all sites with improving trends and its 95% confidence interval.

Results

Sufficient data for calculation of state were available for between 695 and 992 sites, depending on water quality variable. Sites were well distributed across the country and across a gradient of human modified land cover.

Water quality cannot be measured precisely when observed data are below an analytical detection limit (left-censored) or above a reporting limit (right-censored). NH_4-N observations were often censored at many sites compared to other variables. For all variables other than MCI and NH_4-N, many sites had a small proportion of censored observations during the state assessment period, and sites that had a large proportion of censored observations were rare. MCI data are not censored.

Many sites (64%) were graded D or E for *E. coli* combined numeric attribute state. Very few sites were graded D for NH_4-N (toxicity) or NNN (toxicity). For suspended fine sediment, 39% of sites were graded D for clarity, including 30% of sites classified as having low levels (i.e., 0–25%) of human modified land cover. For MCI, 23% of sites were graded D, including over 47% of sites with high levels (i.e., 75–100%) of human modified land cover.

Variation in state for nutrient and *E. coli* concentrations and clarity was partly distinguished by human modified land cover classes. Median concentrations of each nutrient and *E. coli* increased, and MCI scores and visual clarity decreased, with increasing proportions of human modified land cover in the upstream catchment. However, water quality state varied considerably within sites at either end of the human modified land cover gradient.

For 10-year trends, all variables were improving at some sites and worsening at other sites. Clarity, MCI, *E. coli*, and NH_4-N had more sites that were worsening than improving. Turbidity, DRP, and TP had more sites that were improving than worsening. Improving and worsening of water quality were not systematically associated with the proportion of human modified land cover in the upstream catchment. Results for 20- and 30-year trends were broadly comparable to those for 10-year trends despite being applied to different sets of sites due to more sites being available in recent years. The aggregate trends in DRP and TP were very likely improving for all trend periods. The aggregate trend in *E. coli* was very likely worsening for the 10- and 20-year trend periods and likely worsening for the 30-year period. The aggregate trend in MCI was very likely worsening for the 20- and 30-year trend periods and likely worsening for the 10-year period. More complex results were produced for the aggregate trend in TN which was likely worsening for the 30-year trend period and likely improving for the 10-year trend period.

Analysis of rolling 10-year trend windows provided insights into changes in aggregate trends over time. The national-scale proportions of improving trends was not constant or monotonic for any variable. Quasi-periodic fluctuations in proportions of improving trends were particularly strong for clarity, DRP, and TP, and less strong for MCI and NNN.

Discussion

We produced state and trend results by applying a prescribed set of methodological procedures to available water quality SoE data. We then discuss the possibility that results are sensitive to methodological choices (e.g., analysis period), monitoring network design (e.g., site positions and longevity), and observation data features (e.g., detection limits, data precision, observation frequency). One example is the confidence in trend direction at a single site being sensitive to the proportion of censored data and changes in data precision. A second example is the proportion of improving trends calculated for different time periods being sensitive to duration of analysis period and closing or opening of sites between periods. This report explains how our analysis applied robust procedures to mitigate, or explore, some undesirable but unavoidable characteristics of SoE data. For example, transient decreases and increases within rolling trends of water quality data indicated that caution must be applied when interpreting causes and permanence of calculated trends.

We explain that the results presented in this study differ from those produced for previous studies of water quality state and trend because the present study used up-to-date data, improved methods of data preparation (e.g., filtering by quality control codes), and improved statistical methods.

1 Introduction

1.1 Background

State-of-environment (SoE) monitoring for river water quality involves collecting, analysing, and reporting on physical, chemical, and ecological data to assess health and trends over time. SoE monitoring involves measuring a suite of standard variables at representative sites to provide a comprehensive and unbiased understanding of water quality condition (state) and its changes (trends).

River water quality data measured for SoE purposes is typically collected from strategically selected sites at regular, but relatively infrequent, intervals for long durations (e.g., monthly for a duration of years). River water quality data can also be measured at SoE sites or elsewhere at frequent intervals for short durations (e.g., daily or hourly for a single month) for non-SoE purposes such as bespoke scientific studies, validation or calibration of instrumentation, monitoring for consent compliance, or to issue public health warnings.

Monthly or quarterly monitoring has been carried out at approximately 1,000 SoE river sites across New Zealand by New Zealand local government authorities including regional councils and unitary councils, hereafter referred to as “councils”. This monitoring has been carried out for at least five years and continues to the present. A variety of physical, chemical and biological indicators of water quality (referred to in this report as “variables”) are measured at these sites. In addition, water quality and biological monitoring had been carried out by ESNZ (and one of its former institutions; the National Institute of Water and Atmospheric Research, NIWA) starting in 1989 at the sites that make up the National River Water Quality Network (NRWQN).

The New Zealand Ministry for the Environment (MfE) and Stats NZ Tatauranga Aotearoa (Stats NZ) use analyses of state and trends derived from SoE river water quality data 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 physical, chemical, and biological variables that are often measured as part of river SoE monitoring programmes.

1.2 Study purpose

Stats NZ are required to produce up-to-date environmental statistics in the form of environmental reporting indicators under the Environmental Reporting Act 2015 (ERA). Stats NZ relies on access to data collected by various organisations to subsequently produce surface water quality indicators required by the ERA. Under the ERA, Stats NZ decides what statistics will be used to measure environmental reporting topics. Under the Data and Statistics Act 2022, Stats NZ also has the sole responsibility for deciding on the production of statistics by the Statistician, including the selection of methods. Under both Acts, Stats NZ also has a duty to act independently. In addition to the requirements of Stats NZ, MfE use results from national-scale analyses of river water quality originating from SoE monitoring to inform freshwater policy development. The purpose of this report was collation, standardisation, and preliminary analysis of available river water quality data for subsequent use in environmental reporting and to inform freshwater policy.

1.3 General approach

Council and NRWQN river monitoring data have been periodically acquired and compiled into databases for preparation of national-scale SoE reports and to investigate monitoring performance. Earlier reports for MfE provided river water quality state and trends based on monitoring data with ending dates of December 2017 (Larned et al. 2018) and December 2020 (Whitehead et al. 2022). For the current report, we newly acquired all river water quality data. To assess data completeness, we compared the coverage of the newly acquired data to that used for the preceding national-scale report (Whitehead et al. 2022). We then undertook systematic analyses to report updated states and trends to cover the period up to December 2024. We communicated results using graphs, tables, and statistical summaries. We provided all raw results as appendices to this report. The detailed methods and tabulated, site-specific results will enable MfE to use the results for a wide range of purposes (e.g., mapping, inter-comparisons between environmental classes or geographic domains, estimation of reference conditions) that are all based on a single comprehensive methodology.

The 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 classes representing a gradient in human modified land cover, and supplementary files with site-specific results and spatial data for each site.

This work consisted of several major steps:

1. Compile SoE river water quality monitoring data along with associated quality QC information, site metadata, censoring information, and measurement methods from regional councils and ESNZ.
2. Post-process and collate the data into a national database that uses a common format, including error correction, application of reporting conventions, and links to site spatial data.
3. Assess whether there is sufficient data for statistical analyses for each of nine physical, chemical, microbial, and ecological variables by applying site inclusion rules.
4. Carry out analyses of water quality state at a site level, including comparison of site water quality state with numeric attribute states defined in the National Policy Statement for Freshwater Management 2020.
5. Investigate patterns in the water quality state with regard to catchment characteristics by aggregation of site water quality state by human modified land cover classes, and graphing relationships with a gradient of human modified land cover.
6. Carry out trend analyses for 10-, 20- and 30-year periods ending in December 2024 at a site level.
7. Investigate patterns in the water quality trends with regard to catchment characteristics by aggregation of site trends by human modified land cover classes.

8. Evaluate trends for each variable at each 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 2024 (i.e., a total time period of 35 years).
9. Assess water quality trends at the national scale using categorical levels of confidence and a statistical analysis of the proportions of improving trends.

2 Data

In this section we describe the water quality variables, data sources, and organisation of the river water quality 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). Physical, chemical, and microbiological variables are typically measured monthly or quarterly, whereas MCI is typically measured annually in the summer.

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

Variable type	Variable	Abbreviation	Units
Physical	Visual clarity	clarity	m
	Turbidity	turbidity	NTU or FNU
Macroinvertebrate	Macroinvertebrate Community Index	MCI	unitless
Microbiological	*Escherichia coli*	*E. coli*	cfu or MPN 100 mL^{-1}
Chemical	Dissolved reactive phosphorus	DRP	mg L^{-1}
	Total phosphorus (unfiltered)	TP	mg L^{-1}
	Ammoniacal nitrogen	NH_4-N	mg L^{-1}
	Nitrate + nitrite nitrogen	NNN	mg L^{-1}
	Total nitrogen (unfiltered)	TN	mg L^{-1}

Visual water 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 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). The sensor reading is not absolute light scattering, but light-scattering relative to the standard solution, in “nephelometric turbidity units” (NTU) or “formazin nephelometric units” (FNU) depending upon the light source wavelength and calibration standard being used. However, different instruments using the same methods can result in numerically different results with the same water quality (Davies-

Colley et al. 2021). 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, NH_4-N, DRP, TN and TP) were included in our analysis 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 animals, including 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). Water quality in rivers also of interest because it is associated with delivery of nutrients to receiving environment such as lakes and estuaries (Dudley et al. 2020; Plew et al. 2020).

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 following the methods of Whitehead et al. (2022) which did not apply temperature correction due to insufficient temperature data.

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

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, using presence/absence data which are widely available. 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). Most MCI data were supplied by the collecting agency as calculated scores rather than raw invertebrate data.

2.2 Censoring information

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” (DL) for that variable; for very high values of a variable, the minimum acceptable precision corresponds to the “reporting limit” (RL) 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" (i.e., below detection limit) and ">RL" (i.e., above reporting limit) respectively, where DL and RL are the laboratory detection limit and reporting limit, respectively. Water-quality datasets from New Zealand rivers often include DRP, TP, and NH_4-N measurements that are below detection limits, and *E. coli* and clarity measurements that are above reporting limits. 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). Changes in laboratory procedures or equipment can impact observed water quality time-series through changes in DL, RL, and general data precision. Changes in DL and RL therefore need to be acknowledged within trend assessment methods. We quantified the proportion of censored observations for each site × variable combination for each of our analyses. In this study, repeatable procedures were used to handle censored data in the state (see Section 3.1.2) and trend (see Section 3.2.4) analyses.

2.3 Water quality data

River water quality monitoring data have been acquired periodically from regional councils and ESNZ/NIWA for previous national scale analyses for MfE (Ballantine et al. 2010; Unwin et al. 2010; Unwin and Larned 2013; Larned et al. 2015, 2018; Whitehead et al. 2022). For each successive analysis, data were used to update or regenerate a database comprising site information, sampling dates, and measurements for several monitoring variables. Previously derived databases also contained metadata (e.g., methods, alternative variable labels, analytical detection limits).

We used traceable procedures to acquire river water quality observation data. MfE first contacted councils to notify them of the data request and to request a suitable contact for each data provider. We then sent the same data request to all data providers. Data and information describing observed water quality, censoring information, site metadata, and descriptions of measurement methods were requested. We received data directly from data providers via email. We used the data acquired through these procedures to generate a dataset similar in format to that used for the previous national-scale analysis.

After initial data collation, we determined that some data providers had pre-filtered their data to supply data solely collected at regular intervals at SoE monitoring sites for SoE monitoring purposes, whereas other providers had supplied all available water quality data regardless of purpose or frequency of data collection. We therefore requested that data providers supply information that identified the purpose for which each observation was collected or confirm that data supplied to us were solely associated with routine SoE monitoring rather than for other purposes. We used the information supplied to us by data providers regarding the purpose of data collection to maintain spatial and temporal consistency within our analysis by limiting the analysis to SoE data. For example, we removed frequent observations made for non-SoE purposes to maintain consistency between sites experiencing standard SoE sampling regimes and sites experiencing irregular sampling. We also removed observations made at non-SoE sites to maintain consistency between regions who supplied data only at SoE sites and regions who supplied all data including from site strategically placed to detect the effects of specific point sources for consent monitoring purposes.

2.4 Quality control information

We aimed to acquire a dataset covering the period up to 2024 that was broadly comparable to the dataset covering the period up to 2020 that was collated by Whitehead et al. (2022). Quality Control (QC) standards and codes have been applied by data providers since implementation of National Environmental Monitoring Standards (NEMS) QC standards in 2019 (NEMS 2019). We therefore attempted to acquire QC information to allow data labelled as being of poor quality to be identified and removed from our analysis. We requested that data providers supplied NEMS QC codes (Table 2-2) alongside each observed water quality value.

Table 2-2: NEMS QC codes and their named quality zones (source: NEMS 2019).

QC code	Named quality zone
QC 600	Highest quality code
QC 500	Fair
QC 400	Compromised
QC 300	Synthetic
QC 200	No quality
QC 100	Missing record
QC 0	Non verified

2.5 Measurement method information

River water quality measurements can be made with a variety of laboratory and field methods. These methods have evolved over time, driven by changes in technology, instrumentation, test availability, economic considerations, and evolving monitoring requirements, such as the need to detect attributes at low levels.

The NEMS aim to ensure that water sample collection, measurement, and data processing are consistent and reliable across New Zealand. This includes standardised monitoring, including the choice of measurement methods, to ensure robust, high-quality and comparable water quality monitoring and recording results between regional councils and ESNZ/NIWA.

Though different measurement methods are intended to give comparable results, systematic differences in observed values can arise from changes in measurement methods (Davies-Colley and McBride 2016). Different methods being applied to monitor the same variable at different sites could result in systematic between-site or between-region differences in the estimated state. “Step changes” in values associated with changes in method at a site may confound assessment of trend (Wood 2024).

We requested information about laboratory methods as part of our data requests. Our intention was to investigate possible influences of measurement method within our analysis and only retain observations measured using a standard set of methods where necessary. We received some useful information about measurement methods from data providers. One example of an inconsistency in measurement method was that some councils estimated TN using the indirect method[1], which involves summing up the measurements of TKN and NNN, rather than the direct method[2], alkaline persulfate digestion. However, descriptions of measurement methods supplied to us by data providers used a wide range of naming conventions and most method data did not give precise descriptions of which methods had been applied at which sites through time. The information we received was therefore not sufficiently comprehensive to determine a measurement method for each observation contained within our database. Given the available information about measurement methods, we were unable to explicitly account for unknown differences between methods within our analysis.

The use of filtered samples for water quality variables labelled TN and TP is a potential issue raised in previous reports (e.g., Whitehead et al. 2022). If samples are filtered, we would expect these variables to be labelled as total dissolved nitrogen (TDN) or total dissolved phosphorus (TDP) or similar instead of TN and TP, and these samples would provide non-comparable results (Horowitz 2013). The method information we received was therefore not sufficiently comprehensive to consistently identify the presence of mis-labelled filtered “TN” and “TP” samples across water quality observations and data providers, and we assume variables are correctly labelled and thus that TN and TP are unfiltered. We note that unfiltered samples are used for TN and TP methods described in NEMS.

2.6 Landscape-scale data

Digital Networks (DN) comprise representations of surface flow pathways (segments), areas contributing to each surface flow pathway (watersheds), and connections between surface flow pathways (routing). Unique “nzsegment” identifiers are used to identify each segment stored in a DN. The catchment upstream of each segment can be linked to a wide range of spatial data associated with the DN. For example, the River Environment Classification (REC) is a deductively defined hierarchical classification system that combines a DN and various spatial environmental data to classify river segments based on dominant characteristics that control spatial patterns in river hydrology, geomorphology, and ecosystems (Snelder and Biggs 2002).

To be consistent with Whitehead et al. (2022), we accessed national DN version 2.4 and its associated geodatabases (Whitehead and Booker 2019). Upstream catchment area, and categorical REC classes (climate, topography, geology which are used to define NOF suspended sediment classes) projected onto the DN were obtained from existing sources. Land Cover Database Version 6.0 (LCDB6 as supplied to us by MfE) comprises proportional cover of 33 land-cover classes, generated from satellite imagery collected in 2023. We projected LCDB6 data onto the DN to calculate the proportion of each land cover class occupying the catchment upstream of each nzsegment.

Whitehead et al. (2022) aggregated sites by dominant land cover categories, as defined by REC land cover categories. Dominant land cover categories are initially assigned based on the land cover class (simplified from LCDB) that occupies the greatest proportion of the upstream catchment. An exception is made for this rule if *Pastoral* and *Urban* land cover account for greater than 25% and 15% of the upstream catchment, in which case the REC Land Cover class

[1] Total Nitrogen – Indirect (TN-K) - Calculated from the measurement of Total Kjeldahl Nitrogen (TKN, measured via APHA 4500-N_{org} D) + NNN. This method may be preferable when TSS is high.

[2] Total Nitrogen – Direct (TN-A) - APHA 4500-N C OR APHA 4500-P J potassium persulfate digestion then analysis by APHA 4500-NO3 I. This method is preferable for low-nutrient river waters, having a lower detection limit than the TN-K.

is *Pastoral* or *Urban*, respectively (Snelder and Biggs, 2002). A recent review of the use of the REC land cover categories in environmental reporting identified two important issues with application of these categories (Fraser and Snelder, 2021). First, land cover is only one of many factors that affect water quality. REC Land Cover categories leave considerable between-site variation in water quality unexplained. Second, REC Land Cover categories are based on assigning a single dominant land cover to a catchment, which can obscure the contribution of other land cover types to water quality and may give the impression that observed water quality is attributable to a single land cover, such as pastoral, when there are other types of land cover that are contributing to the observed conditions.

We applied an alternative method recommended by Fraser and Snelder (2021) that retains the simplicity of land cover categories but avoids the need to assign a dominant land cover, which is to use a categorical subdivision of a gradient in land cover that represents human modified land cover. A gradient in human modified land cover was defined by adding the proportion of catchment area occupied by Pastoral (LCDB6 classes 40, 30, 33), Urban (LCDB6 classes 1, 2, 5, 6), and Exotic Forest (LCDB6 classes 71, 64) land cover types and subdividing this gradient into four evenly spaced categories. Metadata provided with the state and trend outputs include both the proportion of upstream human modified land cover, the human modified land cover category and the REC dominant land cover category.

2.7 Water quality data processing

River water quality data were processed to ensure that data collated from different sources were represented in a consistent format, that the data were as error-free as possible given available information, that consistent measurement procedures were used where these were known, and that site information was complete and accurate.

Unless otherwise stated, we made no distinction between data collected at regional council sites and NRWQN sites. Where regional councils and NIWA/ESNZ both monitor at the same location for an overlapping period, we treated the data as being associated with separate sites because two legitimate time-series were available.

Development of the data processing steps required additional queries to data providers to ensure issues were resolved appropriately. Common issues included incomplete censoring information, missing data that were present in previous rounds of reporting, missing site metadata, clarification regarding monitoring purpose, and clarification regarding potentially erroneous values such as zero or negative values.

Step 1. Reporting conventions.

The water quality data received from data providers 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 for water quality variable names, sample date, observed values, units, censor information, detection limits (where present), quality control code, and site identifier. Where water quality variable names were ambiguous or did not match the variables we requested, we sought clarification with data providers. Observed values were converted to equivalent units where inconsistent measurements units were indicated (e.g., from µg L^{-1} to mg L^{-1}). In cases where data providers indicated that observed values were raw with no censoring applied, we determined the detection limit associated with each raw value and applied censoring to raw values lower than their associated detection limit. When establishing equivalent units for Turbidity, both FNU and NTU were included. If both FNU and NTU were present, NTU was used preferentially. Quality control codes were represented as NEMS codes where possible. Where no quality control code was present, or a code system could not be equated to a NEMS code, then QC 200 was used. We resolved any inconsistencies in spelling or capitalisation (e.g., transcription errors) of site identifiers to ensure internal consistency. We used time-series plots and quantile plots to identify potential data transcription errors and missing censoring information.

Step 2. Calculation of composite and adjusted observations.

Where appropriate, composite water quality variables were calculated by summing separate constituent components from the raw observed data. NNN was calculated as the sum of NO_2-N and NO_3-N where they were available and where a NNN observation was not already present. TN was calculated as the sum of TKN and NNN for Waikato Regional Council data, for which TN data was not provided. The effect of censoring of constituent components was considered when they were summed to calculate a composite observation. If only one constituent component was censored because it was at or below the detection limit, then the value of the censored constituent was halved before it was added to the other constituent. If both constituent components were censored, then the raw constituent components were summed and the calculated composite observation was flagged as being at or below its detection limit. The lower quality code of the constituents was associated with the newly calculated composite observation.

We applied a pH correction to NH_4-N 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 “NH_4-N (Adj.)”.

Step 3. Data exclusion.

We excluded data which was not relevant or it was not possible to use in our analyses. Data not relevant to our analysis included water quality variables not included in those we requested, data for areas not included in our request (i.e., Chatham Islands), and data collected for non-SoE purposes. We received data for sites on the Chatham Islands, but these data were excluded from our analysis because they have not been included in previous environmental reporting (Whitehead et al. 2022) and landscape-scale data (i.e., river network and land cover) is not available for them. Data that could not be used in our analysis included observations with missing observed values, dates, or associated variable name. We excluded overlapping data where we could identify that it was provided by more than one data provider or had been supplied as overlapping datasets from a single data provider.

We further excluded data identified as potentially erroneous. Zero and negative values were excluded, following censoring of raw values where appropriate. We included zero values in cases where data providers indicated that these were legitimate observations. We further identified a single pH value of 1024 which was excluded as it was deemed to be infeasible.

Our default method was to include observations coded as QC 600 (highest quality), 500 (fair), 200 (no quality), or 0 (non-verified). After consultation with MfE, we decided to include observations labelled as QC 200 because this label was associated with most observations collected before the introduction of NEMS QC standards in 2019. We further included QC 0 (virtually no observations) in recognition that the unknown quality of these data is comparable to QC 200. Our default method equated to omitting observations labelled as QC 400 (compromised). The collated dataset contained virtually no observations labelled QC 300 (synthetic) or QC 100 (missing), and these were excluded.

Approximately 6000 observations from rivers in the years 2021 to 2024 (25% of all observations in these years) supplied by Bay of Plenty Regional Council (BOPRC) were labelled as QC 400. Further discussion with BOPRC staff revealed that approximately 80% of observations (from rivers and lakes) were labelled as QC 400 because BOPRC's interpretation of the NEMS 2019 procedures has resulted in an approach that automatically drops the QC code to 400 if samples are not verified as arriving to the lab below 10°C (or less than the sample collection temperature where samples are delivered within two hours of collection), rather than assuming this temperature condition was met. We noted that none of the observations labelled as QC 200 could be verified as arriving to the lab below 10°C from the information available to us. We also noted that our default approach to exclude observations labelled as QC 400 would have greatly reduced the number of sites for which state and trends could be applied within the Bay of Plenty region. We therefore included observations labelled as QC 400 by BOPRC. Observations labelled as QC 400 by all other data providers were excluded.

Step 4. Monitoring site spatial information.

The following spatial data were associated with each river monitoring site: LAWA (Land Air Water Aotearoa) identifier, site location, regional council identifier, coordinates, and nzsegment identifier. Missing monitoring site metadata such as LAWA identifiers or coordinates were identified from previous LAWA water quality datasets and previous environmental reporting datasets where possible. Coordinates were converted to a consistent format and mapped to identify potential errors. Nzsegment identifiers were used to link each site to metadata including land cover, and REC classes. Nzsegment identifiers supplied from data providers were used if the segment was located within 500m of the site location, otherwise an automatically generated segment was used. Automatically assigned nzsegment identifiers were identified from site coordinates as the highest order river segment within 200m, or within 500m if there are no river segments within 200m. Nzsegments and site locations were mapped and site nzsegment identifiers were manually adjusted where required.

Step 5. Replacing repeat observations.

Repeat observations of the same variable on the same day may have been present in our data despite our intention not to collate frequent observations made for non-SoE purposes. Where there was more than one observation on a day for a given variable at a site, the multiple observations were replaced by the median of the observations.

Step 6. Assigning observations to periods.

We assigned each water quality observation metadata describing the date, year, month, bi-month, quarter, and bi-annual periods the observation belonged to. These periods represent the range of monitoring frequencies that exist across all variables and sites in the water quality dataset. A common issue in monthly SoE water quality monitoring arises when sampling is not conducted on the same calendar day each month. For example, if sampling is avoided on weekends or holidays, the timing can shift so that one month ends up without a sample and the next has two. If this happens frequently, the site may not comply with filtering rules (i.e., minimum data requirements for a state or trend assessment to be judged robust – see Section 3.1.3). We identified cases where months with two or more observations were preceded or followed by months missing an observation, and where the time between these two observations is at least two weeks (to ensure independence). Where the previous month has a missing observation, the first observation is assigned metadata associated with the previous month, and where the following month has a missing observation, the last observation is assigned metadata associated with the following month. The observations retain the original date for use in evaluating the trend rate. This data filtering step is a new addition since the last SoE reporting.

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 their associated 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).

For attributes with more than one numeric attribute state, an overall grade was also evaluated. This is recommended in the NPS-FM for the *E. coli* attribute, where an overall grade is taken as the lowest grade across all four of the numeric attribute states. The same approach was applied for the other attributes with more than one numeric attribute state.

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 (NH_4-N), and Turbidity. For these variables, we report the median of the observations.

Following the methods applied by Whitehead et al. (2022), we used median site NNN concentrations rather than NO_3-N 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. 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 NO_3-N. This approach makes best use of the available nitrate toxicity data and aligns with the approach taken by LAWA and Whitehead et al. (2022).

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	(Annual) Median concentration of Ammoniacal-N	mg l^{-1}	NH_4-N Median
		(Annual) 95th percentile concentration of Ammoniacal-N	mg l^{-1}	NH_4-N 95th percentile
A2A; Table 6 – Nitrate (toxicity)		(Annual) Median concentration of NNN	mg l^{-1}	NNN Median
		(Annual) 95th percentile concentration of NNN	mg l^{-1}	NNN 95th percentile
A2A.; Table 8 - Suspended fine sediment	Median of 5 years of at least monthly observations (at least 60 samples); grades dependent on REC class	Median visual clarity	m	Clarity
A2A; Table 9 - *Escherichia coli*	minimum of 60 observations over a maximum of 5 years	% exceedances over 260 100 mL^{-1}	%	*E. coli* % >260
		% exceedances over 540 100 mL^{-1}	%	*E. coli* % >540
		Median concentration of *E. coli*	number of colony forming units (cfu) per 100 ml	*E.coli* Median
		95th percentile concentration of *E. coli*	cfu per 100 ml	*E.coli* 95th percentile
A2B; Table 14 - Macroinvertebrates	State calculated as 5-year median based on observations between Dec-Mar	Median MCI score	-	MCI
A2B; Table 20 - DRP		Median concentration of DRP	mg l^{-1}	DRP Median
		95th percentile concentration of DRP	mg l^{-1}	DRP 95th percentile

3.1.2 Handling censored values

Following the methods applied by Whitehead et al. (2022), 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 reporting 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 the detection limit multiplied by 1.1. 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 standard 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. 2021a).

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 2024. For MCI, this time period was shifted by 6 months (the 5-year period to end of June 2024), to align with water years (time period 1 July 2020 – 30 June 2025), in order to prevent splitting summer observations into two calendar years. The start and end dates for this period were determined by the availability of quality assured data (see Section 2), reporting time periods defined by MfE, 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 periods used in previous national water quality state analyses (Larned et al. 2018; Whitehead et al. 2022). 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, 27 of 30 bi-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 four (with samples in at least four 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 2021 to 2025 (inclusive), with the exception of MCI, where we used water years defined by the time period 1 July 2020 – 30 June 2025 (to prevent splitting summer samples into two calendar years). All percentiles were calculated using the Hazen method.[3]

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

Our analysis procedure was designed to account for monitoring records that exhibited irregular sampling frequencies, e.g., multiple observations within a sampling interval (e.g., a month), or changes in sampling frequency over the state period (e.g., from quarterly to monthly monitoring). To avoid a bias in the state assessments towards the period with high sampling frequency, the data was downsampled so that there was only one observation per sampling interval. The sampling interval was selected as the highest frequency sampling interval that would comply with the minimum data requirements. The downsampling procedure involved filtering for the observations closest to the middle of each of the sampling intervals in the state time 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

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

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 in 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 Relationships between water quality state and catchment land cover

We tallied sites by each of the evenly spaced categories of human modified land cover. For each water quality variable, we produced box and whisker plots of the median of each site in each evenly spaced category of human modified land cover. For *E. coli*, we produced box and whisker plots of the 5^{th}, 20^{th}, 50^{th}, 80^{th}, and 95^{th} percentiles of each site in each evenly spaced category of human modified land cover. For *E. coli*, we also calculated exceedance measures as the percentage of observations that exceeded 260 and 540 100 mL^{-1}, respectively. For each variable, we plotted the median of each site against human modified land cover. We coloured the points of the scatterplot by dominant land cover (as used by Whitehead et al. 2022) to allow interpretation of which land cover types are contributing to the gradient of human modified land cover. We used linear regressions to relate water quality state to a gradient in human modified land cover. All water quality variable values were log-transformed to improve the normality of residuals within linear regressions.

3.2 Water quality trend analyses

3.2.1 Sampling dates, seasons, and time periods for analyses

In trend assessments, there are several reasons why it is important to define the trend period and sampling frequency and to assess whether the observations are adequately distributed over time. First, because variation in many water quality variables is associated with the time of the year, the robustness of trend assessment is likely to be diminished if the observations are biased to particular times of the year. Second, a trend assessment will represent observed patterns from within a predefined time period; essentially that time period demarcated by the first and last observations. The assessment's characterisation of the change in the observations over the time period will not be robust if the observations are not reasonably evenly distributed within the time period. For these reasons, important steps in the data compilation process include specifying the sampling frequency, the time period, and ensuring observations are adequately distributed throughout the time period.

Monitoring programmes are typically designed to sample with a set frequency, (e.g., monthly, quarterly). The trend analysis "time increment" is specified to match this sampling typical frequency (e.g., time increment are months, bi-months or quarters). There is therefore generally an observation for each sample interval (i.e., each time increment 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) a change in sampling interval within the time period; and (2) the collection of more than one observation in a sample interval (e.g., two observations within a month). Both these deviations can occur in the data supplied to us but were mitigated in our analysis after we were able to use information supplied by data providers to identify and remove high-frequency observations that were not undertaken for routine SoE monitoring purposes as described above. Nevertheless, our procedure applied the following methods. For type (1) deviations, we coarsened the sampling interval to define the trend time increment for the part of the record with the higher sampling frequency. This was achieved by taking the observation in the higher frequency part of the record that was closest to the midpoint of the time increments defined by the coarser part of the record. A similar treatment is applied for type (2) deviations, where only the observation closest to the midpoint of the time increment was retained. The reason for not using a median value is that it will induce a trend in variance, which will invalidate the distributional assumptions of the Mann Kendall S test statistic (Helsel et al. 2020).

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 2024.

MfE also requested trends be 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 2024 (i.e., a total time period of 34 years). Aggregate rolling trends allowed inspection of transience in proportions of improving trends. The purpose of producing aggregated rolling trends was to investigate temporal patterns with 10-year aggregate trends. We therefore inspected the results of rolling trends to assess the degree to which the proportions of improving trends were either persistent or contained quasi-periodic fluctuations.

For a national study that aims to allow robust comparison of trends between sites and to provide a synoptic assessment of trends across the whole country, such as the present study, it is important that the trends evaluated at each site are commensurate in terms of their statistical power and representativeness of the time period. In these types of studies, it is general practice to define consistent time periods (i.e., trend duration and start date) so that all sites are subjected to the same conditions (i.e., equivalent political, climate, economic conditions). 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 comparable data. The 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 this is also termed “site screening criteria” and “completeness criteria”.

Universally applied data requirements or filtering rules have not been developed for trend assessments performed over many sites and variables. The definition of filtering rules is complicated by a trade-off: more restrictive rules, which increase the robustness of the individual trend analyses but will generally exclude numerous sites thereby reducing spatial coverage. In general, this trade-off is also affected by the duration of trend period. Steadily increasing monitoring effort over time means that shorter and more recent trend 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) and Whitehead et al. (2022), which set X and Y to 90%. Further, the definition of the time increments specified was flexible to maximise the number of sites that were included. If the site failed to comply with filter rule (2) when time increment was set as months, a coarsening of the data to bi-monthly (once every two months) or quarterly time increments was applied and the filter rule (2) was reassessed. Trends were evaluated based on the highest frequency coarser time increment that complied with filter rule (2). It is noted that the filtering rules imply a tolerance of variable levels of statistical power and temporal representativeness across the sites that were included in the analysis.

For MCI, we allowed both annual and bi-annual sampling intervals, as some 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 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 of 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 3.2.3).

In previous national scale assessments, the season assigned to a site × variable combination was limited to the sampling interval. However, for our trend assessment, season can be defined separately from the time increment of the assessment. The possible options for season are the selected time increment, and any other increments (i.e., monthly, bi-monthly, quarterly, biannually) that are whole multiples of the time increment. We evaluated seasonality for each potential season using the Kruskal-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 season × site × variable combinations as being seasonal based on the *p*-value from the Kruskal-Wallis test with $\alpha=0.05$. At this point, a check was also performed to determine whether each season option would have sufficient data variability within each season to undertake the trend assessment (>=5 non-censored values and/or >=3 unique non-censored values) (Section 3.2.5). The selected season is the time increment that meets minimum variability requirements and has a p value <0.05 or, in the case that more than season increment meets these thresholds, has the largest Kruskal-Wallis test statistic. For these sites × variable combinations, subsequent trend assessments followed the “seasonal” variants, described in Section 3.2.3.

Our approach varies slightly from that used in the previous national assessment, where season and time increment were required to be the same interval, and if a seasonal assessment did not meet minimum within season variability requirements, both the time increment and season were coarsened, and the seasonality test repeated. The approach used in this study reduces unnecessary loss of data in the coarsening process.

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 Kruskal-Wallis test for *p*-values values larger than 0.05 (Helsel et al. 2020).

3.2.3 Analysis of trends

The purpose of trend assessment is to evaluate the direction (i.e., improving or worsening) and rate of the change in the central tendency of the observed water quality values over the period of analysis (i.e., the trend). Because the analysed data represent observations 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 evaluated trends using the LWPTrends functions (Fraser and Snelder 2025) that are implemented in the R statistical computing software (R Core Team 2023). A brief description of the theoretical basis for these functions is described below.

Assessments of trend directions

The trend direction and the confidence in trend direction 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 is recommended, rather than Sen slope regression, because the former more robustly handles censored values. However, Sen slope regression is the recommended method for assessing the trend.

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.

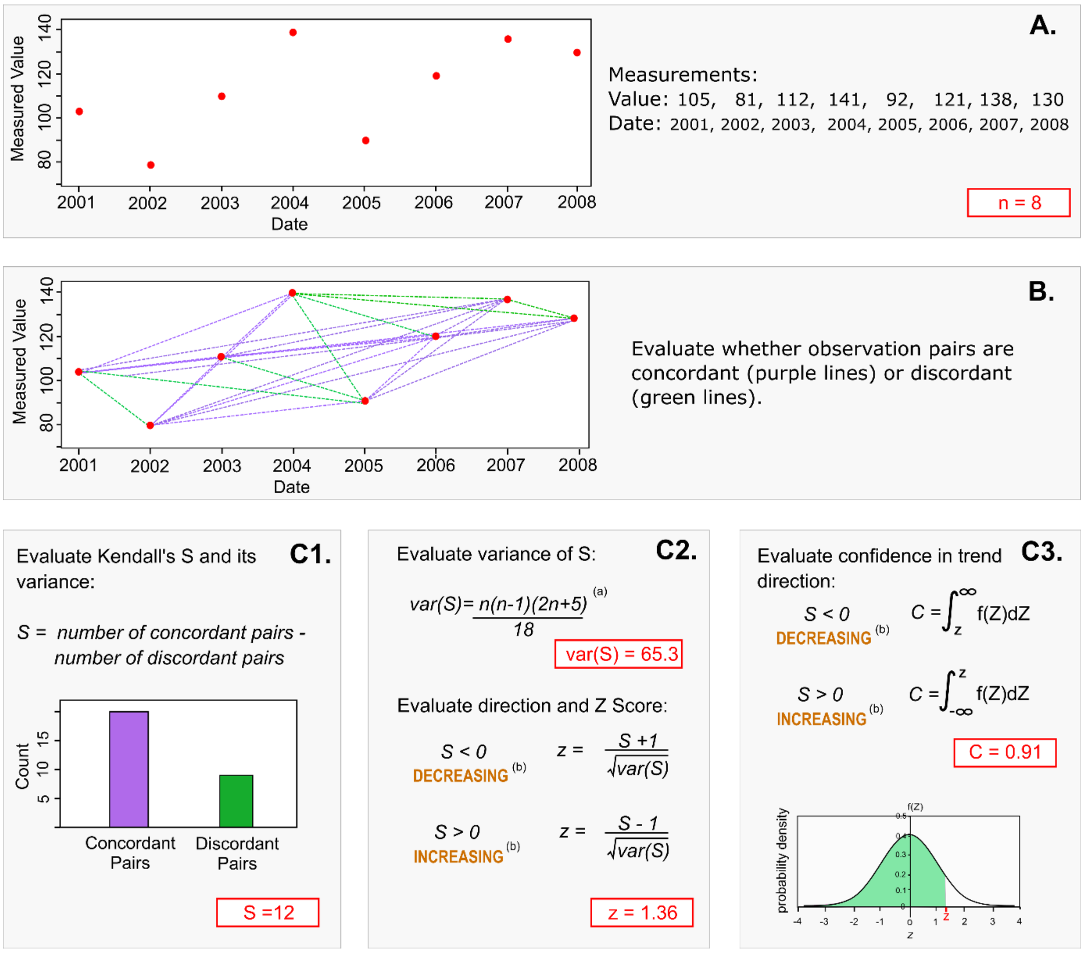

Figure 3-1: Schematic diagram demonstrating how the Kendall *S* statistic and confidence in trend direction (C) is calculated. *Notes: [a] the calculation of the variance in S has some adjustments to account for ties (numerically equal values, or ties in time) and censored values. Details of these adjustments can be found in (Gilbert 1987; Helsel 2005, 2012). [b] There is a third alternative, where S=0. In this case C is 0.5, and the trend direction is classified as "indeterminate". Values of S equal to -1 or 1 will also result in a Z value of 0, and a C value of 0.5 and the trend direction is similarly classified as "indeterminate".*

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). Pairs of observations that are tied in value or tied in time[4] (i.e., have the same time increment-Year) are assigned differences of zero (Gilbert 1987, p. 243). 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).

[4] Accounting for ties in time is an addition to the analysis since Whitehead et al. (2022). This addition is to accommodate seasonal analyses where the Season time increment is larger than the data TimeIncr (time increment).

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). A continuous measure of confidence in the assessed trend direction can be determined based on the posterior probability distribution of S, the true (i.e., population) difference in concordant and discordant pairs (Snelder et al. 2022). The posterior probability distribution of *S* is given by a normal distribution with mean of *S* and variance of *var(S)*. The confidence in assessed trend direction can be evaluated as the proportion of the posterior probability distribution that has the same sign as S.

In practice, confidence can be calculated by first transforming the value of $S = 0$ on the posterior probability distribution into a standard normal deviate, Z (panel C2). *C* is then calculated as area under the standard normal distribution to the left (Z>0) or right (Z<0) of the value of Z, using the quantile function for the normal distribution.

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 sign of S is equally likely to be in the opposite direction to that indicated by the true trend, to 1, indicating complete confidence that the sign of *S* is the same as the true trend.

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 rates 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. 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).

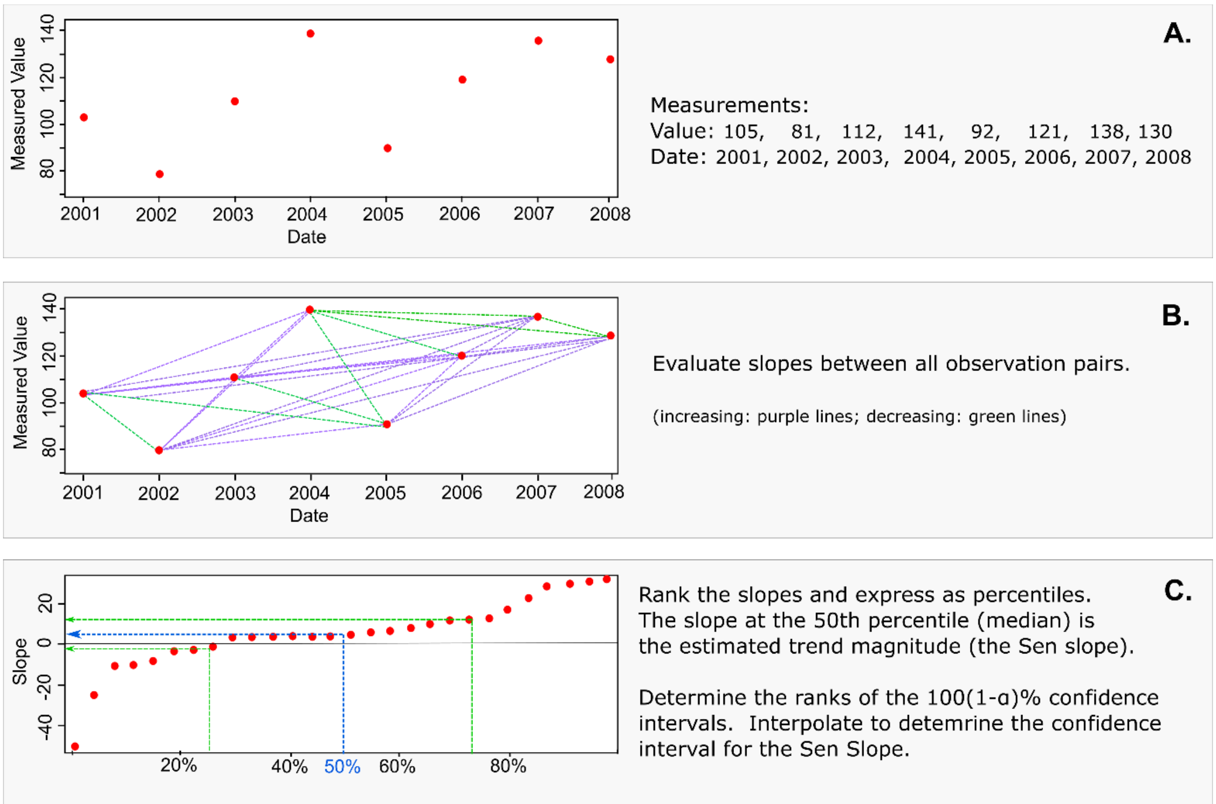

Figure 3-2: Schematic diagram of the calculation of the Sen slope, which is used to characterise trend rate. (from Snelder et al. 2021a).

The seasonal version of the SSE is used in situations where there are significant (e.g., $p \leq 0.05$, as evaluated using a Kruskal Wallis test) differences in water quality measurements between "seasons". Seasons are defined primarily by sampling intervals, which are commonly monthly or quarterly for water quality monitoring but can also be defined as whole multiples of the sampling interval (e.g., monthly sampling interval and quarterly season). 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 but for data pertaining to observations in each individual season. When the time increment is smaller than the season time increment (i.e. there is generally more than one observation per season by year) the inter-observation slopes are excluded for pairs of observations that are in the same season-year (they are treated as ties in time), but all slopes in the season of one year are compared with all slopes in the season of the other years (Gilbert 1987, p. 218). Second, SSSE is the median of the seasonal values. 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.

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.

3.2.4 Handling censored values

Censored values are managed differently to non-censored values by the non-parametric trend assessment methods described in Section 3.2.3. 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.

Censored values in the data used to calculate Kendall's S and its p-value are robustly handled in the manner recommended by Helsel (2005, 2012). Briefly, for left-censored data, increases and decreases in a water quality variable are identified whenever possible. Thus, a change from a censored data entry of <1 to a measured value of 10 is considered an increase. A change from a censored data entry of <1 to a measured value 0.5 was considered a tie, as is a change from <1 to a <5, because neither can definitively be called an increase or decrease. Similar logic applied to right censored values. The information about ties is used in the calculation of the Kendall S statistic and its variance following Helsel (2012) and this provides for robust calculation of the p-value associated with the Kendall test. This approach is implemented in LWPTrends using source code from the NADA package. The method is robust to changes in detection limit over time.

Note that as the proportion of censored values increases, the proportion of ties increases and the confidence in the trend direction decreases. Therefore, confidence in direction tends to be low when trends are calculated from data with high proportions of censored observations.

When calculating Sen slopes, the inter-observation slope cannot be definitively calculated between any combination of observations in which either one or both are censored. Therefore, when SSE and SSSE (i.e., Sen slopes) are calculated by the LWPTrends library, the censored data entries are substituted with their corresponding raw values (i.e., the numeric component of a censored data entry) multiplied by a factor (0.5 for left-censored and 1.1 for right-censored values). This ensures that any measured value that is equal to a raw value is treated as being larger than the censored value if it is left-censored value and smaller than the censored value if it is right-censored. The inter-observation slopes associated with the censored values are therefore imprecise (because they are calculated from the substituted values). However, because the Sen slope is the median of all the inter-observation slopes, the Sen slope is unaffected by censoring when a small proportion of observations are censored. As the proportion of censored values increase, the probability that the Sen slope is affected by censoring increases. 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.

In previous versions of LWPTrends used to evaluate national water quality trends (Whitehead et al. 2021), the influence of changes in detection limits over time on the Sen slope were handled through a high censor filter. Briefly, this filter replaced all observations below the highest detection limit (or some user specified value) to the highest detection limit and marked these observations as censored. We now implement a different approach to handle changes in detection limit when calculating trend rates that follows the logic used in the calculation of Kendall's S. Inter-observation slopes are considered to be ties and set to zero, regardless of their values, when: (1) both observations were either left or right censored, (2) when one observation is left censored and larger than the other non-censored observation; (3) when one observation is right censored and smaller than the other non-censored observation. This approach has the advantage that the user decisions around it and how to implement a high censor filter are removed.

3.2.5 Interpretation of trends

The trend assessment procedures used here allow a more nuanced inference than the categorical measure of confidence associated with Null Hypothesis Significance Testing (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.

To assist with interpretation of confidence in trend direction, trends are categorised based on the confidence that a trend was improving. This involved firstly converting C_d into a confidence that a trend was improving (C_i). Improvement is indicated by decreasing trends for all the water quality variables in this study ($C_i = C_d$) except for visual clarity and MCI (for which increasing trends indicate improvement). For these variables, C_i is the complement of C_d (i.e., C_i=1-C_d).

The trend for each site × variable combination was assigned a categorical level of confidence that the trend was improving 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 improving.

Categorical level of confidence trend was improving	Value of C_i (%)
Very likely improving	$0.90 < C_i \leq 1.00$
Likely improving	$0.67 < C_i \leq 0.90$
Low confidence	$0.33 < C_i \leq 0.67$
Likely worsening	$0.10 < C_i \leq 0.33$
Very likely worsening	$0.00 \leq C_i \leq 0.10$

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.

3.2.6 Aggregation of site trends

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

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 improving (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 improving 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. Results for rolling time windows were only shown where there were at least 200 sites within the 10-year time window. This was an arbitrary cut-off point selected to minimise bias that might be associated with a small sample size but maximise the number of time windows that were reported.

The second approach also utilises the confidence that the true trend was improving to provide a probabilistic estimate of the proportion of improving site-specific trends (P_i) within a geographic or environmental domain. For a given water quality variable, the trends at multiple monitoring sites distributed across a domain of interest can be assumed to represent independent samples of the population of trends, for all sites within that domain.

The statistic P_i is calculated by letting the sampled sites within this domain be indexed by m, so that $m \in \{1, \ldots, M\}$ and letting I be a random Bernoulli distributed variable which takes the value 1 with probability $p = C_i$ and the value 0 with probability $q = 1 - C_i$ (where C_i is the confidence that the trend was improving, as described in Section 3.2.3). Therefore, $I_m = 1$ denotes an improving trend at site $s \in \{1, \ldots, M\}$ when the estimated $p_m \geq 0.5$ and a worsening trend as 0 when $p_m < 0.5$. Then, the estimated proportion of sites with improving trends in the domain is:

$$P_i = \sum_{m=1}^{M} I_m / M$$

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_i is:

$$\boldsymbol{Var(P_i) = \frac{1}{M^2} \sum_{m=1}^{M} Var(I_m) = \frac{1}{M^2} \sum_{m=1}^{M} p_m(1 - p_m)}$$

P_i and its variance represent an estimate of the population proportion of improving trends, within a spatial or environmental domain, and the uncertainty of that estimate. It is noted that the proportion of worsening trends is the complement of the result (i.e., 1 - P_i).

If there is spatial correlation in the assessed trend directions at the individual sites, the assumption of independence, when calculating P_i and its variance as above, is violated (Douglas et al. 2000). Spatial correlation means that the effective sample size of the dataset is less than the number of sites and this results in under-estimation of the variance and over-estimation of confidence in the P_i statistic. The method of Douglas et al. (2000) can be used to calculate an adjusted version of the variance of P_i as:

$$\boldsymbol{Var(P_i) = \frac{1}{M^2}\left[\sum_{k=1}^{M} Var(I_k) + 2\sum_{k=1}^{M-1}\sum_{l=1}^{M-k} Cov(I_k, I_{k+l})\right]}$$

where the covariance between monitoring sites is calculated as:

$$\boldsymbol{Cov(I_k, I_{k+l}) = \sqrt{Var(I_k)Var(I_{k+l})\rho^c_{k,k+l}}}$$

and where $\rho^c_{k,k+1}$ is replaced by the sample cross-correlation coefficient $r_{k,k+l}$ (Yue and Wang 2002), which is computed from the observation time series at site *k* and *k+l* as:

$$\boldsymbol{r_{k,k+l} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(x_{k,i} - \overline{x_k}\right)\left(x_{k+l,i} - \overline{x_{k+l}}\right)}{\sqrt{Var(x_k)Var(x_{k+l})}}}$$

where x^i_k and x^i_{k+l} represent one set of N concurrent observations at the sites.

The estimated variance of P_i can be used to construct 95% confidence intervals[5] around the P_i statistics as follows:

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

We calculated P_i and its confidence interval for all water quality variables and for domains of interest defined by the entire country, and by the four human modified land cover classes defined in Section 2.6.

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

4 Results – river state

4.1 State summary statistics

Between 695 and 983 river monitoring sites met the filtering rules for the state analysis of nutrients, *E. coli*, clarity, and MCI; the number of qualifying sites varied by water quality variable and by human modified 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. Only a small proportion of observations were censored for many site × variable combinations (Figure 4-2). A large proportion of observations were censored at a small number of sites for some variables, except for NH_4-N which exhibited a larger proportion of censored observations at more sites compared to the other variables. The median observation was below the detection limit for 42% of sites for NH_4-N and 6% of sites for DRP. The proportion of data with QC information has increased at a rapid rate since introduction of QC standards in 2019, although the majority of recent observations were not associated with QC information (Figure 4-3).

Table 4-1: Number of river monitoring sites by percentage of catchment human modified land cover and water quality variable that were included in the state analyses of nutrients, *E. coli*, clarity, turbidity and MCI. The site numbers shown refer to sites that met the data requirements outlined in Section 3.1.3. Note NH_4-N (Adj.) is pH-adjusted ammoniacal-N.

Variable	Number of sites				
	Total	0–25%	25–50%	50–75%	75–100%
Clarity	695	137	123	146	289
Turbidity	940	202	171	183	384
MCI	874	253	144	168	309
E. coli	965	217	179	185	384
DRP	992	225	183	190	394
TP	990	223	183	190	394
NH_4-N	992	225	183	190	394
NH_4-N (Adj.)	974	221	179	186	388
NNN	991	225	183	189	394
TN	990	223	183	190	394

Figure 4-1: River water quality monitoring sites used for state analyses of nutrients, *E. coli*, clarity, turbidity and MCI.

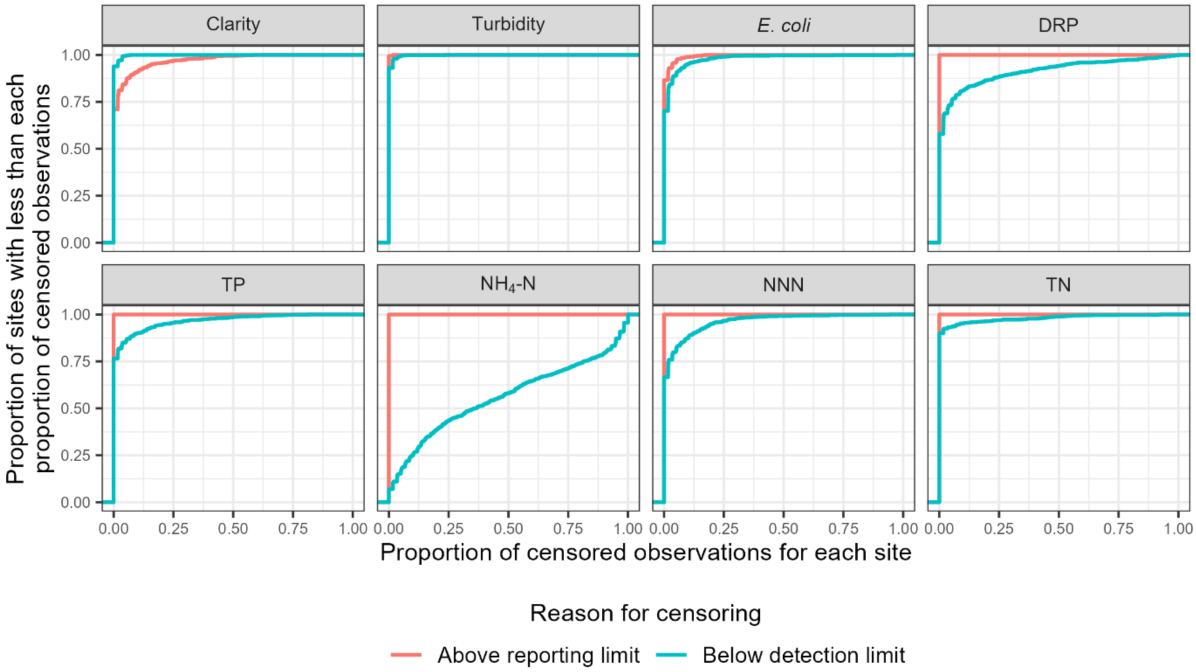

Figure 4-2: Proportion of censored data for each site × variable combination for which state was assessed. Note no MCI observations were censored.

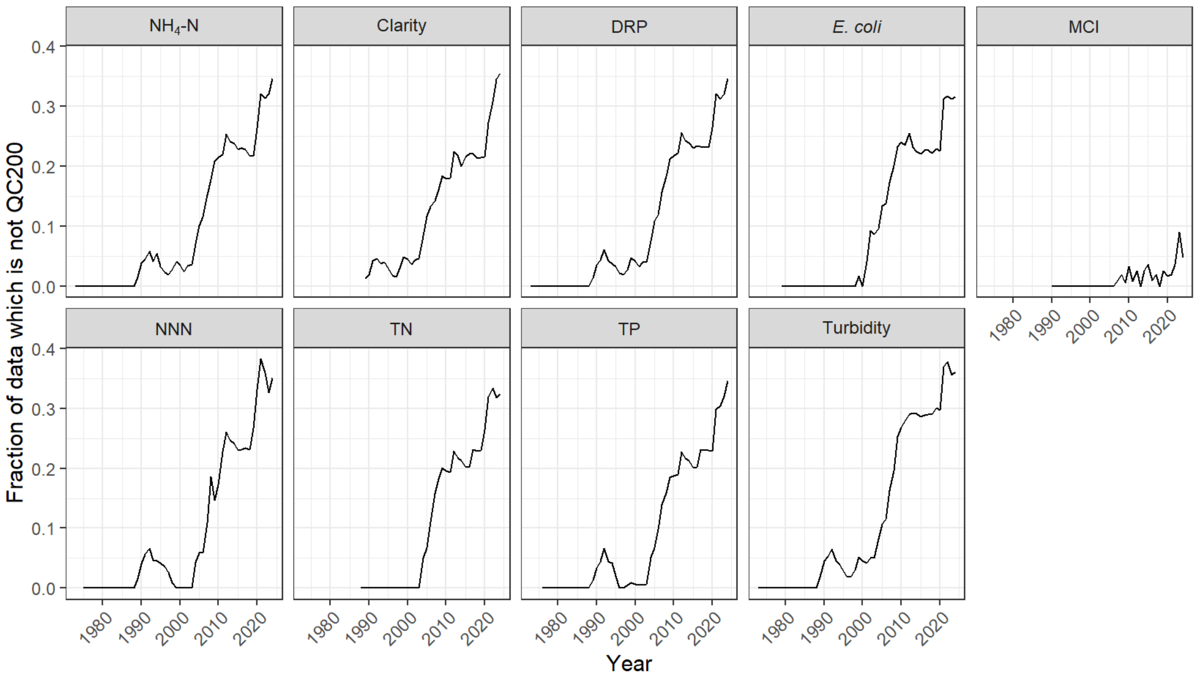

Figure 4-3: For each variable, the proportion of analysed data that was not labelled QC 200. The remaining proportion was either labelled as QC 200 (i.e., no quality) by the data provider or had missing QC information.

The distributions of site-median values of the nine water quality variables for the 2020–2024 period are summarized as box-and-whisker plots, with sites grouped by human modified land cover classes (Figure 4-4). The plots in Figure 4-4 indicate that water quality state (i.e., site medians for nutrients, *E. coli*, MCI and clarity) was highly variable, with some of the variation explained by the degree of catchment human modified land cover. There is a consistent pattern of higher water quality (lower values for nutrients and *E. coli,* and higher values for MCI and clarity), as indicated by the class median, with lower areal proportions of catchment human modified land cover.

The distribution of *E. coli* concentration percentiles (5th, 20th, 50th, 80th, and 95th) are shown in Figure 4-5, and the distribution of the *E. coli* exceedance measures (the percentage of observations that exceeded 260 and 540 100 mL^{-1}, respectively) are shown in Figure 4-6. All percentiles and exceedances showed category median values that increased with increasing proportions of catchment human modified land cover. However, there is considerable overlap in the distributions for each of the classes.

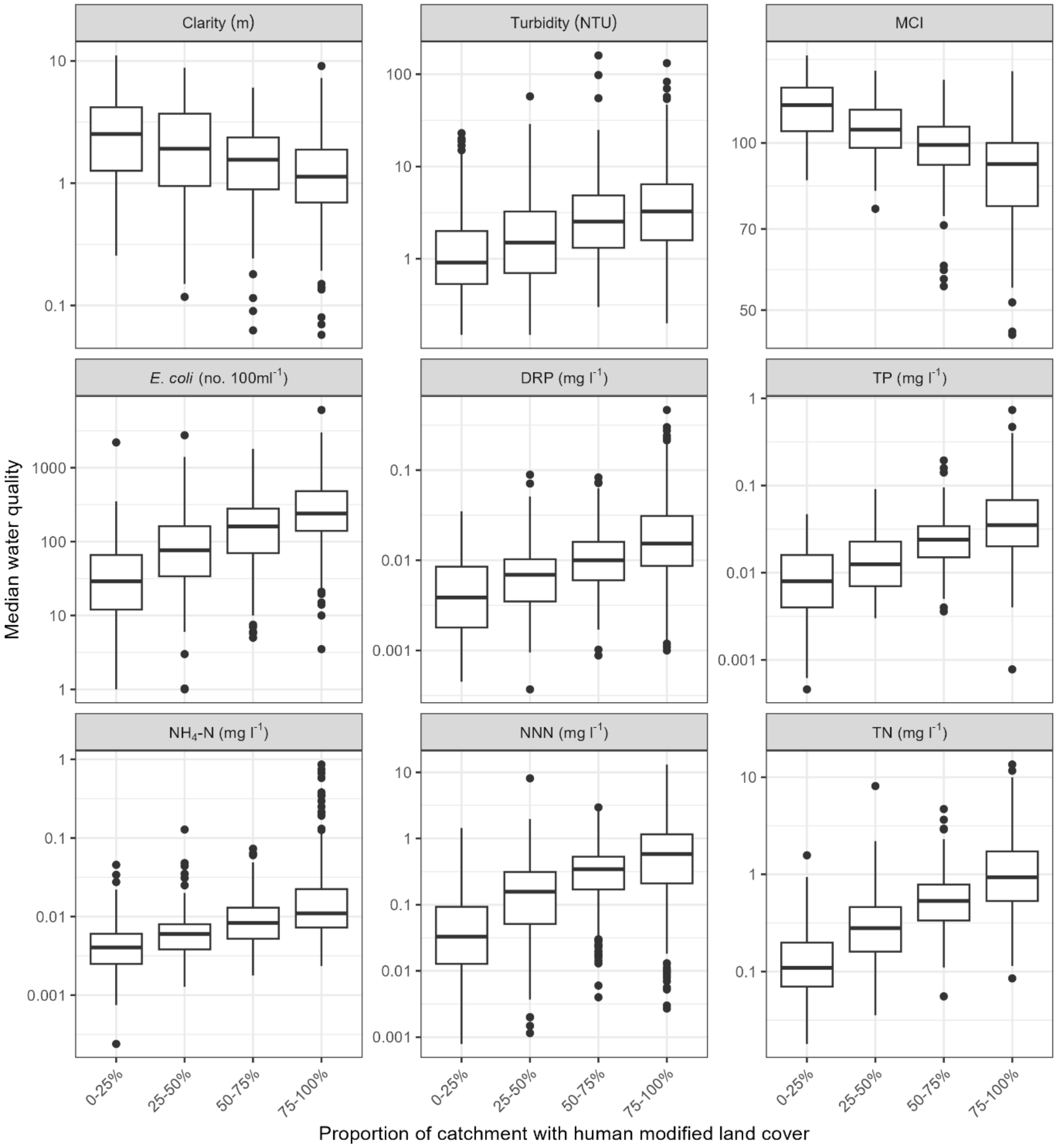

Figure 4-4: River water quality state in human modified land cover classes. Box-and-whisker plots show the distributions of monitoring site medians within classes. For y-axes units of measure refer to Table 2-1. Black horizontal line in each box indicates the median of site medians, and the box indicates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

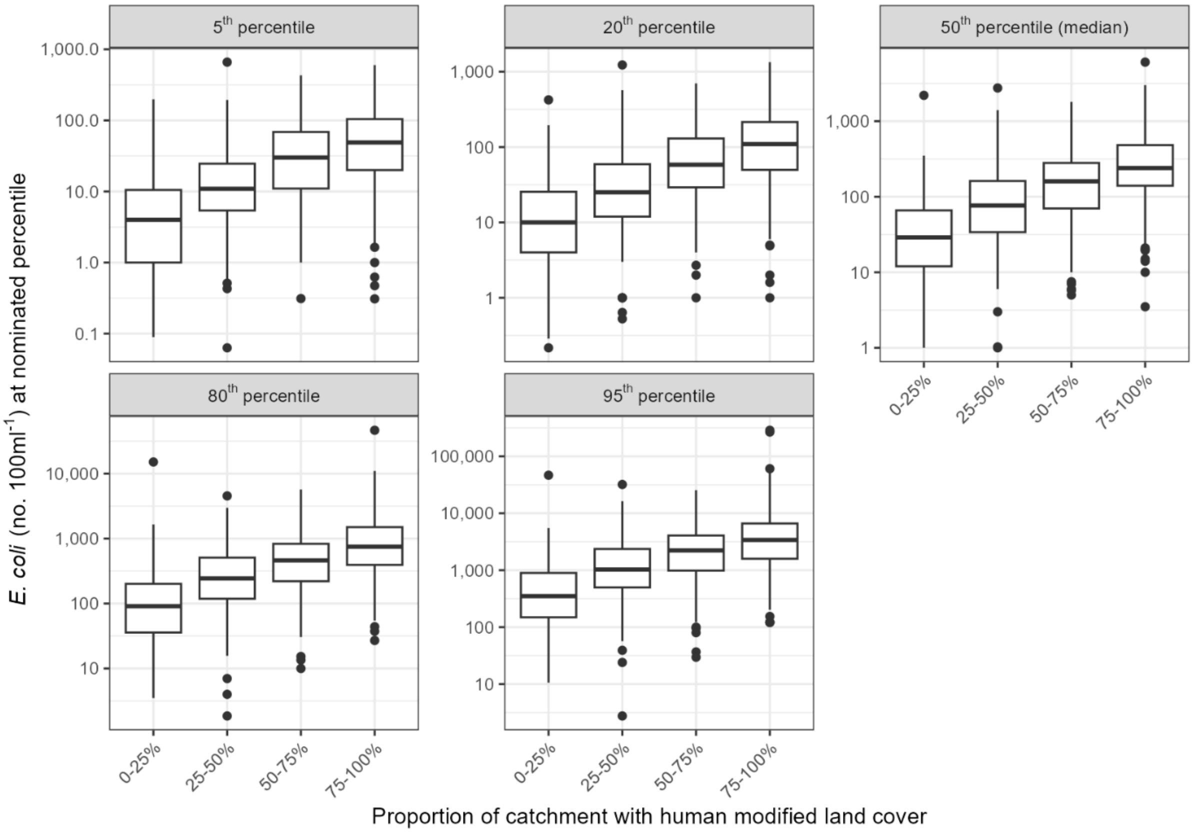

Figure 4-5: ***E. coli*** **concentrations in human modified land cover classes.** Box-and-whisker plots show the distributions of monitoring site percentiles within classes. Black horizontal line in each box indicates the median of site percentiles, and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5×IQR from the box. Data beyond the whiskers are shown as black circles. Note log-scale on Y-axes.

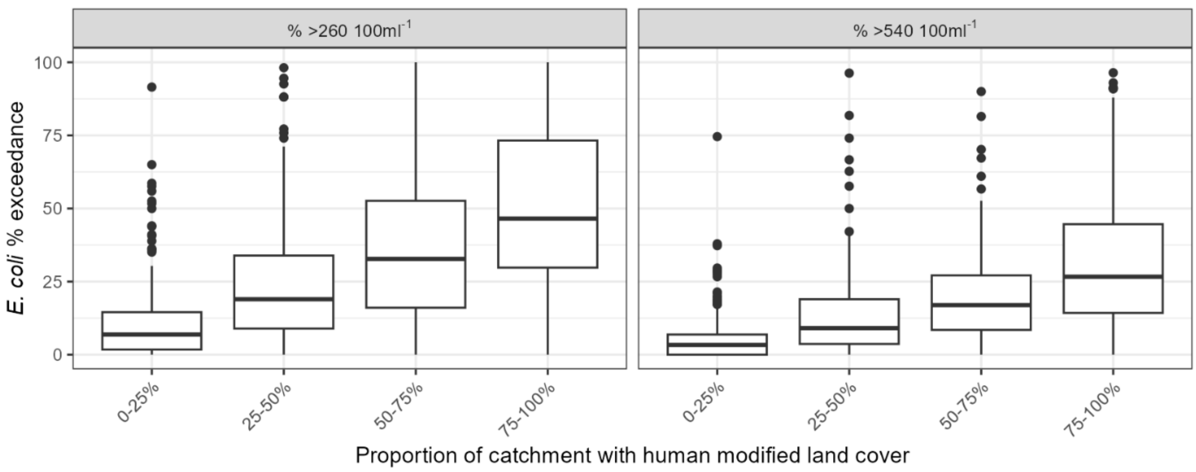

Figure 4-6: ***E. coli*** **percent exceedance in human modified land cover classes.** Box-and-whisker plots show the distributions of percentage exceedance over 260 100 mL^{-1} (% >260 $100mL^{-1}$) and 540 100 ml-1 (% >540 $100mL^{-1}$) at river monitoring sites within classes. Black horizontal line in each box indicates the median of percent exceedances and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5×IQR from the box. Data beyond the whiskers are shown as black circles.

4.2 NOF grades

Table 4-2 and Figure 4-7 provide summaries of water quality grades for each NPS-FM attribute, demonstrating the number and percentage of sites that are classified in each NOF grade. Figure 4-8, Figure 4-9, and Figure 4-10 provide maps for each attribute showing the sites coloured by their evaluated NOF grade. Figure 4-11, Figure 4-12, and Figure 4-13 show the percentage of sites belonging to each grade, by land cover category and variable.

The majority of sites (64%) 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). More than 88% of sites with 75–100% upstream human modified land cover were below the bottom line. Very few sites (<1–5%) were below the bottom line for the ammonia (toxicity) and nitrate (toxicity) attributes. For the suspended fine sediment attribute (Clarity), 39% of sites were below the bottom line, including 30% of sites with 0–25% upstream human modified land cover. For the macroinvertebrate attribute (numeric attribute state of median MCI), 23% of sites were below the bottom line, including over 47% of sites 75–100% upstream human modified land cover. There is no national bottom line for the DRP attribute, but 27% of sites received D grades for overall attribute grade. Many of the lowest DRP grades were located in Taranaki and Bay of Plenty, which may in part reflect local geological conditions.

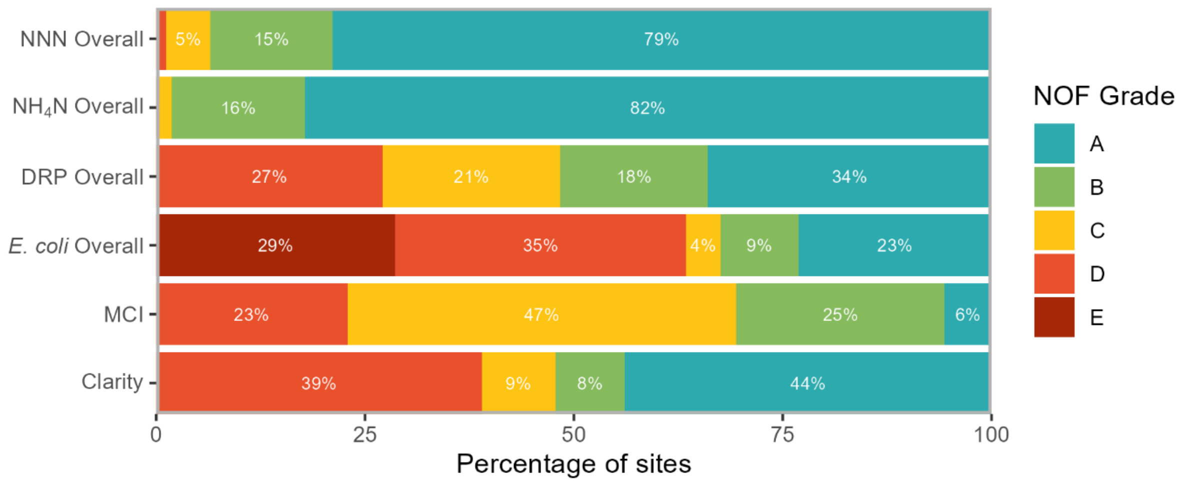

Figure 4-7: Stacked bar chart showing the percentage of sites assigned each NOF grade overall NOF attribute.

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
Clarity	305 (44%)	58 (8%)	61 (9%)	271 (39%)	
MCI	49 (6%)	218 (25%)	407 (47%)	200 (23%)	
E. coli Overall	223 (23%)	90 (9%)	40 (4%)	336 (35%)	276 (29%)
E. coli Median	488 (51%)	0 (0%)	0 (0%)	224 (23%)	253 (26%)
E. coli 95th percentile	239 (25%)	122 (13%)	40 (4%)	564 (58%)	0 (0%)
E. coli % >260	382 (40%)	128 (13%)	36 (4%)	174 (18%)	245 (25%)
E. coli % >540	231 (24%)	157 (16%)	212 (22%)	145 (15%)	220 (23%)
DRP Overall	337 (34%)	175 (18%)	211 (21%)	269 (27%)	
DRP Median	353 (36%)	190 (19%)	210 (21%)	239 (24%)	
DRP 95th percentile	517 (52%)	138 (14%)	157 (16%)	180 (18%)	
NH_4-N Overall	801 (82%)	155 (16%)	17 (2%)	1 (0%)	
NH_4-N Median	914 (94%)	55 (6%)	5 (1%)	0 (0%)	
NH_4-N 95th percentile	802 (82%)	155 (16%)	16 (2%)	1 (0%)	
NNN Overall	782 (79%)	145 (15%)	52 (5%)	12 (1%)	
NNN Median	856 (86%)	91 (9%)	32 (3%)	12 (1%)	
NNN 95th percentile	786 (79%)	143 (14%)	54 (5%)	8 (1%)	

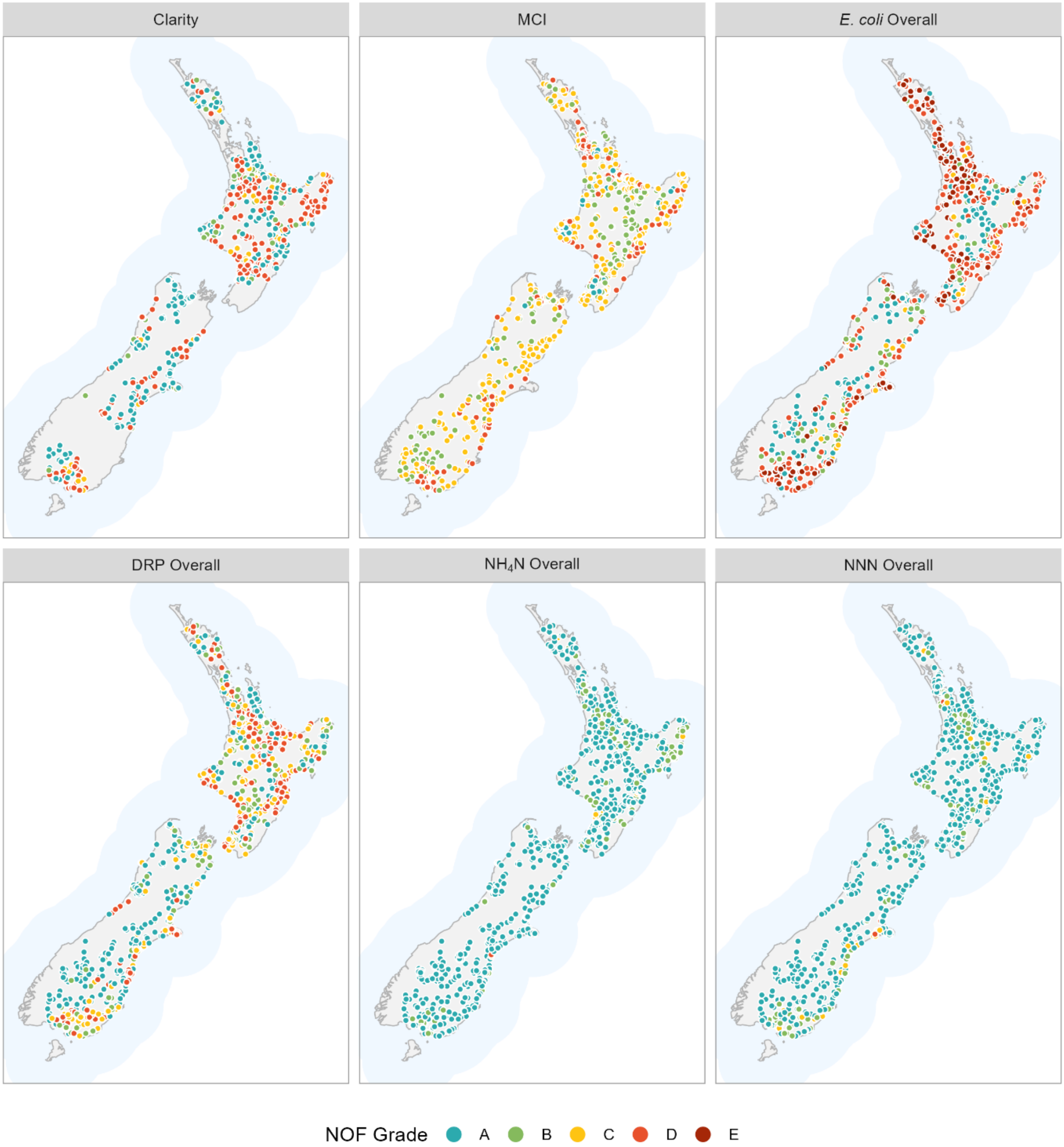

Figure 4-8: Maps showing NOF overall attribute grades.

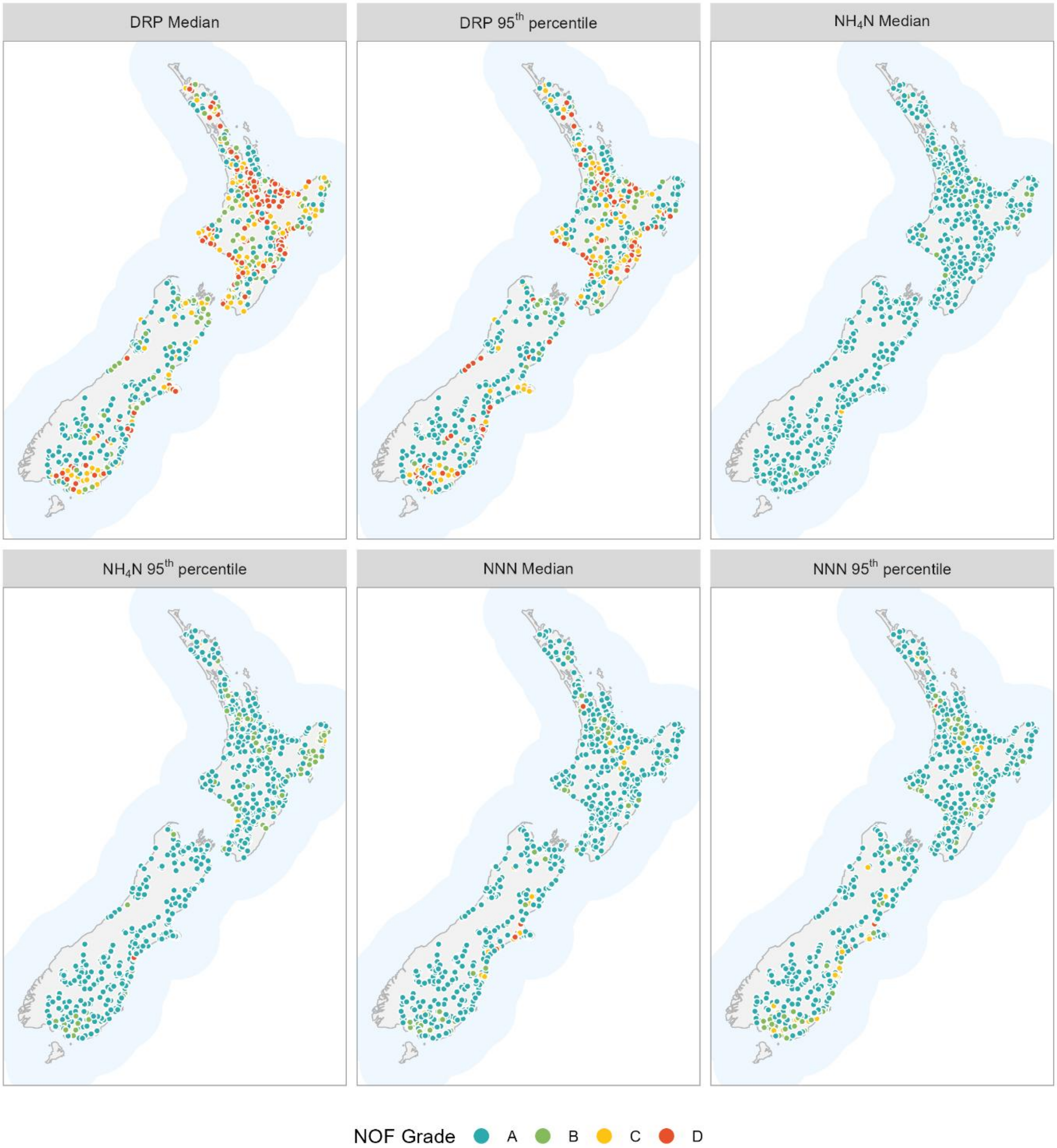

Figure 4-9: Maps showing NOF grades for chemical attribute components.

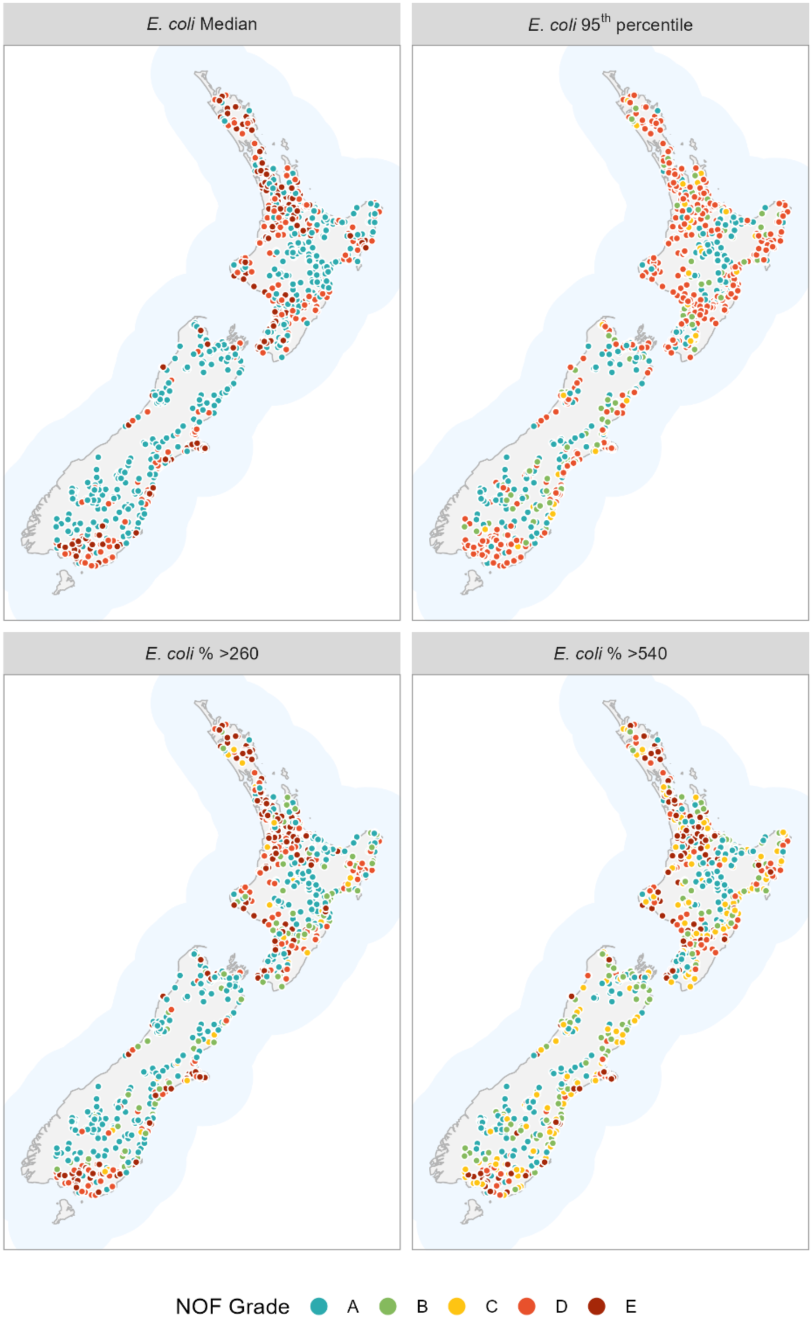

Figure 4-10: Maps showing NOF grades for the *E. coli* attribute components.

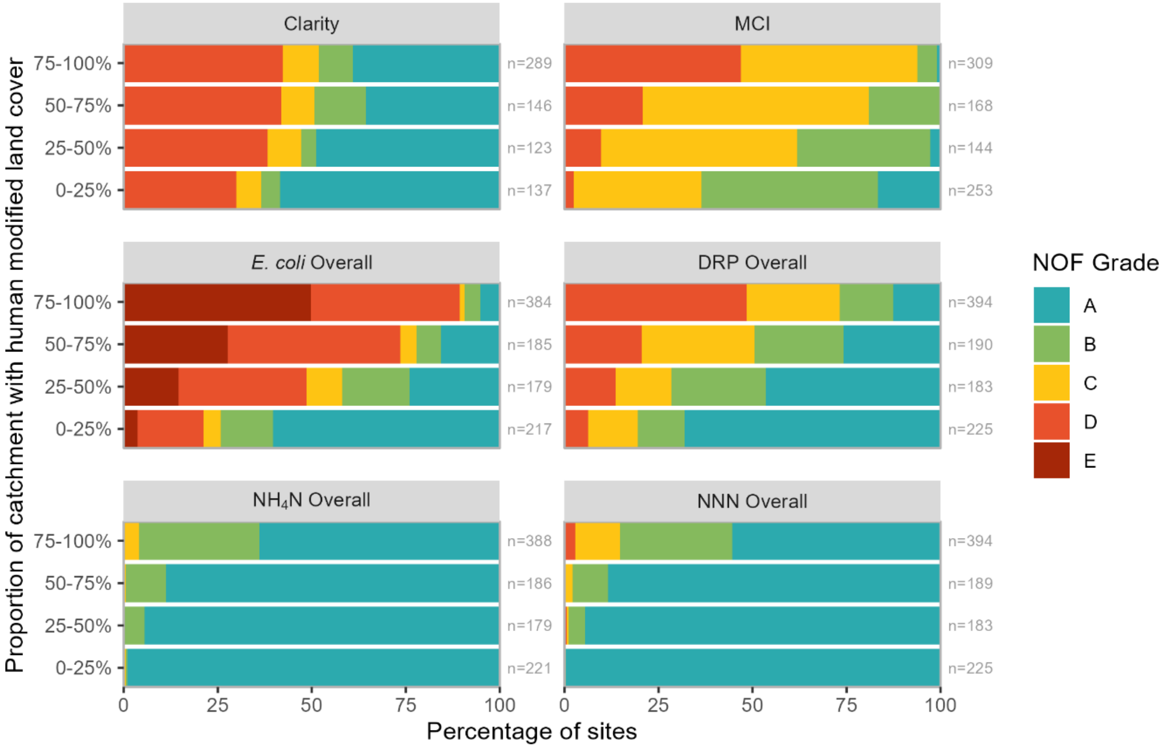

Figure 4-11: Stacked bar charts showing the percentage of sites assigned each NOF grade, by human modified land cover class and overall NOF attribute. n indicates the number sites in each group.

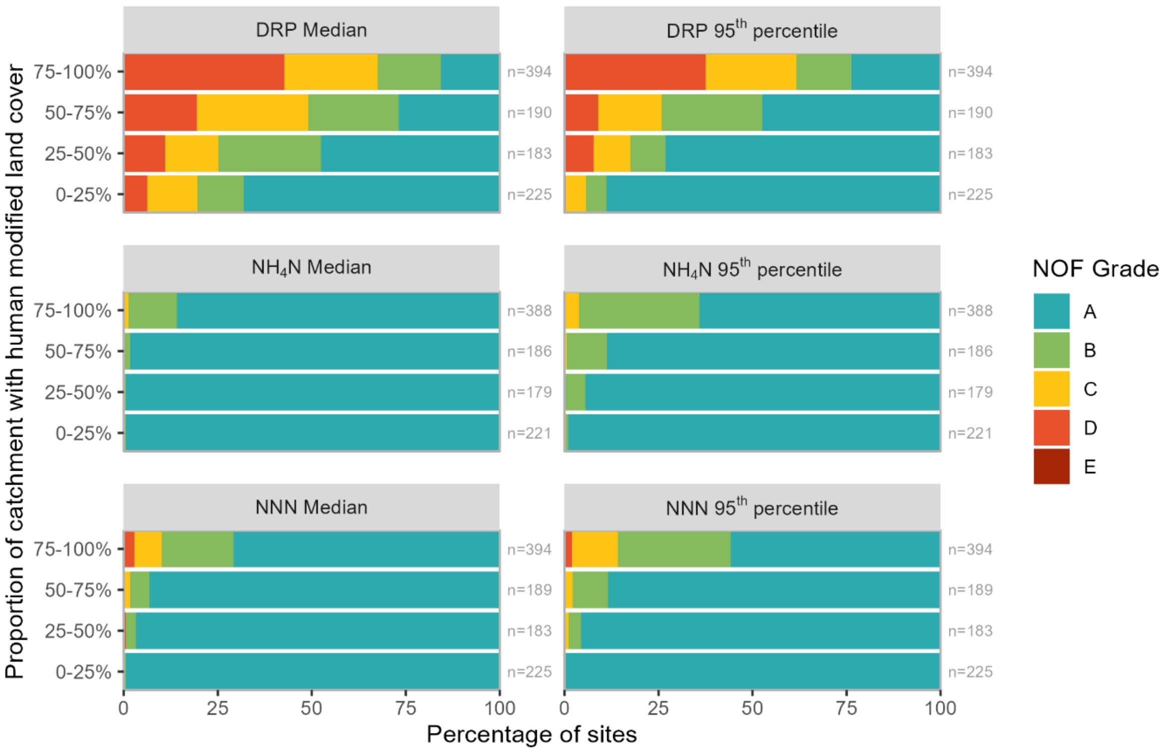

Figure 4-12: Stacked bar charts showing the percentage of sites assigned each NOF grade, by human modified land cover class and NOF chemical attribute component. n indicates the number sites in each group.

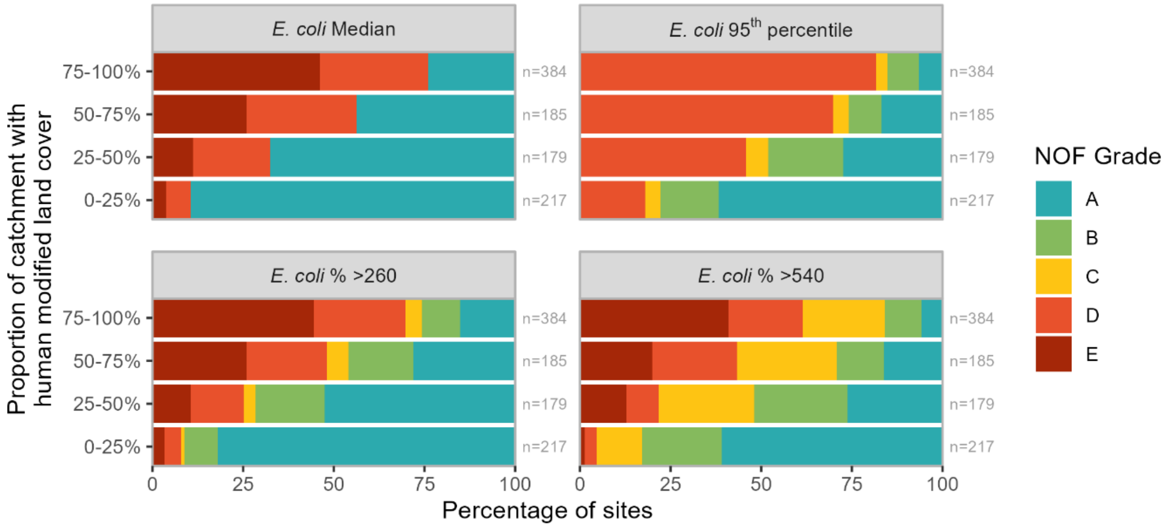

Figure 4-13: Stacked bar charts showing the percentage of sites assigned each NOF grade, by human modified land cover class and NOF *E. coli* attribute component. n indicates the number sites in each group.

4.3 Relationships between water quality state and catchment land cover

The regression results indicated that the concentrations of each nutrient and *E. coli* increased, and MCI scores and visual clarity decreased, with increasing proportions of upstream human modified land cover (Figure 4-14). Human modified land cover explained 9%–56% of the variation in log-transformed water quality variables; these relationships were strongest for median TN, NNN, TP and *E. coli* concentrations and MCI scores. Colouring of the points in Figure 4-14 indicated a tendency for sites with dominant urban land cover to be associated with low MCI and sites with dominant exotic forest land cover to be associated with high MCI. There was also a tendency for sites with dominant urban land cover to be associated with high *E. coli* concentrations and sites with dominant exotic forest land cover to be associated with low *E. coli* concentrations.

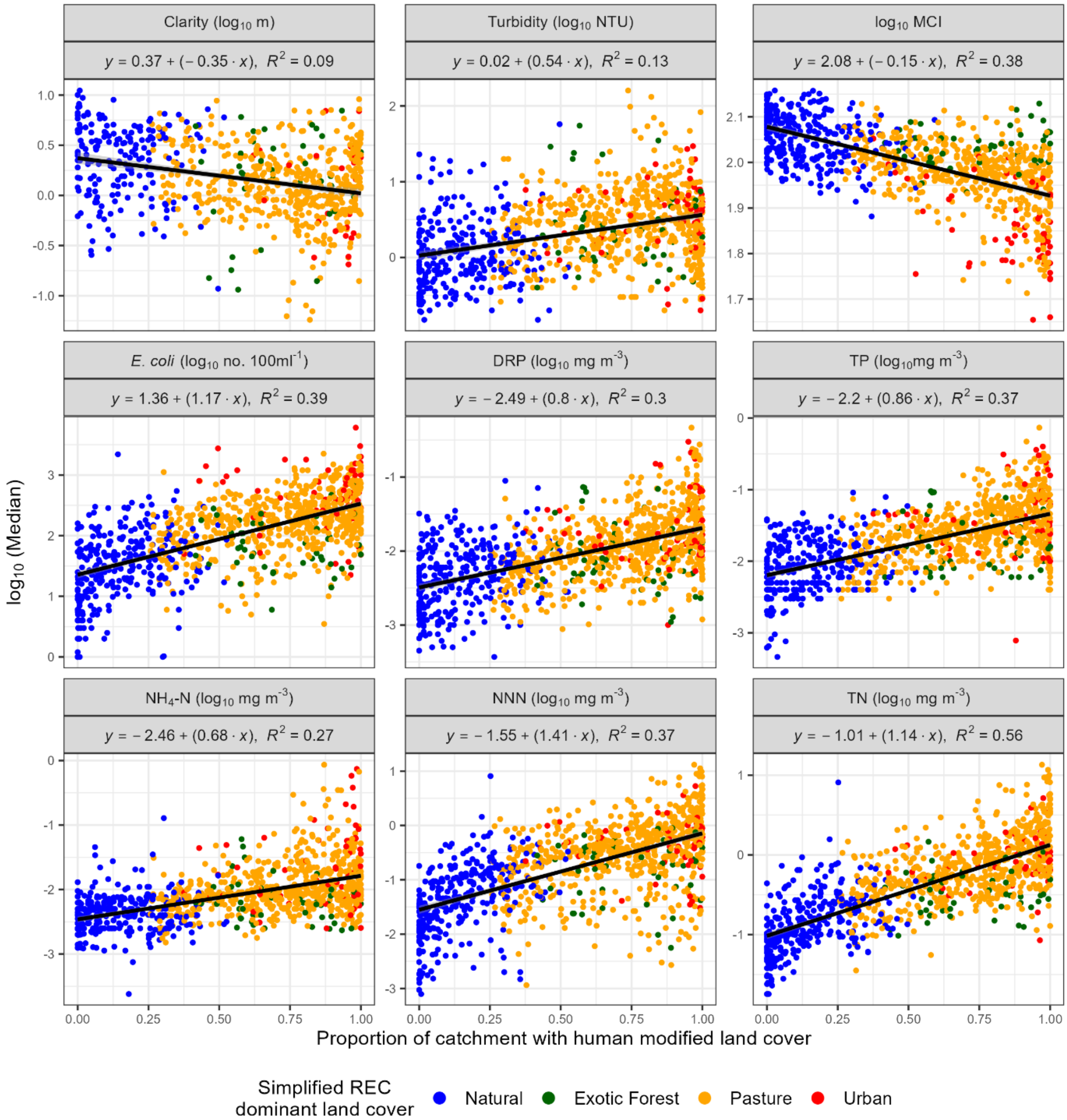

Figure 4-14: Relationships between median water quality state and proportion of upstream catchment with human modified land cover. Sites are coloured based on their simplified REC dominant land cover class. Solid lines indicate least squares linear regression models.

5 Results – river trends

5.1 Ten-year trends (2011–2020)

Between 624 and 865 river monitoring sites met the filtering rules for the 10-year trend analysis of nutrients, *E. coli*, Turbidity, MCI and clarity (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. No clarity data was available from Auckland Council or Otago Regional Council, as such the only sites in these regions for which trends in clarity could be calculated are those from the NRWQN network. For the 10-year trend period, a large proportion of observations were censored at a small number of sites for some variables, except for NH_4-N which exhibited a larger proportion of censored observations at more sites compared to the other variables (Figure 5-2).

Table 5-1: Number of river monitoring sites by human modified land cover class and water quality variable included in the 10-year trend analyses of nutrients, *E. coli*, clarity, turbidity 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	0–25%	25–50%	50–75%	75–100%
Clarity	624	133	122	128	241
Turbidity	814	162	157	158	337
MCI	834	229	137	159	309
E. coli	864	181	168	166	349
DRP	860	178	166	166	350
TP	826	166	156	158	346
NH_4-N	865	180	167	167	351
NNN	861	179	166	166	350
TN	826	166	156	158	346

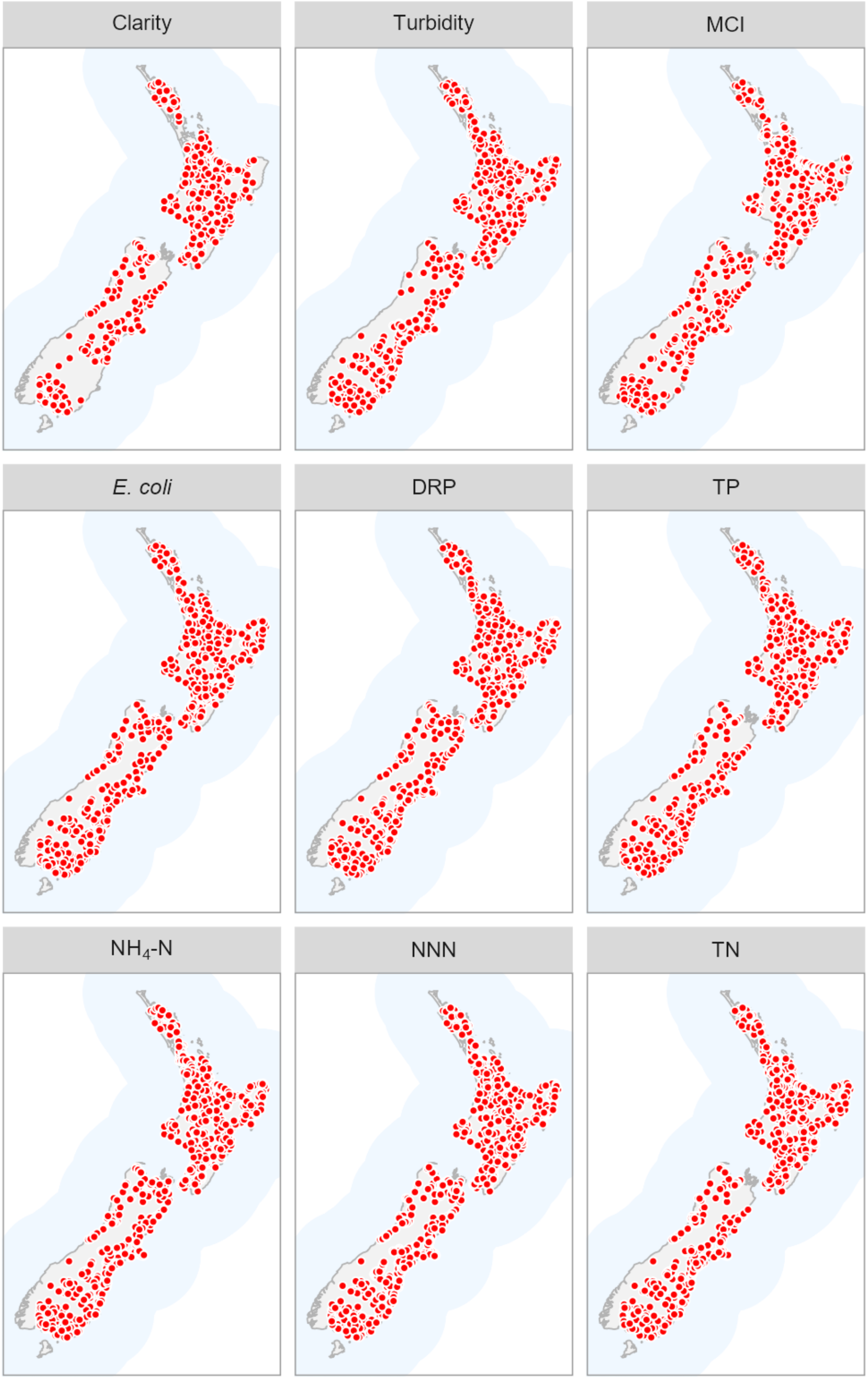

Figure 5-1: River water quality monitoring sites used for 10-year trend analyses of nutrients, *E. coli*, clarity, turbidity and MCI.

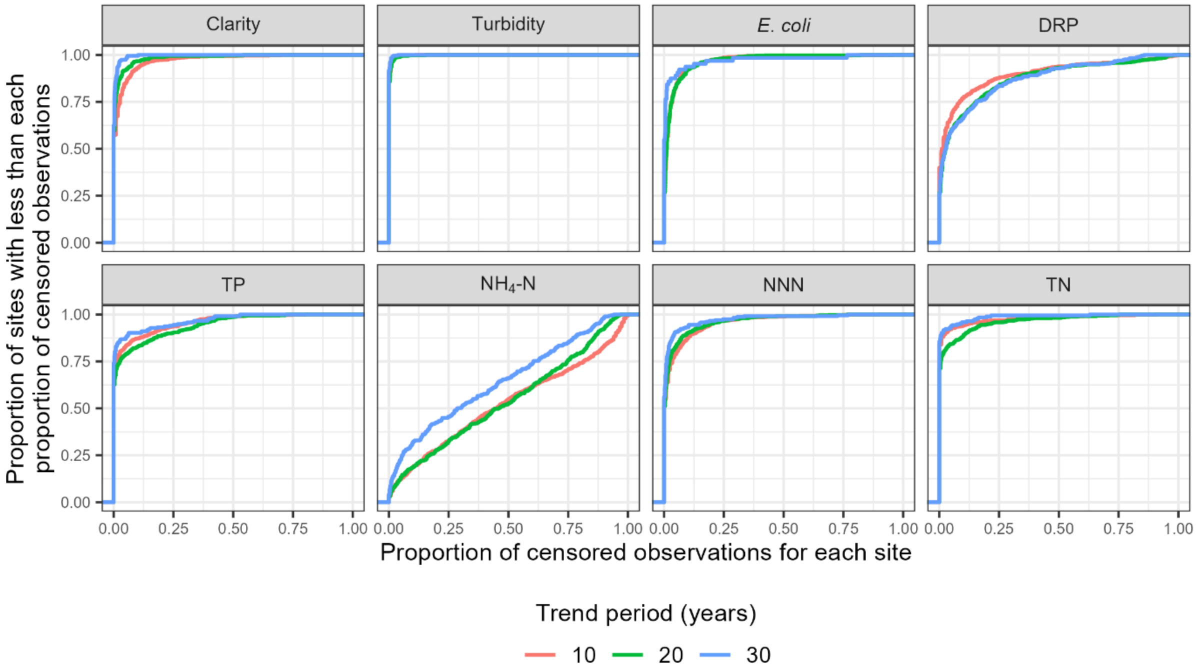

Figure 5-2: Proportion of censored data for each site × variable combination for which trend was assessed. Note no MCI observations were censored.

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 2015 – 2024 across the four human modified land cover classes (Figure 5-3). All estimated trend rates are included in these plots, irrespective of the level of confidence in the assessment (as defined in Section 3.2.3). These plots indicate that differences in the proportional area of upstream human modified land cover 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 human modified land cover classes (Figure 4-4, Figure 4-5, Figure 4-6).

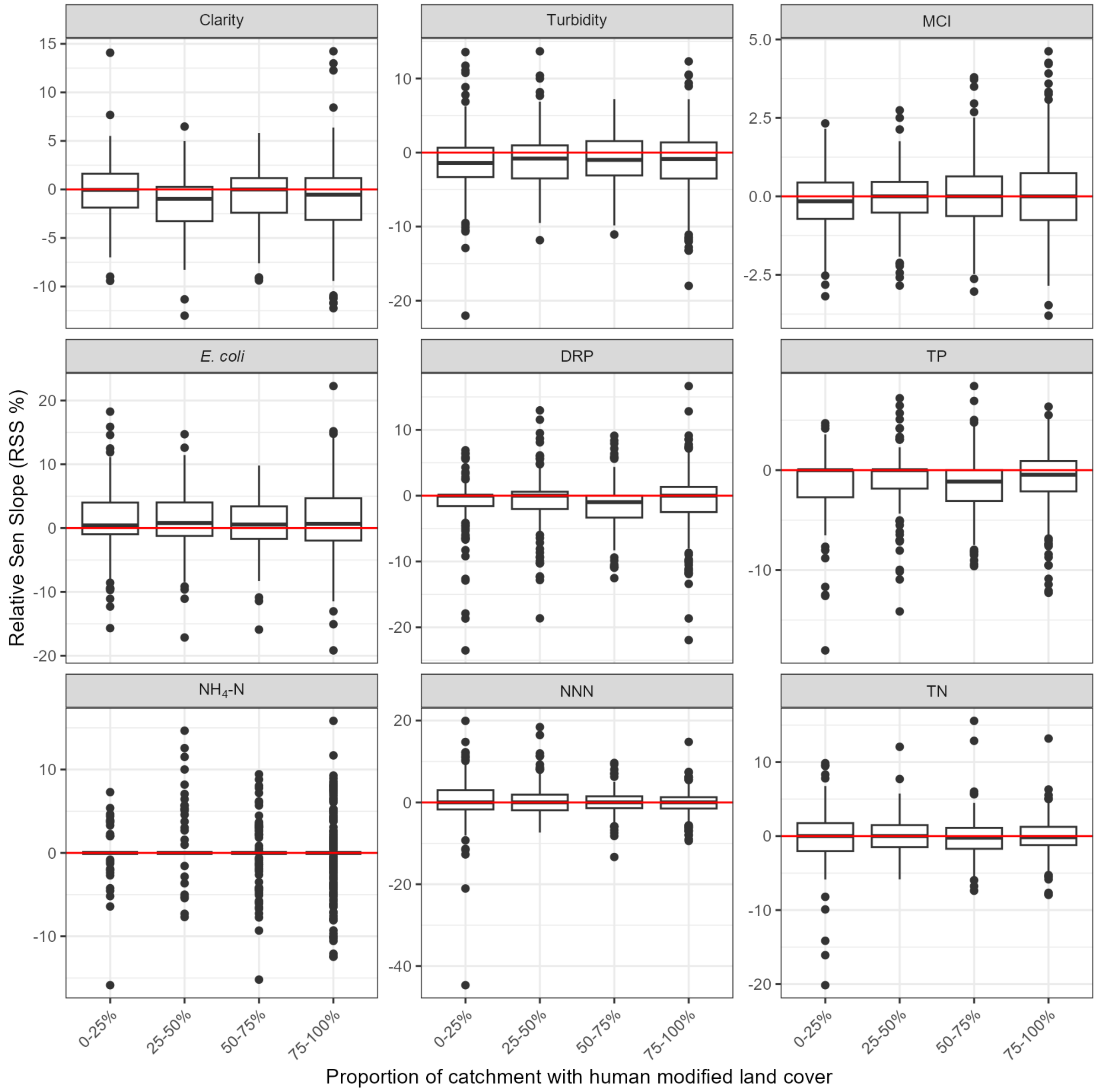

Figure 5-3: Summary of 10-year raw trend rates. Box-and-whisker plots show the distributions of site trend rates within human modified land cover classes. Black horizontal line in each box indicates the median of site trend rates and the box indicates the inter-quartile range (IQR). Whiskers extend from the box to the largest (or smallest) values no more than 1.5×IQR from the box. Data beyond the whiskers are shown as black circles.

5.1.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence of an improving 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 5-4.

Figure 5-4: Water quality monitoring sites categorised by the confidence that the 10-year trend is improving (C_i) for each variable. C_i 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 5-5 shows the proportions of sites belonging to each of the five categorical levels of confidence for C_i 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 improving trends (P_i) and their confidence intervals are summarised in Table 5-2. The 10-year P_i statistics ranged from 40–63%. *E. coli* had a majority (i.e., P_i <50%) of worsening trends at the 95% confidence level. Two of the variables had a majority of improving (i.e., P_i>50%) trends at the 95% confidence level (DRP and TP). The remaining variables had 95% confidence intervals for the P_i that included 50% (turbidity, clarity, MCI, NH_4-N, NNN and TN) and we cannot infer widespread improvement or worsening for these variables.

The 10-year P_i statistics and 95% confidence intervals for each water quality variable and land-cover class are shown in Figure 5-6. The majority of P_i statistics were below the 95% confidence level (confidence intervals spanning 50%). Exceptions were: worsening *E. coli* and MCI and improving turbidity and TP at sites in 0–25% human modified land cover class; worsening NH_4-N in the 25–50% human modified land cover class; worsening NH_4-N and improving DRP and TP in the 50–75% human modified land cover class.

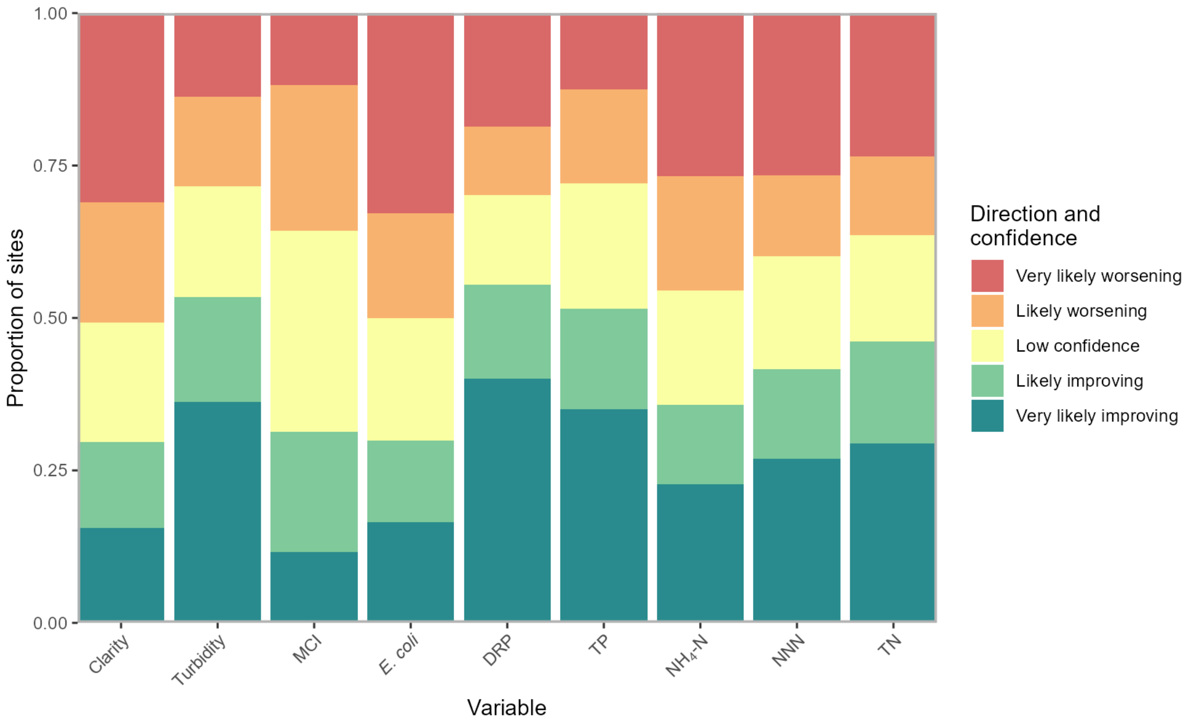

Figure 5-5: Summary plot representing the proportion of sites with improving 10-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with improving trends at levels of confidence defined in Table 3-2.

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

Variable	Number of sites	P_i (%)	95% confidence interval for P_i (%)
Clarity	624	39.6	19.7 – 59.5
Turbidity	814	61.9	49.4 – 74.4
MCI	834	46.5	39.9 - 53.0
E. coli	863	39.4	29.5 – 49.3
DRP	847	63.0	53.1 – 72.9
TP	826	62.6	52.4 – 72.8
NH_4-N	769	42.7	34.7 – 50.7
NNN	861	51.0	23.4 – 78.6
TN	826	55.3	33.9 – 76.6

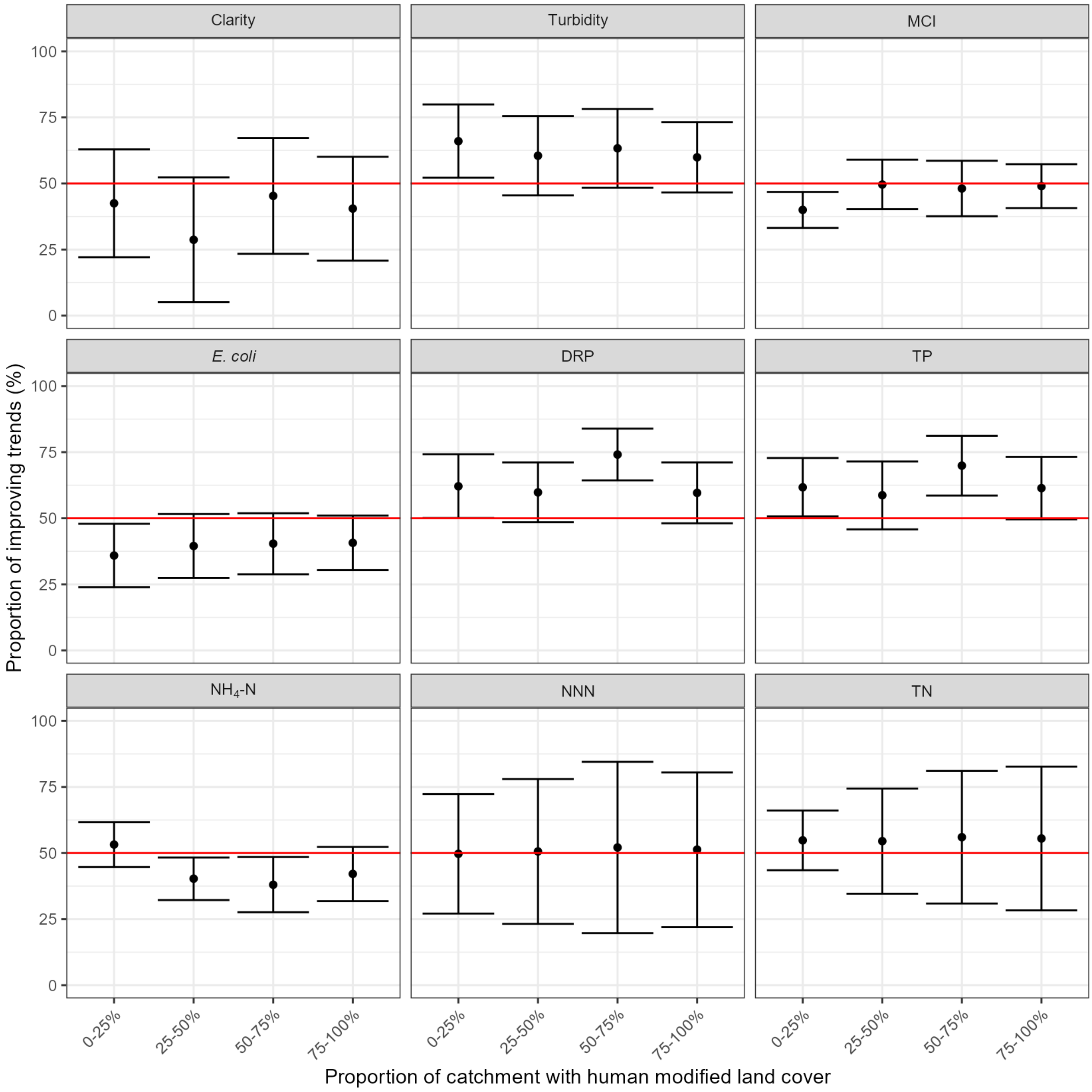

Figure 5-6: Proportions of improving trends (P_i) within human modified land cover classes for 10-year trends. Error bars are 95% confidence intervals.

5.2 Twenty-year trends (2005–2024)

Between 353 and 591 river monitoring sites met the filtering rules for the 20-year trend analysis of nutrients, *E. coli*, turbidity, MCI and clarity (Table 5-3). Site-coverage for 20-year trend analysis was reduced compared to 10-year trend analysis. However, the qualifying sites were reasonably well-distributed geographically (Figure 5-7). There was limited coverage of sites for clarity in the South Island. The proportion of observations that were censored for each variable was similar for the 20-year period compared to the 10-year period (Figure 5-2).

Table 5-3: Number of river monitoring sites by human modified land cover class and water quality variable included in the 20-year trend analyses of nutrients, *E. coli*, clarity, turbidity 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	0–25%	25–50%	50–75%	75–100%
Clarity	362	81	69	76	136
Turbidity	574	108	114	108	244
MCI	422	132	76	72	142
E. coli	544	119	113	102	210
DRP	569	111	111	109	238
TP	495	98	95	93	209
NH_4-N	591	117	116	112	246
NNN	506	101	93	94	218
TN	453	92	90	76	195

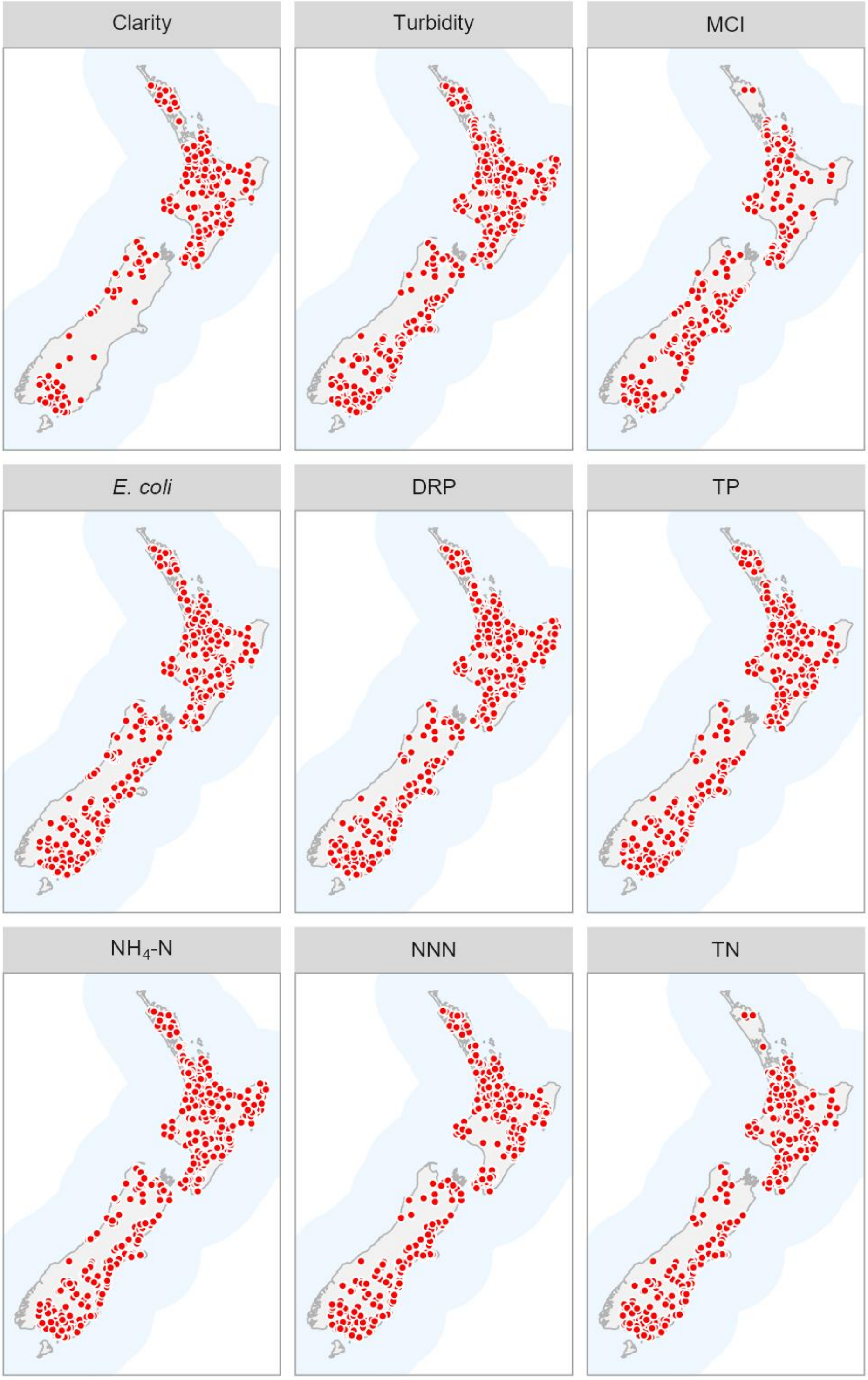

Figure 5-7: River water quality monitoring sites used for 20-year trend analyses of nutrients, *E. coli*, clarity, turbidity and MCI.

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 2004 – 2024 across the four human modified land cover classes (Figure 5-8). All estimated trend rates are included in these plots, irrespective of the level of confidence in the assessment (as defined in Section 3.2.3). These plots indicate that the proportional area of upstream human modified land cover did not account for a substantial amount of the variation in trend rates for any variable. However, there are some consistent relationships of greater improvement associated with increasing proportion of catchment human modified land cover for MCI, DRP, TP and NH_4-N. This contrasts with the state analyses of river variables, where water quality state clearly varied between human modified land cover classes (Figure 4-4, Figure 4-5, Figure 4-6).

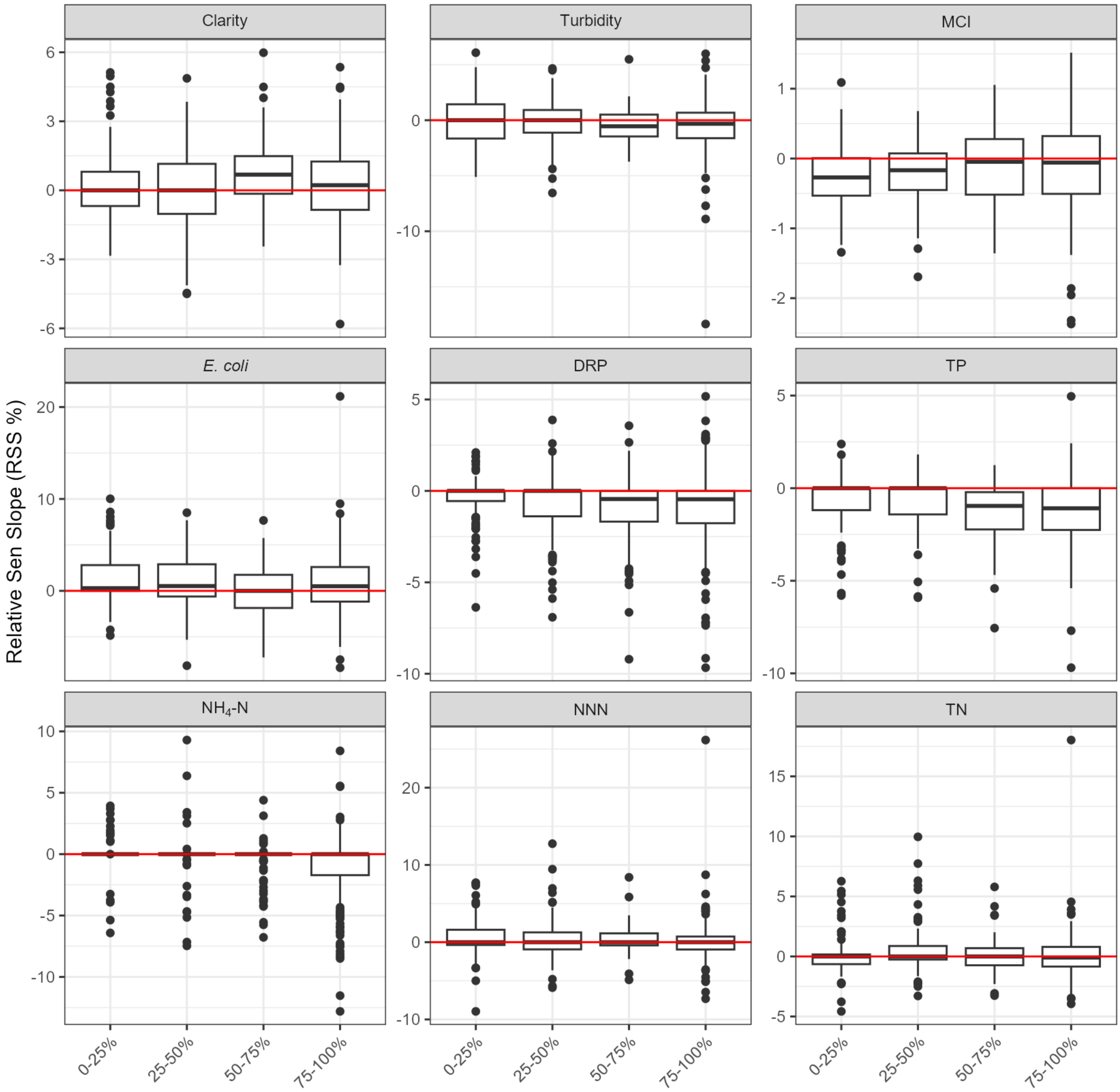

Figure 5-8: Summary of 20-year raw trend rates. Box-and-whisker plots show the distributions of site trend rates within human modified land cover classes. Black horizontal line in each box indicates the median of site trend rates, and the box indicates the inter-quartile range (IQR). Whiskers extend 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 black circles.

5.2.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence of an improving 20-year, raw trend for each site × variable combination. The spatial distributions of categorised individual sites are shown in Figure 5-9.

Figure 5-9: Water quality monitoring sites categorised by the confidence that the 20-year trend is improving (C_i) for each variable. C_i 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 5-10 shows the proportions of sites belonging to each of the five categorical levels of confidence for C_i 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 improving trends (P_i) and their confidence intervals are summarised in Table 5-4. The 20-year P_i statistics ranged from 41–81%. *E. coli* and MCI had a majority (i.e., P_i<50%) of worsening trends, at the 95% confidence level. Three of the variables had a majority of improving (i.e., P_i>50%) trends, at the 95% confidence level (DRP, NH_4-N, and TP). The remaining four variables had 95% confidence intervals for the P_i that included 50% (NNN, TN, clarity and turbidity), and we cannot infer widespread improvement or worsening for these variables.

The 20-year P_i statistics and 95% confidence intervals for each water quality variable and human modified land cover class are shown in Figure 5-11. The P_i statistics for all human modified land cover classes were below the 95% confidence level (confidence intervals spanning 50%) for clarity, NNN and TN. There are a majority of improving trends at the 95% confidence level for all land cover classes for DRP, TP and NH_4-N, and for the 50–75% land cover class for turbidity. There are a majority of worsening *E. coli* and MCI trends at the 95% confidence level for the 0–25% and 25–50% human modified land cover classes.

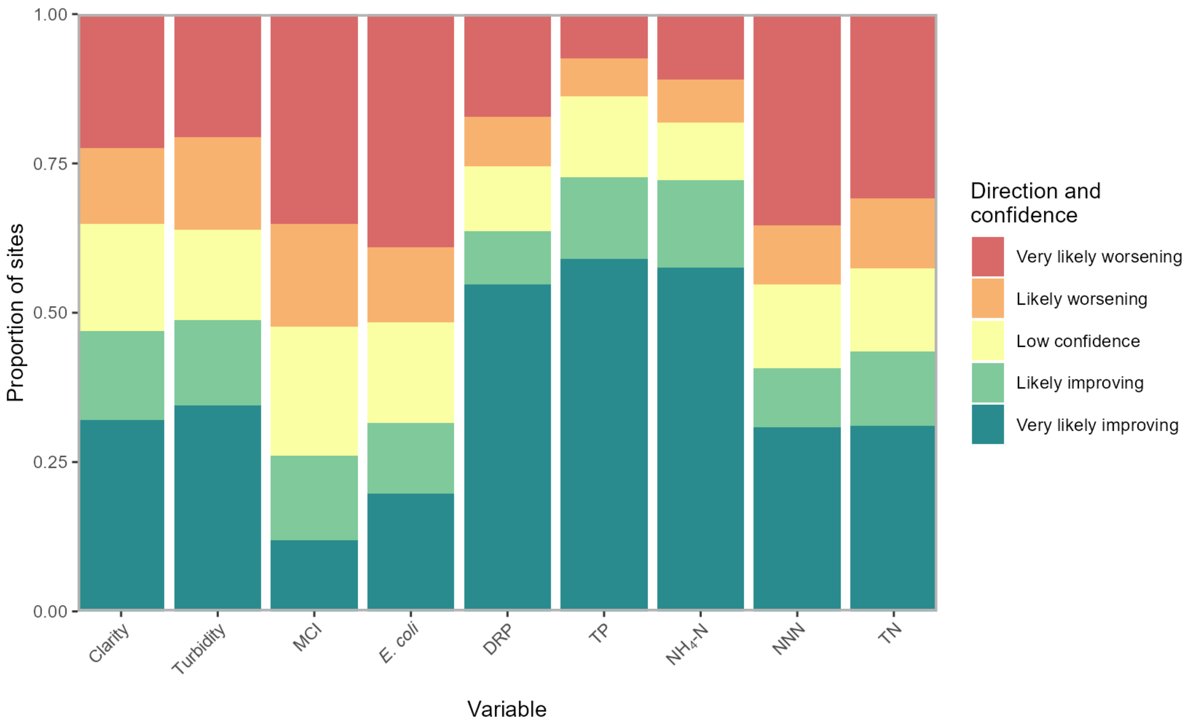

Figure 5-10: Summary plot representing the proportion of sites with improving 20-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with improving trends at levels of confidence defined in Table 3-2.

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

Variable	Number of sites	P_i (%)	95% confidence interval for P_i (%)
Clarity	362	57.5	38.8 – 76.1
Turbidity	574	57.0	46.4 – 67.6
MCI	422	36.1	27.7 - 44.6
E. coli	544	40.9	32.9 – 48.9
DRP	565	68.5	61.5 – 75.5
TP	495	81.2	74.5 – 87.9
NH_4-N	578	77.7	70.9 – 84.5
NNN	506	49.0	27.5 – 70.5
TN	453	51.5	33.3 – 69.8

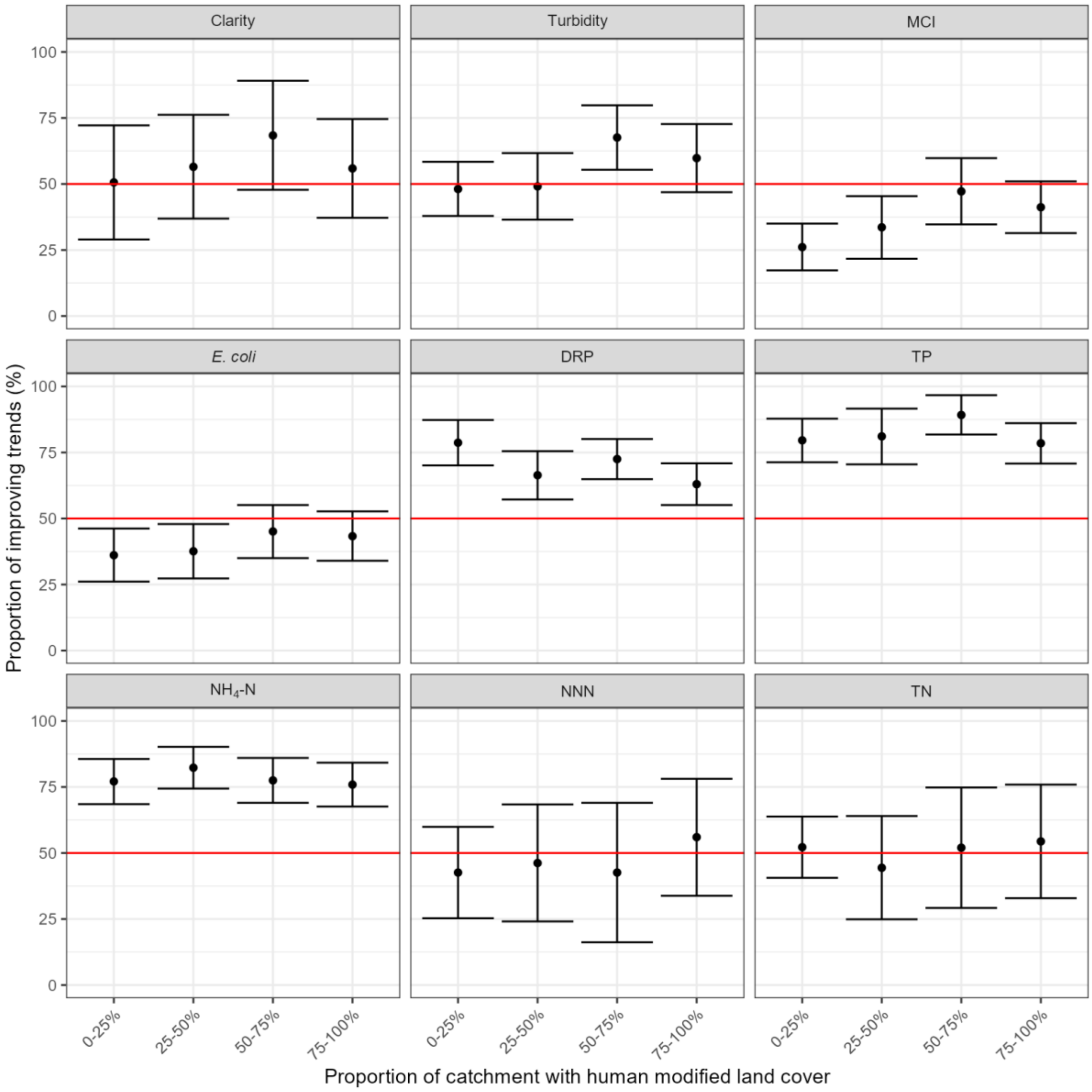

Figure 5-11: Proportions of improving trends (P_i) within human modified land cover classes for 20-year trends. Error bars are 95% confidence intervals.

5.3 Thirty-year trends (1995–2024)

Between 64 and 307 river monitoring sites met the filtering rules for the 30-year trend analysis of nutrients, *E. coli*, turbidity, MCI and clarity (Table 5-5). Site-coverage for 30-year trend analysis was considerably reduced compared to 10-year trend analysis. The qualifying sites were reasonably well-distributed geographically for some variables (Figure 5-12), although *E. coli* was largely limited to sites in Waikato. The proportion of observations that were censored for each variable was similar for the 30-year period compared to the 20-year period (Figure 5-2).

Table 5-5: Number of river monitoring sites by human modified land cover class and water quality variable included in the 30-year trend analyses of nutrients, *E. coli*, clarity, turbidity 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	0–25%	25–50%	50–75%	75–100%
Clarity	195	39	32	37	87
Turbidity	256	48	46	49	113
MCI	81	28	16	14	23
E. coli	64	12	12	18	22
DRP	303	56	52	54	141
TP	236	41	42	45	108
NH_4-N	307	56	52	56	143
NNN	236	42	38	46	110
TN	194	36	35	39	84

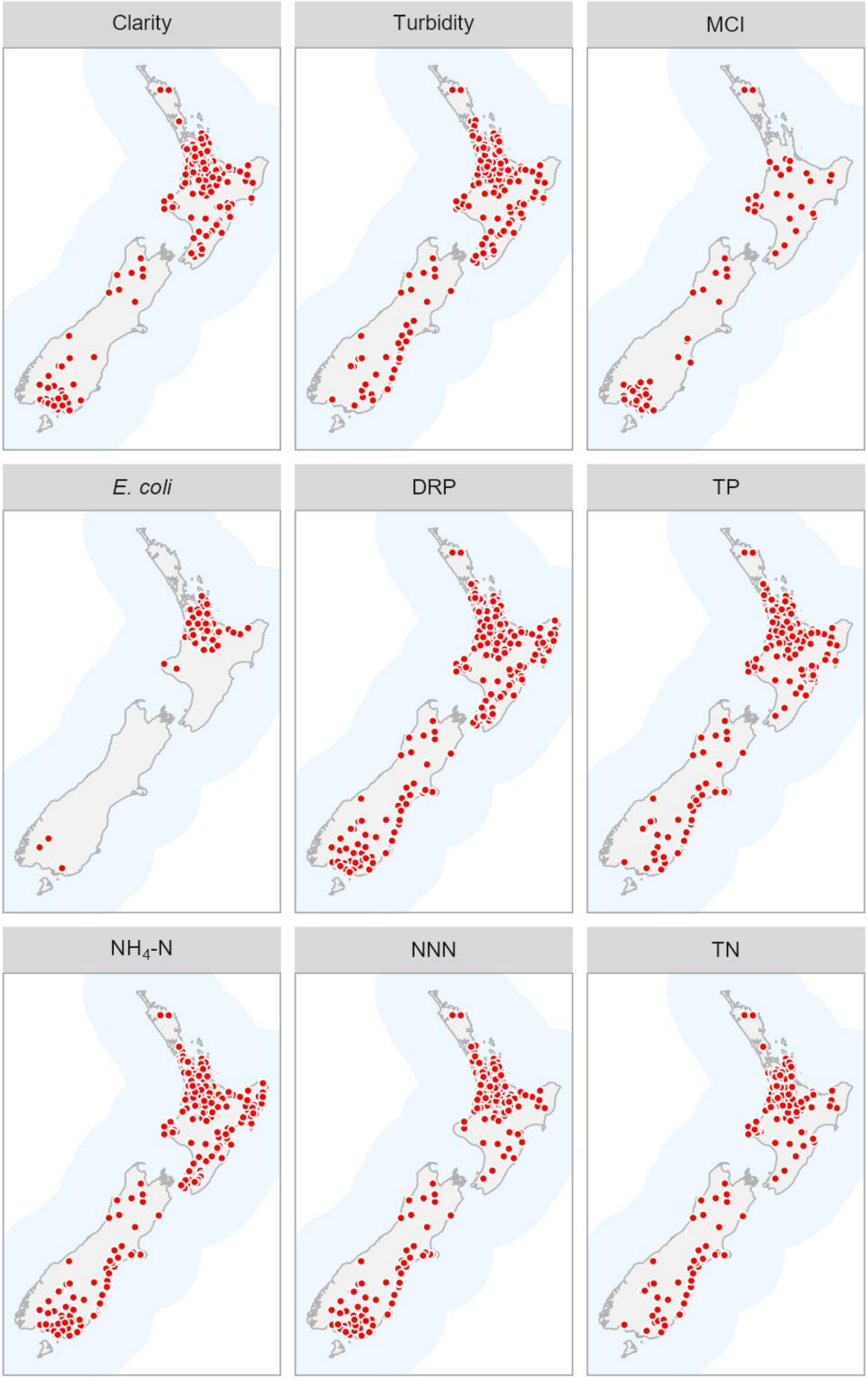

Figure 5-12: River water quality monitoring sites used for 30-year trend analyses of nutrients, *E. coli*, clarity, turbidity and MCI.

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 1995 – 2024 across the four human modified land cover classes (Figure 5-13). All estimated trend rates are included in these plots, irrespective of the level of confidence in these assessments (as defined in Section 3.2.3). These plots indicate relatively consistent patterns of greater predominance of worsening trends with greater upstream human modified land cover for *E. coli,* NNN and TN. In contrast, MCI, DRP, TP and NH_4-N demonstrate a greater predominance of improving trends with greater upstream human modified land cover.

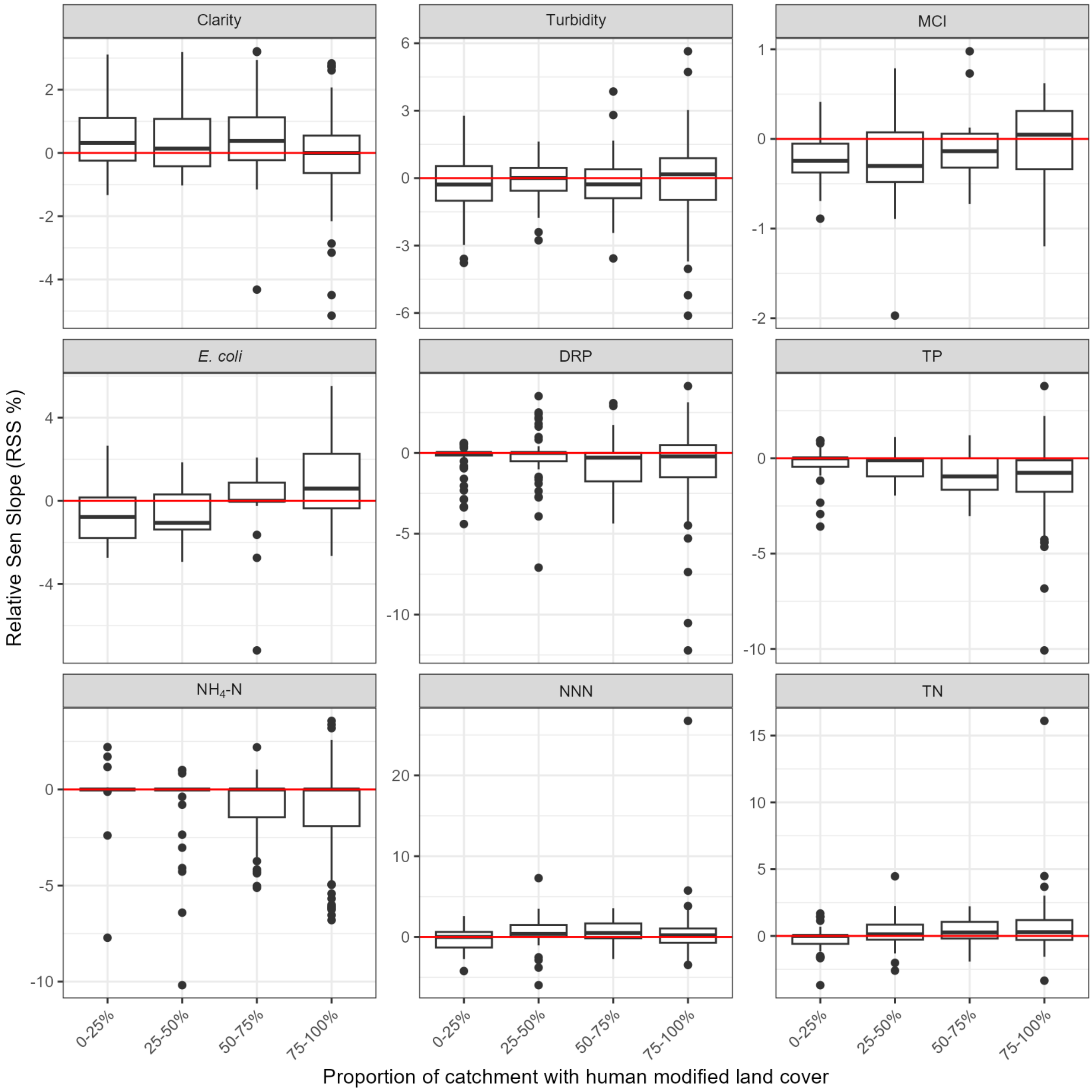

Figure 5-13: Summary of 30-year raw trend rates. Box-and-whisker plots show the distributions of site trend rates within human modified land cover classes. Black horizontal line in each box indicates the median of site trend rates, and the box indicates the inter-quartile range (IQR). Whiskers extend 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 black circles.

5.3.2 Trend direction

The levels of confidence listed in Table 3-2 were used to categorise the confidence of an improving 30-year, raw trend for each site × variable combination. The spatial distributions of categorised individual sites are shown in Figure 5-14.

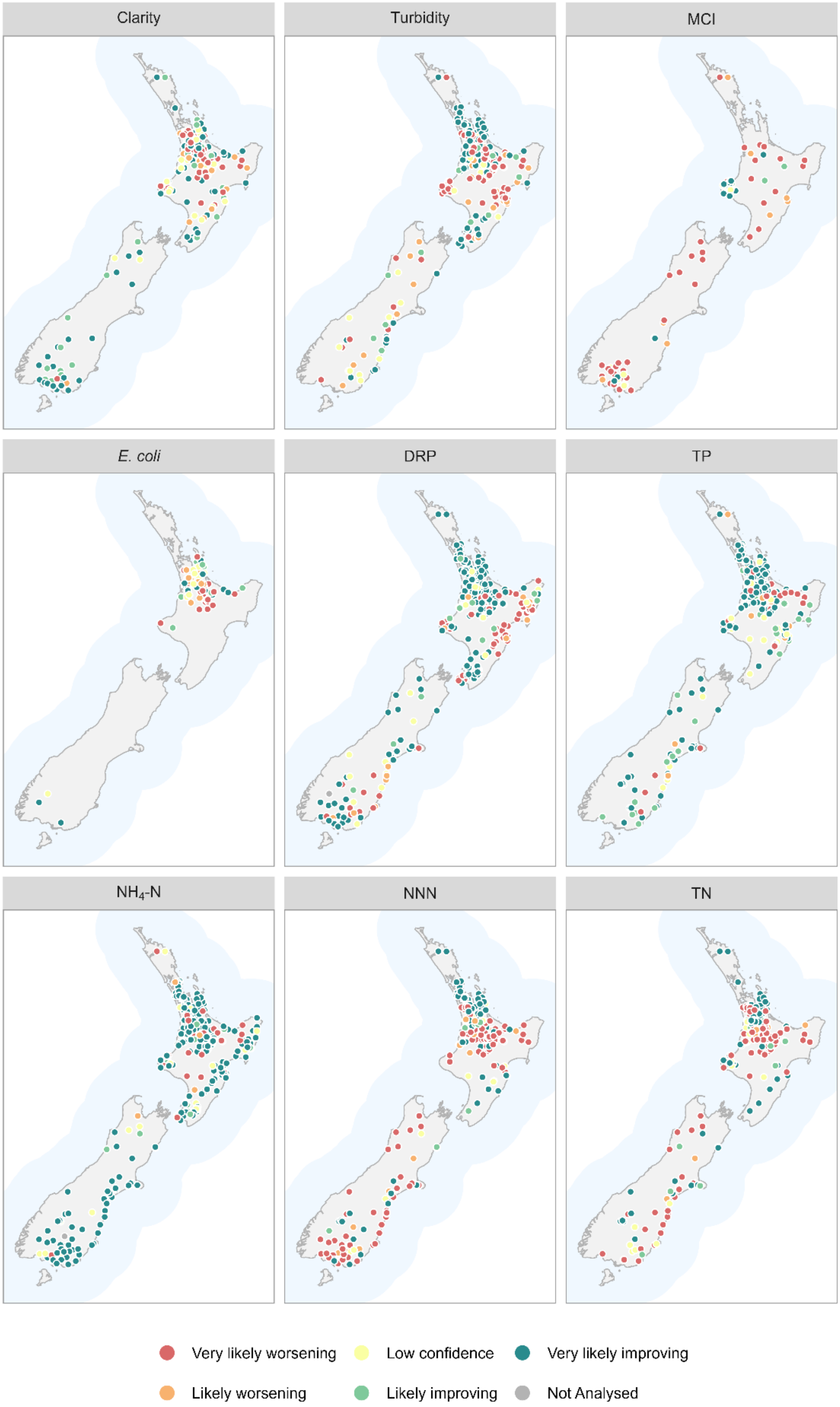

Figure 5-14: Water quality monitoring sites categorised by the confidence that the 30-year trend is improving (C_i) for each variable. C_i is expressed using the confidence categories Table 3-2. Only sites that met the sampling requirements outlined in Section 3.2.1 are shown in the figure.

5.3.3 Aggregate Trends

Figure 5-15 shows the proportions of sites belonging to each of the five categorical levels of confidence for C_i 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 improving trends (P_i) and their confidence intervals are summarised in Table 5-6. The 30-year P_i statistics ranged from 40–84%. MCI had a majority (i.e., P_i<50%) of worsening trends, at the 95% confidence level. Three of the variables had a majority of improving (i.e., P_i>50%) trends, at the 95% confidence level (DRP, TP, NH_4-N). The remaining five variables had 95% confidence intervals for the P_i that included 50% (clarity, turbidity, NNN, TN, *E. coli*), and we cannot infer widespread improvement or worsening for these variables.

The 30-year P_i statistics and 95% confidence intervals for each water quality variable and human modified land cover class are shown in Figure 5-15. These plots indicate relatively consistent patterns of greater predominance of worsening trends with greater upstream human modified land cover for *E. coli,* NNN and TN. In contrast, MCI and NH_4-N demonstrate a greater predominance of improving trends with greater upstream human modified land cover.

The interpretation of Figure 5-15, Table 5-6 and Figure 5-16 should take into account the site numbers described in Table 5-5, which shows that numbers of sites by human modified land cover class for the 30-year period were low for some classes × variable combinations.

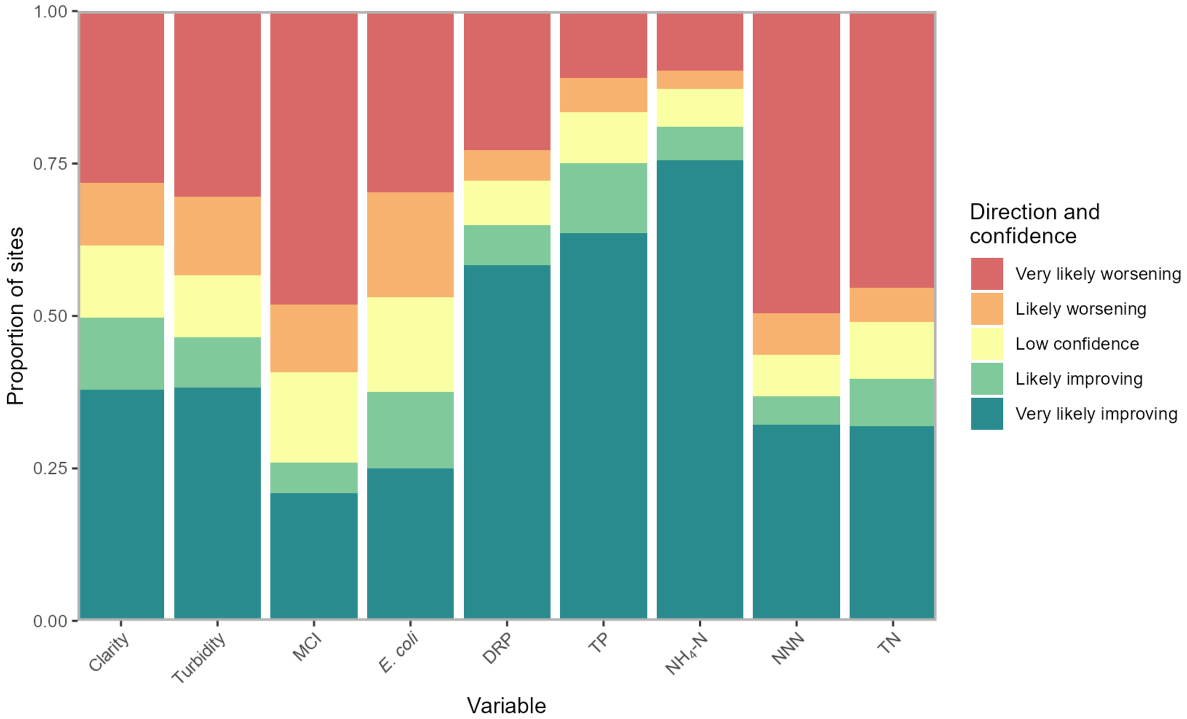

Figure 5-15: Summary plot representing the proportion of sites with improving 30-year time-period trends at each categorical level of confidence. The plot shows the proportion of sites with improving trends at levels of confidence defined in Table 3-2.

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

Variable	Number of sites	P_i (%)	95% confidence interval for P_i (%)
Clarity	195	55.4	38.7 – 72.1
Turbidity	256	52.3	43.9 – 60.8
MCI	81	33.3	25.3 - 41.4
E. coli	64	46.9	34.5 – 59.2
DRP	302	67.9	64.0 – 71.8
TP	236	79.7	74.1 – 85.2
NH_4-N	306	84.0	79.7 – 88.3
NNN	236	40.3	26.1– 54.4
TN	194	44.8	31.7– 58.0

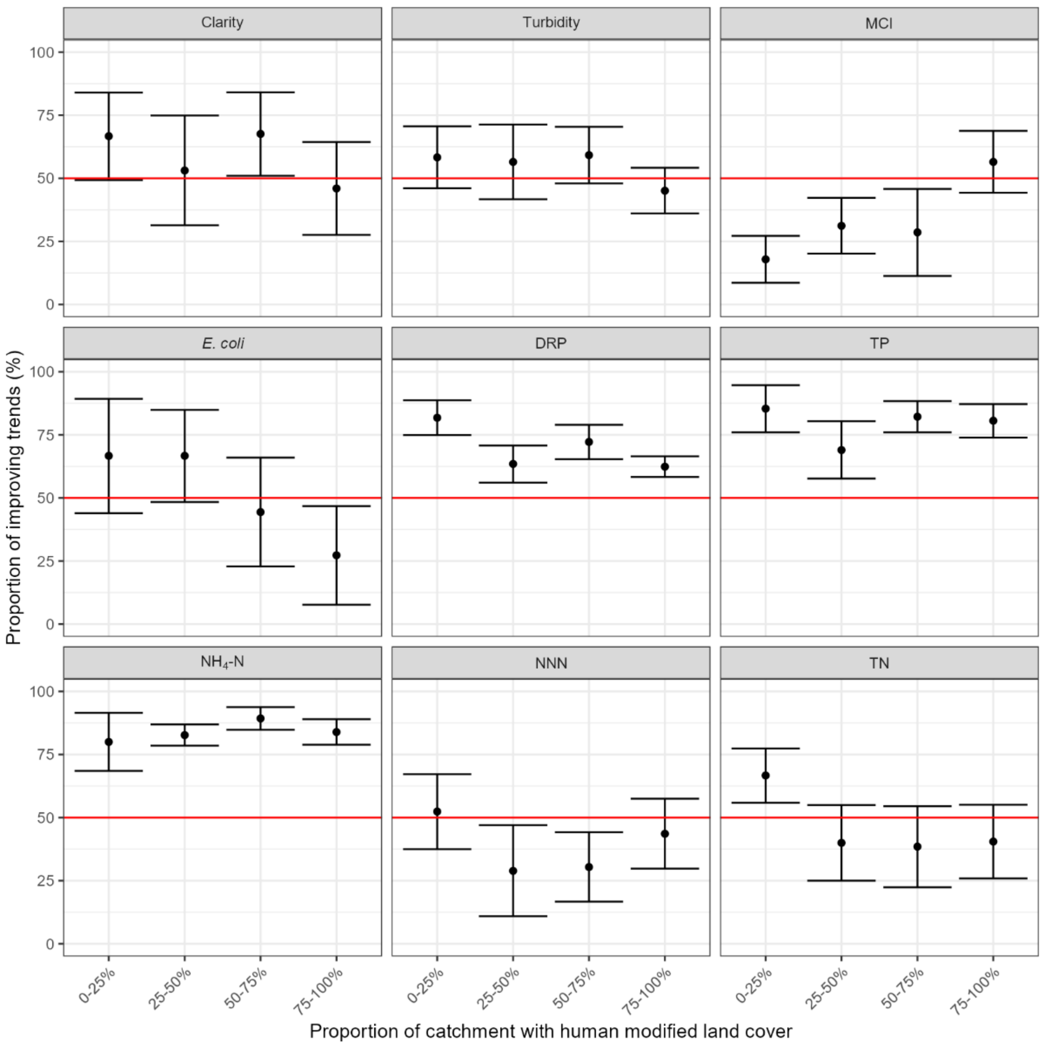

Figure 5-16: Proportions of improving trends (P_i) within human modified land cover classes for 30-year trends. Error bars are 95% confidence intervals.

5.4 Trend directions for 10-, 20- and 30-year periods

The national scale P_i 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 improving and worsening trends at the sites incorporated into the analysis of each period:

1. a predominance of worsening 30-year trends in TN, shifted to roughly equal proportions of improving and worsening 20-year trends and then to a predominance of improving 10-year trends;
2. a predominance of worsening 30-year trends in NNN, shifted to roughly equal proportions of improving and worsening 20-year and 10-year trends;
3. a predominance of improving 20- and 30-year trends in clarity and NH_4-N shifted to a predominance of worsening 10-year trends.

In contrast to the changes above, some trends have persisted between trend periods, including predominantly worsening trends in MCI and *E. coli* and improving trends in turbidity, DRP, and TP. However, it should be noted that there are differences in both the numbers and the positions of sites being aggregated into the 10-, 20-, and 30-year periods. This means that care should be taken when comparing between the columns of Table 5-7. For example, for *E. coli*, the 863 sites incorporated into the 10-year trend period were more evenly distributed across the country (Figure 5-4) whereas the 64 sites incorporated into the 30-year trend period were mostly located in the Waikato region (Figure 5-14).

Table 5-7: National-scale P_i statistics. Values are estimated percentages of river sites with improving trends across New Zealand. Small grey text indicates number of sites.

	10-year trend (2015-2024)		20-year trend (2005-2024)		30-year trend (1995-2024)	
	P_i	sites	P_i	sites	P_i	sites
Clarity	39.6	624	57.5	362	55.4	195
Turbidity	61.9	814	57	574	52.3	256
MCI	46.5	834	36.1	422	33.3	81
E. coli	39.4	863	40.9	544	46.9	64
DRP	63	847	68.5	565	67.9	302
TP	62.6	826	81.2	495	79.7	236
NH_4-N	42.7	769	77.7	578	84	306
NNN	51	861	49	506	40.3	236
TN	55.3	826	51.5	453	44.8	194

Confidence and direction
Very likely worsening
Likely worsening
Low confidence
Likely improving
Very likely improving

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 4 to 82 for the first time window, and from 617 to 852 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 5-17. The number of time windows for each site × variable combination that complied with the filtering rules are mapped in Figure 5-18. The proportion of observations that were censored remained broadly similar over time for some site x variable combinations. However, the effect of changes in detection limits as well as changes in sites through time was evident for DP, DRP, NNN, to a lesser extent for TN, and to a greater extent for NH_4-N (Figure 5-19).

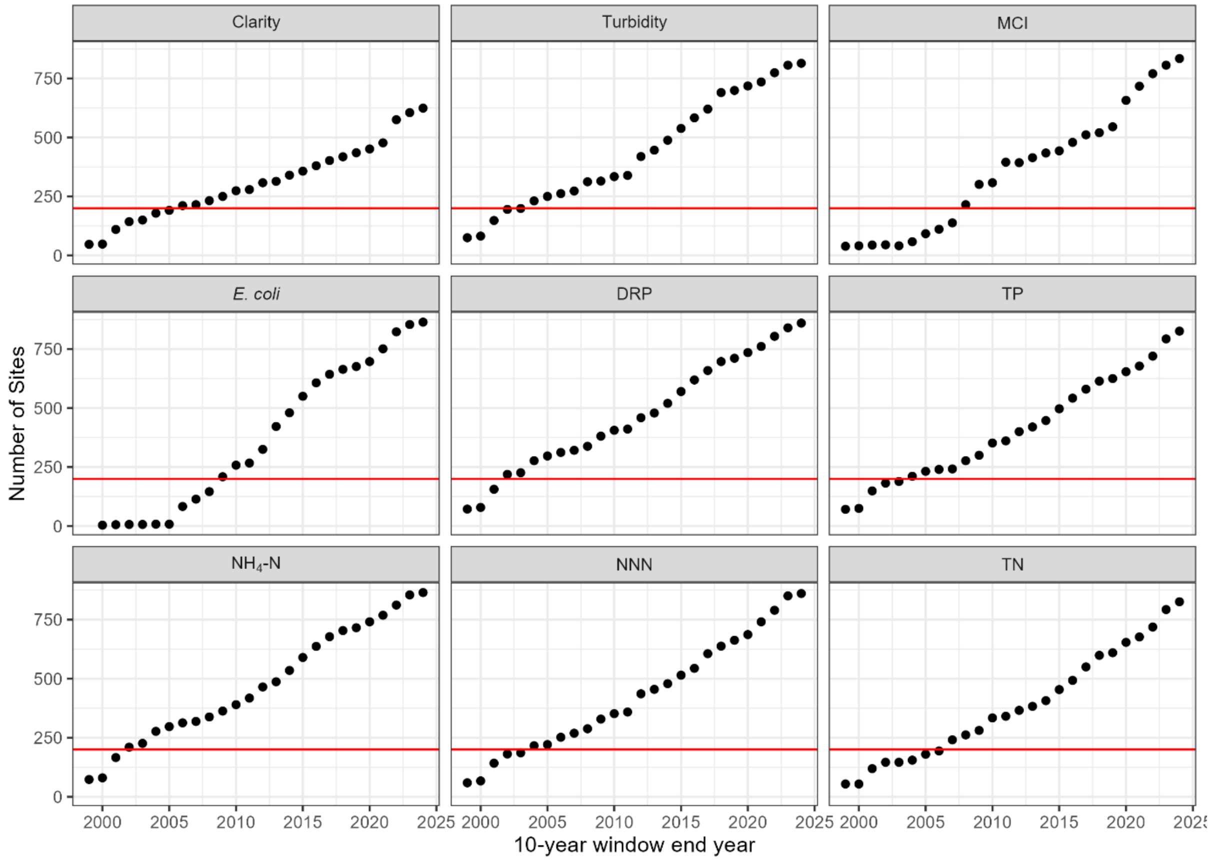

Figure 5-17: Number of sites that were included for each 10-year window included in the rolling trends analysis by variable.

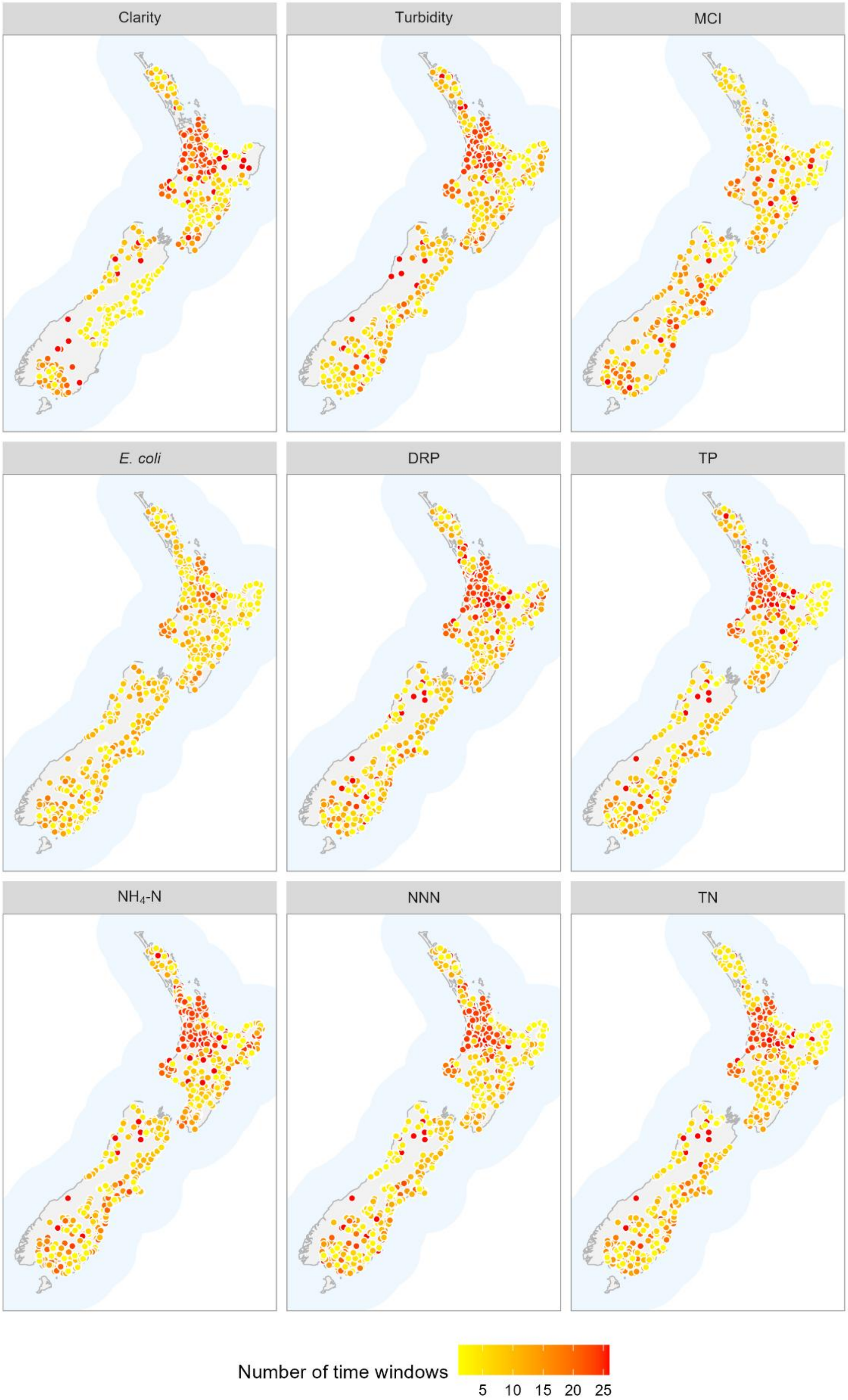

Figure 5-18: Number time windows that each river water quality monitoring site was included the 10-year rolling trend analyses of nutrients, *E. coli*, clarity, turbidity and MCI.

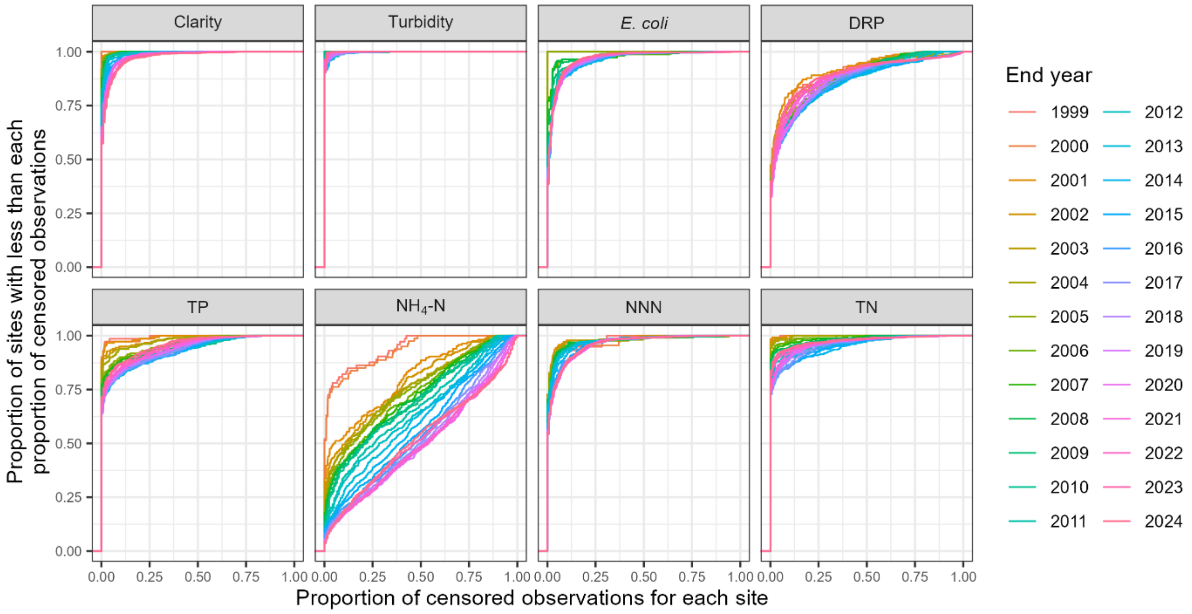

Figure 5-19: Proportion of censored data for each site × variable combination for which 10-year rolling trends were assessed. Note no MCI observations were censored.

5.5.1 Trend rate

Box and whisker plots summarising the trends assessed for each 10-year time window and each water quality variable are provided in Figure 5-20. All assessed trends are included in these plots provided that there were at least 200 sites within the 10-year time window, irrespective of the level of confidence in the assessment (see Section 3.2.3). The plots show transient patterns in in median RSS for all variables.

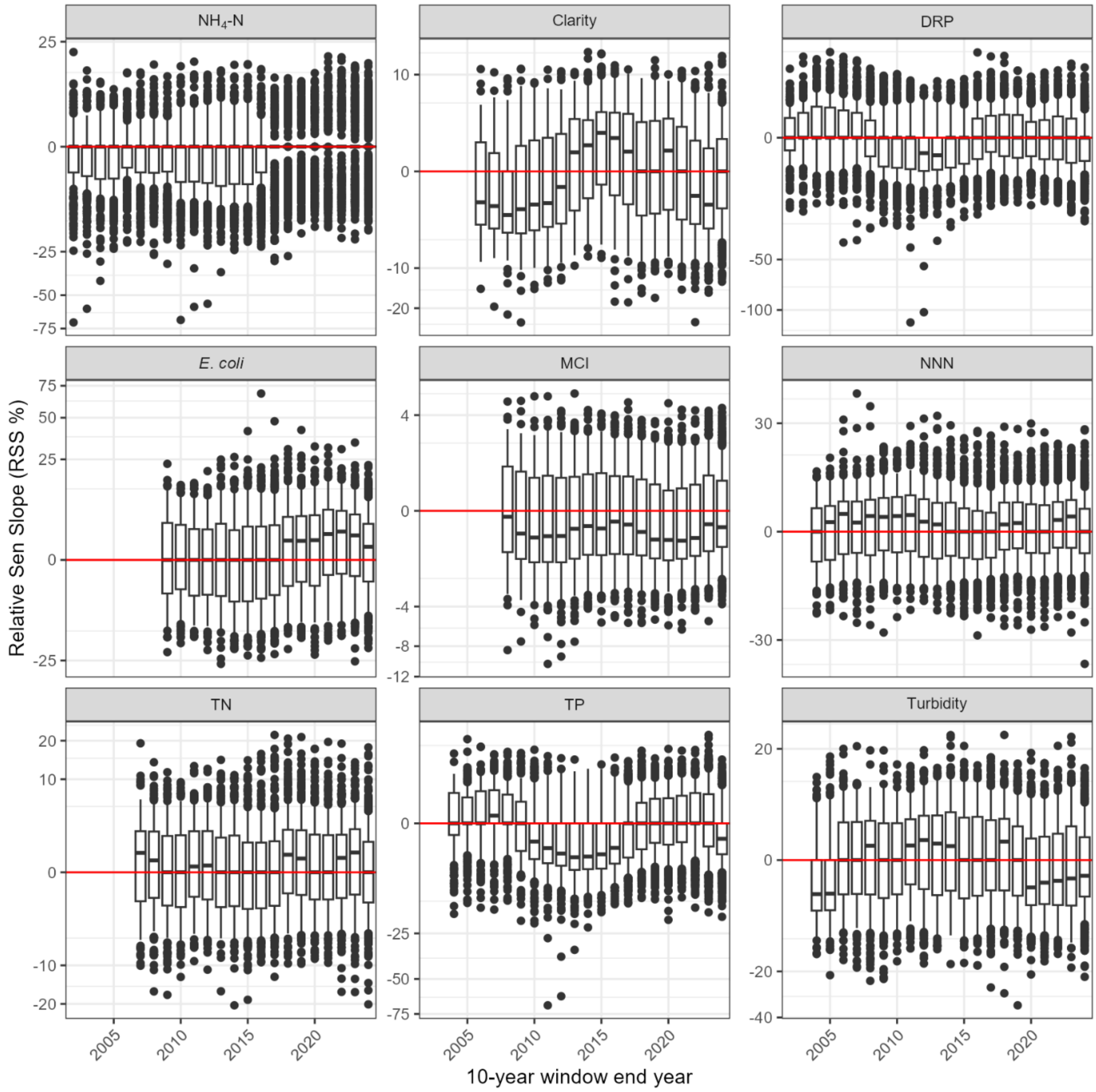

Figure 5-20: Summary of raw trend rates for rolling 10-year windows. Box-and-whisker plots show the distributions of relative trend rates (i.e., Sen slopes) within each ten-year window. Black horizontal line in each box indicates the median of site trend rates, and the box indicates the inter-quartile 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 5-21 shows the proportions of site trends assigned to each of the five categorical levels of confidence for C_i 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 there were at least 200 sites within the 10-year time window.

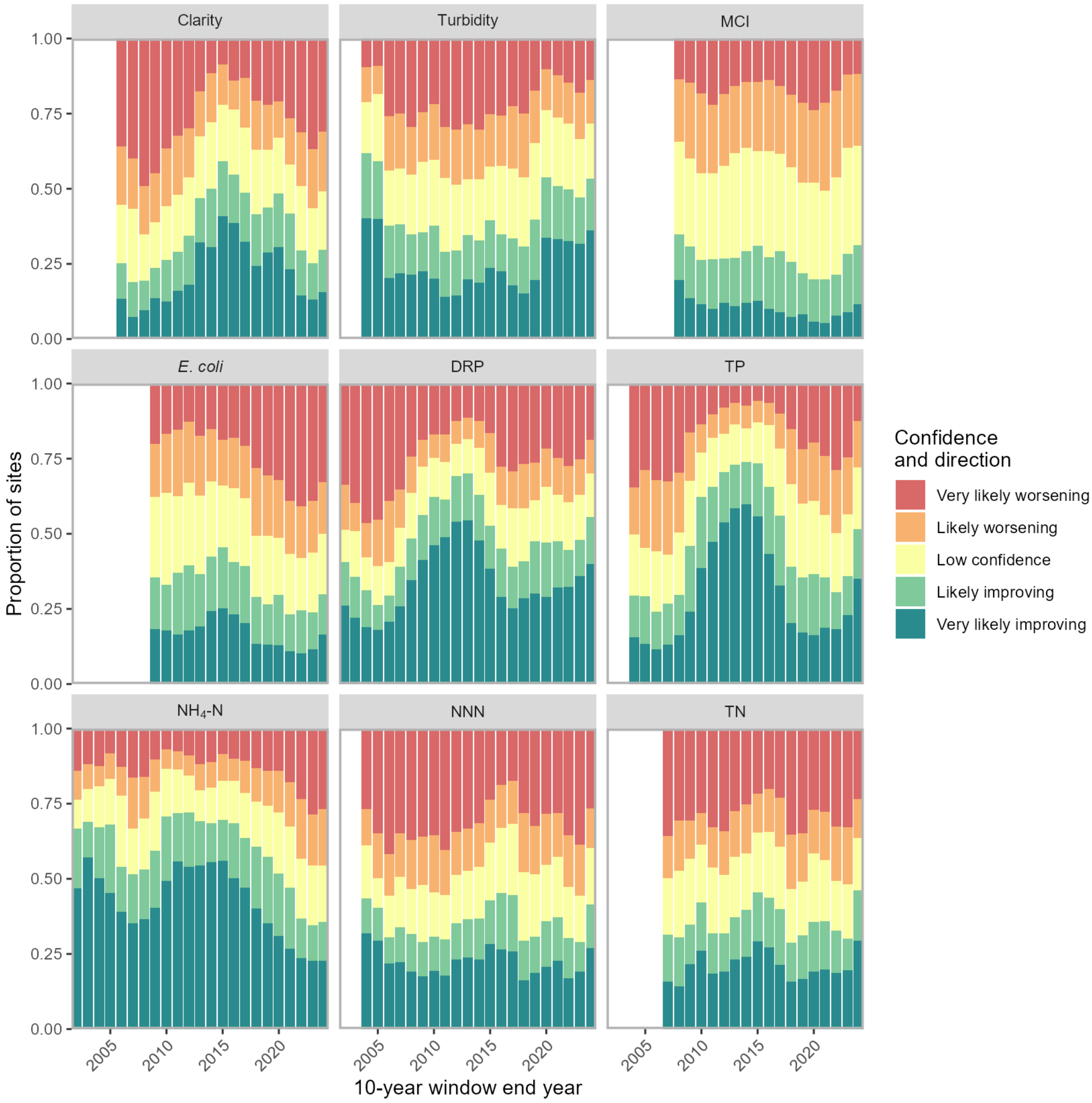

Figure 5-21: Summary plot representing the proportion of improving site trends at each categorical level of confidence for each of the rolling 10-year windows. The plot shows the proportion of sites with improving trends at levels of confidence defined in Table 3-2.

The national-scale proportions of improving trends (P_i) for each time window with more than 200 sites are summarised in Figure 5-22. No variables exhibited monotonic changes in the P_i score. There were quasi-periodic fluctuations in P_i that varied between variables. The magnitude of the fluctuations was greatest for TP (ranging from 34% to 82%) and smallest for MCI (ranging from 33% to 49%). MCI consistently had a majority of sites (i.e., >50%) that had worsening trends. Prior to 2021, NH_4-N demonstrated consistently improving trends, but over the last three time windows trends are predominantly worsening.

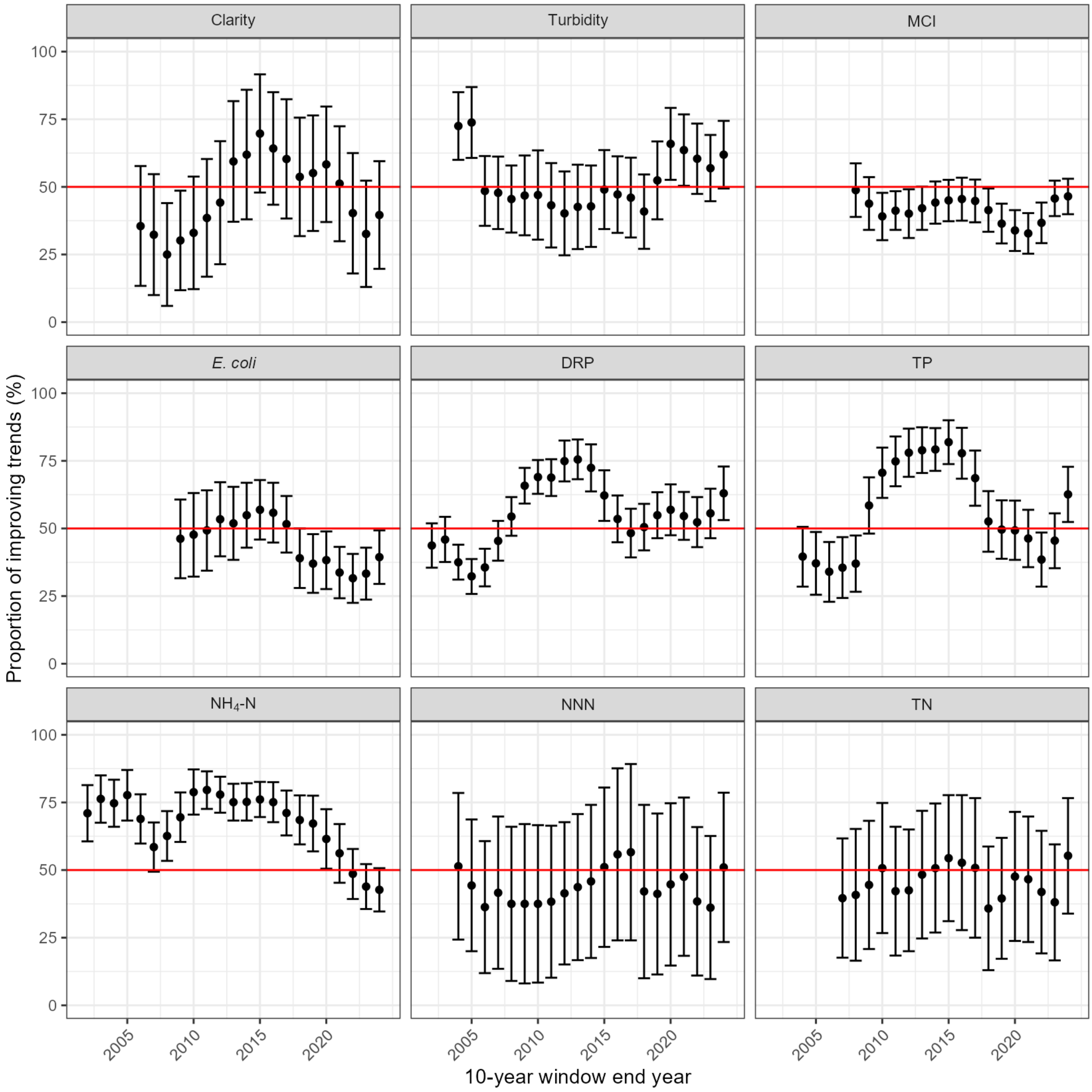

Figure 5-22: Summary plot representing the proportion of improving site trends (P_i) for each of the rolling 10-year windows. Error bars are 95% confidence intervals for P_i.

6 Discussion

The purpose of our analyses was to provide information required for national environmental reporting and to assist freshwater policy development. We overcame several challenges to fulfil this purpose by collating and analysing a large volume of data for several water quality variables collected at many sites over a long period by a diverse set of providers.

It is impossible for state and trend results to precisely represent true, but unknown, water quality conditions with unequivocal certainty because the observed data are samples of a complete population. Furthermore, the results of our analyses will be sensitive to various features of the water quality observation data, some methodological choices, and the specifics of the monitoring network that generated the data.

We discuss factors to consider when interpreting results in Section 6.1 and describe some of the measures we applied to mitigate or better understand their effects, but emphasise that users of our results should be cognisant of these factors when interpreting results.

Several additional analyses and interpretive approaches could further enhance the utility of the information presented in this report. We acknowledge some of these in Section 6.2 but emphasise that these were outside the scope of the current work.

6.1 Factors to consider when interpreting results

Operational, methodological, and strategic aspects relating to the water quality data and analysis techniques have the potential to influence our results. In this section we outline examples of how these three aspects can be influential and their possible consequences for results of our state and trend analyses. First, we highlight some technical methods applied to mitigate the potential influence of operational factors on state and trend results. Second, we clarify why methodological choices of the analysis can influence state and trend results and describe how our rolling trend analyses can aid interpretation of results. Third, we describe how strategic factors can influence state and trend results but were not investigated in this report. The degree to which results are sensitive to the various influencing factors will depend on water quality variable, site, time period, and type of analysis.

6.1.1 Operational

Features within water quality observation data that can arise at the operational level for a site × variable combination and have the potential to influence our results, include the following.

- The proportion and position of missing data.
- Uneven observation frequency.
- The proportion of censored data.
- Observation data precision.
- Changes in detection limits within the assessment period.
- Changes in data precision within the assessment period.

Our analysis applied transparent assumptions and robust technical procedures designed to maximise the degree to which results for a site × variable combination reflect true, but unknown, water quality conditions while accounting for the features listed above. When producing results for each site × variable combination, our approach applied the following technical procedures designed to mitigate the effects of (sometimes time-varying) features that can be contained within observation data.

- The irregular or systematic changes in sampling frequency were robustly handled by downsampling before calculating state and as part of the trend assessment.
- The censored data were robustly handled when calculating state by applying imputation based on the distribution of uncensored data.
- The censored data were robustly handled when calculating both trend rates and confidence in trend direction.
- The effect of seasonality was detected and accounted for by the trend analyses.

The proportion of censored data was relatively low at most sites for most variables except for NH_4-N which contained considerably more censored data at more sites across all time periods compared to other variables (Figure 5-2). Therefore, median values were not influenced by data censoring for most variables at most sites. However, our trend analyses produced trend rates of zero for NH_4-N for many sites (i.e., ≥50%) due to the high proportion of censored data for NH_4-N (Figure 5-3). A trend rate of zero indicates that any underlying trend is too small to be reliably detected given the precision of the measured observations. We note that results showing trend rates of zero are possible for any site × variable combination with a high proportion of censored values.

6.1.2 Methodological

Methodological choices required for water quality state and trend analysis that have the potential to influence our results include selection of analysis periods and rules for data inclusion. Estimated state is influenced by length of assessment period because of differences in sample size and an implicit assumption of stationarity of the population within an assessment period. Trend direction is specific to the trend assessment period for which it is calculated. Trend direction can change between adjacent periods as demonstrated by our analysis of rolling trends.

Our analysis applied a 5-year assessment period for state and 10-, 20-, and 30-year assessment periods for trend across all variables to maintain consistency with previous SoE reporting conducted for MfE and Stats NZ. We applied rules for inclusion within these analysis periods to ensure adequate sample sizes and seasonal coverage (where applicable) for each analysis.

We compared trends across rolling 10-year windows to aid interpretation of results for 10-year trends ending 2024. For example, transient decreases and increases in proportions of improving trends within rolling aggregate trends indicate that caution must be applied when interpreting causes and permanence of calculated trends in water quality.

When interpreting rolling trends, it should be acknowledged that there were more sites from rivers available in more recent windows as indicated by counts of sites in different windows of the rolling trend analysis. Analysis of the proportion of censored data at each site also indicated that this was not constant across the rolling windows for some variables. Any further analysis of the drivers of the long-term patterns should be cognisant of these features of the water quality data.

6.1.3 Strategic

Strategic factors relating to the water quality SoE monitoring network that have the potential to influence our results include the following.

- Definition of SoE monitoring being applied by data providers.
- Application of QC codes by data providers, including retrospective application of QC codes to historic data.
- The number of sites.
- Site locations being over- or under-representative of a type of river or catchment (e.g., lowland rivers, larger rivers, high proportion of upstream human modified land cover) compared to other types (e.g., mountain rivers, smaller rivers, low proportion of upstream human modified land cover).
- Duration of monitoring at each site.
- Closing or opening of sites through time.
- Sampling frequency (e.g., monthly versus bi-monthly versus quarterly versus annually).

We provided the fullest possible picture of water quality state and trends by incorporating all available SoE data that passed rules for inclusion into each analysis. This approach meant that many sites were available for calculation of 5-year states and 10-year trends for each variable for the period ending 2024. However, fewer sites were available for calculation of longer trend assessment periods. Thus, aggregate trends calculated for each variable for 10-, 20-, and 30-year trends were not commensurate because they were not calculated for the same set of sites. Arguably, comparison between aggregate trends for different periods (10-, 20-, and 30-year periods) should use a consistent set of sites in each period. Maps and counts of sites in different periods indicated that this was not the case because more sites were available in more recent times and longer running sites were biased towards regions that have maintained consistent monitoring procedures. Furthermore, maps of sites included in the different trend periods indicated that the longer the trend assessment period the less well spread, and therefore less nationally representative sites included in any analysis became.

For most variables, sites for which state (for the period ending 2024) was calculated were well spread across a gradient of human modified land cover and well spread geographically. An exception to this was water clarity, for which data was not available or did not pass rules for inclusion for three regions for 5-year states and two regions for 10-year trends. Furthermore, there was an absence of water quality sites (i.e., monitoring for any variable) in some locations that are dominated by natural land cover such as the spine of the Southern Alps and Fiordland.

6.2 Additional analyses and interpretive approaches not applied in this report

The following additional analyses would assist interpretation of our results but were outside the scope of this report.

- We used predefined NOF thresholds to grade water quality state at each site with respect to some of the analysed water quality variables. We did not compare NOF thresholds with detection limits in the observed data. It is possible that in some cases, detection limits preclude the accurate assignment of sites to NOF grades. However, we note that generally NOF thresholds are considerably larger than most detection limits (e.g., those typically applied for clarity, NNN, and *E. coli*) therefore, NOF grading should not be influenced by data censoring. For example, in the 5-year state assessment period ending 2024, for DRP the largest detection limit was 0.005 and the NOF A-band threshold is 0.006, and for NH_4-N, the largest detection limit was 0.01 and the NOF A-band threshold is 0.03.

- We relied on data providers to interpret our data request consistently (e.g., we accepted data providers interpretation of monitoring that was carried out for SoE purposes). We did not assess the consistency with which the definition of SoE monitoring was interpreted between data providers. Differences in this definition could influence patterns of water quality state that are inferred from the results of the state analysis.

- The monitoring networks operated by each data provider was outside of our control. We did not conduct any analyses of the representativeness of the SoE monitoring networks with respect to any target population (e.g., all river water quality within a specified period whether defined by catchments, river segments, environmental classes etc).

- We did not make comparisons with any other nationwide analysis of water quality including past analyses carried out for MfE and Stats NZ for environmental reporting purposes or those associated with LAWA. We note that methodological differences between the current study and previous studies should be controlled for such a comparison to be commensurate.

- We applied linear regressions between site-medians and a gradient in human modified land cover. This approach had the advantage of being simple and transparent, but does not account for other factors that influence water quality such as geology, slope, altitude, rainfall, or type of human modification.

- Trends can be caused by various drivers including climate variability, high disturbance events such as storms/cyclones or earthquakes and changes in land use and land management. As discussed above, trend results can also be influenced by operational factors (e.g., data completion, data precision, and detection limits), methodological choices (e.g., analysis duration, algorithm), and strategic factors (e.g., site representativeness). We did not undertake any analyses to disentangle these potential drivers of water quality trend, or attribute trend results to these drivers. We note that there is evidence that river water quality is partly associated with the ENSO climate cycle (Scarsbrook et al. 2003; Snelder et al. 2021b,c; Snelder and Fraser 2025).

- We used all available data to calculate rolling aggregate trends. We did not control for changes in data availability between 10-year periods.

6.3 Differences to previous studies

Methods used in the current report are broadly comparable with those used in the previous analyses carried out for MfE and Stats NZ by Whitehead et al. (2022). Differences between this and the previous study include differences in raw data availability, availability of measurement method information, availability of quality control (QC) information, improved methods for assigning observations to sampling periods, minor updates to statistical methods, and omission of flow adjusted trends. Changes in trend assessment methodology and terminology used in the report were largely made to align the reporting with recently published guidance (Helsel et al. 2020; Snelder et al. 2021a; Fraser and Snelder 2021; Snelder et al. 2022). Differences between this report and Whitehead et al. (2022) are summarised below.

- As agreed with MfE and Stats NZ, five trend confidence and direction categories were used in this report instead of the nine previously applied.
- As agreed with MfE and Stats NZ, in this report we use terminology of “improving and worsening” and “proportion improving, P_i” rather than “increasing and decreasing” and “proportion decreasing, P_d”.
- As agreed with MfE and Stats NZ, we applied linear regressions between site-medians and a gradient in human modified land cover rather than high-intensity agricultural land cover, which was a method previously applied by Whitehead et al. 2022). Preliminary analysis (results not shown) and comparison of our results with those of Whitehead et al. (2022) indicated that regressions against human modified land cover explained more variation than regressions against specific types of land cover such high-intensity agricultural land cover. Human modified land cover was used to demonstrate how water quality state can be related to various independent variables. Data accompanying this report allow similar analysis using various independent variables depending on user needs.
- For this report we shifted observation date metadata to an adjacent sample interval when a monthly monitoring regime has two observations in a month and a missing observation in the previous/following month.
- We used an updated method to account for spatial correlation in the assessed trend directions at individual sites when calculating P_i and its variance.
- We provide an overall grade for attributes with more than one numeric attribute state.
- This report uses “downsampling”, which involved selecting the observation closest to the middle of each sample interval in the state time period, rather than calculating state from all available observations.

- For trend assessment, season was here defined separately from the time increment of the assessment, rather than the season assigned to a site × variable combination being limited to the sampling interval.
- When calculating the Kendall S statistic for this report, pairs of observations that were tied in time were assigned differences of zero.
- When calculating trend rates for this report, inter-observation slopes were treated as ties and set to zero depending on the detection level and observation value within each pair, rather than applying a high censor filter.
- Flow adjusted trends were not calculated for this report, whereas the previous report provided results for flow adjusted trends but did not show or discuss those results.

6.4 Overview

In this report we present straightforward summaries of state and trend results using tables and plots. Our results indicate patterns in space and time. For example, there was an overall tendency for water quality to be worse in sites with higher human modified land cover compared to sites with lower human modified land cover for all variables. However, our results also indicated complicated patterns in state and trend between water quality variables, across sites, and through time. This report is accompanied by detailed information for each river monitoring site, such that bespoke future analysis can aggregate results for sites and their water quality conditions in various ways to meet different information requirements (e.g., grouped by region or environmental class, distributed along environmental gradients, associated by combinations of upstream catchment characteristics).

7 Acknowledgements

Many thanks are due to the numerous staff across the various organisations who provided water quality data.

8 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.

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., Hughes, A.O., Vincent, A.G., Heubeck, S. (2021) Weak numerical comparability of ISO-7027-compliant nephelometers. Ramifications for turbidity measurement applications. *Hydrological Processes*, 35(12): 12. 10.1002/hyp.14399

Davies-Colley, R., McBride, G. (2016) Accounting for changes in method in long-term nutrient data: recommendations based on analysis of paired SoE data from Wellington rivers. *NIWA Client Report*, HAM2016-070: 34.

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.

Douglas, E.M., Vogel, R.M., Kroll, C.N. (2000) Trends in floods and low flows in the United States: impact of spatial correlation. *Journal of hydrology*, 240(1-2), 90-105.

Dudley, B.D., Burge, O., Plew, D., Zeldis, J. (2020). Effects of agricultural and urban land cover on New Zealand's estuarine water quality. *New Zealand Journal of Marine and Freshwater Research*, *54*(3), 372-392.

Fraser, C., Snelder, T. (2025) The LWP-Trends Library; V2502 March 2025. LWP Ltd Report.

Fraser, C., Snelder, T. (2021) Update to REC Land Cover categories and review of category membership rules (LWP Client Report 2021-18). LWP Ltd, Christchurch, New Zealand.

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

Gilbert, R.O. (1987) *Statistical methods for environmental pollution monitoring*. John Wiley & Sons.

Helsel, D.R. (2005) More than obvious: better methods for interpreting nondetect data. *Environmental science & technology*, 39, 419A-423A.

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

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.

National Environmental Monitoring Standards (2019) National Environmental Monitoring Standards Water Quality; Part 2 of 4: Sampling, Measuring, Processing and Archiving of Discrete River Water Quality Data. Version: 1.0.0. March 2019. https://www.nems.org.nz/documents/water-quality-part-2-rivers

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

Plew, D.R., Zeldis, J.R., Dudley, B.D., Whitehead, A.L., Stevens, L.M., Robertson, B.M., Robertson, B.P. (2020) Assessing the eutrophic susceptibility of New Zealand estuaries. *Estuaries and Coasts*, *43*(8), 2015-2033.

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Scarsbrook, M R., McBride, C.G., McBride, G.B., Bryers, G.G. (2003). Effects of Climate Variability on Rivers: Consequences for Long Term Water Quality Analysis. *JAWRA Journal of the American Water Resources Association*, 39 : 1435–47.

Snelder, T.H., Fraser, C. (2025) River Water Quality Trends and Climate Variation in the Canterbury Region. Including Implications for Assessing Policy Effectiveness. LWP Client Report Nos 2025–01. Christchurch, New Zealand: LWP Ltd.

Snelder, T.H., Fraser, C., Whitehead, A. (2022) *Continuous Measures of Confidence in Direction of Environmental Trends at Site and Other Spatial Scales*. Environmental Challenges 9: 100601.

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., Monaghan, R., De Malmanche, S. Whitehead, A.L. (2021c) Attribution of river water-quality trends to agricultural land use and climate variability in New Zealand. *Marine and Freshwater Research*, 73: 1-19.

Snelder, T.H., Fraser, C., Larned, S.T., Whitehead, A.L. (2021a) *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., Fraser, C., De Malmanche, S. (2021b) Effect of climate variability on water quality trends in New Zealand rivers. *Marine and Freshwater Research*, 73: 20-34.

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.

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.

Whitehead, A., Fraser, C., Snelder, T.H., White, R., Walter, K., Woodward, S., Zammit, C. (2022) Water quality state and trends in New Zealand rivers analyses of national rivers data ending in 2020, 2021296CH: 73.

Whitehead, A.L., Booker, D.J. (2019) Communicating biophysical conditions across New Zealand's rivers using an interactive webtool. New Zealand Journal of Marine and Freshwater Research, 53(2), 278-287.

Wood, D. (2024) Assessing and accounting for the influence of changes in laboratory measurement methods on the interpretation of long-term time-series data. *NIWA Client Report*, 2024055CH: 54.

Yue, S., Wang, C.Y. (2002) Regional streamflow trend detection with consideration of both temporal and spatial correlation. International Journal of Climatology: A *Journal of the Royal Meteorological Society*, 22: 933-946.

Appendix A Output data

The complete set of state analysis results is provided in the supplementary file "RiverState_2020to2024_250926.csv". The metadata for each water quality and invertebrate monitoring site (including human modified land cover classes) are provided in the supplementary files "RiverMetaData_WQ_v250926.csv" and "RiverMetaData_Inverts_v250926.csv", respectively. Trend results for all time periods are provided in the supplementary file "RiverTrends_to2024_v250926.csv".