Relationships between flow and river water quality monitoring data and recommendations for assessing NPS-FM attribute states and trends

For Auckland Council

October 2022

Prepared By:

Ton Snelder

Tim Kerr

For any information regarding this report please contact:

Ton Snelder
Phone: 0275758888
Email: ton@lwp.nz
LWP Ltd
PO Box 70
Lyttelton 8092
New Zealand

LWP Client Report Number:	2022-3
Report Date:	October 2022

Quality Assurance Statement

Version	Reviewed by
Final	Doug Booker (NIWA)
	Ned Norton (LWP)

Recommended citation

Snelder, T and Tim Kerr (2022). Relationships between flow and river water quality monitoring data and recommendations for assessing NPS-FM attribute states and trends. Prepared for Auckland Council by LWP, Land Water People.

ISBN 978-1-99-106030-3 (PDF)

Approved for publication by:
Jacqueline Lawrence-Sansbury, Acting Head of Research, Evaluation and Monitoring
Sietse Bouma, Manager Water Quality

Auckland Council, 12 December 2022

Table of Contents

Figures

Tables

Executive Summary

Assessment of water quality state and trends are requirements of Section 35 of the Resource Management Act (RMA: NZ Government 1991) and the National Policy Statement for Freshwater Management 2020 (NPS-FM). The NPS-FM defines certain compulsory water quality variables to be attributes of the values of ecosystem health and human contact, the details of which are set out in the National Objectives Framework (NOF). NOF numeric attribute states are evaluated from water quality observations obtained by water quality monitoring and are generally percentiles of water observations (e.g., 50th percentile (median), 95th percentile). For each attribute, the NPS-FM also defines categorical attribute states, which are derived by assigning numeric attribute states in four (or five) “NOF bands”, which are designated A to D (or E).

The NPS-FM requires that the baseline attribute states are established as an initial step in the planning process and that attributes are used as a basis for setting target attributes states. The NPS-FM also requires that the condition of water bodies (the current attribute state) is systematically monitored and reported, and that action is taken where monitoring indicates deteriorating trends.

Flow influences water quality in rivers and streams across a range of timescales and therefore has an impact on attribute state and trends. Therefore, flow is often monitored at water quality monitoring sites and flow data is important supplementary information in the analysis of water quality state and trends. It is important for Auckland Council (AC) to consider how flow influences attribute state and trends, how to incorporate information about flow into assessment of state and trends and to ensure that continued monitoring of water quality and flow is consistent with the requirements of the NPS-FM. It is also important for AC to consider whether assessment of state and trends is consistent with the requirements of the NPS-FM and to be clear about the uncertainties associated with these assessments. AC therefore commissioned this study to provide guidance with respect to how to account for the influence of flow in both the sampling and analysis of water quality data and the implications of this specifically in relation to the requirements of the NPS-FM.

This study undertook a series of analyses of water quality and flow data associated with 15 sites and 10 water quality variables (of which five are NOF attributes) in the Auckland region. These analyses and the salient results are as follows:

1. Attribute state is estimated from water quality observations pertaining to an “assessment period”. AC samples water quality monthly, and in this study, the assessment period was five years, which are consistent with current practice. We assessed the current attribute state from the observations and estimated the precision of these assessments. Limited precision means the assessed state is not exact and arises because the observations represent a finite sample of the population (i.e., a subset of all possible water quality observations). The 95% confidence interval for the assessed state of some NOF attributes often extended over two, three or even four NOF bands.

2. We refer to instantaneous flow as the flow in a river or stream at the time the water quality was sampled. In this study the instantaneous flow was quantified by the mean daily flow observed at a nearby river flow gauging station. The flow regime refers to the characteristics of flows at longer than instantaneous timescales, including weeks, months and years. Flow regimes can be characterised by many statistics such as mean and median flows, variability, and seasonality. Variation in these types of statistics at timescales of weeks, months and years are all measures of flow regime variability. A

drought year, for instance, will feature more frequent low flows than the long-term flow regime. This study showed that there is marked flow regime variation associated with different five-year assessment periods. Because flow influences water quality in rivers and streams, flow regime variation can cause differences in attribute states between assessment periods. This leads to uncertainty in state assessments that is additional to imprecision, and which is unquantified. We refer to this uncertainty as type A. Unquantified uncertainty of type A means there will be differences in assessments between assessment periods (such as between baseline state and current state) that is driven by flow regime variability that would occur even if there were no changes in the anthropogenic pressure in a catchment.

3. Water quality measures in rivers and streams are influenced by the instantaneous flow rate (i.e., discharge at the time the sample was taken). The strength of the relationship between observations and instantaneous flow differs across variables and sites and is associated with differences in the underlying mechanisms of mobilisation ("wash-off") and dilution of the contaminants. Water quality observations need to be unbiased with respect to instantaneous flow if they are to represent the true attribute state. We found that the distribution of flows associated with AC's water quality observations were not significantly different to the full flow distribution and can therefore be regarded as unbiased with respect to flow.

4. The relationships between water quality observations and instantaneous flow are commonly represented by bivariate (i.e., observation - flow) statistical models. These models are used in trend assessment in a statistical treatment known as flow-adjustment. The purpose of flow-adjustment is to remove the confounding effect of flow so that the pattern of interest (the relationship between the observed water quality observations and time (i.e., the trend) can be more confidently inferred. This study showed that the definition of models describing observations - instantaneous flow is subjective and therefore there are unquantified uncertainties that arise due to procedural choices around flow adjustment that are likely to be made by individual analysts.

5. Trend analyses for all site variable combinations were used to demonstrate that procedural choices made in association with flow-adjustment have an appreciable influence on the assessment. These differences represent an unquantified uncertainty that is associated with trend assessment that we refer as unquantified uncertainty of type B.

6. We undertook rolling trend analyses with assessment periods of differing duration (5, 10 and 20-years) and the starting year incrementing by one year. These analyses indicate that site trend direction tends to oscillate (i.e., trend directions can reverse from increasing to decreasing over short time periods). The length of the reversal time decreases with decreasing trend period duration. This indicates that short-term trends (e.g., 5 and 10-year duration) are likely to be strongly influenced by flow regime variation. This was true even when trends were flow-adjusted. We showed that these reversals are associated with flow regime variability by employing a model that combines the entire flow record with the water quality observation know as Weighted Regression on Time, Discharge, and Season (WRTDS). This result indicates that water quality trends are at least partly associated with flow regime variation. The oscillations in the trends are evidence that the flow regime variation, and associated water quality variation, is partly attributable to quasi-periodic climate variation such as the El Niño-Southern Oscillation (ENSO). It is important to emphasise that this study has shown a

link between water quality state and flow regime variation but did not directly investigate climate variation. However, there is direct evidence of the link between ENSO and water quality variation in New Zealand at interannual timescales. Therefore, when this report refers to the impact of "flow regime variability" on water quality, it is appropriate to consider that climatic variation is at least partly involved.

The important conclusions and some of our recommendations from these findings are as follows:

1. Water quality in rivers and streams is linked to flow, including the instantaneous flow at the time of sampling and the longer-term flow regime. Flow data should therefore be regarded as important additional information that assists the analysis and interpretation of attribute state and trends. Continuous flow time series can be used to characterise flow regime variation, which in turn, provides insights into the generation, mobilisation, storage and transport of contaminants in catchments. Therefore, we recommend that where possible, water quality monitoring should be associated with continuous flow measurements preferably from flow recorders or alternatively derived synthetically (i.e., modelled).
2. Assessed attribute states are associated with both quantified uncertainty (imprecision) and unquantified uncertainty of type A. We recommend that the assessed state be regarded as the "best information at the time" as defined by NPS-FM Section 1.6(1) but that AC is transparent about both types of uncertainty and that uncertainty is given due consideration when using and publishing data describing attribute states.
3. We recommend that the impact of unquantified uncertainty of type A on assessment of baseline and current state is considered when setting target attribute states and developing actions to improve water quality. This could take the form of sensitivity analyses that test the extent to which planned actions may fail to achieve target attribute states in future assessment periods due to foreseeable influence of flow regime variability on future attribute states.
4. We recommend that water quality trend assessments are always represented as model outputs that are unavoidably uncertain. To the extent possible, AC should be transparent about the uncertainties associated with water quality trend assessment, particularly when reporting trends. Because flow-adjustment introduces additional unquantified uncertainty of type B and does not remove the influence of flow regime variation on trends, we recommend that only raw (un-adjusted) trends are reported under S3.30(2)(c).
5. Because trend assessments are uncertain, we recommend a cautious and staged approach with respect to taking action when deteriorating trends are detected. The staged approach would be triggered by an observed deteriorating trend that is reported under NPS-FM S3.30(2)(c) requirements. Stage 1 should include more detailed analysis of the available data including consideration of flow-adjusted trends or potentially the use of more sophisticated models such as WRTDS or AC's process-based Freshwater Management Tool (FWMT). The appropriate action at Stage 1 may include taking cautious and proportionate action on the ground and/or potentially increased monitoring effort and ongoing surveillance of possible water quality pressures action. If deteriorating trends continue and/or confidence in the causes of these trends is judged sufficiently high, then stage 2 would be triggered that would involve significant intervention in the catchment to halt and reverse deterioration.

6. The requirement under NPS-FM S3.30(2)(d) to assess causes of deteriorating trends was not explicitly considered by this study. However, flow-adjustment and flow normalisation (a particular output of the WRTDS model), as undertaken in this study, can be regarded as statistical approaches to removing the influence of flow so the influence of other factors can be more robustly inferred. AC should strive to undertake robust attribution of cause(s) in seeking to carry out the requirements of NPS-FM S3.30(2)(d). However, this is extremely challenging for two reasons. First, suitable data characterising spatio-temporal variation in environmental drivers of water quality are scarce and fragmented. We therefore recommend gathering data describing possible causes of trends, such as changes in land use practices and intensity, changes in point source discharge loads, and adoption of actions in the catchments of monitoring sites and across the Region in general. The second reason NPS-FM S3.30(2)(d) presents a challenge is that water quality is generally influenced by multiple environmental drivers, including anthropogenic drivers such as land use and natural drivers such as climate variability and its impact on flow regimes. There may be additive, compensatory or synergistic interactions among these drivers, making it difficult to reliably attribute water quality responses to specific water quality pressures. The influences can only be elucidated by modelling and models are dependent on there being sufficient sites for the signals (i.e., causes) to rise above the noise. We therefore recommend that monitoring and modelling are treated as equal and mutually informative processes that must work together to fulfil AC's functions and duties under the RMA and NPS-FM.

Glossary

The table below defines the terms used in this report.

Term	Definition
Assessment period	A specific time period over which the available observations pertaining to a site and water quality variable are used to assess state or trends. In this report, the state assessment period is always 5-years and trends are assessed over periods of 5, 10 and 20 years.
Confidence	The confidence in the assessed direction of a trend, which is limited due to sample error. Confidence in trend direction indicates the probability that the assessed direction is the same as the true (i.e., population) direction.
Flow-adjustment	A statistical process that attempts to remove the influence of instantaneous flow on water quality observations. The purpose is to remove the confounding influence of instantaneous flow in trend analysis.
Flow regime variability	Variation in characteristics of flows at longer than that of instantaneous flow (see Instantaneous flow). Flow regimes can be characterised by many statistics such as mean and median flows, variability, and seasonality. Variation in these types of statistics at timescales of weeks, months and years are all measures of flow regime variability. Flow regimes vary in response to variation in hydrological processes including precipitation, evaporation and associated storage and release of water from the catchment.
Instantaneous flow	Flow at a specific point in time, such as when water quality is sampled. In this study the instantaneous flow was quantified by the mean daily flow observed at a nearby river flow gaging station.
Precision	The exactness of a quantified current or baseline state, which is limited due to sample error. Precision is quantifiable and indicates the range over which we could expect the state to vary if there had been multiple independent sets of samples collected over the same assessment period.
Unquantified uncertainty of type A and type B	Uncertainty associated with state and trend assessments that is not quantified by statistical analyses pertaining to evaluation of state and trends from water quality monitoring data. This report identifies two types of unquantified uncertainty, type A and type B. Type A uncertainty pertains to variation that would arise if the assessment was repeated in the future. For state, type A uncertainty is the difference in assessments between five-year assessment periods that is driven by flow regime variability and would occur even if there were no changes in the anthropogenic pressure in a catchment. For trends, type A uncertainty is variation in a trend assessment between assessment periods that occur due to influence of flow regime variability on water quality. type B uncertainty occurs due to differences in assumptions and choices made in the trend modelling process.
Sample error	Sampling error is the difference between a statistic that is calculated from a sample (e.g., a series of water quality observations) and the actual but unknown true value of that statistic (the population parameter). Sampling error is due to the variability inherent among data taken from a population (a statistical sample).
Baseline state	The state of compulsory NPS-FM attributes as of the 7 September 2017, which is assessed from observations for the preceding five-year period. For the purposes of this report, the period of 1 January 2013 to 31 December 2017 is referred to as the baseline state period although AC may utilise different time frames.

Term	Definition
Current state	The state of an attribute at the time of current reporting, which is based on the observations for the preceding five-year period. For the purposes of this report the period of 1 January 2016 to 31 December 2020 is referred to as the current state.
Attribute	A statistic calculated from the distribution of observations pertaining to an assessment period that is used to represent the state of freshwater systems in relation to specific values. Several compulsory attributes are defined in Appendix 2A and 2B of the NPS-FM 2020 and the potential numeric range is expressed as four bands (A, B, C and D).
Model	A representation of reality; cartoons, diagrams, graphs, computer simulations, and statistics and relationships derived from observations are all types of models.
National bottom line (NBL)	A minimum state set for several NPS-FM attributes. States below the NBL are considered degraded (if not due to natural causes) and councils must include actions in their plans that will improve these waterbodies to the NBL or a better state through time.

1 Introduction

Assessment of water quality state and trends are requirements of Section 35 of the Resource Management Act (RMA: NZ Government 1991) and the National Policy Statement for Freshwater Management (NPS-FM; MFE 2020). The NPS-FM requires that the baseline state of certain water quality variables (called attributes) is established as an initial step in the National Objectives Framework (NOF) process (S3.10(3)). Baseline states are a basis for setting water quality targets to achieve forward-looking environmental outcomes and objectives. Targets must be set at or above the expressed baseline state (NPS-FM S3.11) (or at or above national bottom lines if the baseline state is below this threshold (S3.11(4) not withstanding several exceptions) (NPS-FM S3.11). Baseline states for attributes defined by the NOF are derived from statistics, such as median values, which are calculated from water quality observations[1].

The NPS-FM also requires that the condition of water bodies is systematically monitored over time (NPS-FM S3.18), and action is taken where monitoring indicates deteriorating trends (NPS-FM S3.19/S3.20). Councils must publish annually data describing attributes and the associated uncertainty of those data (NPS-FM S3.30). In addition, the NPS-FM requires councils to publish comparisons of current and target attribute states (S3.30(2)(b)), assessments of whether target attribute states are being achieved, and if not, whether they are likely to be (S3.30(2)(c)), and assessments of trends and their causes (S3.30(2)(d)).

Auckland Council (AC) continues to monitor stream and river water quality at 37 sites across the Auckland region. Water samples are taken on a monthly basis and are analysed for up to 26 water quality variables. The values of the water quality variables that are observed on these sample occasions (i.e., the observations), supplemented with additional monitoring by NIWA at one river water quality site, are the basis for assessment of the current water quality state and trends for streams and rivers in the region (Ingley 2021a; Ingley and Groom 2022).

Flow influences water quality in rivers and streams, variation in flow across sampling occasions can be expected to impact to some degree on the assessment of water quality state and how this is changing over time (i.e., trends). For this reason, as well as others (such as water allocation and flood management), flow continues to be monitored at 15-minute intervals at 15 water quality monitoring sites. The flow data provided by this monitoring is important supplementary information in the analysis of water quality state and trends. It is important for AC to consider flow influences attribute state and trends, how to incorporate information about flow into assessment of state and trends, and to ensure that continued monitoring of water quality and flow is consistent with the requirements of the NPS-FM. It is also important for AC to consider whether assessment of state and trends is consistent with the requirements of the NPS-FM and to be clear about the uncertainties associated with these assessments. AC therefore commissioned this study to provide guidance with respect to how to account for the influence of flow in both the sampling and analysis of water quality data and the implications of this specifically in relation to the requirements of the NPS-FM.

The study has proceeded in two steps. First, we undertook a series of analyses to describe flow variation and the influence on state and trend assessment in the Auckland region. These analyses were to:

[1] In this study, we are concerned with observed water quality data. We acknowledge that assessments of baseline, or current state could be conducted in other ways. We also consider that it is likely that estimates of attribute states at unmonitored locations may be needed to inform NPS-FM processes and that modelling will necessarily be part of assessments.

1. Assess attribute state for selected attributes based on observed data.
2. Assess variation in flows between state assessment periods and longer time periods,
3. Assess the representation of flow variation by water quality observations
4. Assess the relationship between water quality variables and instantaneous flow rate.
5. Assess the impact on trend assessments of flow-adjustment and alternative plausible flow-adjustments.
6. Assess the evolution of water quality state at sites using a modelling approach that elucidates the impact on water quality of flows at timescales longer than that represented by instantaneous flows (referred as the flow regime)

The second step was to combine the results of the above analyses with our experience and expertise in state and trend assessment to provide guidance for:

1. Monitoring practice to obtain unbiased estimates of state and assessing and considering uncertainties associated with evaluation of attribute state.
2. Understanding the causes, and being transparent about, unquantified uncertainty associated with the evaluation of attribute state.
3. Understanding the causes of, and being transparent about, uncertainties associated with trend assessments.
4. Understanding, and being transparent about, the need to use models and modelling to make sense of water quality data and to carry out the requirements of the NPS-FM effectively and robustly.
5. Considerations pertaining to assessments of trends and the requirement to take action where monitoring indicates deterioration and the implications for NPS-FM requirements.

We also provide commentary on wider issues that are raised by this study with respect to potential changes to sampling frequency of monitoring and to improve spatial coverage.

2 Background

This section outlines four concepts that are used in this study, and which are important to the methods and discussion sections that follow.

2.1 Assessment of current state

In this study we assess river water quality state for 10 variables including five (of 22) compulsory National Objectives Framework (NOF) attributes defined by the of the NPS-FM (MfE 2020) and two additional urban water quality attributes proposed for the Auckland region - soluble zinc and copper (Ingley 2021b; Ingley and Groom 2021). The approach we have taken to state assessment is set out below.

Each table in Appendix 2 of the NPS-FM (2020) represents a NOF attribute (hereafter "attribute) that provides for a particular environmental value (either individually or in combination with other attributes). For example, Appendix 2A, Table 6, defines the nitrate toxicity attribute, which is defined by nitrate-nitrogen concentrations that will, in part so far as concerns nitrate toxicity, ensure an acceptable level of support for "Ecosystem health (Water quality)" values. The state of an attribute at a site is primarily defined by the value of one or more statistics (hereafter numeric attribute state) that are generally percentiles of water observations[2]. For example, for the nitrate-nitrogen attribute there are two numeric attribute states defined by the annual median (i.e., the 50th percentile) and the 95th percentile concentrations. For each attribute, the NPS-FM also defines categorical attribute states in four (or five) "NOF bands", which are designated A to D (or A to E, in the case of the *E. coli* attribute). The NOF bands represent a graduated range of support for achieving environmental values from high (A band) to low (D or E band). Narrative descriptions of the level of support for the values are associated with each categorical attribute band. The ranges of the numeric attribute states that define NOF bands are defined in Appendix 2 of the NPS-FM (2020). For most attributes, the D band represents a condition that is recognised nationally as unacceptable (with the threshold between the C and the D band being referred to as a national bottom line (NBL)). In the case of the nitrate (toxicity) and ammonia (toxicity) attributes in the 2020 NPS-FM, the B/C band threshold is the national bottom line, and for the *E. coli* and DRP attributes, no bottom lines are specified.

In our opinion, the primary aim of the NOF bands is to provide a simple shorthand for communities and decision makers to discuss options and aspirations for acceptable water quality and to define objectives. Categorical attribute bands avoid the need to discuss objectives and targets in terms of technically complicated numeric ranges. Each NOF band is associated with a narrative description of the outcomes for values that can be expected if that NOF band is chosen as the objective. However, it is also logical to use NOF bands to provide a grading of the baseline and current state of water quality; either as a starting point for target setting or to track progress toward objectives.

Water quality observations derived from monitoring are used to assess state. State assessment uses the numeric attribute state (e.g., median or 95th percentile of nitrate-nitrogen concentration) as a compliance statistic. The value of the compliance statistic for a site is calculated from the observations of the relevant water quality variable (e.g., the median value is calculated from the observed nitrate-nitrogen concentrations). Current state can be expressed as a numeric value or a NOF band. A band is assigned by comparing a site's

[2] Note that the *E. coli* attribute includes two additional statistics that are defined by proportions of observations exceeding stated values.

compliance statistic to the numeric ranges associated with each NOF band (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). For attributes with more than one numeric attribute state, bands are defined for each numeric attribute state (e.g., for the nitrate (toxicity) attribute, bands are defined for both the median and 95th percentile concentrations).

An important consideration that arises with assessing numeric attribute states and NOF bands from observations is the issue of sampling variability, often referred to as ‘statistical sampling error’. Statistical sampling error means that we never know the true attribute state because we only ever have a finite sample (i.e., a subset of all possible water quality observations) of the true continuous time-series occurring in the river. We know that the attribute state estimated from the sample is sometimes higher than the true value and sometimes lower than the true value. The precision associated with a numeric attribute state indicates how different the true value is likely to be from the estimate and in this study, we represent this by a 95% confidence interval[3]. In the context of assessing attribute states, the precision of the estimate can be understood as the range over which we could expect the assessed state to vary if there had been multiple independent sets of samples taken, all other things being equal (i.e., samples being taken over the same sample period and at the same site). In some places in this report, we use the term “face value” to mean the evaluated numeric attribute state. This is to remind the reader that this value is imprecise. The precision of statistics (such as those used to define numeric attribute states) will increase as number of observations increases but is dependent on the variability of the observations and the number of observations. As a general rule, the rate of increase in the precision of the numeric attribute states slows for sample sizes greater than 30 (i.e., there are diminishing returns on increasing sample size with respect to precision; McBride 2005).

The NOF is generally not clear whether numeric attribute states apply to percentiles of time or percentiles of samples. McBride (2016) notes that there are significant implications for the number of observations that are required if the former is intended. This is because a percentile of time indicates that the numeric attribute state is regarded to be an estimate of the population statistic. In this case, it is relevant to consider the risk that our estimate is incorrect (e.g., the assigned NOF grade is not exactly the same as the true (population) NOF grade). This risk depends on the precision of the estimated numeric attribute state and whether this is acceptable depends on decisions about the burden-of-proof (i.e., the level of evidence required to demonstrate that the grade is correct). McBride (2016) shows that if a precautionary approach to burden-of-proof is taken (i.e., high confidence the true grade is not worse than the assessed grade) then an appreciably larger sample is required than if an even-handed (i.e., use the face value of the assessed grade) approach is taken. However, for some attributes, the NOF specifies the sampling frequency and duration (e.g., the *E. coli* attribute state is defined by four statistics that are calculated from three years of monthly observations). This suggests that we can assume that numeric attribute states apply to percentiles of samples, or equally, that we are taking an even-handed approach to the burden-of-proof.

A further complication that arises with assessing attribute states is that for some attributes, the NOF specifies “annual statistics” (i.e., annual median; annual maximum) for assessing state (e.g., Nitrate and Ammonia Toxicity). This appears to indicate that assessments are made from one year of observations. However, if monitoring was monthly, this would result in

[3] The standard error of the mean is the most generally understood example of the precision of a statistical estimate. The standard error of the mean indicates how different the population mean is likely to be from a sample mean. The standard error of the mean quantifies how much the sample mean (i.e., an estimate of the population mean derived from a sample) would vary if you were to repeatedly sample the population.

only 12 observations and therefore very imprecise estimates of the median and 95th percentile (see below for discussion of assessment precision). In this study, therefore, we have assessed attribute states by calculating the statistics prescribed by the relevant NOF attribute table (e.g., median, 95th percentile) from records of observations of 5-years duration (hereafter a “state assessment period”) as recommended by McBride (2016) and as generally undertaken by national environmental reporting studies (Larned et al. 2018; Whitehead et al. 2021). We note that this is the approach implemented by AC for previous reporting, and is consistent with the approach taken by Land, Air, Water Aotearoa (LAWA), and analyses undertaken on behalf of the Ministry for the Environment (MFE) and Statistics New Zealand (StatsNZ) (e.g., Larned et al. 2018; Whitehead et al. 2021)

In the state assessments undertaken in this study, we have evaluated numeric attribute states and their precision (i.e., the uncertainty associated with the sample error). We note that the reported precision is relevant whether the numeric attribute states are regarded as percentiles of time or percentiles of samples, but with different interpretations. If the numeric attribute state applies to the population (i.e., percentiles of time), the precision describes the uncertainty of the assessment of the true state. Alternatively, if the numeric attribute state applies to the sample (i.e., percentiles of samples), the precision can be interpreted as the range over which the sample statistic could be expected to vary if sampling had occurred on different days within the same sampling period.

2.2 Influence of instantaneous flow and flow regime variability on water quality observations

Many water quality measures in rivers and stream are influenced by the instantaneous flow rate (i.e., discharge at the time the sample was taken). In this study, we have used the mean daily flow to represent the discharge at the time the sample was taken. Water quality observations can vary systematically with instantaneous flow due to two kinds of physical processes. Observations may decrease systematically with increasing flow due to the effect of dilution of the contaminant, or increase with increasing flow due to mobilisation (“wash-off”) of the contaminant (Smith et al. 1996). The relationship between water quality measures and instantaneous flow may also depend on the location on the hydrograph when the sample was taken such that concentrations at a given flow rate can differ between the rising and falling limbs. In urban contexts this mechanism is referred to as “first flush” where at the initiation of the rising limb of the hydrograph, concentrations are higher compared to later when the sources of contaminants have been depleted or “washed off” (Lee et al. 2004).

Different mechanisms may dominate at different sites so that the same water quality variable can exhibit positive or negative relationships with increasing instantaneous flow. Some water quality variables can be associated with a combination of dilution and wash off with increasing flow. For example, a portion of the suspended sediment load may come from point source discharges such as sewage treatment plants (dilution effect), but another portion may be derived from surface wash-off. Increasing flow in this situation may result in an initial dilution at low flow rates, followed by an increase at higher flow rates (Helsel et al. 2020).

Relationships between water quality variable observations and instantaneous flow are commonly represented by bivariate (i.e., observation - flow) statistical models (e.g., Helsel et al. 2020; Snelder et al. 2021a). These models are used, in a process called “flow-adjustment”, to remove the confounding influence of instantaneous flow so that the trend in the observations can be more confidently assessed (see Section 4.4). An important assumption underlying these models is that the relationship between the observations and instantaneous flow is constant in time. This assumption simplifies the definition of the model but is likely to be violated in situations where catchment processes associated with the generation, storage, transport and transformation of contaminants are changing (see Section 4.6).

Notwithstanding the simplifying assumptions underlying bivariate observation - flow models, selecting the most appropriate statistical model to represent the relationship is complicated for several reasons. First, observation - flow relationships differ between variables at individual sites and between sites for a variable. Second, there is often a trade-off between the goodness-of-fit and the physical plausibility of the relationships represented by models. Simple models such as linear regression will represent physically plausible monotonic increasing or decreasing relationships between observations and flow but these may have poor goodness-of-fit. More complicated models allow for non-linear relationships between observations and flow. These models can represent more complex relationships that may have plausible mechanistic explanations. For example, non-linear models can represent large increases in concentrations of contaminant with increasing flow that could be expected where a threshold of movement of a contaminant is crossed. Non-linear models can represent local maxima that occur if initially increasing concentration with flow is followed by source depletion or dilution at very high flows. However, non-linear models may also represent physically implausible relationships between observations and flow such as multiple local maxima (Snelder et al. 2021a). Care and expert judgment in selecting observation - instantaneous flow models is therefore required and are discussed in Section 4.3.

Flow also varies at longer than timescales “instantaneous flows” in response to variation in hydrological processes including precipitation, evaporation and associated storage and release of water from the catchment (Sofi et al. 2020). This hydrological variability is manifested as flow regime variation, but also water quality variability because contaminant mobilisation, transport, storage and dilution is affected by the same hydrological processes (Gascuel-Odoux et al. 2010). In this study, we refer to the “flow regime” to mean characteristics of river flows at longer than the daily timescale that we use to indicate instantaneous flow. Flow regimes can be characterised by many flow statistics such as mean and median flows, flow variability and seasonality (Snelder and Booker 2013). Like instantaneous flows, the flow regime at a site varies over time. For example, periods of uncharacteristically low or high flows can occur at various timescales (e.g., weeks, months and years).

Flow regimes are influenced by anthropogenic activities occurring within the catchment such as abstraction and land use changes and are also by influenced by natural processes such ecological succession of land cover from scrub to forest (Best 2019; Chen et al. 2019; Margariti et al. 2019). Flow regimes are also strongly controlled by climatic processes such as precipitation and evaporation (Sofi et al. 2020). For example, effective rainfall (precipitation minus evaporation) drives water storage and release from catchments at timescales of days to years (Wilusz et al. 2017). Irrespective of the cause, flow regime variability is linked to variation in water quality and in study, we undertake analyses to show that link. The link between flow regime variability and water quality variability is important because it impacts on assessments of water quality state and trends.

In this study, we do not undertake any analysis of the causes of the flow regime variability. However, some of the analyses we present indicate that water quality oscillates at interannual timescales. These oscillations are evidence that water quality variation is partly attributable to quasi-periodic climate variation such as the El Niño-Southern Oscillation (ENSO; Mullan 1996; Salinger and Mullan 1999). Evidence for the link between ENSO and water quality variation in New Zealand at interannual timescales has been provided by studies of fluctuations in water quality trend assessments at time scales from 5 to 15 years to (Scarsbrook et al. 2003; Snelder et al. 2021b). Therefore, when we refer to the impact of "flow regime variability" on water quality, it is appropriate to consider that climatic variation is at least partly involved. The relevance of climatic variation is that it is factor that cannot be managed but may influence state and trends and can confound determination of the anthropogenic causes of water quality changes.

2.3 Consideration of flow with respect to water quality sampling, state and trend assessment

Because the specification of the NOF attributes in the NPS-FM makes no mention of flow, it is reasonable to assume that attribute states, and therefore monitoring data, should represent the full flow range. This makes sense if we consider that the purpose of NOF attributes is to manage the effect of water quality on values such as ecosystem health and human health risk. Because these values are not specific to certain flow states[4], it is logical that attribute states, and monitoring data, should represent the full flow range.

In this regard, we consider that there are different considerations associated with the flow at the time of sampling for state assessments compared to trend analysis. For assessment of attribute state, we assume that the monitoring data should represent the full flow distribution. Therefore, the question is whether the flows represented in the observation data are a reasonable representation of the flow distribution. In principle, this will be true if sampling is punctual. Punctual sampling involves setting the sample frequency and occasion (i.e., date) in advance and then not deviating from this schedule. This will ensure that sampling is random with respect to flow and, as the number of observations increase, the sampled flow distribution will increasingly closely correspond to the actual flow distribution[5].

A complication that arises with assessing attribute states is that there is likely to be hydrological variation between state assessment periods that is manifested as flow regime variation. Because water quality observations are generally influenced by the same processes that influence flow regimes, state assessments can be expected to vary between assessment periods (e.g., between baseline and current state) in association with flow regime variation. This produces a component of uncertainty in state assessments that is in addition to precision, which we refer to as "unquantified uncertainty of type A". In attribute state assessments, unquantified uncertainty of type A can be understood as the variation in state between five-year assessment periods that is associated with flow regime variability and would occur even if there were no changes in the anthropogenic pressure in a catchment.

Trend analysis seeks to quantify the relationship between the water quality observations and time. In this context, flow can be considered a "covariate"; a variable that is also related to the water quality observations but whose influence is confounding the water quality – time relationship that trend analysis seeks to expose. The process of flow-adjusting is used to remove the influence of instantaneous flow on the water quality observations prior to trend analysis. Flow-adjustment has two purposes. First, it theoretically increases the statistical power of the trend assessment (i.e., increase the confidence in the estimate of direction and rate of the trend) by removing some of the variability that is associated with flow. Second, it removes any component of the trend that is attributable to a trend in instantaneous flow (e.g., a trend in the flow on sample occasions such as increasing or decreasing flow with time). However, whether it is appropriate to undertake flow-adjustment depends on the objectives of

---

[4] Although certain flow states may be associated with greater human contact with water and therefore greater risk.

[5] We note that practically, observations at very high flow are not achievable for health and safety reason so that some rescheduling of sampling is inevitable, and observations cannot be truly random with respect to flow.

the trend assessment. If the aim is to understand whether a management action has affected water quality over time, then the contribution of flow to the trend is a confounding factor and flow-adjustment is promoted as a means to increase the confidence in the trend assessment (Snelder et al. 2021a). In contrast, if the aim of the assessment is to quantify the water quality trend that actually occurred, then flow-adjustment may not be applicable. An example where quantification of the actual (unadjusted) trend might be desired is where a biological change has occurred in a stream and there is interest in whether this was associated with changes in water quality variables.

In our opinion, irrespective of the purpose of trend assessment and whether this indicates flow-adjustment is required or not, it is desirable that monitoring data should represent the full flow distribution to avoid biased trend assessments. If the purpose is to quantify the water quality trend that actually occurred, then the sample should represent the population as well as possible and therefore samples should be representative of the flow distribution. On the other hand, if flow is regarded as a confounding factor whose influence is to be removed, there is a need to first model the relationship between instantaneous flow and observations (see Section 2.2). In this case it is also desirable that monitoring data represents the full flow distribution. A complication that arises in this case is that, for some water quality variables, there are rapid changes in the flow – observation relationship at high flows. If sampling is punctual, there will be relatively few high flow observations, which impacts on the accuracy of the model at high flows. In addition, there is a strong tendency for observations at high flow to have high variability, which also impacts on the accuracy of the model at high flows. This combination introduces considerable uncertainty and subjectivity into instantaneous flow - observations models, which is described later.

2.4 Use of models

Section 1.6 of NPS-FM directs councils “to use best information available” and “in the absence of complete and scientifically robust data, the best information may include information obtained from modelling as well as partial data…”. However, trend analysis involves fitting statistical (regression) models and characterisation of population by calculating summary statistics, such as the median, from a sample involves making statistical assumptions. Therefore, the requirements to assess water quality state and trends are reliant on models and, accordingly, the outputs should be interpreted as being uncertain and influenced by the associated modelling procedures and assumptions.

From a technical perspective, numeric attribute states are models of some characteristic of the distribution (e.g., a median describes the characteristic “central tendency” of a data). It is important to recognise the assessed attribute state is a model (i.e., the assessed attribute state is a representation of reality) because that clarifies that the assessment is uncertain; irrespective of whether it is interpreted as applying to the population (i.e., percentiles of time) or the sample (see section 2.1). In this report, we illustrate the uncertainty that is associated with numeric attribute states, provide some commentary on how this impacts the ability to detect change and attribute change to causes and suggest how to respond to the uncertainty.

Trend assessment is a process of building a statistical model of the behaviour of a variable over time from a series of observations (Helsel et al. 2020). The model is built from observations pertaining to a site/variable combination that represent a sample of the population (i.e., a sample of the actual conditions over the entire period of interest). The trend is an estimate of what actually occurred, which is subject to ‘statistical sampling error’. Therefore, trend assessments are always associated with quantifications of uncertainty that are analogous to the quantification of precision in assessments of state.

Like all statistical models, a trend assessment is a simplification of reality that aims to expose the most important features of the true relationship or pattern of interest. The simplest and most salient features of the relationship between a variable and time are the direction and the rate of change in the variable. Although there are accepted methods for trend assessment (e.g., Snelder et al. 2021a), their application automatically implies making assumptions and simplifications. It is important to be aware that quantifications of uncertainty accompanying trend assessments express the combination of the statistical sampling error' interacting with the mathematical description of the trend represented by the model. This quantified uncertainty does not include the impact of model assumptions and necessary methodological or procedural judgements made by the analyst in performing the trend assessment. These aspects introduce additional uncertainties that are unquantified by the trend assessment. We refer to these uncertainties as "unquantified uncertainty of type B". In this report we show that unquantified uncertainties associated with trend assessments may be consequential and therefore need to be kept in mind when interpreting and using the results of trend analysis.

3 Data

In this study we used stream flow and water quality data pertaining to river and stream state of environment water quality monitoring sites in the Auckland region (Figure 1). Stream flow data (hereafter "flow") and water quality observation data were provided by the Research and Evaluation Unit (RIMU), Auckland Council) in six files as listed in Table 1.

Table 1. Flow and water quality data sources.

File Name	Description	Format
20211210_RiverWQMaster.csv	Water quality observations.	comma separated variable
20211701_Master_Land Cover_WQ.xlsx	Land cover (LCDB 5) breakdown of each water quality site catchment through time.	Excel
Daily mean flow data currently paired sites.xlsx	Mean daily flow for multiple flow sites.	Excel
6604 Matakana Gaugings.xlsx	Flow measurements from a single site	Excel
44603 Cascades Gaugings.xlsx	Flow measurements from a single site	Excel
Makarau at Coles mean daily flows.xlsx	Flow measurements from a single site	Excel
Sites and Coordinates_2.xlsx	Water quality measurement site metadata	Excel

The water quality observations file contained data for 37 unique sites that had been collected by AC over the period (1985 to 2020, although individual sites had varying record length). Two sites (Hoteo River (site 45703) and Rangitopuni River (site 7805)) had additional data collected by NIWA (site names "AK1" and "AK2") that supplemented the Auckland Council data.

The water quality data comprised observations of 37 water quality variables on discrete occasions defined by dates. The ten water quality variables that were the focus of this study are shown in Table 2. The observation frequency was generally/approximately monthly at all sites. Hereafter, we refer to discrete observation of a water quality variable as an "observation" and to each date on which an observation occurred as an "observation-date".

The analysis of samples for total oxidised nitrogen concentrations (NNN) has been carried out for many more years than nitrate-nitrogen concentrations (which were initiated by AC in December 2018). Following Ingley (2021b) NNN values have been used as a proxy for nitrate-nitrogen. This assumes that nitrite concentrations are low compared to nitrate concentrations and can be ignored.

Clarity observations were provided by AC as "Clarity (converted)" values for the purposes of comparison to the suspended fine sediment NOF attribute. These values were calculated from turbidity (NTU) based on Franklin et al. (2019). This national adjustment has not been validated for the Auckland region or specific sites and state assessment is considered provisional until such verification is undertaken. AC also provided pH adjusted ammonia observations as the ammonia toxicity NOF attribute is intended to apply to these adjusted values, rather than ammoniacal N. Following Ingley (2021b), we used pH adjusted ammoniacal nitrogen observations to assess state but the non-adjusted ammoniacal nitrogen in all other analyses. We note that these conversions introduce uncertainty to the provided values (in addition to the measurement and analysis uncertainties). No attempt has been made to include these uncertainty contributions. This study included the metals copper and zinc at some, but not all, sites because monthly water quality samples have only been analysed for metals at all sites since 2018. When reported in this study, metals have not been adjusted for dissolved organic carbon or hardness.

Table 2. Water quality variables analysed in this study.

Water quality variable	Unit	Auckland Council Data name	Alias	Comment
Clarity	m	Clarity	CLAR	This is referred to as the Suspended fine sediment attribute in Table 8 of the NPS-FM. Clarity was calculated from turbidity (NTU) by AC except for analyses that were uniquely performed on data collected by NIWA (Hoteo River and Rangitopuni River) where measured clarity data was used.
Escherichia coli	CFU/100 mL	*E. coli*	ECOLI	*E. coli* is included as an attribute for all lakes and rivers in Table 9 of the NPS-FM. This differs from requirements for primary contact sites set out in Table 22 of the NPS-FM.
Ammoniacal nitrogen	mg/L	Ammonia as N	NH4N	A measured and not corrected for pH
	mg/L	pH adjusted NH4	NH4_adj	Corrected for pH by AC regarding Ammonia (toxicity) attribute Table 5 in the NPS-FM.
Dissolved inorganic Nitrogen	mg/L	DIN	DIN	DIN is specifically identified in NPS-FM S3.13 but its monitoring and assessment requirements are unspecified.
Nitrate-Nitrite-Nitrogen	mg/L	Total Oxidised N	NNN	Used as a proxy for the Nitrate (toxicity) attribute in Table 6 of the NPS-FM.
Dissolved reactive phosphorous	mg/L	DRP	DRP	Dissolved reactive phosphorus Table 20 of the NPS-FM.
Total Nitrogen	mg/L	Total N	TN	Not included as a compulsory attribute for river water quality in the NPS-FM.
Total Phosphorous	mg/L	Total P	TP	Not included as a compulsory attribute for river water quality in the NPS-FM.
Soluble Copper	mg/L	Dissolved Copper	CU	Draft proposed river water quality attribute for consideration by Auckland Council.
Soluble Zinc	mg/L	Dissolved Zinc	ZN	Draft proposed river water quality attribute for consideration by Auckland Council.

Continuous mean daily flow data was provided for 21 of the water quality sites. Flows were measured at the water quality site, or on the same river segment and within 2 km of the sample location. Two sites were discarded. The Matakana site, (site 6604) was discarded because the flow data were derived from a limited number of gauging observations. The Paerata Rise site, (site 43968) was discarded because it had less than 10 years of flow observations. The Newmarket Stream site (site 10814) was discarded because water quality monitoring was only established in 2018. The Okura and Kumeu River sites (7502 and 45313) were discarded because they were closed in 2015. The Oakley Creek site (8110) was discarded because the flow site location was >6km from the water quality site and was therefore considered not sufficiently representative.

The remaining 15 sites have been used for analysis (see Figure 1 for locations). The site details, including the names of the related flow measurement sites are listed in 0, Table 8. Auckland Council water quality and flow sites names are often, but not always, the same. For consistency in this report, the water quality site names are used rather than the flow site names. The relationship between the two is provided in 0, Table 8.

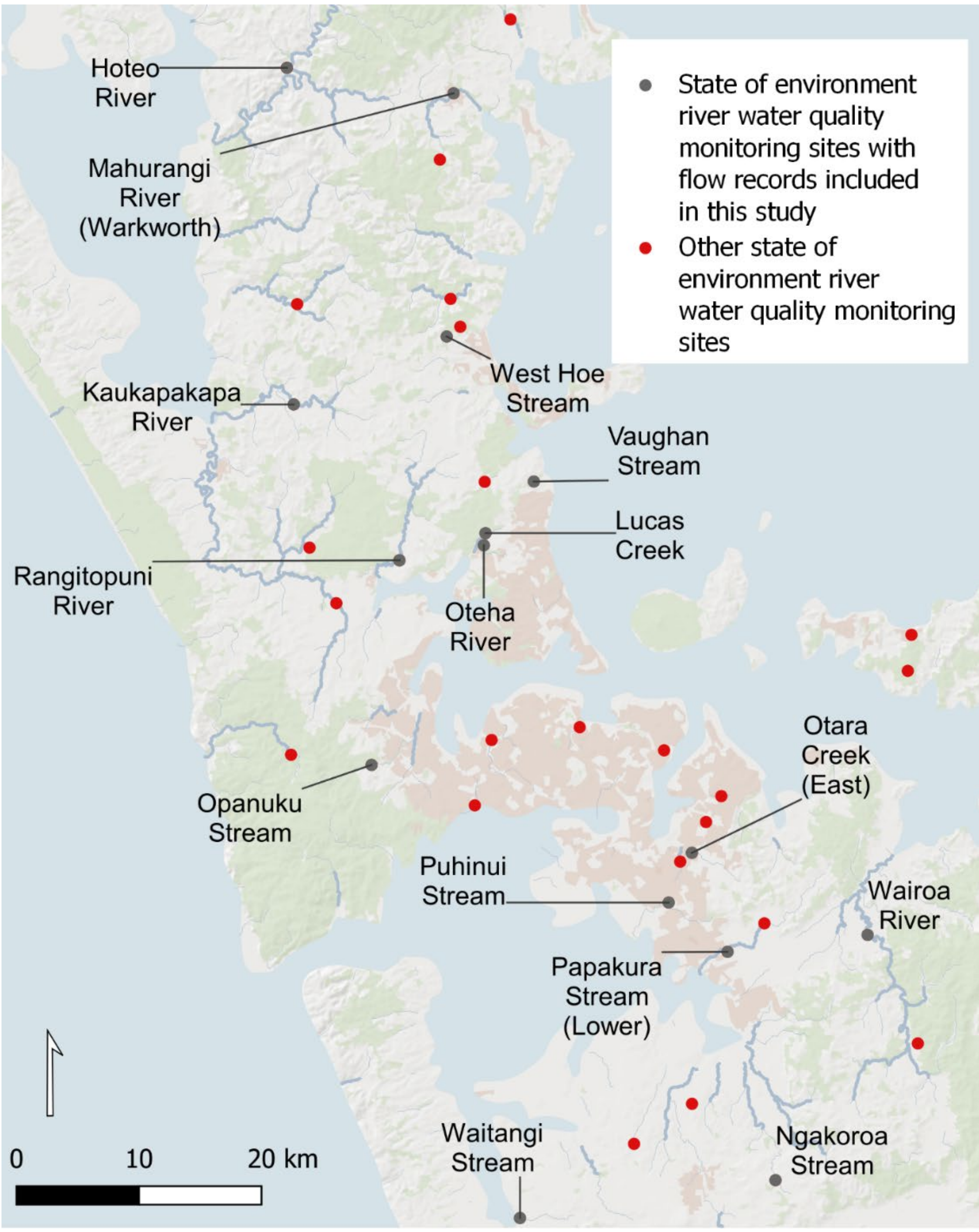


Figure 1. Locations of all of Auckland region's state of environment river water quality monitoring sites including the 15 paired water quality and flow sites.

4 Methods

In this study, we undertook six sets of analyses to describe flow variation and its influence on state and trend assessment in the Auckland region. The methods used in these analyses are set out in the following six subsections.

1. Assess current attribute states and the precision of these assessments based on the water quality variables shown in Table 2 for the 5-year period ending 2020.
2. At each site, compare variability in flow regimes between each of four five-year state assessment periods between 2013 and 2020 and to the long-term flow record.
3. At each site, compare the instantaneous flow conditions sampled on water quality monitoring occasions for each of four five-year state assessment periods between 2013 and 2020 and describe any bias flow associated with the water quality monitoring observations.
4. At each site, for each variable, consider flow-adjustment using alternative functional forms for flow-concentration models. Use expert judgement to select the most appropriate model to undertake flow-adjustment and compare the results to those obtained using an alternative "default" flow-concentration model.
5. Calculate and compare flow-adjusted vs non-adjusted trends for the period 2007-2017 and rolling trends to 2011-2020.
6. Use an alternative modelling approach to describe observation – flow relationships and trends to those derived using the above methods.

The details of these analyses are explained in the following sections.

4.1 State assessment

We assessed current numeric states for the 5-year period ending 2020 for the following variables, using the following listed statistics and as set out in the remainder of this section:

- *E. coli*, using four statistics (exceedances of 540 and 260 *E. coli*/100 mL, median and 95th percentile).
- DRP, using the median and 95th percentile.
- Nitrate, using the median and 95th percentile.
- Ammonia, using median after adjusting for pH. Because it is not possible to estimate the uncertainty for a maximum value, we have not included it here.
- Clarity (converted from turbidity (NTU) representing the suspended fine sediment attribute), using the median.
- DIN, reported as the median and 95th percentile
- Dissolved Zinc, following Ingley (2021b), numeric attribute state was defined by median and 95th percentiles and categorical attribute state was defined based on Gadd *et al.* (2019). There was no adjustment for dissolved organic carbon or hardness.
- Dissolved Copper, as for Zinc

For each water quality variable at each site, the numeric attribute state (i.e., face value of the relevant statistics described above), NOF band (where applicable) and the 95% confidence interval for the numeric attribute state was calculated for rolling 5-year state assessment periods between 2013 and 2020 (inclusive). An assessment period of five years was used for all variables in the state assessment. This is consistent with the methods used by AC for environmental reporting (Ingley and Groom 2022).

Medians and 95th percentiles were calculated using the Hazen method. Values at or outside detection limits were retained at the detection limit without imputation. Precision for both percentiles and proportions (i.e., for *E. coli*, exceedances of 540 and 260 *E. coli*/100 mL) were estimated based on the method of Wilson (1927) as recommended by Brown et al. (2001) and are expressed as the 95% confidence interval.

When grading sites based on NPS-FM attributes, it is general practice to define the acceptable proportion of missing observations (i.e., data gaps) and how these are distributed across sample intervals so that site bands are assessed from comparable data (Whitehead et al. 2021). 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). In this study, a period of five years was used for state assessment. Some NOF attributes require 5-years of monthly data but more generally, 5 years represents a reasonable trade-off for grading assessments because it yields a sample size of more than 30 observations. The five-year period for the NPS-FM state analyses is consistent with previous national water-quality state analyses (Larned et al. 2018; Whitehead et al. 2021).

Because water quality data tend to fluctuate seasonally, it is also important that each season is well-represented over the period of record. AC has sampled river water quality at a monthly frequency for the period under consideration in this analysis and therefore we defined seasons by months. We applied a filtering rule that restricted site and variable combinations in each state assessment period to those with measurements for at least 90% of the sampling intervals in that period (i.e., at least 54 of 60 months). We did not place a restriction on the allowable proportion of gaps for individual seasons because this first requirement (90% of sample intervals) means the potential impact of missing seasonal observations is minimal. Site × variable combinations that did not comply with these rules were excluded from the state analysis. We note that 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. This is consistent with the methods used by AC for environmental reporting (Ingley and Groom 2022)

4.2 Variability of flow regimes between assessment periods

The flow regime for a five-year state assessment period is unlikely to perfectly represent the flow regimes of alternative five-year state assessment periods or the long-term flow regime. An assessment period that is associated with a drought, for instance, will feature more frequent low flows than the long-term flow regime. An assessment period in which rainfall was higher will have higher mean flows and more frequent high flows that the long-term flow regime and the assessment period that was associated with a drought.

We quantitatively assessed differences in flow regimes at the time scale of assessment periods and compared these to the long-term flow regime in three steps. First, at each site we calculated the mean flow in every month of record. Second, we calculated the mean of mean monthly flows in each month of the entire record and used this to represent the long-term flow regime. Third, we calculated the mean of mean monthly flows in each 5-year state assessment period between 2013 and 2020 (i.e., four assessment periods ending 2017, 2018, 2019 and 2020) and used these to represent that period's flow regime.

We plotted these data to visualise the variability of flow regimes between assessment periods and the deviation of flows in each assessment period to the long-term flow regime. We did not undertake formal statistical testing of the significance of the differences in flow regimes because it is not clear what type of test would be relevant or what the benefit of statistical confidence in differences would be. However, we performed analyses that combined the water quality data and flows to demonstrate that differences in flows between assessment period contributes to variation in water quality assessments (see Section 4.6).

4.3 Representation of instantaneous flow by water quality observations

The instantaneous flow at the time of water quality observations were assessed graphically by timeseries plots that depict the super-position of the observations on the flow hydrograph. These plots indicate the distribution of observations over the flow range and variability of flows over time. For water quality variables that are affected by streamflow it is ideal if, for any state assessment period, the distribution of flows on observation dates matches the distribution of all flows in the assessment period. If this is the case, then observations are unbiased with respect to flow. We used flow duration curves (FDC) to graphically assess whether the distribution of flows on observation dates matched the distribution of all flows through the state assessment period. A FDC is a cumulative frequency distribution that shows the percent of time specified discharges were equalled or exceeded during a given period. FDCs indicate the flow characteristics of a stream over the full range of discharge, but with no indication of the sequence of occurrence of the flows. The y-axis of an FDC indicates the flow and the x-axis represents either the rank or the percentage of time that the flow has been exceeded (by scaling the flow ranking from 0% to 100%).

We assessed whether the distribution of flows on observation dates matched the FDC for rolling 5-year assessment periods between 2013 and 2020 in two steps. First, for each site, we produced FDCs from the daily flows in each 5-year period. Second, for each site, we overplotted these assessment period FDCs with points representing the observation dates.

The portion of the FDC outside the ranges of the observation-date flows are of particular interest. Flows that are not represented by observations limit the ability to accurately assess and model the relationship between water quality variables and instantaneous flow (see Section 4.4). In general, we can expect that modelled relationships are less reliable near the end and outside the range of the observation-date flows. Inaccurate representation of water quality observation – instantaneous flow relationships have implications for flow-adjustment (see Section 4.4) and for calculation of contaminant loads (e.g., Snelder et al., 2017). The percentage of the FDC that is within the observation-date flow range provides a measure of the proportion of the time a site's flow was represented by the observations. The area under the FDC represents the total volume of water passing a flow measurement site. Therefore, the area under the FDC that is represented by observations provides a measure of the proportion of the total volume of flow that is represented by the observations. For each site, we also evaluated the proportion of time, and the proportion of volume, that was represented by the observations for rolling 5-year state assessment periods between 2013 and 2021.

We also quantitatively assessed whether there was bias in the representation of the flow distribution at each site by the observations. This assessment used the Kolmogorov-Smirnov test to assess whether the distribution of flows on observation dates matched the distribution of all flows within each five-year assessment period. The Kolmogorov-Smirnov test assesses whether two sets of data are drawn from the same cumulative frequency distribution (CFD). The Kolmogorov-Smirnov test statistic is the maximum discrepancy between the CFDs derived from the two datasets. For our data, this means the largest difference in the proportion of the time a given flow is equalled or exceeded. This statistic can vary between zero (indicating the distributions are exactly the same) or one (indicating there is no overlap between the two distributions). The significance of the test statistic is assessed based on the null hypothesis (H_o) that the two datasets are from the same distribution. If the *p*-value for the test is less than alpha (which we set at 0.05) the null hypothesis is rejected, and it can be concluded that the two distributions are significantly different. If the *p*-value for the test is greater than alpha, the null hypothesis is not rejected, and it can be concluded that the two distributions are consistent with the same population distribution. Note that the test can only determine that two distributions are different—it does not indicate whether the change is an increase or decrease in the mean or due to a change in the variance or extremes (Kundzewicz and Robson 2000).

A complication arises in our analysis due the non-independence of daily flows (i.e., their autocorrelation) that are used to characterise the distribution of all flows. The non-independence of daily flows violates an assumption of the Kolmogorov-Smirnov test that the observations are independent (Lanzante 2021). This results in under-estimation of the variance and therefore under-estimation of the *p*-value. There are corrections that can be made to adjust for autocorrelation of data in Kolmogorov-Smirnov tests (e.g., Lanzante 2021). However, in this study we were primarily interested in the strength of the evidence that the two distributions are consistent with the same population distribution. We were therefore not concerned about under-estimation of the *p*-value because this leads to a more conservative assessment of the evidence for the two distributions being consistent with the same population (i.e., we are more likely to reject this). We therefore did not apply any correction to the Kolmogorov-Smirnov test *p*-value. For each site, we therefore calculated the Kolmogorov-Smirnov test statistic and its significance (Massey 1951) for rolling 5-year state assessment periods between 2013 and 2021. We interpreted the *p*-values >0.05 for these tests as strong evidence that the distribution of observation-date flows was an unbiased sample of all flows within each five-year assessment period.

4.4 Flow-adjustment

We first assessed the general level of association between water quality observations and instantaneous (daily mean) flow using the non-parametric Kendall rank correlation coefficent (known as Kendall's tau; τ). Kendall's tau is a measure of rank correlation; the similarity of the orderings of the data when ranked by each of the quantities (Zar 1999). We used τ to quantify the level of monotonic association between the water qualtiy observations and their associated instantaneous flows. Kendall's τ takes values between -1 and +1; a positive value indicating that the observations increased with increasing instantaneous flow and vice versa. For each site and variable combination, we calculated Kendall's τ and plotted the distributions of results over sites for each variable as box and whisker plots to indicate the general level of association between water quality observations and instantaneous flow and the inter-site variability of this association within each variable.

Flow-adjustment builds a statistical model of the relationships between water quality variable observations and instantaneous flow that is subsquently used to remove the confounding influence of instantaneous flow in trend analysis (Helsel et al. 2020; Snelder et al. 2021a). As mentioned in Section 2.2, expert judgment is required in selecting observation - instantaneous flow models because there are many alternative models and selection of the most appropriate model requires striking a balance between physical plausibility and goodness-of-fit. This means that there is more than one plausible model, which introduces unquantified uncertainties into trend analysis that we refer to in Section 2.4 as unquantified uncertainty of type B.

The purpose of the analyses of flow-adjustment in this study was to assess the extent to which selection of model representing the relationship between water quality observations influences the results of trend analysis. We used alternative plausible instantaneous flow - observation models to produce two sets of flow-adjusted water quality observations. We then undertook two sets of trend analyses based on the alternative data and compared the results.

There are a wide range of statistical regression methods (linear and non-linear) that have been, or could be, used to model the observations - instantaneous flow relationships. In this study we used a range of statistical regression models that have been used in previous studies to represent the observation – instantaneous flow relationship (e.g., Smith et al. 1996; Ballantine and Davies-Colley 2014; Larned et al. 2016). In each case the log (base 10) of flow was used as the independent model variable. The eight models were as follows:

1. linear model of untransformed water quality and log of stream flow (LinLog),
2. linear model of log of water quality and log of stream flow (LogLog),
3. locally estimated scatterplot smoothing (LOESS) with a span of 0.7 applied to untransfomed water quality and stream flow (LOESS 0.7),
4. as for 3, but with log of water quality(LOESS 0.7-Log),
5. LOESS with a span of 0.9 (LOESS 0.9),
6. as for 5 but with log of water quality (LOESS 0.7-Log),
7. generalised additive model (GAM) with smoothing spline local fitting ,
8. as for 7 but with log of water quality (GAM-Log).

LOESS and GAM models allow more fexible fit to the data and can represent non-linear relationships. For LOESS models, the span refers to the proportion of points that are considered when calculating the weighted local regression at each point. A large span produces a smoother more global fit than a smaller span and a smaller span produces a model that conforms more to the local data. We did not trial LOESS models with spans less than 0.7 because our experience is that these almost always result in implausible modeled relationships.

Only site and variable combinations with at least 54 observations were included in this analysis. This limit was chosen because it represents the minimum observations required to conduct a water quality trend assessment for a 5 year period with monthly sample intervals and with the commonly used requirement that observations are available for at least 90% of sample intervals (Whitehead et al. 2021).

The model goodness-of-fit was assessed using the coefficient of determination (r^2) and the model *p*-value. The model r^2 indicates the proportion of the variability in the water quality observations that is explained by (log of) the flow. The model *p*-value indicates the degree of evidence that the fitted relationship is consistent with the population. Low *p*-values indicate that the fitted relationship would be unlikely were there no relationship between the observations and flow in the population.

To keep all comparisons to site and variable combinations for which an instantaneous flow - observation relationship was considered objectively robust, combinations for which none of the eight possible statistical models had $r^2 > 20$ % and $p < 0.01$ were discarded. To enable comparison between the models, we always calculated r^2 using the raw water quality values and the log of the flow, irrespective of whether the relationship was derived between the log of the water quality variable and the log of the stream flow. The r^2 and p thresholds were subjectively chosen. Flow-adjustment of data that does not achieve these thresholds is unlikely to have a noticeable effect on the trend assessment.

For all site and variable combinations that met the above criteria, all eight models were considered by the expert. This was aided by producing scatter plots of the data (observation - instantaneous flow) with all eight models super imposed on the plot. In addtion, the r^2 and p values for each model were provided. The expert selected the "most suitable" model based on three considerations recommended by Snelder *et al.* (2021) including: (1) homoscedasticity (constant variance) of the regression residuals, (2) model goodness-of-fit measures and (3) plausibility of the shape of the fitted model. It is noted that homoscedasticity of the regression residuals indicates that the model fits through the central tendency of the data and is not overly influenced by particular values.

We used the LOESS 0.9 model to provide an alternative "default model" to that selected by the expert. The purpose of the default model is to provide an alternative flow-adjustment to indicate how sensitive trend analysis results are to this choice. The choice of the LOESS 0.9 model is subjective. In our experience LOESS 0.9 is a reasonable compromise between the purely linear model (i.e., LinLog), which is often unable to represent increasing rates of change in observations at high flows and more flexible models (e.g., LOESS 0.7), which can result in implausible modeled relationships.

Flow-adjustment was carried out by subtracting the observed water quality values from the corresponding values predicted by the using the expert-selected and LOESS 0.9 models (i.e., to obtain the model residuals). The model residuals are used as input for flow-adjusted trend assessment.

To visualise the impact of flow-adjustment and the temporal variation in the observations, we produced time series plots of two sets of flow-adjusted data and the raw (i.e., non-flow-adjusted). To compare the flow-adjusted and raw time series, we offset the residuals of the adjusted values by the median of the raw observations so that their magnitudes were consistent with the raw values. Note that the offset-residuals may include negative values where the flow-derived water quality estimate is larger than the observed water quality.

4.5 Trend analysis

We assessed trends for all 81 site and variable combinations that we had flow-adjusted (see Section 4.4). Trends were assessed for rolling assessment periods of 10 years that increment by one year, starting from 2007-2017 (inclusive) and finishing with 2011-2020 (inclusive). The trend assessment was carried out with both sets of flow-adjusted data (i.e., based on a model selected by an expert and based on the default LOESS 0.9 model see Section 4.4) and the raw (i.e., unadjusted) data.

To assess variability in trend assessments of differing time-period and duration over the longest possible timescale, we performed rolling 5- 10- and 20-year trend analyses for selected variables for two sites with the longest monitoring records in the Region: Hoteo River (NIWA) and Rangitopuni River (NIWA). We used these sites due to the continuity of record between 1989 and 2020 but note that NIWA has discontinued monitoring at Rangitopuni from 2021 and the continuation of these sites is represented by the AC site Rangitopuni River @ Walkers / 7805. We calculated trends for four relevant water quality variables (NNN, DRP, NH4N, and CLAR) for which there was a continuous record of monthly observations from 1989 to 2020 (inclusive). We note that in these assessments we used measured water clarity rather than calculated from turbidity (see Table 2). Trend assessments were performed for time-periods of 5, 10 and 20 years duration starting in 1990 with time-periods incrementing by one year until the end of 2020 was reached. This resulted in 27, 22 and 12 trend assessment periods of 5, 10 and 20 years duration, respectively. These trend assessments were performed with one set of flow-adjusted data (i.e., based on models selected independently by the expert, see Section 4.4) and the raw (i.e., unadjusted) data.

Trends were assessed using the methods set out in Whitehead et al. (2021), which for brevity have not been reproduced in their entirety in this report. Briefly, trends were analysed in four steps.

- First, we filtered the sites and variable combinations that had been flow-adjusted and, for each trend assessment period, retained those combinations for which there was observations for at least 90% of the sampling intervals in that period.
- Second, for each site, variable and assessment period combination, we assessed the seasonality of observations using the Kruskal Wallis test. Where there was a statistically significant difference in the observations grouped by month (Kruskal Wallis test $\alpha \leq 0.05$), we categorised the data as seasonal.
- Third, for each site, variable and assessment period combination we calculated the trend direction (D) and confidence in this evaluation (C) using either the Seasonal Kendall statistic or the Mann–Kendall statistic; depending on whether the data were seasonal or non-seasonal, respectively (Snelder et al. 2022).
- Fourth, for each site, variable and assessment period combination we calculated the rate of change using the Seasonal Sen slope or Sen slope, depending on whether the data were seasonal or non-seasonal, respectively (Snelder et al. 2021a). Sen slopes were expressed as annual values and where appropriate, were expressed as relative annual Sen slopes by dividing by the median value of the observations. Confidence in the assessed rates of change were expressed as the 90% confidence interval of the Sen slope.

The impact of choices made in the flow-adjustment process on the evaluated trends were assessed by comparing the two sets of flow adjusted trends for the 2011 – 2020 assessment period. The trends assessed from the raw data (i.e., not flow-adjusted) were also compared to the flow-adjusted trends that were based on models selected independently by the expert to assess the impact of flow-adjustment on the trend assessments.

Trend assessment results were compared graphically using scatter plots of confidence the trend direction was decreasing and Sen slopes. Confidence the trend direction was decreasing[6] was calculated from the assessed confidence in trend direction as:

[6] For all 10 variables shown in Table 2, other than clarity, decreasing trends indicated 'improving' water quality. Decreasing clarity trends indicate 'degradation'.

$$C_d = \begin{cases} C & if\ D = \ Decreasing \\ 1 - C & if\ D = \ Increasing \end{cases}$$

In graphical reporting of some trend assessments in this report, we express C_d categorically using the same simplified classification system (outlined in Table 3) used by Land Air Water Aotearoa (LAWA[7]):

Table 3. Level of confidence categories used to convey confidence trend direction was decreasing.

Categorical level of confidence the trend was decreasing	Range in C_d
Very likely	0.90 – 1.0
Likely	0.67 – 0.90
As likely as not	0.33 – 0.67
Unlikely	0.10 - 0.33
Very unlikely	0 – 0.1

4.6 Alternative models of the relationships between water quality observations, time and flow

The methods for describing observation – flow relationships and trends (i.e., observation – time relationships) described above make simplifying assumptions. While these methods are accepted practice, the impact of these simplifications need to be kept in mind when the outputs are used in decision-making. Three of the most obvious simplifications are:

1. the models fitted to the observation – flow data for each variable-site combination represent a bivariate characterisation of an assumed general water quality-flow relationship that remains equally applicable throughout the period of record regardless of season or antecedent flow conditions,
2. the trend assessment method assumes that the seasonal pattern in the data is repeated annually and remains the same throughout the period of record,
3. the trend assessment method assumes a specific (monotonic) functional form for the relationship between observations and time.

The first and second simplifications imply that the mathematical shape or form of the observation – flow relationship remains constant over the record. If we consider that conditions in the catchment have changed over time and that these changes have influenced water quality (e.g., urbanisation, changes in land management, changes in flow regime due to increased water use), then this assumption may not be reasonable and may contribute to uncertainty in assessments made using the model.

The third simplification implies that trend analyses can only detect monotonic (i.e., increasing or decreasing) trends over the time-period being assessed. This limitation will mean that cyclic temporal patterns and trend reversals will not be detected.

[7] https://www.lawa.org.nz/explore-data/auckland-region/river-quality/

To provide alternative descriptions of observation – flow relationships and trends to those derived using the methods described above, we fitted models to selected water quality data at two sites in the Region using the Weighted Regression on Time, Discharge, and Season (WRTDS) method (Hirsch et al. 2010). The WRTDS method provides for considerable flexibility in representing the long-term trend, seasonal components, and discharge-related components of the behaviour of the water-quality variable of interest. However, this flexibility comes at the expense of requiring more data. Fitting a WRTDS model requires that the number of samples collected at the sampling site is more than 200 and the period of sample collection is at least 20 years. In addition, model fitting requires a complete record of daily flow values for the site over the entire period being modelled.

The WRTDS method expresses concentration as a function of time, discharge, and season with the following form:

$$ln(\widehat{C}) = \beta_0 + \beta_1 ln(Q) + \beta_2(t_i) + \beta_3 sin(2\pi t) + \beta_4 cos(2\pi t) + \varepsilon$$

where, $\hat{C}$ is the predicted concentration of the water quality variable, Q is the flow rate, the β values are fitted parameters, t is the time in years and ε is the unexplained variation. The functional form is linear in t, linear in ln(Q), and sinusoidal on an annual period (i.e., season). However, the method of fitting the model means that the parameter values are not constant throughout the entire domain of the data but vary over the explanatory variable space defined by Q and t. This is achieved by weighting the observations based on their relevance to the point in the explanatory variable space being considered (referred to by Hirsch et al. 2010 as an estimation point Q_0, T_0). Thus, observations that are close to Q_0, T_0 have a strong influence on the parameter values at that point in the explanatory variable space and the influence decreases the further the observation is from the estimation point. This approach has the following advantages over the methods described above:

1. The observation – flow relationship is allowed to change smoothly over time.
2. The trend component is not constrained to be any particular functional form and is allowed to change smoothly over time.
3. There is no assumption that the seasonal pattern repeats but rather the shape of the seasonal pattern is allowed to change smoothly over time.

These advantages mean that a WRTDS model can detect and fit both long term (secular) trend, short term fluctuations, as well as cyclic seasonal variability that evolves over time. Collectively this allows for more realistic representation of how water quality changes and increases the potential to understand the drivers of change.

A WRTDS model includes a ‘‘flow-normalisation’’ procedure that has a similar motivation to flow-adjustment; to remove the association between water quality and flow regime variation that happen to have occurred during the monitoring period and thereby describe the water quality outcome that would have occurred under “average” flow regime. The weighted regression approach to fitting a WRTDS model means “flow-normalised” predictions are not simply adjustments for instantaneous flows but account for flow regime variability in water quality at longer timescales.

The performance of a fitted WRTDS model is assessed using “leave-one-out cross validation” (Hirsch and De Cicco 2015). This procedure leaves one observation out of the fitting dataset, fits a model to the remaining observations and uses that model to estimate the concentration for the left-out observation. This step is repeated for all observations in the dataset producing a set of independent predictions for each observation. These independent predictions can be used to quantify various measures of model performance (Hirsch and De Cicco 2015). In this study, we fitted a linear regression of the observations against the predictions and used the R^2 value of this regression to describe the performance of the model. We note that WRTDS can also be used to assess trends by assessing the magnitude and significance of differences in predicted concentrations between dates of interest (Hirsch et al. 2015). However, we did not make use of this capability of WRTDS in this study.

We fitted WRTDS models to the two sites with the longest and most consistent monitoring records in the Region: Hoteo River (NIWA) and Rangitopuni River (NIWA). We fitted models to four relevant water quality variables (NNN, DRP, NH4N, and CLAR) for which there was a continuous record of monthly observations from 1989 to 2020 (inclusive). These assessments used measured water clarity rather than calculated from turbidity (see Table 2). Note that we also assessed rolling 5- and 10-year trends for these site variable combinations as described in Section 4.5.

We produced several types of graphical output from these models and examined these to assess whether:

- the modelled evolution of water quality state indicates a combination of long term (secular) trend and short-term fluctuations?
- the modelled observation –flow relationships change over time?
- the modelled observation –flow relationships differ between seasons?

In addition, we examined whether short term fluctuations were reduced when the model was used to generate flow-normalised concentration predictions and whether assessments of water quality state made using the monitoring data were consistent with the flow normalised predictions made by the WRTDS models

5 Results

5.1 State assessment

Assessments of current state for attributes considered here for the five-year period ending 2020 and the precision of these estimates expressed as 95% confidence intervals are shown Figure 2 to Figure 5. Supplementary data with the results of trend analyses pertaining to rolling 5-year state assessment periods between 2013 and 2020 are provided in WhatFile.xlsx.

The width of the confidence intervals, compared to the difference between adjacent NOF attribute band thresholds, differed by variable and statistic. For some variables and statistic combinations, the width of the confidence intervals was small compared to the difference between adjacent NOF band thresholds. For example, for the nitrate median metric, the 95% confidence interval was entirely contained within the A Band for most sites (Figure 2). In contrast, for DRP 95th percentile, the 95% confidence interval often extended over two, three or even four bands (Figure 3). Other than for pH adjusted ammonia and nitrate (Figure 2), at most sites, the 95% confidence intervals for most combinations of statistics and variables, extended over more than one NOF band.

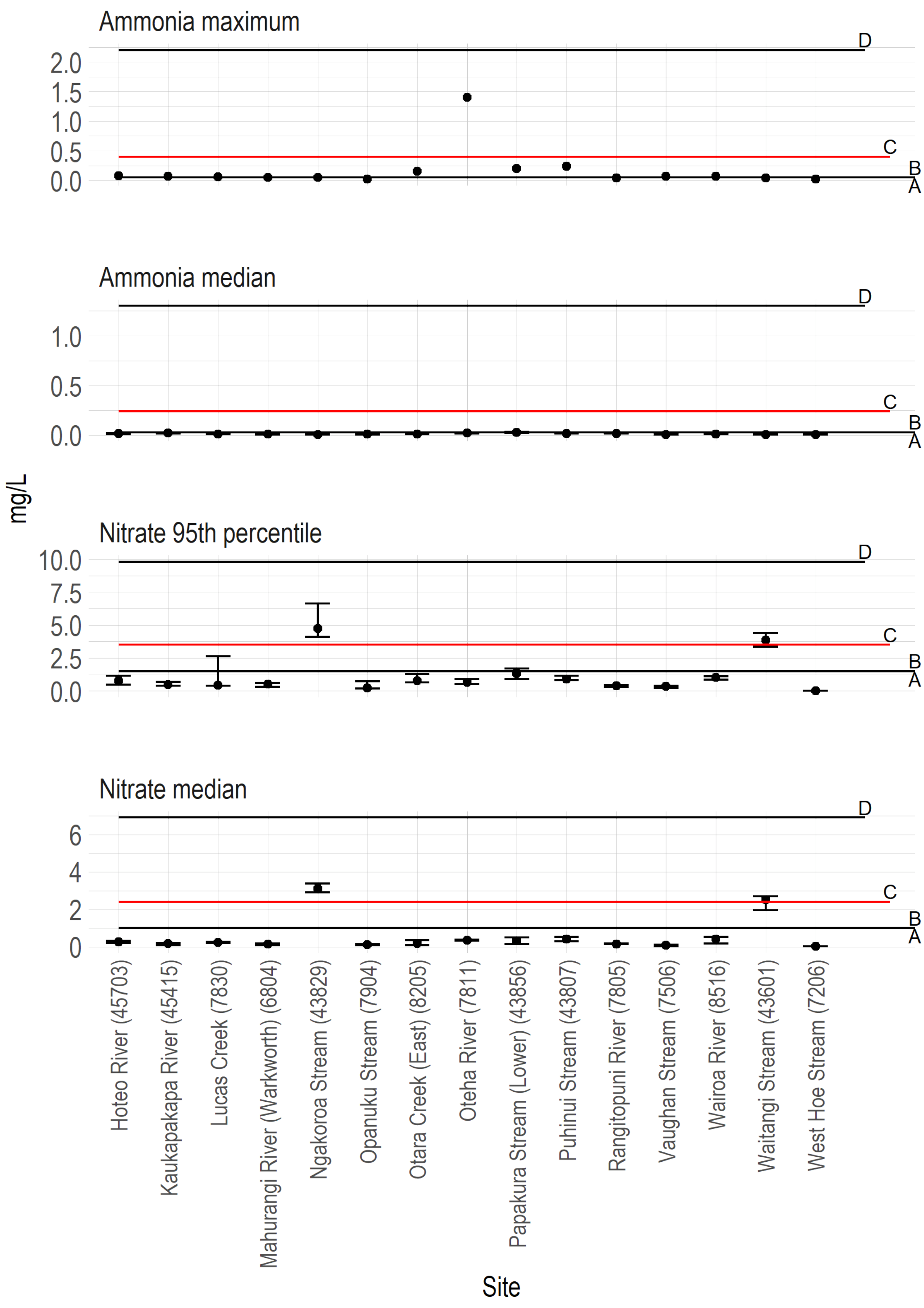


Figure 2. Attribute state of pH adjusted ammonia and nitrate concentrations for the 2016-2020 period. The precision of these assessments is indicated by the width 95% confidence intervals. Lettered horizontal lines denote the NOF attribute band thresholds. The red lines are reminders that the NBL for these attributes are the B/C threshold and are high concentrations in contrast to low values for Clarity in Figure 3.

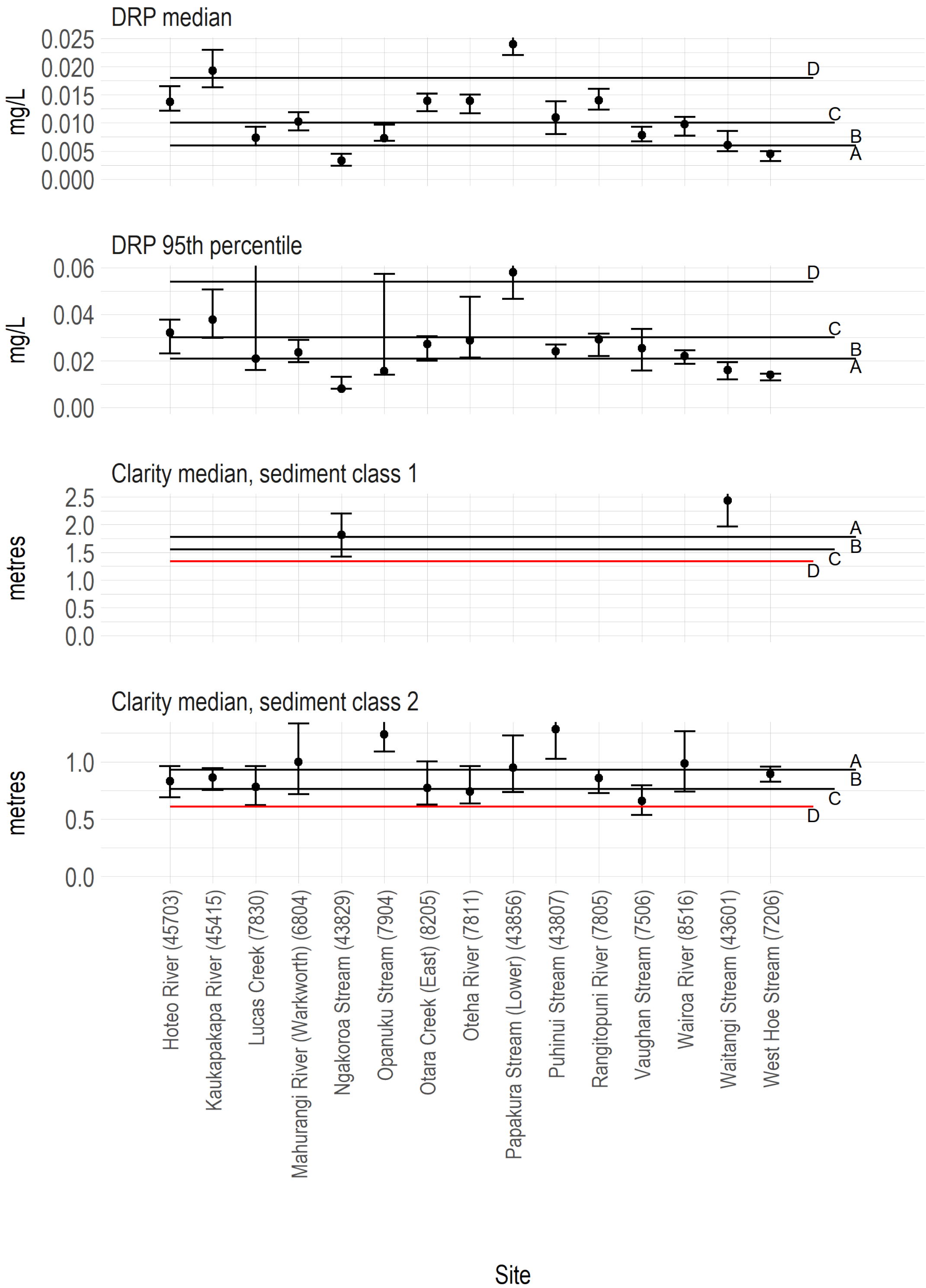


Figure 3. Attribute state of DRP concentrations and clarity for the 2016-2020 period. The precision of these assessments is indicated by the width 95% confidence intervals. Lettered horizontal lines denote the NOF attribute band thresholds. The red lines are reminders that the NBL for Clarity is the C/D threshold, which is a is a low value.

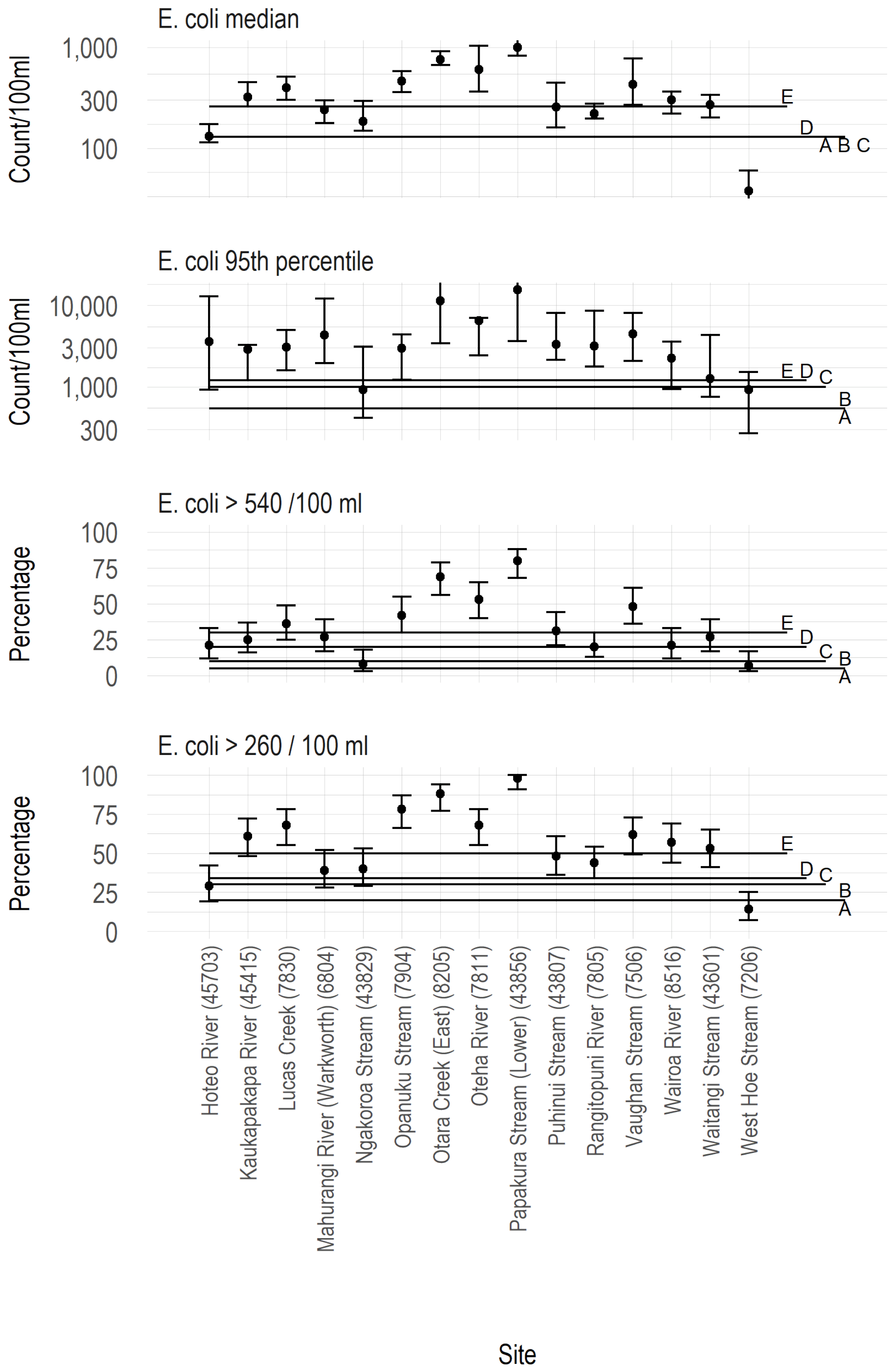


Figure 4. Attribute state of E. coli for the 2016-2020 period. Note log-10 scale for the Count attributes. The precision of these assessments is indicated by the width 95% confidence intervals. Lettered horizontal lines denote the NOF attribute band thresholds. Where data are not shown there was insufficient to assess state.

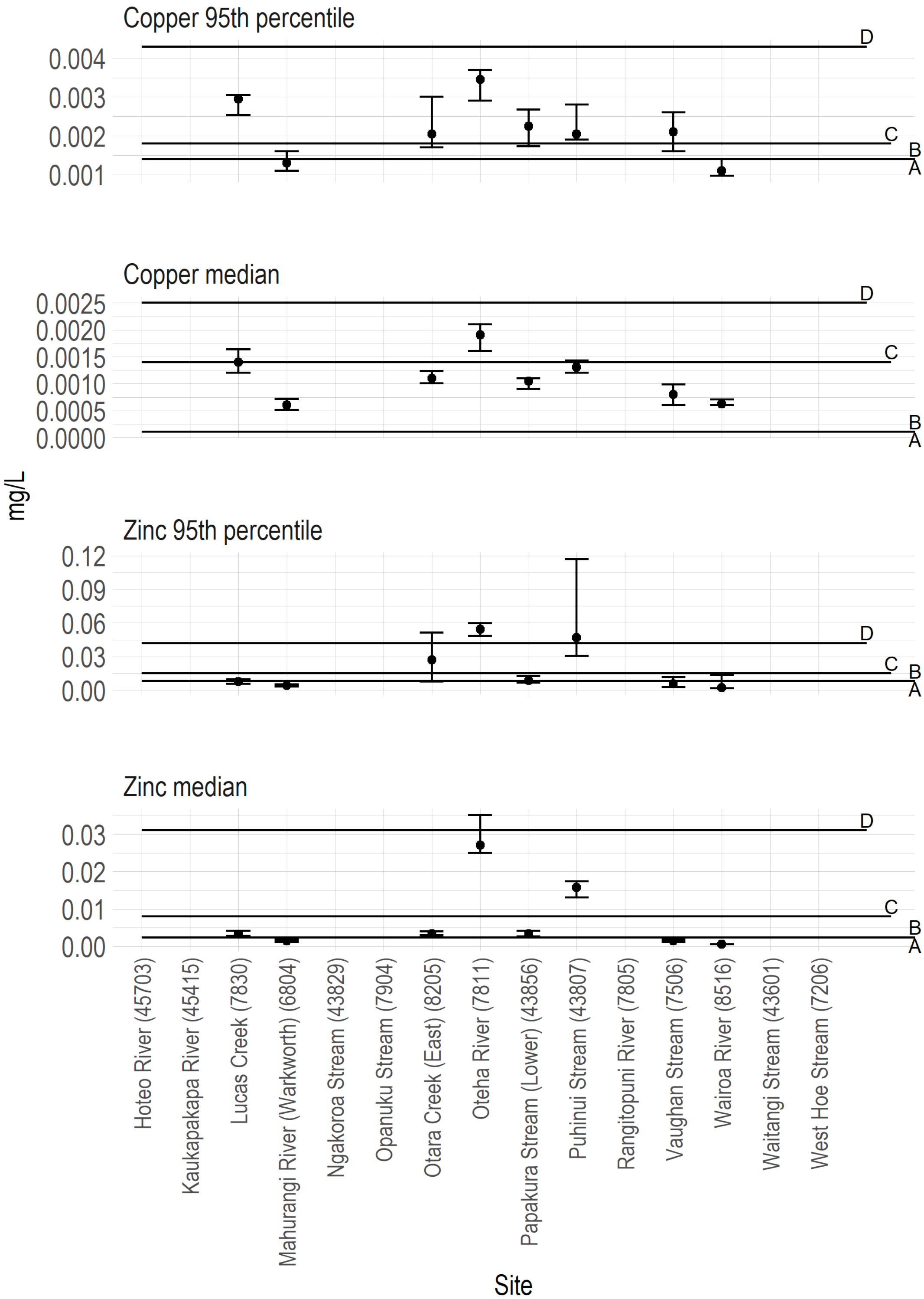


Figure 5. Attribute state of Copper and Zinc concentrations for the 2016-2020 period. The precision of these assessments indicated by the width 95% confidence intervals. Lettered horizontal lines denote the proposed regional attribute band thresholds (Gadd et al. 2019).

5.2 Variability of flow regimes

At all 15 sites, flow regimes differed between assessment periods and were different to the long-term flow regime (Figure 6). At many sites, mean flows in the four assessment periods were higher than the long-term flow regime between approximately February and May and flows between approximately September and December were lower than the long-term flow regime. However, patterns of differences between the four assessment periods and the long-term flow regime varied between sites. For example, mean flows in the four assessment periods were lower than the long-term flow regime between approximately February and May for the Vaughan Stream, West Hoe Stream and Kaukapakapa River. There were also marked differences in flow regimes between assessment periods within sites. For example, at many sites, mean flows were higher in September for the assessment period ending 2017 than other assessment periods and this assessment period was generally associated with lower flows than other assessment periods between approximately January and May (Figure 6). Overall, the plots indicate that there is considerable variability in mean monthly flows between assessment periods.

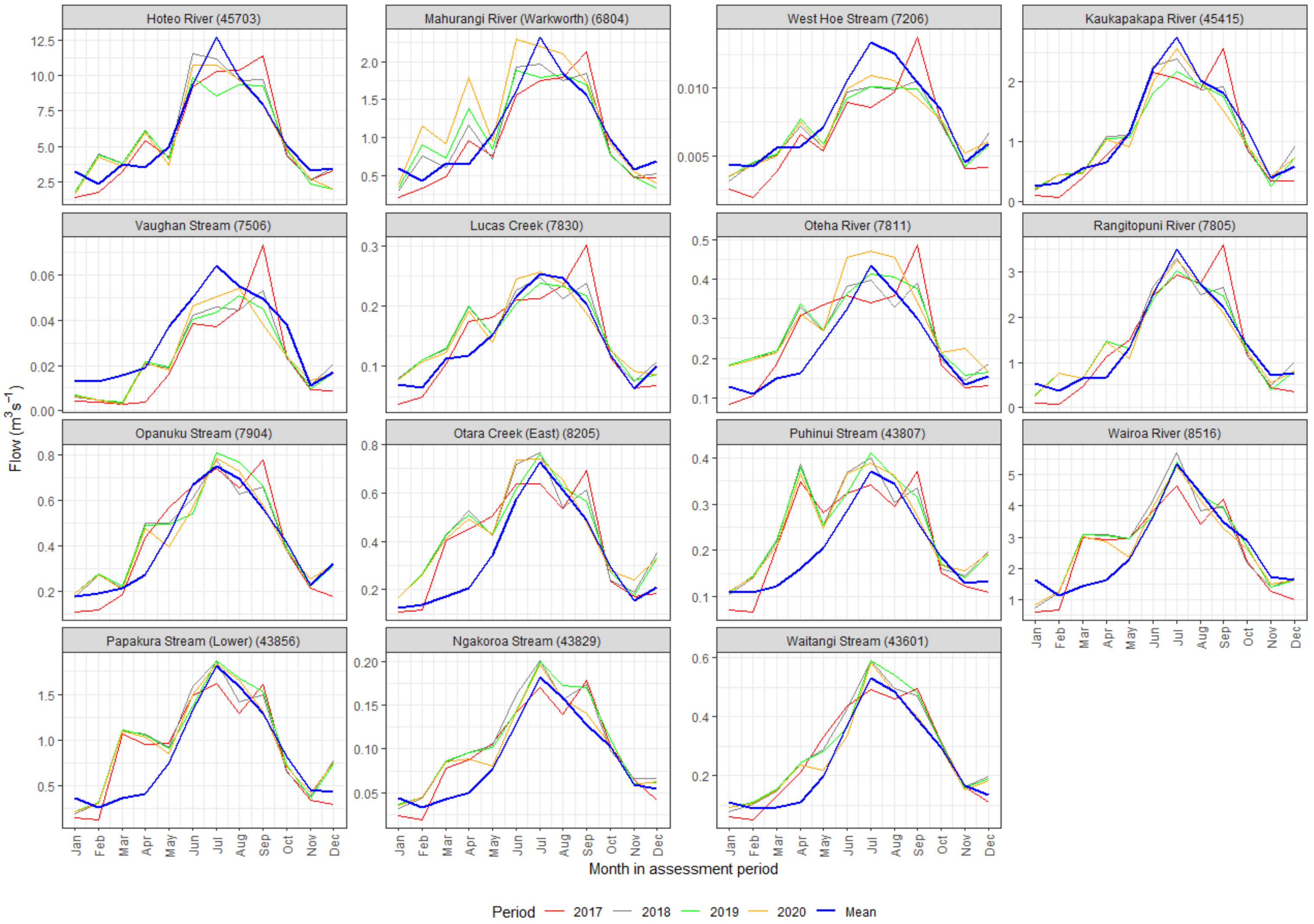


Figure 6. Flow regimes in assessment periods and the long-term flow regime at each site. The plotted data represents the mean flows in each month for the four assessment periods and the over the entire flow record for the site (see Table 8 for length of record at each site).

5.3 Representation of instantaneous flow by water quality observations

5.3.1 Graphical representation of flows and observations

Figure 7 is an example flow timeseries (hydrograph) for the Oteha Stream between 2013 and 2021 with the water quality observations super-imposed. Most observations occur at low flows but some represent higher flows, especially between 2019 and 2020. Hydrographs for all the sites are provided in Appendix B, Figure 32.

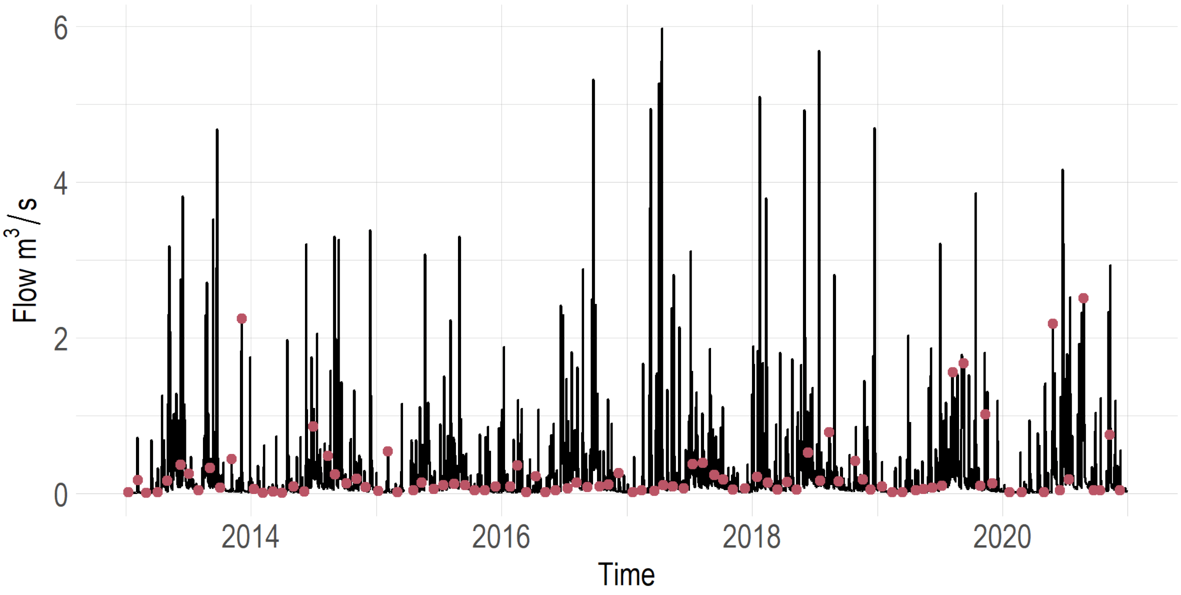


Figure 7. Oteha Stream hydrograph based on mean daily flow (black line) and superimposed observation-dates (maroon dots).

The flow distribution curve (FDC) for Oteha Stream between 2013 and 2020 is shown in Figure 8 with the corresponding water quality observations superimposed. For the Oteha Stream site, observations generally occurred across the flow range but 83% and 41% of the observations occur at flows that are lower than the mean and median flows; respectively. Therefore, a relatively small proportion of the observations represent high flows (i.e., > mean flow).

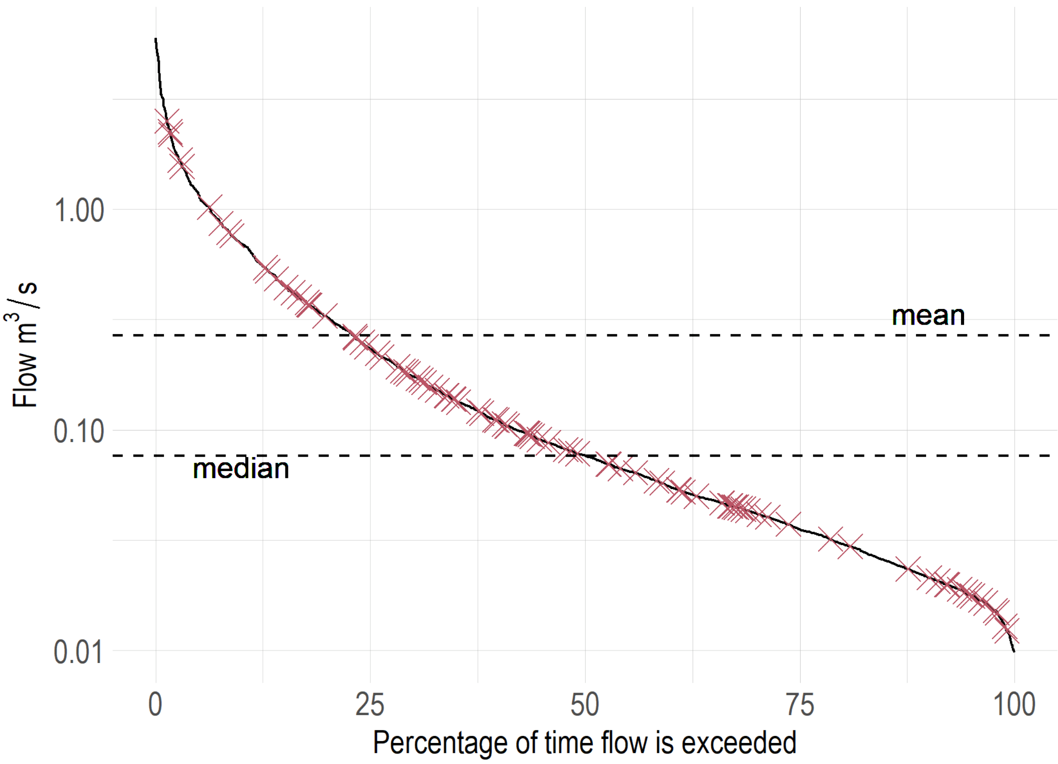
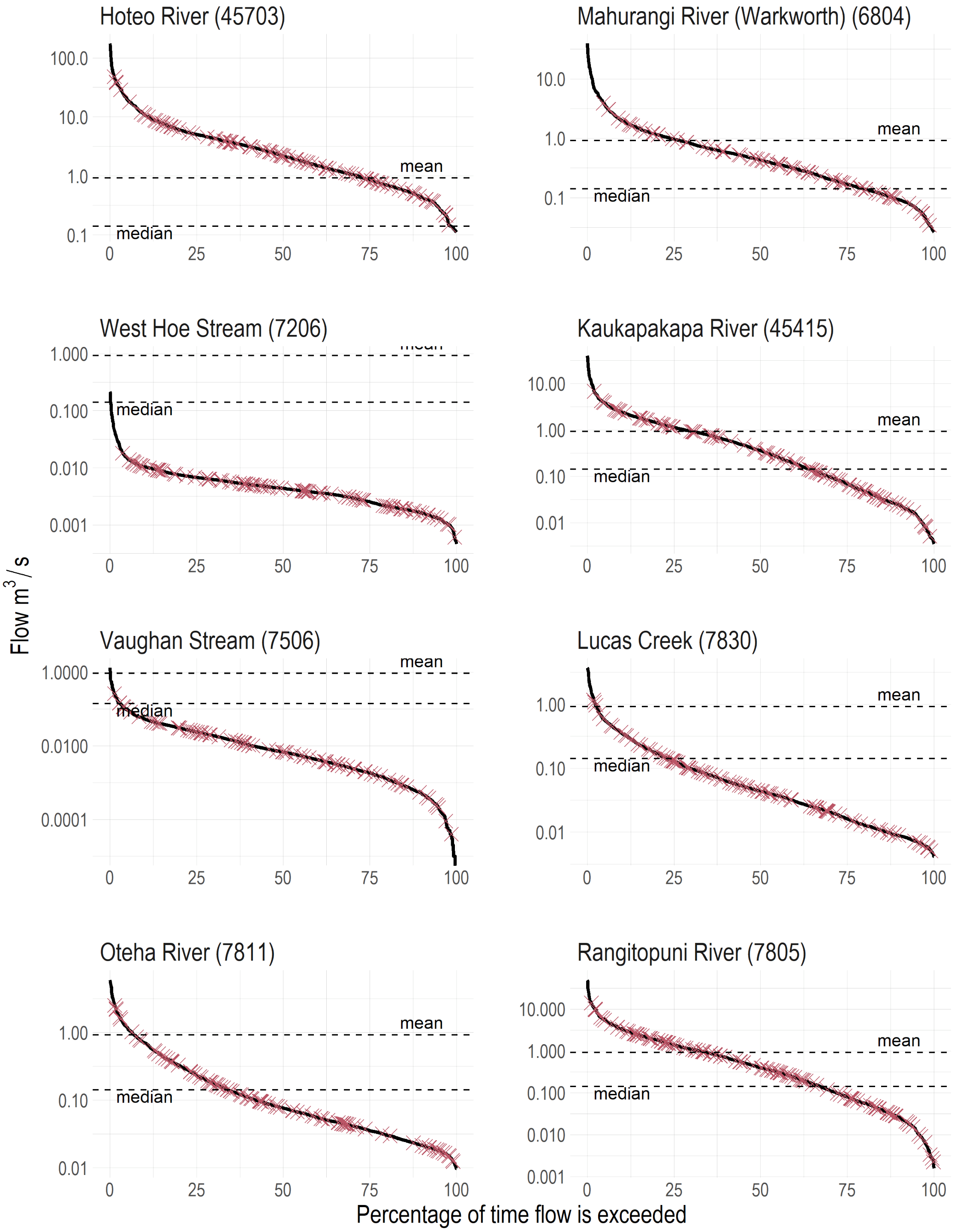


Figure 8. Oteha Stream flow duration curve (black line) with observation-date flows (maroon crosses). Note that the Y-axis has also been log transformed to improve the discrimination of flow differences.

Figure 9 shows FDCs with overplotted water quality observations for each of the sites for the 2013 to 2020 period. These plots indicate that there is between-site variation in how well the water quality observations represent the full distribution of flows. Some sites (e.g., Opanuku Stream) have observation-date flows covering most of the flow distribution. Many of the sites do not have observations at high flows as indicated by the exposed (black line) at the high flow (i.e., left-hand side) of the FDC.


Hoteo River (45703)
Mahurangi River (Warkworth) (6804)
West Hoe Stream (7206)
Kaukapakapa River (45415)
Vaughan Stream (7506)
Lucas Creek (7830)
Oteha River (7811)
Rangitopuni River (7805)
mean
median
Flow m^3/s
Percentage of time flow is exceeded

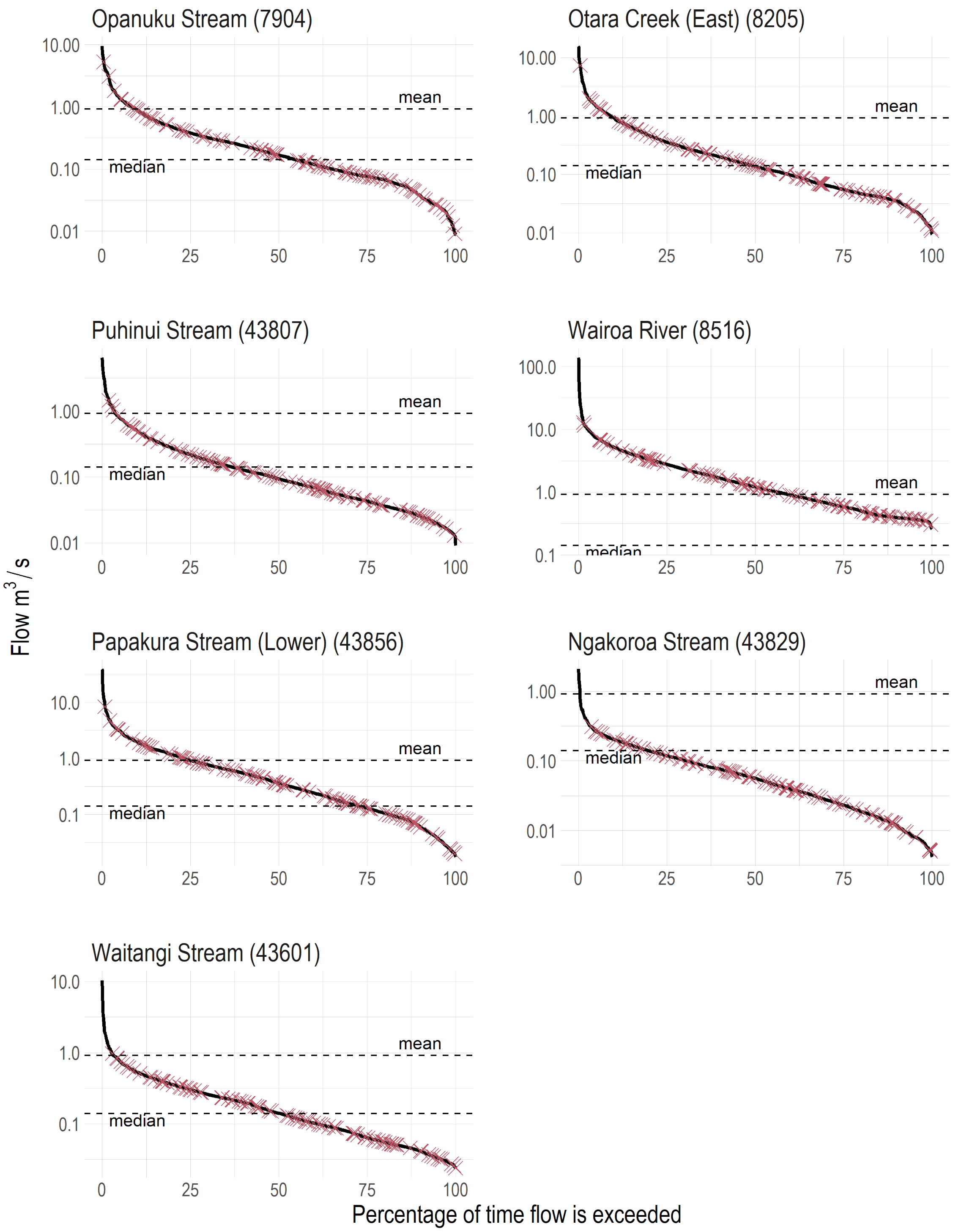


Figure 9. Flow duration curves for each site for the 2013 to 2020 period. The black lines indicate the flows and the corresponding water quality observations are shown as maroon crosses. Note the log scale on the y (flow) axes.

5.3.2 Assessment of water-quality-observation-date flow distributions

Figure 10 shows FDCs for the Oteha Stream for 5-year assessment periods between 2013 and 2020 with the corresponding water quality observations super-imposed. For each assessment period, the water quality observations generally represent most of the flow distribution.

Table 4 lists the percentage of flow range and volume represented by the observations for each site and 5-year assessment period. On average, flows on water quality observation-dates in the five-year assessment periods represented 96% of the range of flows in the period. Water quality observation-dates best represented the flow range in the 2013 to 2017 period when, on average observations represented 97% of the flow range. Several sites had 5-year periods when observations represented 99% of the flow range. The least representative water quality observations were for West Hoe Stream for the 2014 to 2018 and 2015 to 2019 periods when observations represented only 91% of the flow range.

On average, 76% of the stream flow volume is within the observation-date flow range (Table 4). This is lower than the proportion of the flow range that is represented by the observations because the volume attributable to a given range of flows is equal to the area under the FDC that pertains to that range. Therefore, the high-flows that are not represented by the observations, account for a disproportionately large amount of total volume of water passing a flow station. Water quality observation-dates best represented the flow volume for the 2016 to 2020 period when 78% of the volume was represented by observations. The least representative water quality observations were for 2014 to 2019 when only 63% of the flow volume was represented by observations. Note that the worst representative period from a volume perspective can be different to the worst from a time perspective if the missing high flows from each period are of different magnitudes.

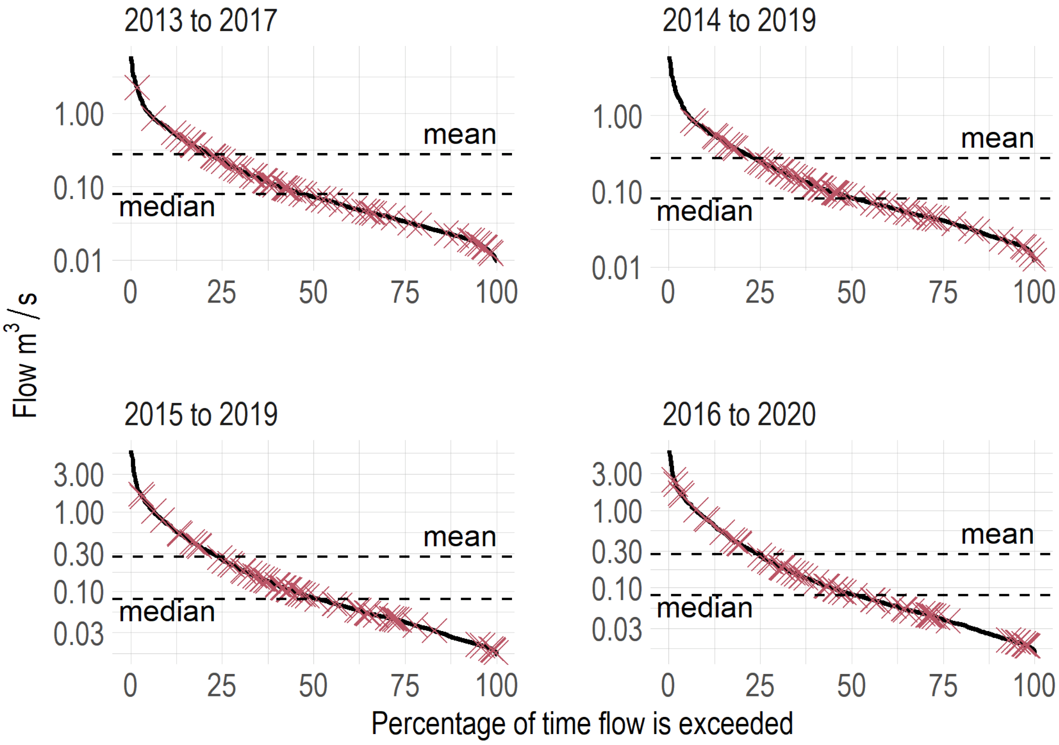


Figure 10. Oteha Stream flow duration curves for each 5-year period. The black lines indicate the flows and the corresponding water quality observations are shown as maroon crosses. Note the log scale on the y (flow) axes.

Table 4. Proportion of time or volume of streamflow that was represented by water quality observations for rolling 5-year periods between 2013 and 2020. Periods without at least 90 % of flows (>1642 days) or samples (> 53) are omitted

Site	2013 to 2017		2014 to 2018		2015 to 2019		2016 to 2020	
	Time	Vol	Time	Vol	Time	Vol	Time	Vol
Average of all sites	**97**	**76**	**96**	**73**	**96**	**77**	**96**	**78**
Hoteo River (45703)	97	71	97	77	98	82	95	81
Mahurangi River (Warkworth) (6804)	94	61	95	59	-	-	-	-
West Hoe Stream (7206)	95	71	91	63	91	67	93	67
Kaukapakapa River (45415)	97	73	98	74	95	62	93	61
Vaughan Stream (7506)	95	58	-	-	-	-	-	-
Lucas Creek (7830)	97	72	92	50	96	65	98	73
Oteha River (7811)	97	77	93	54	97	71	97	82
Rangitopuni River (7805)	97	70	99	82	99	82	99	83
Opanuku Stream (7904)	99	93	99	92	99	94	99	93
Otara Creek (East) (8205)	99	91	98	85	98	86	97	85
Puhinui Stream (43807)	97	78	94	74	96	75	97	76
Wairoa River (8516)	98	82	97	80	96	81	98	82
Papakura Stream (Lower) (43856)	99	87	98	81	98	82	99	82
Ngakoroa Stream (43829)	96	79	95	76	95	75	94	72
Waitangi Stream (43601)	96	76	96	75	95	75	95	73

The results of the Kolmogorov-Smirnov tests of differences between the flow distributions (i.e., FDCs) pertaining to state assessment periods to the corresponding distribution of flows associated with observations are shown in Table 5. These tests indicate that in all cases the *p*-value is greater than 0.05, which conservatively indicates that the null hypothesis is not rejected (because any autocorrelation in the data leads to under-estimation of the *p*-value). The conclusion from this is that the distribution of flows on sampling occasions are always consistent with the full flow distribution and therefore the samples are not biased with respect to flow.

Table 5. Kolmogorov-Smirnov statistics comparing five-year observation-date flow distributions to the FDC for the whole the assessment period.

Site	2017	2018	2019	2020
Hoteo River (45703)	0.08	0.07	0.06	0.06
Mahurangi River (Warkworth) (6804)	0.10	0.11	0.08	0.10
West Hoe Stream (7206)	0.08	0.09	0.10	0.09
Kaukapakapa River (45415)	0.13	0.12	0.09	0.09
Vaughan Stream (7506)	0.07	0.09	0.09	0.08
Lucas Creek (7830)	0.08	0.09	0.09	0.09
Oteha River (7811)	0.11	0.10	0.13	0.11
Rangitopuni River (7805)	0.06	0.07	0.07	0.05
Opanuku Stream (7904)	0.16	0.14	0.09	0.09
Otara Creek (East) (8205)	0.10	0.09	0.12	0.11
Puhinui Stream (43807)	0.08	0.10	0.10	0.09
Wairoa River (8516)	0.09	0.07	0.06	0.07
Papakura Stream (Lower) (43856)	0.06	0.07	0.04	0.07
Ngakoroa Stream (43829)	0.06	0.06	0.06	0.06
Waitangi Stream (43601)	0.06	0.07	0.06	0.06

5.4 Relationships between water quality observations and flow

The 10 water quality variables (Table 2) were generally monotonically associated with instantaneous flow as indicated by Kendall's τ (Figure 11). Observations of forms of nitrogen (NNN, and TN) were most strongly associated with instantaneous flow and DRP was least associated. Observations could be positively or negatively associated with flow but were predominantly positive for concentrations and negative for CLAR. The associations reflect generally increasing concentrations with flow. For CLAR, the association is reversed because visual clarity decreases with increasing concentrations of particulate material in the water column.

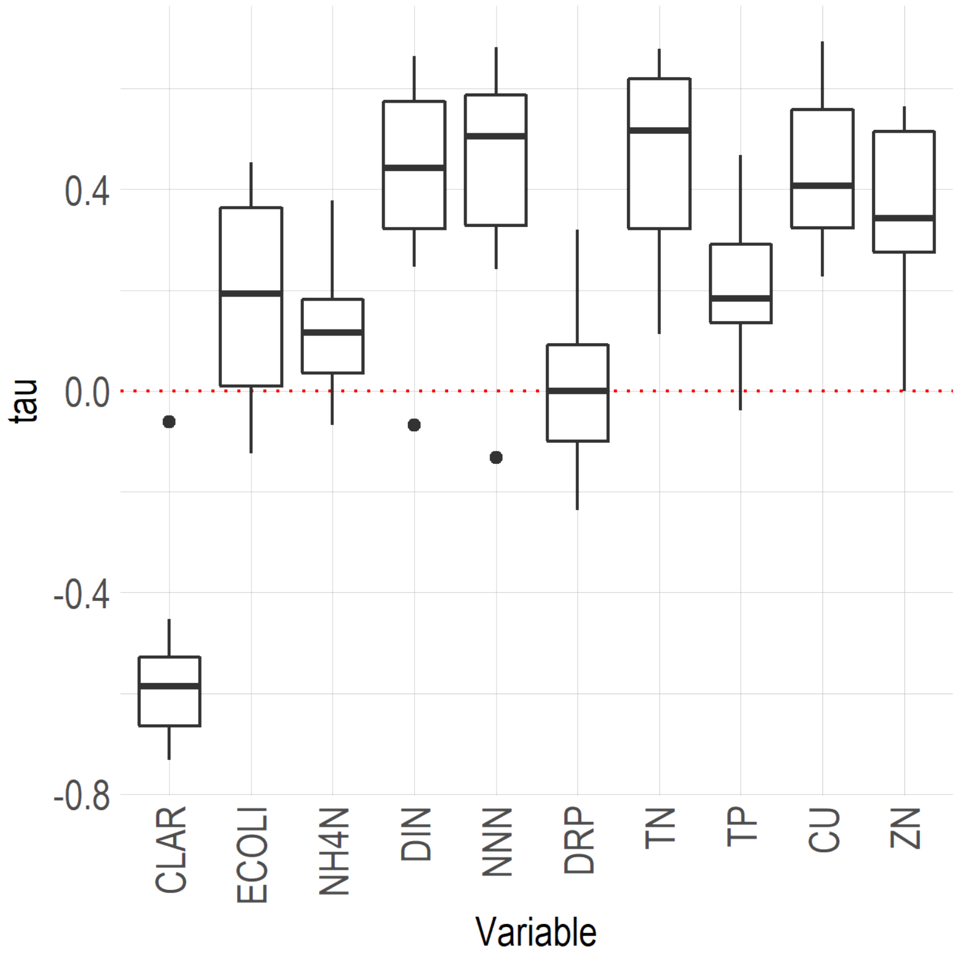


*Figure 11. Box and whisker plots showing the distributions of Kendall's τ measuring the correlation between observations and instantaneous flow at each site by variable. The black horizontal line in each box indicates the median of site Kendall's τ, and the box indicates the inter-quartile range (IQR). 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. The horizontal red line indicates a Kendall's τ value of zero.*

Of the 202 site-variable combinations that had a sufficient number of observations for flow-adjustment, 81 combinations were considered objectively robust (i.e., had observation – instantaneous flow models with r^2 values greater than 0.2 and p values < 0.01). These 81 combinations are detailed in Appendix D Table 9, and were investigated further.

Flow and water quality data from the Wairoa River (site 8516) are used below as an example of modelled relationships between water quality and flow. Plots of instantaneous flow - NNN concentration relationship for the Wairoa River are shown in Figure 12. Observations are predominantly at low flows with few observations at higher flows (e.g., > 10 m^3/s). Note that a log (base 10) transformation of the flow axis spreads the sample values more evenly across the range of flows and displays the relationship between concentration and flow in a more linear fashion (Figure 12, right hand side). For the remainder of plots presented in this section, the flow axes are log transformed.

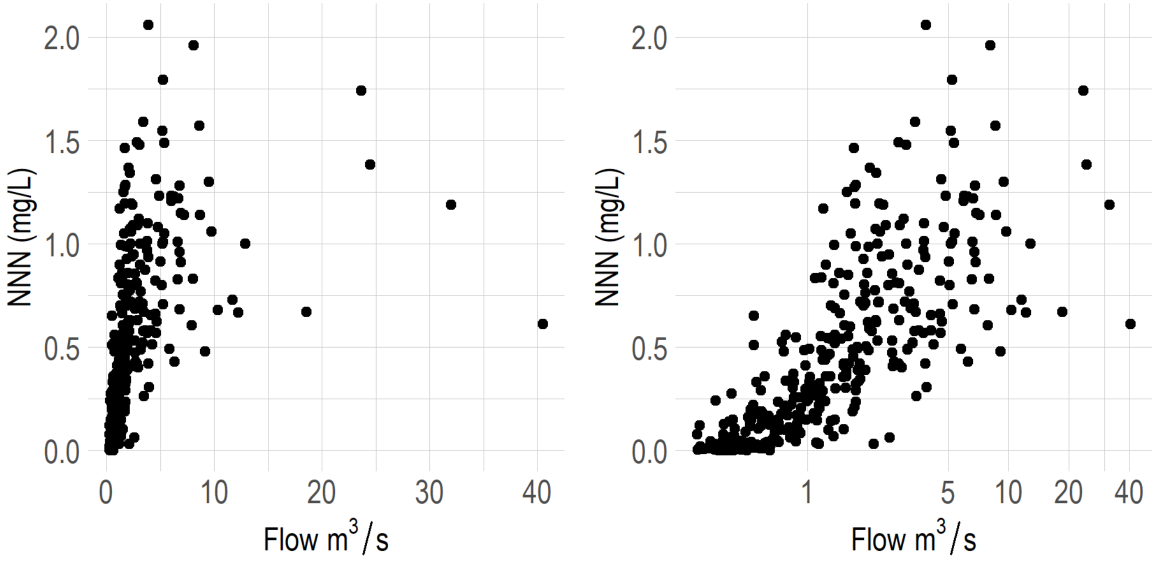


Figure 12. Plots of NNN vs flow for Wairoa River with linear scale for flow (left hand side) and log (base 10) transformed flow (right hand side).

The r^2 values for each of the eight potential instantaneous flow - observation models and variable (Table 2) combinations for the Wairoa River site are shown Table 6. Some of the models pertaining to ZN and two DRP models had $r^2 < 20\%$ and *p*-values > of 0.01. This indicates that we consider that there is insufficient statistical support for the instantaneous flow - observation relationships represented by these models.

Table 6. Fitted r^2 for instantaneous flow - observation for the Wairoa River site. Model fits with $p < 0.01$ are indicated by bold text.

Fitting models	**CLAR**	***E. coli***	**NH4N**	**DIN**	**NNN**	**DRP**	**TN**	**TP**	**CU**	**ZN**
GAM	**0.65**	**0.61**	**0.11**	**0.61**	**0.61**	**0.03**	**0.67**	**0.19**	**0.2**	0.03
GAM-Log	**0.65**	**0.74**	**0.11**	**0.61**	**0.6**	**0.03**	**0.67**	**0.19**	**0.2**	0
LOESS 0.7	**0.65**	**0.6**	**0.11**	**0.62**	**0.61**	**0.03**	**0.67**	**0.19**	**0.2**	0.02
LOESS 0.7-Log	**0.65**	**0.75**	**0.11**	**0.61**	**0.61**	**0.03**	**0.66**	**0.19**	**0.2**	0
LOESS 0.9	**0.65**	**0.56**	**0.11**	**0.61**	**0.61**	**0.03**	**0.66**	**0.19**	**0.2**	0.02
LOESS 0.9-Log	**0.65**	**0.75**	**0.1**	**0.61**	**0.6**	**0.03**	**0.65**	**0.19**	**0.19**	0
LinLog	**0.59**	**0.14**	**0.09**	**0.58**	**0.57**	0.01	**0.64**	**0.1**	**0.16**	0
LogLog	**0.65**	**0.21**	**0.1**	**0.21**	**0.16**	0.01	**0.63**	**0.12**	**0.16**	0

Figure 13 shows the eight models fitted to the NNN concentration observation - instantaneous flow data for Wairoa River. All models are highly significant and all but one have $r^2 > 60\%$.

Although the fitted relationships are reasonably similar, they deviate from each other appreciably at high flows. For example, the LOESS models indicate that NNN decreases at high flows, representing dilution. In contrast the linear models indicate continuing increases in NNN at high flows and the GAM models indicate a plateau occurs. The differences in these models would produce appreciable differences in the residual values for the high flows and these differences would therefore impact on the results of trend assessments. The difficulty in choosing the “right” model is that most are plausible but confidence in the model fits at high flows is low due to the limited numbers of observations and increasing variability in concentration at high flows.

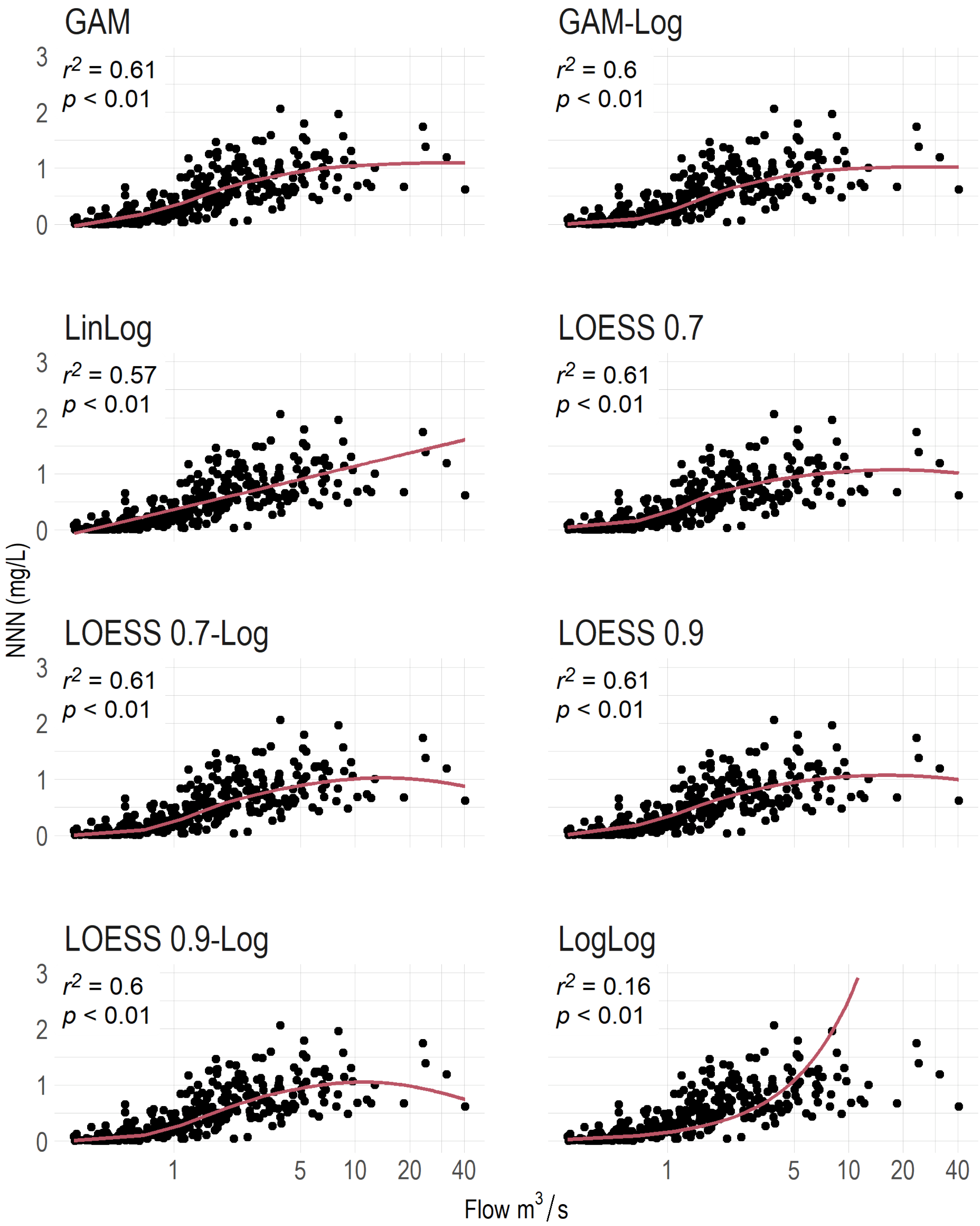


Figure 13. Fitted models representing the instantaneous flow - NNN concentration relationship for the Wairoa River site.

Figure 14 shows the observation - instantaneous flow relationships for ten water quality variables (Table 2) for the Wairoa site. The relationships represented by the eight fitted models are represented as lines in Figure 14. The plots in Figure 14 indicate that DRP and ZN have no obvious relationship with flow (as also indicated by Table 6). For the remaining variables, most of the fitted models appear to be plausible but there are appreciable differences in these models at high flows that would impact on the results of trend assessments. As noted above for the NNN concentration - instantaneous flow model, the difficulty in choosing the “most suitable” model is that most are plausible but confidence in the model fits at high flows is low due to the limited numbers of observations and increasing variability in concentration at high flows.

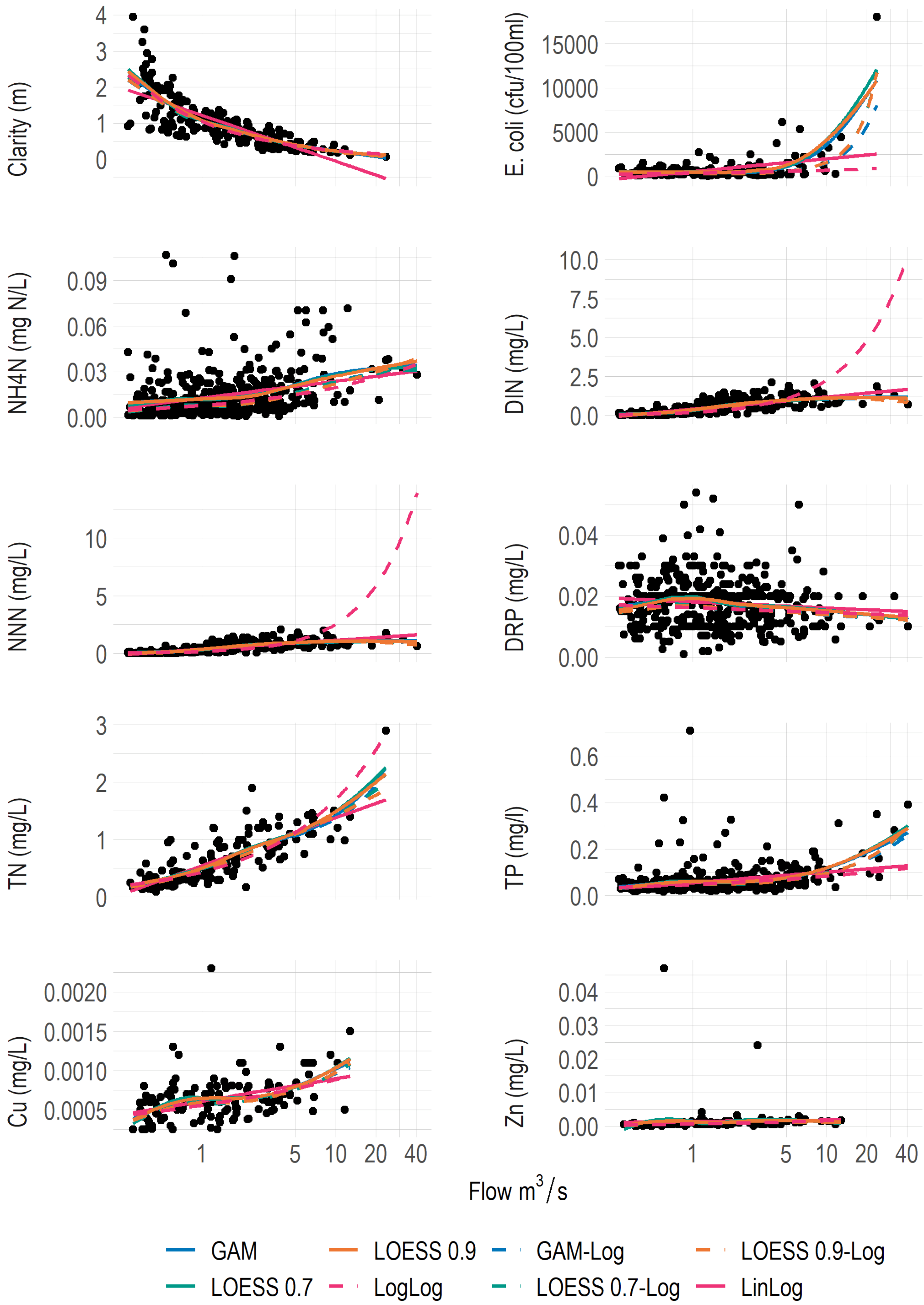


Figure 14. Fitted models representing the observation - instantaneous flow relationships for 10 water quality variables for the Wairoa River site.

Figure 15 summarises the type of model that was identified as the most suitable model for flow-adjustment by expert opinion for each of the 81 site-variable combinations that were considered objectively robust and investigated further. Appendix D Table 9 provides a complete list of the models chosen by expert opinion. The plots used for the subjective model selection are provided in a supplementary information file: ReasonableSiteVariablePlots.pdf.

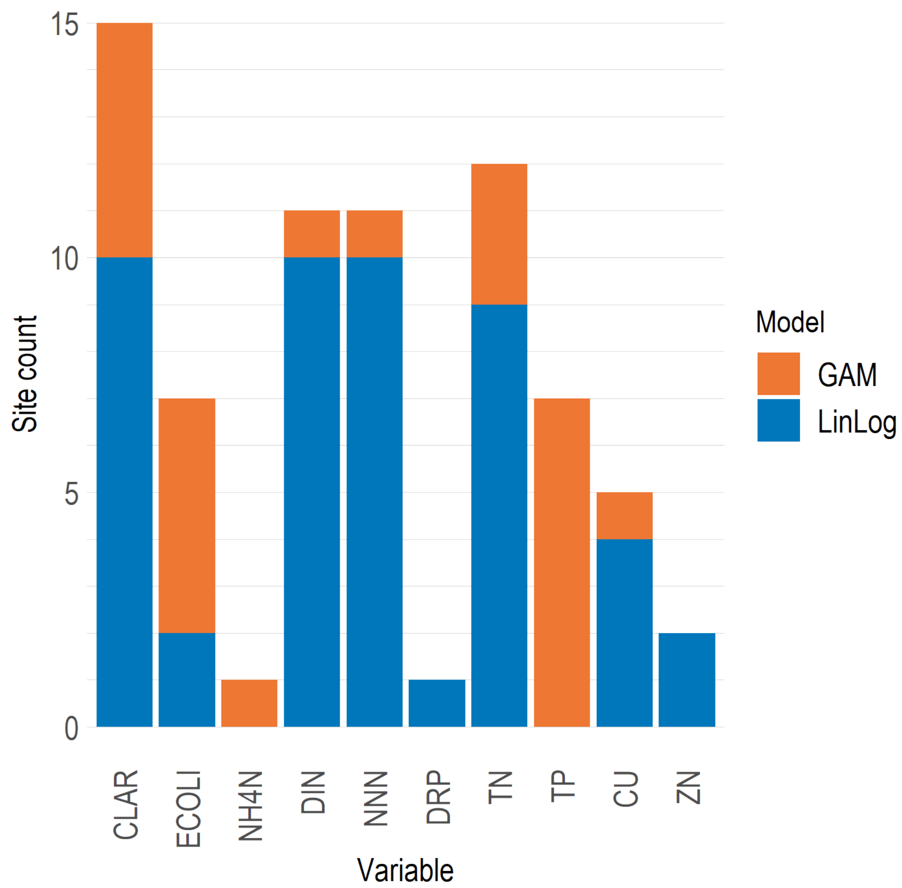


Figure 15. Summary of the choices of model used to represent the observation -instantaneous flow relationship for the 81 site-variable combinations that were considered objectively robust. The bar chart indicates the number of sites for which the differing types of models were selected by the expert.

5.4.1 Flow-adjustment

Figure 16 shows the effect of flow-adjustment based on the Linear-Log (LinLog) model for TN observations for the Wairoa River site. The effect of flow-adjustment on the data is visible as the flow-adjusted values are no longer increasing with flow.

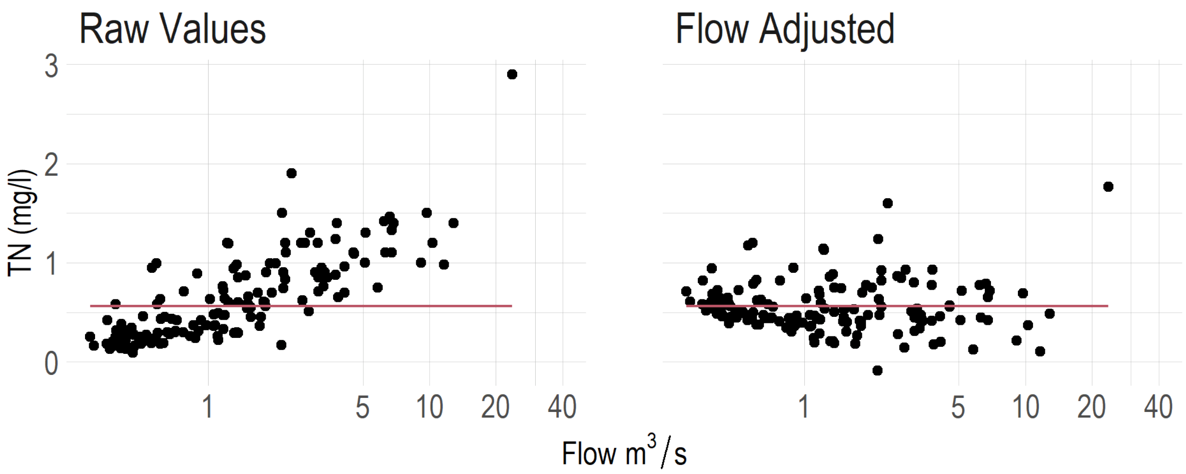


Figure 16. Effect of flow-adjustment of TN concentrations for the Wairoa River site. The right hand side shows the raw observation – instantaneous flow relationship and the left hand side shows the relationship after flow-adjustment using LinLog model. The red horizontal line indicates the median of the raw values.

Figure 17 shows the time series of NNN observations for the Wairoa River sites as measured (i.e., raw values) and after flow-adjustment based on two concentration – instantaneous flow models. Figure 12 shows the concentration –- instantaneous flow model that was used to perform the flow-adjustment seen in Figure 17. Concentrations are more variable in the early part of the time series (i.e., prior to approximately 2010) compared with the later years, but the flow-adjusted concentrations are less variable that the raw data, which is consistent with one of the aims of flow-adjustment.

Figure 17 indicates that overall, there are only small differences between the two sets of flow-adjusted observations. However, these differences are more obvious for high flow observations where the two contrasted models (LinLog and LOESS 0.9) most deviate from each other (Figure 12 note red points indicate flows > 8 m^3/s). The red points indicate that the magnitudes of the flow-adjusted data differ between the two flow-adjustment methods. This means these high flow observations, in particular, will contribute differently to the trend analysis and the trend results can be expected to be differ between the two sets of flow-adjusted data.

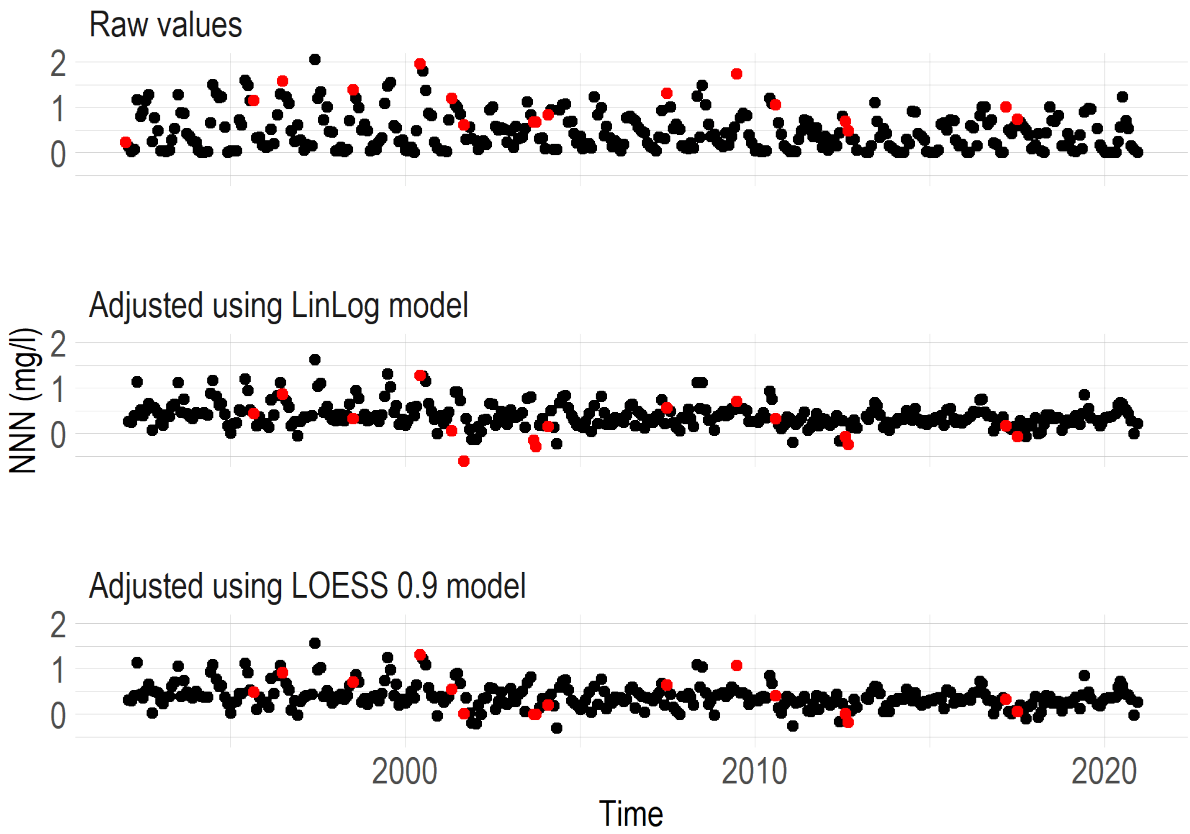


Figure 17. Time series of NNN at Wairoa River as observed and after flow-adjustment. Note that the red points indicate observations that were associated with flows > 8 m^3/s.

5.5 Trend assessments

5.5.1 Comparison of flow-adjusted and raw trends for the 2011 – 2020 assessment period

The impact of the flow-adjustment, where this was considered appropriate based on the criteria set out in Section 4.4, on the trend assessment is presented in Figure 18. Flow-adjusted trend rates differed from their raw counterparts as indicated by the scatter of points away from the one-to-one line in Figure 18. There were also differences in the assignment of sites to categories indicating confidence the trend was decreasing (Figure 19). Figure 18 and Figure 19 indicate that direction of assessed trends can differ between trends based on raw and flow-adjusted data. For example, for TN, a site was categorised as "Very likely" decreasing for the assessment based on the raw observations and "Unlikely" decreasing (and therefore "Likely" increasing) for the assessment based on the flow-adjusted observations (Figure 19). These differences are expected outcomes based on the purpose of flow-adjustment.

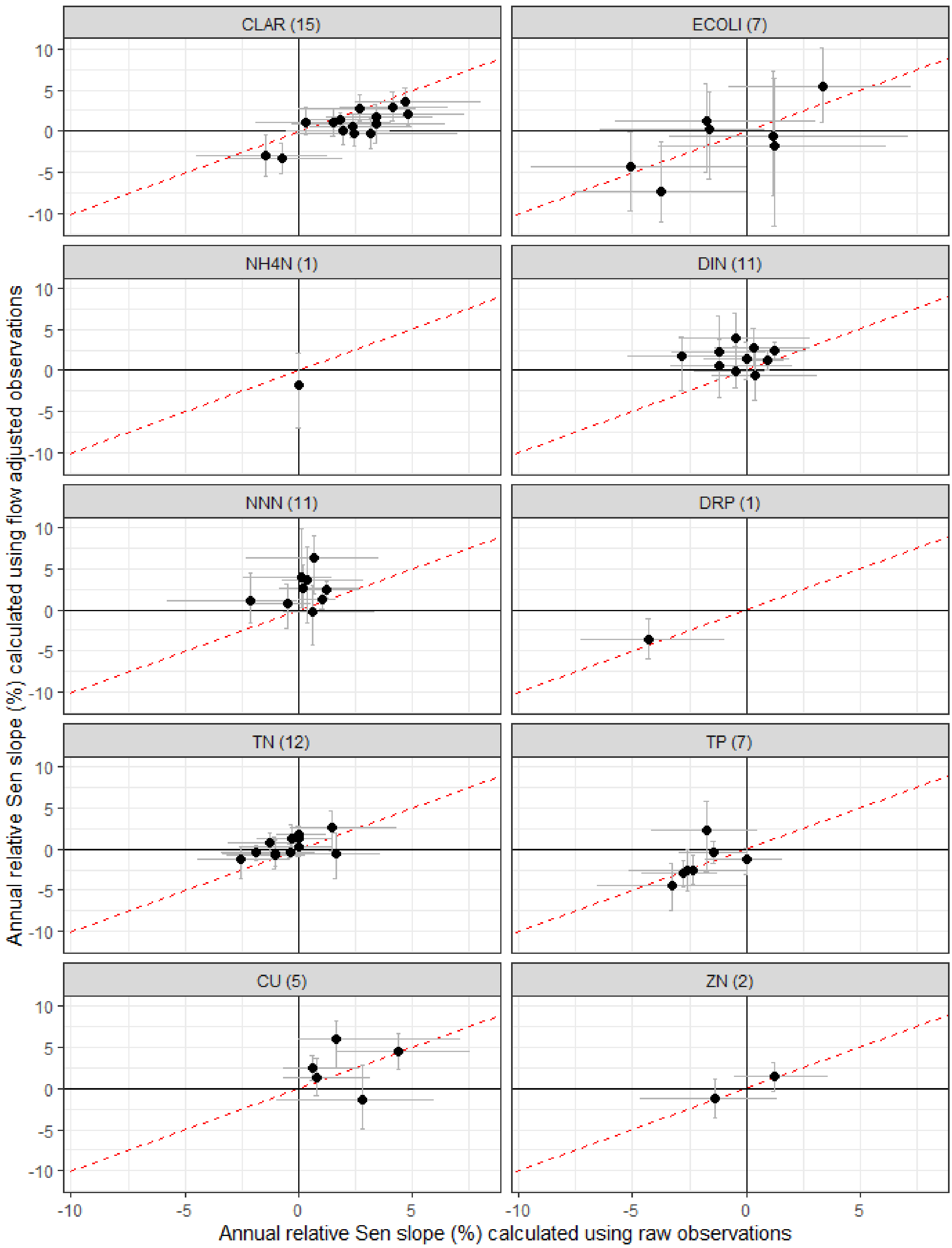


Figure 18. Annual relative Sen slope for trends (2011-2020) calculated from raw and flow-adjusted observations. Each panel represents a different variable. The number in parentheses in the header of each panel indicates the number of sites represented for each water quality variable. Error bars indicate the 90% confidence interval for the annual relative Sen slope. The 1:1 line is shown in red and Sen slopes of zero are shown as black lines.

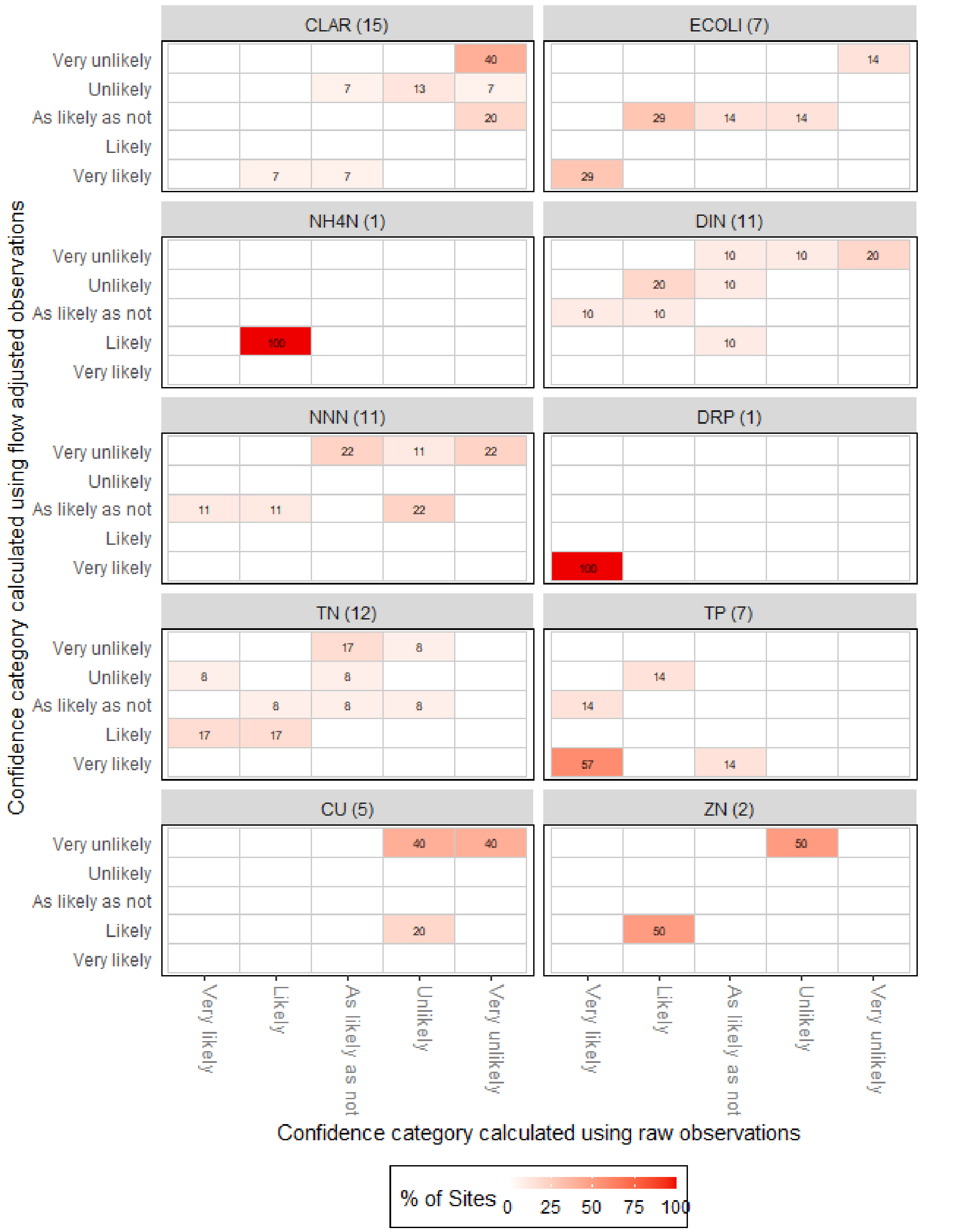


Figure 19. Comparison of confidence that trends (2011 – 2020) were decreasing for trends calculated from raw and flow-adjusted observations. The values in the cells indicate the proportion of sites with the indicated combination of confidence categories. Each panel represents a different variable. The number in parentheses in the header of each panel indicates the number of sites represented for each water quality variable. Note that the diagonal line of cells rising from left to right indicate agreement in the categories and cells away from this line indicate disagreement.

The impact of the flow-adjustment method (i.e., the choice of model used to represent the instantaneous flow – observation relationship) on the assessed trend rate (expressed as an annual relative Sen slope) is shown in Figure 20. The deviation of the points from the one-to-one line in these plots indicate that assessments of trend rate are sensitive to the method of flow-adjustment. The deviation of the two sets of flow-adjusted trends is less than that of the raw versus flow-adjusted trends but is nevertheless appreciable (i.e., compare Figure 20 with Figure 18). This indicates that choices made in the flow-adjustment process impact on the evaluated trends.

Differences in flow-adjustment methods also produced disagreements in the assignment of sites to categories expressing confidence in trend direction (Figure 21). For example, for DIN, two sites that were categorised as "Unlikely" decreasing for the assessment based on the LOESS 0.9 flow-adjusted observations were categorised "As likely as not" and "Very unlikely" based on the expert-based flow-adjusted observations (Figure 19). The disagreements in the assignment of sites to confidence categories produced by comparing trends assessed using different flow-adjustment methods were not as large as those produced by comparing raw and flow-adjusted trends (i.e., compare Figure 19 with Figure 21). However, the differences in confidence categories produced by comparing trends assessed using different flow-adjustment methods were appreciable indicating that choices made in the flow-adjustment process impact of on the evaluated trends.

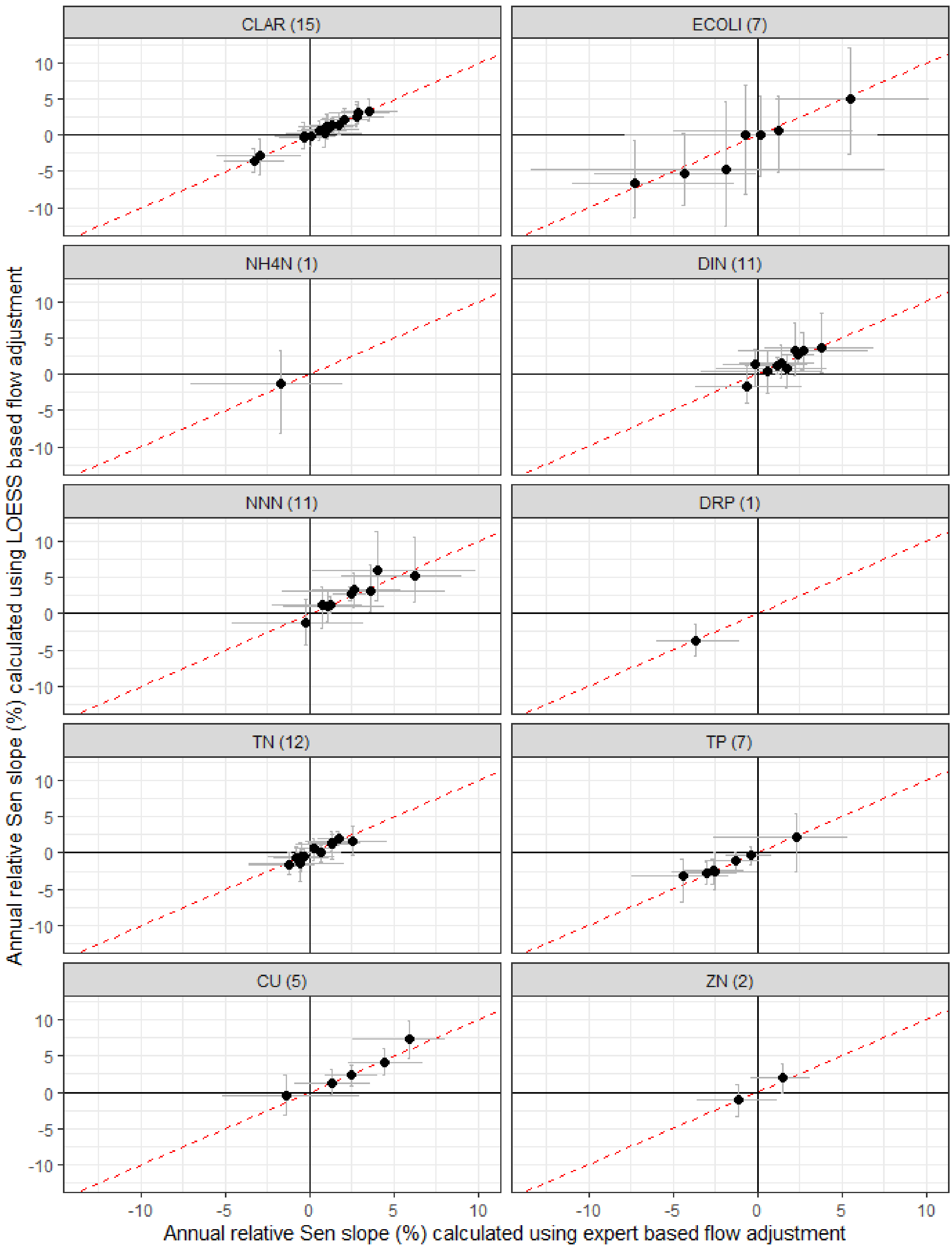


Figure 20. Annual relative Sen slope for trends calculated from flow-adjustment of observations based on expert-selected model versus using a default LOESS 0.9 model. Each panel represents a different variable. The number in parentheses in the header of each panel indicates the number of sites represented for each water quality variable. Error bars indicate the 90% confidence interval for the annual relative Sen slope. The 1:1 line is shown in red and Sen slopes of zero are shown as black lines.

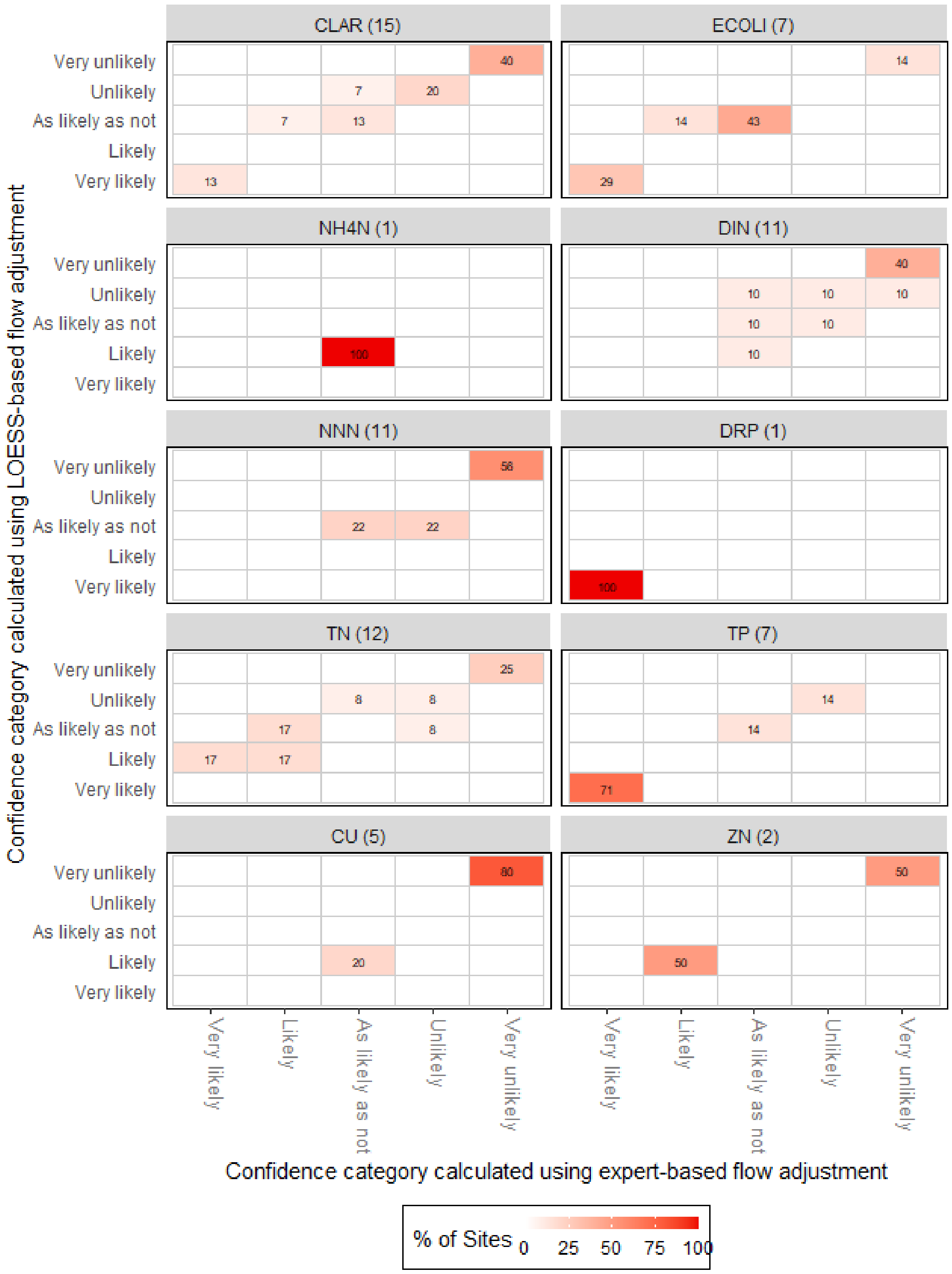


Figure 21. Comparison of confidence that trends were decreasing for trends calculated from using expert-based flow-adjustment and using LOESS-based flow-adjustment. The values in the cells indicate the proportion of sites with the indicated combination of confidence categories. Each panel represents a different variable. The number in parentheses in the header of each panel indicates the number of sites represented for each water quality variable. Note that the diagonal line of cells rising from left to right indicate agreement in the categories and cells away from this line indicate disagreement.

5.5.2 Trend variability over time

Raw and flow-adjusted trend assessments of rolling time-periods of 5- 10- and 20-year duration for the Hoteo River and Rangitopuni River sites and for CLAR, DIN, DRP and NH4N are shown in Figure 22 to Figure 27. For each variable, Sen slope and confidence the trend was decreasing (C_d) tended to oscillate between time periods for all three durations. Within a variable, the magnitude of changes in Sen slopes between adjacent time periods decreased with increasing time window duration (Figure 22, Figure 24, Figure 26). For example, for CLAR at the Hoteo river site, Sen slopes varied between approximately -0.2 and 0.2 m year^{-1} for the 5-year duration, -0.05 and 0.1 m year^{-1} for the 10-year duration and -0.02 and 0.04 m year^{-1} for the 20-year duration. For the 5-year duration, there were frequent changes in the direction of trends between time periods that were separated by only one or two years. For example, for the 27 individual 5-year duration assessments of CLAR trends for the Hoteo river site, there were three groups of end years with “Very likely” decreasing trends that were separated by groups of end years with “Very unlikely” decreasing trends (Figure 23). This oscillation in trend over periods of approximately 5 years is also seen in the Sen slope assessment (Figure 22). Changes in direction of site trends were less frequent for the 10-year time-period duration (Figure 25) and less frequent again for the 20-year time period duration (Figure 27).

Although there were small differences in the assessed Sen slope and confidence the trend was decreasing (C_d) between the raw and flow-adjusted trends, the magnitude and frequency of oscillations were not appreciably different. This indicates that if the driver of the oscillations is hydrological processes, their effect is not removed by flow-adjustment.

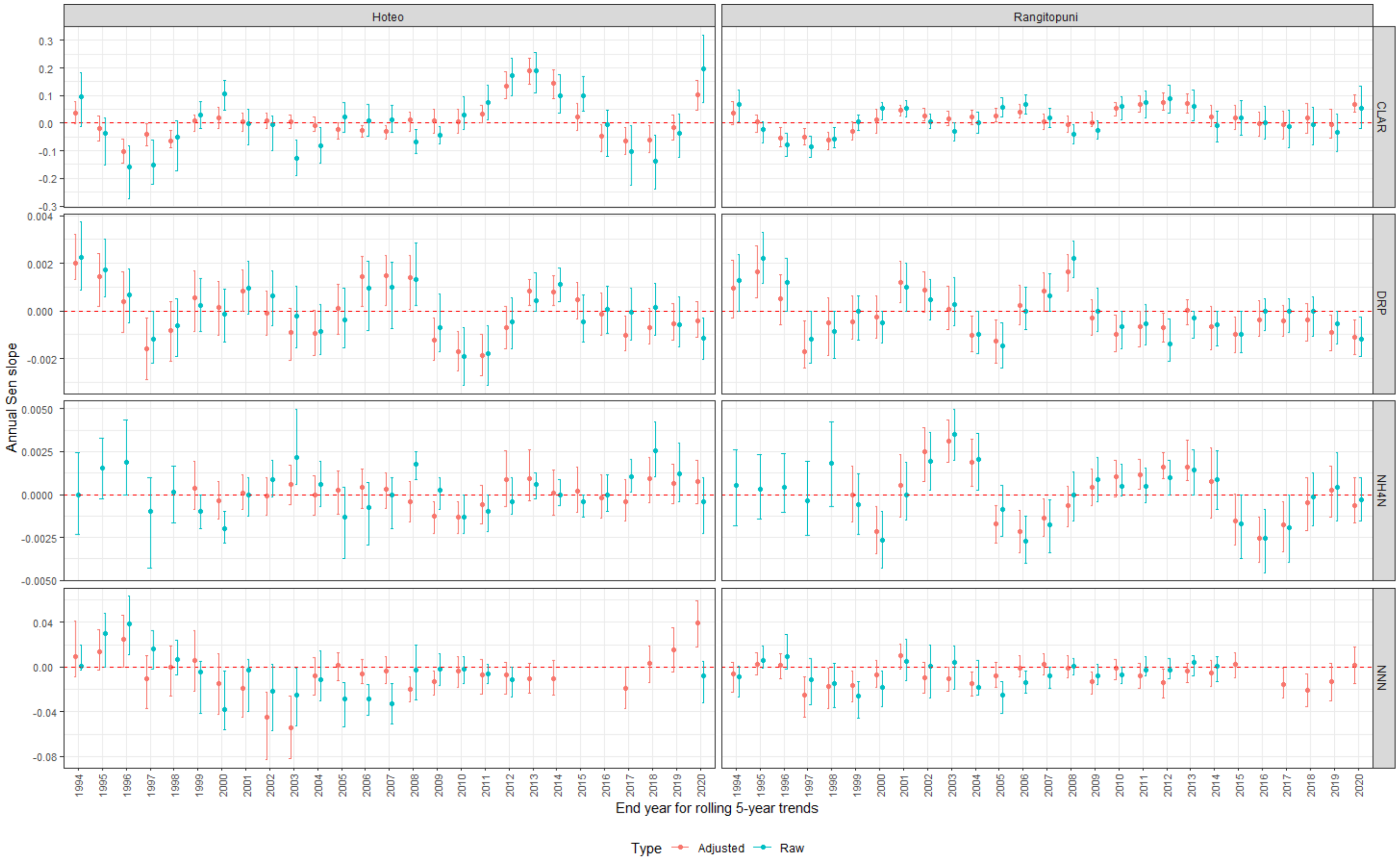


Figure 22. Rolling 5-year Sen slopes for four water quality variables for the Hoteo River and Rangitopuni River sites (NIWA, NWQN) for the period from 1990 to 2020. Each panel represents a site (columns) and a water quality variable (rows). The points indicate the raw and flow-adjusted Sen slopes and the error bars indicate the 90% confidence intervals. The red dotted line indicates a Sen slope zero. The y-axis units are rates of change of the units for each variable shown in Table 2 per year. The x-axis indicates the end year for each trend assessment period. Sen slope values could not be calculated for some trend periods because there was insufficient variability in the water quality data to undertake the trend analysis.

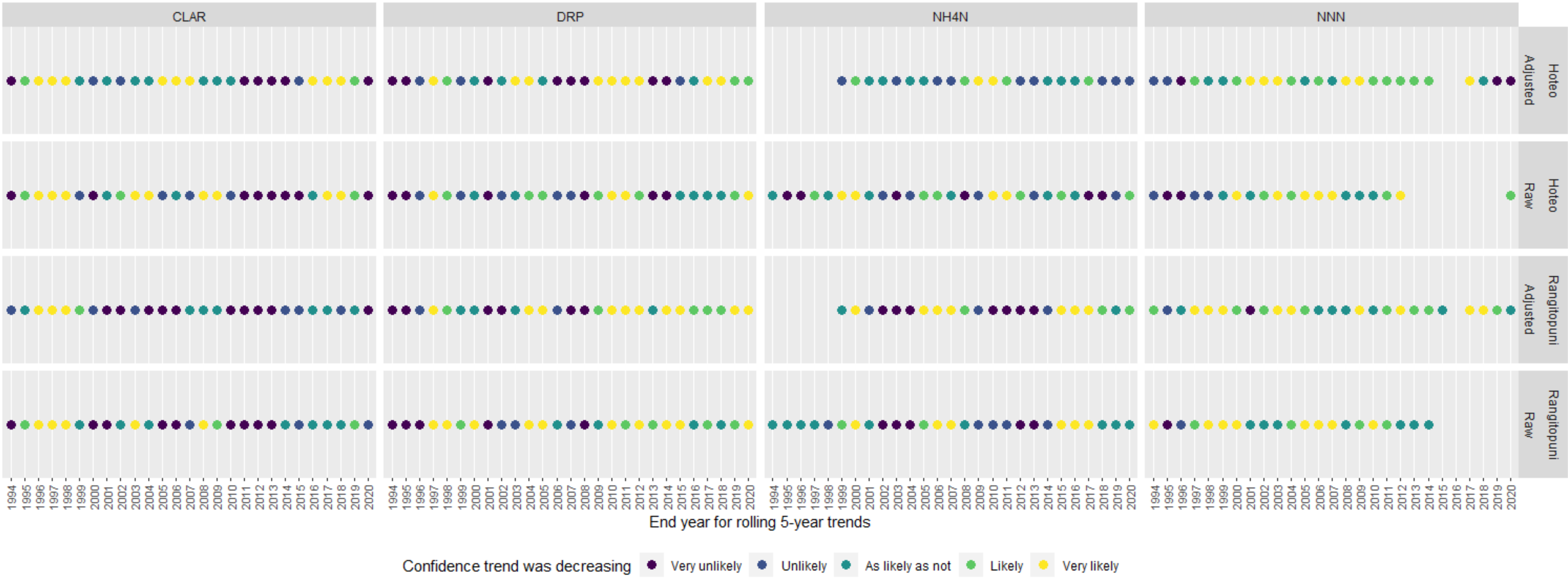


Figure 23. Rolling 5-year confidence in trend direction for four water quality variables for the Hoteo River and Rangitopuni River sites for the period from 1990 to 2020. Confidence (C_d) is indicated using the categories defined in Table 3. The x-axis indicates the end year for each trend assessment period. Confidence values could not be calculated for some trend periods because there was insufficient variability in the water quality data to undertake the trend analysis.

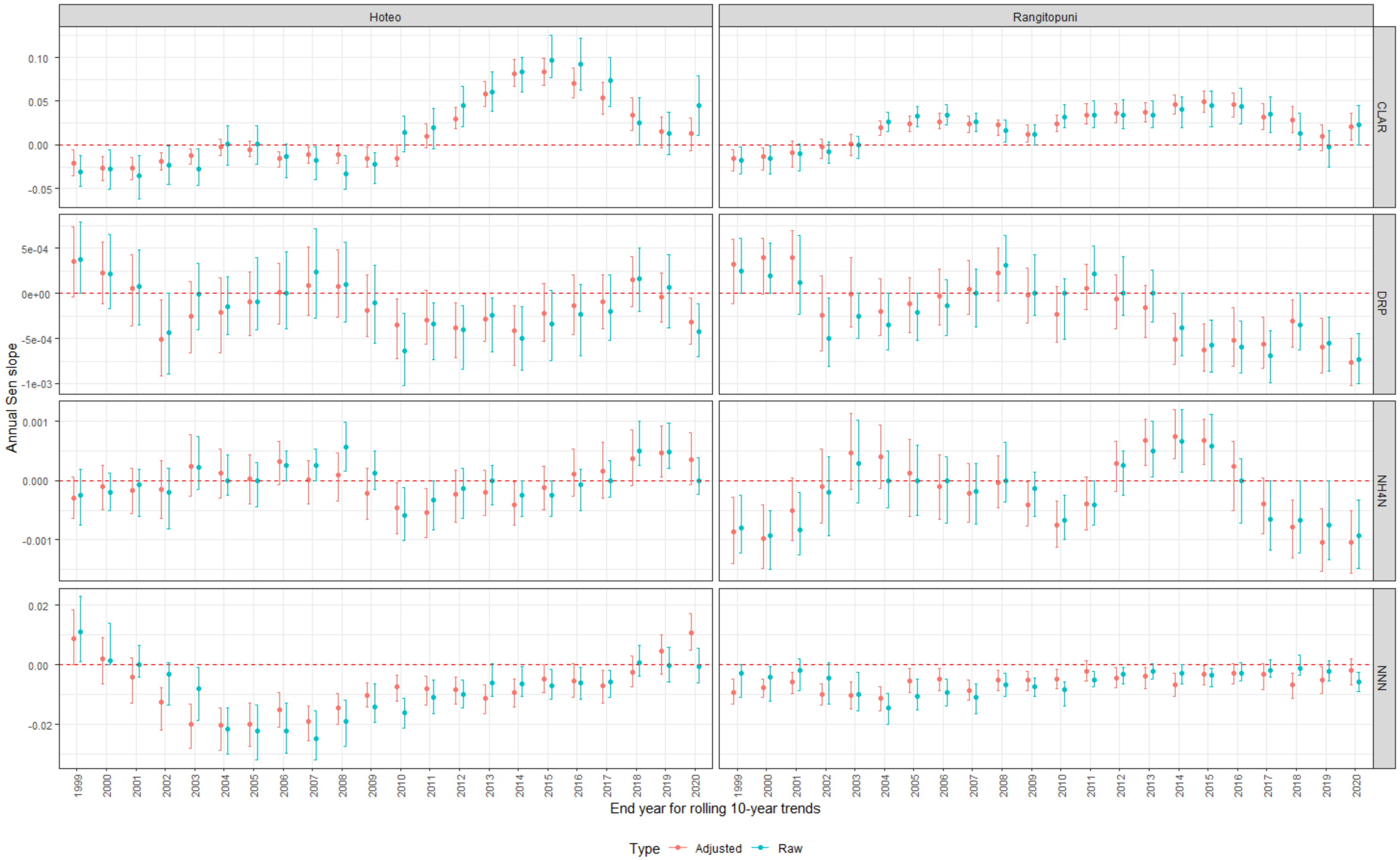


Figure 24. Rolling 10-year Sen slopes for four water quality variables for the Hoteo River and Rangitopuni River sites for the period from 1990 to 2020. Each panel represents a site (columns) and a water quality variable (rows). The points indicate the raw and flow-adjusted Sen slopes and the error bars indicate the 90% confidence intervals. The red dotted line indicates a Sen slope zero. The y-axis units are rates of change of the units for each variable shown in Table 2 per year. The x-axis indicates the end year for each trend assessment period.

Figure 25. Rolling 10-year confidence in trend direction for four water quality variables for the Hoteo River and Rangitopuni River sites (NIWA, NWQN) for the period from 1990 to 2020. Confidence (C_d) is indicated using the categories defined in Table 3. The x-axis indicates the end year for each trend assessment period.

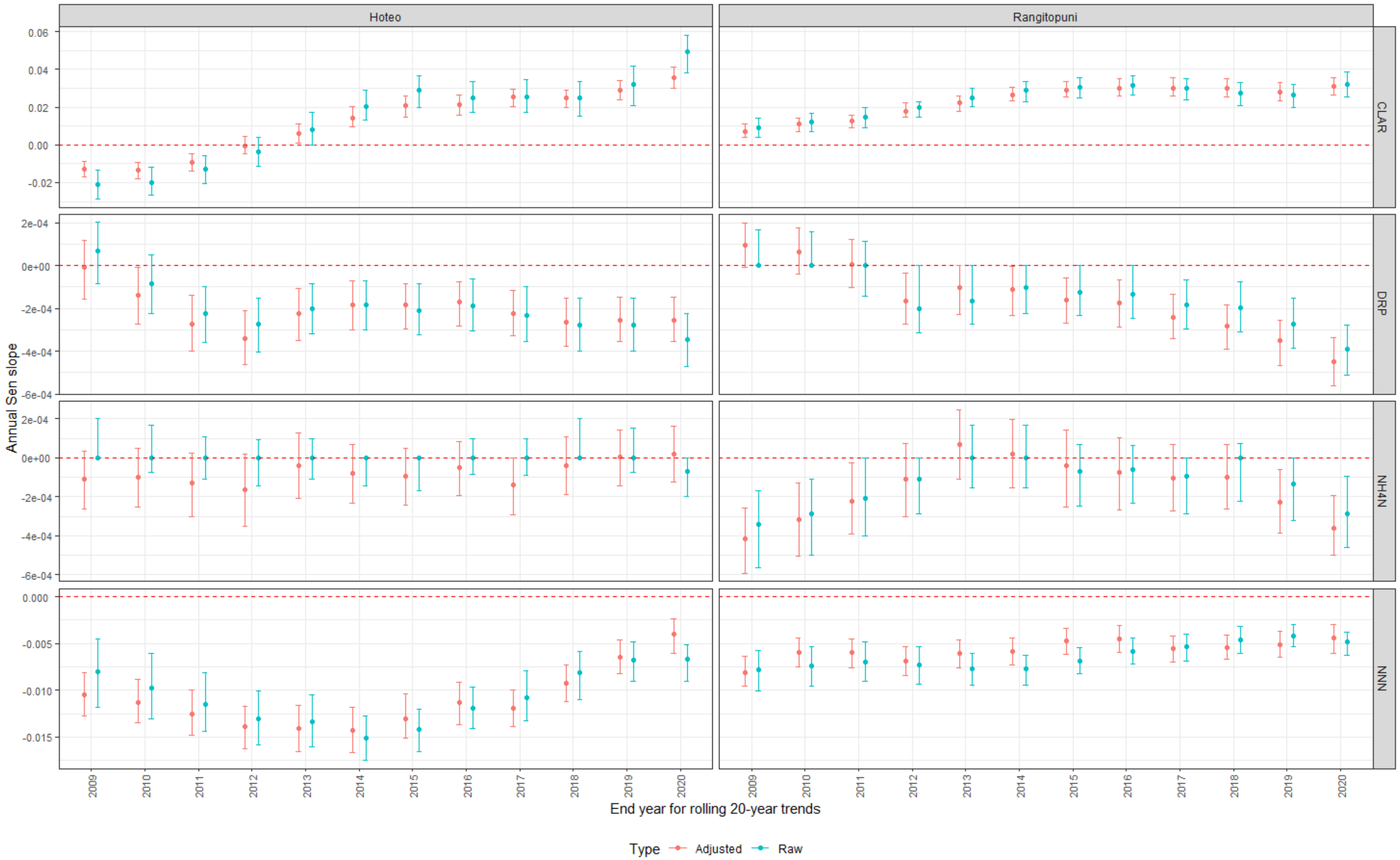


Figure 26. Rolling 20-year Sen slopes for four water quality variables for the Hoteo River and Rangitopuni River sites for the period from 1990 to 2020. Each panel represents a site (columns) and a water quality variable (rows). The points indicate the raw and flow-adjusted Sen slopes and the error bars indicate the 90% confidence intervals. The red dotted line indicates a Sen slope zero. The y-axis units are rates of change of the units for each variable shown in Table 2 per year. The x-axis indicates the end year for each trend assessment period.

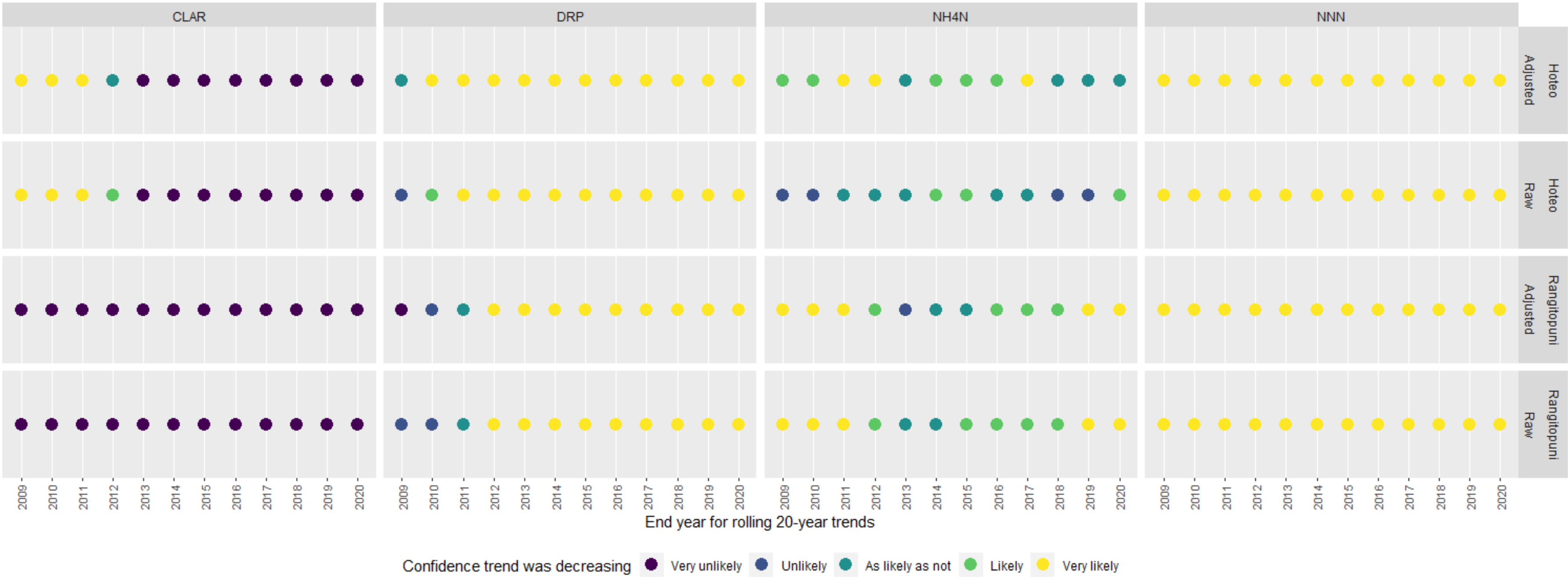


Figure 27. Rolling 20-year confidence in trend direction for four water quality variables for the Hoteo River and Rangitopuni River sites for the period from 1990 to 2020. Confidence (C_d) is indicated using the categories defined in Table 3. The x-axis indicates the end year for each trend assessment period.

5.6 WRTDS models

The weighted regression on time, discharge, and season (WRTDS) models of CLAR, NNN, DRP and NH_4N for the Hoteo and Rangitopuni monitoring sites had R^2 values between 0.5 and 0.83. Based on the evaluation criteria of Moriasi et al. (2015) water quality models with R^2 values of >0.3 are satisfactory, >0.6 are good and >0.7 are very good. Therefore, all models were at least satisfactory, and some were very good (Table 7).

Table 7. Cross validated model R^2 values for WRTDS models of selected variables for the Hoteo and Rangitopuni monitoring sites.

Site	Variable	R^2
Hoteo	CLAR	0.83
	NNN	0.77
	DRP	0.50
	NH4N	0.70
Rangitopuni	CLAR	0.79
	NNN	0.83
	DRP	0.53
	NH4N	0.52

The predicted daily values of the four water quality variables at both sites are shown in Figure 28 along with the observations. The blue line indicates the seasonal rolling mean (i.e., moving average at the seasonal timescale), which smooths some of the daily variability. The seasonal rolling mean highlights two features of the predictions. First, there is generally a strong seasonal pattern in the data, and the amplitude of this pattern is often variable over time. For example, CLAR at the Hoteo site had a large range in 2020 compared to 2018. Second, mean concentrations are highly variable at the annual time scale and there are fluctuations in the mean values of variables at the interannual time scale. For example, NH_4N at the Hoteo site was generally lower in the period from 2004 to 2007 compared to the earlier period 2001 to 2003 and the later period from 2008 to 2010 (Figure 28).

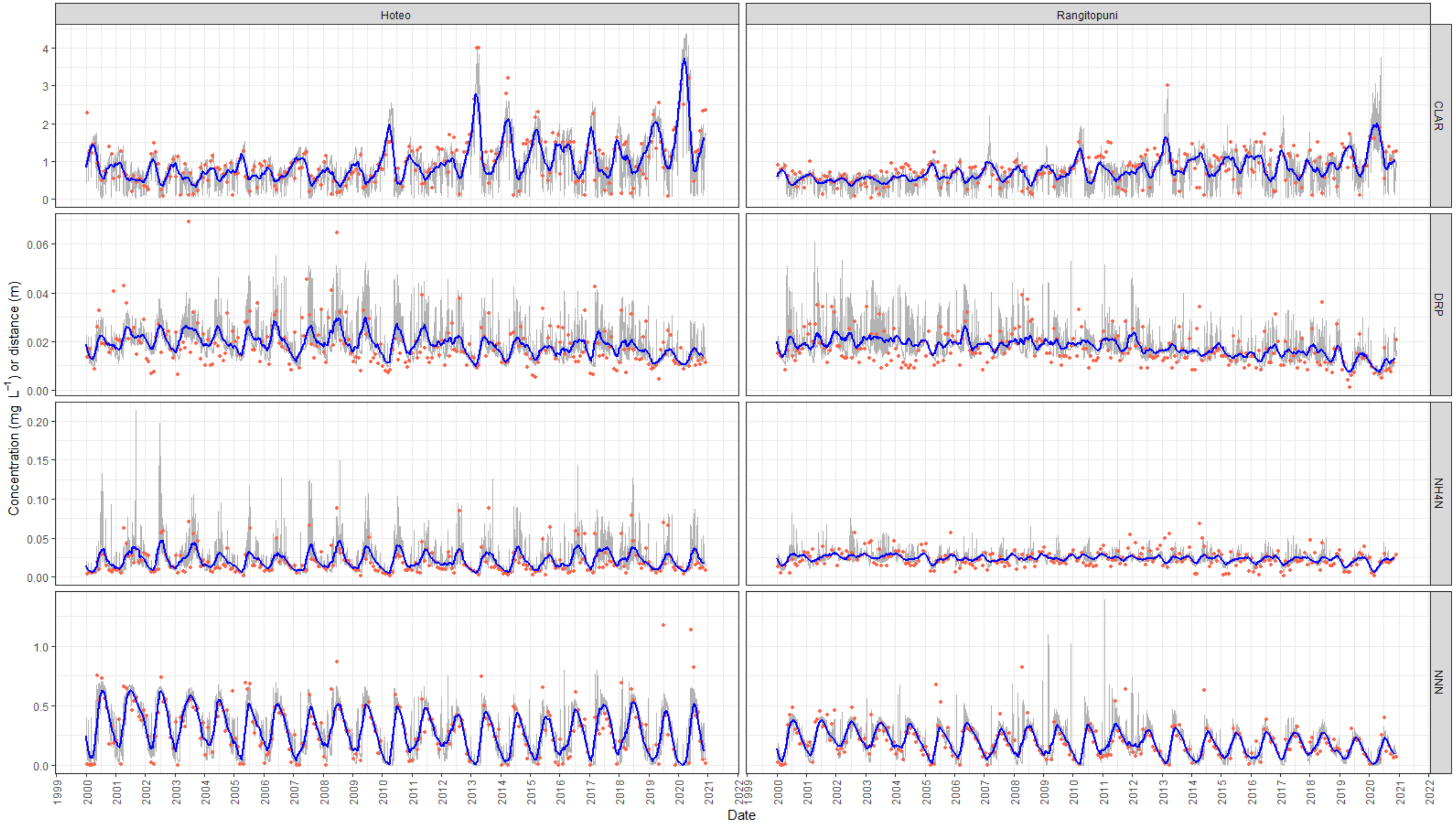


Figure 28. Concentrations of CLAR, DIN, DRP and NH4N for the Hoteo and Rangitopuni monitoring sites predicted by the WRTDS models. The grey lines indicate daily predictions, and the blue line indicates the predicted seasonal rolling mean value. The red dots indicate the observations. The y-axis indicates units of concentration for DIN, DRP and NH_4N and distance for CLAR.

Examples of the observation – flow relationships fitted by WRTDS for different dates are shown in Figure 29. These plots indicate that the WRTDS model has detected and fitted appreciably differing observation – flow relationships for different dates for some site and variable combinations. For example, Figure 29 shows that NNN concentrations at the Hoteo site were consistently higher at a given discharge in 2000 compared to the later years (2009 and 2019).

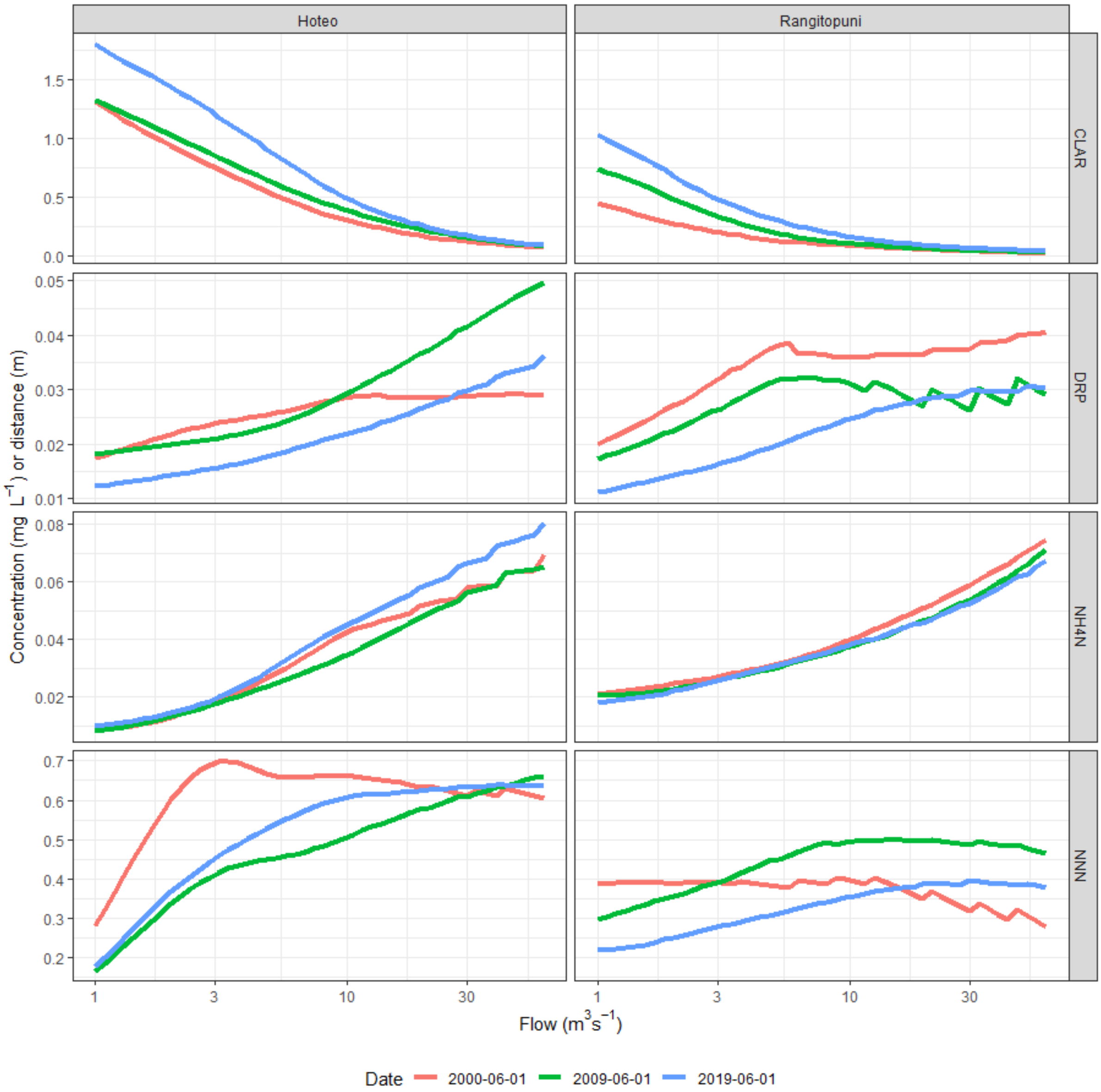


Figure 29. Examples of observation – flow relationships fitted by the WRTDS models for a fixed day in three different years. Note that y-axis indicates units of concentration for DIN, DRP and NH_4N and distance for CLAR.

Examples of the observation – flow relationships fitted by WRTDS for different seasons within the same year are shown in Figure 30. These plots indicate that the WRTDS model has detected and fitted appreciably differing observation – flow relationships for different seasons for some site and variable combinations. For example, Figure 30 shows that NNN concentrations were consistently higher at a given discharge in winter (indicated by 2019-06-01) compared to spring and summer (indicated by 2019-10-01 and 2019-02-01, respectively).

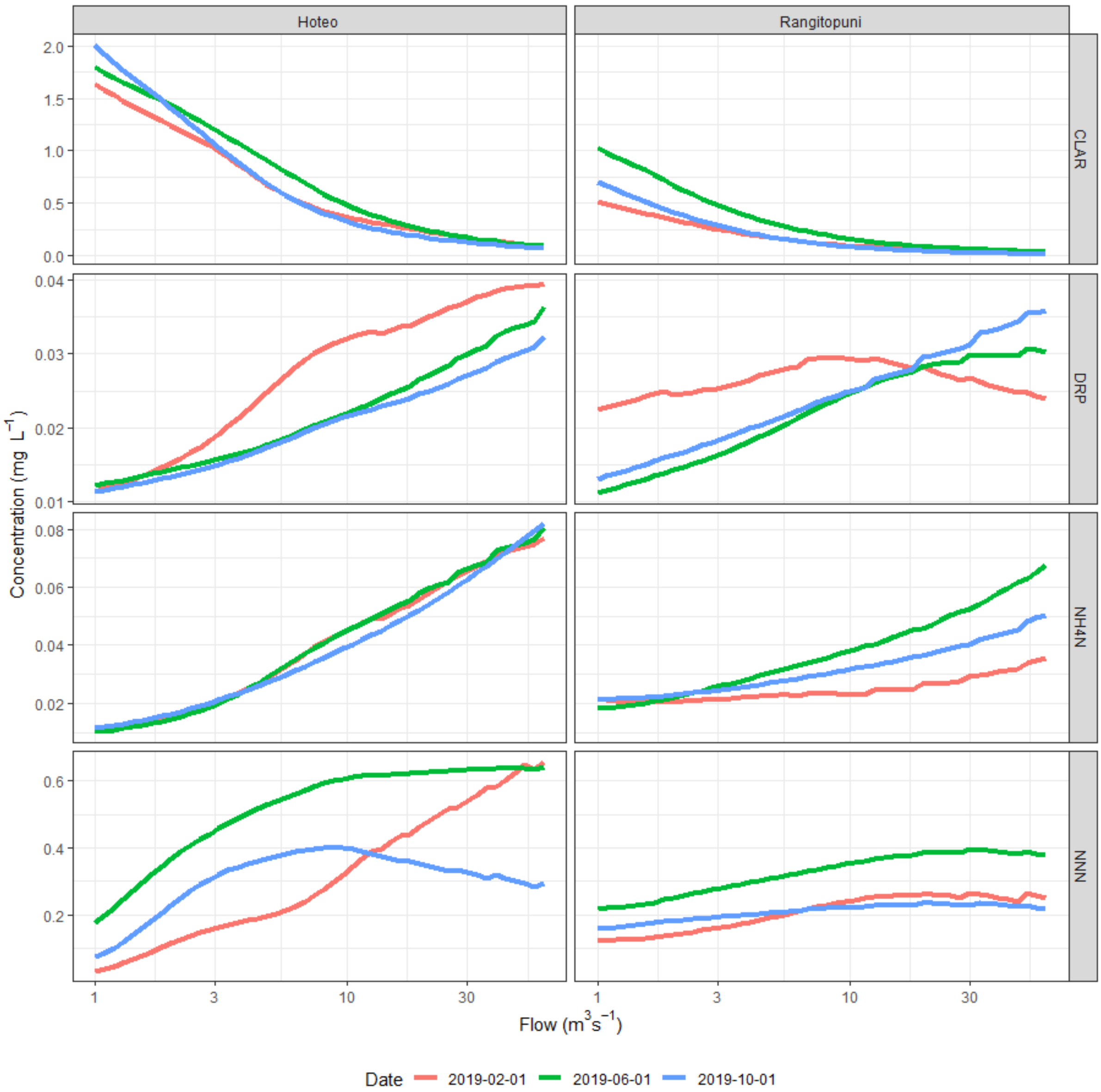


Figure 30. Examples of fitted observation – flow relationships on different days, representing seasons (summer (red), winter (green) and spring (blue)), in the same year (2019).

Comparisons of observed and predicted water quality state (as median values and the precision of those estimates) for rolling five-year periods ending 2020 are shown for the four water quality variables at both sites in Figure 31. The plots show the rolling 5-year median values based on the daily values predicted by the WRTDS model and the rolling 5-year median flow standardised values (also predicted by the WRTDS model). The predictions obtained from the WRTDS models highlight three features of the data. First, most site and variable combinations exhibit secular (i.e., long term) trends through the whole record. For example, over the whole period between 1994 and 2020, CLAR increased, and NNN and DRP decreased at both sites (i.e., improving trends across these variables). The secular trends indicated by the WRTDS models are consistent with the trend assessments for the 20-years duration reported in Section 5.2.2. For example, for most 20-year trend assessment periods, there were increasing trends for CLAR, and decreasing trends for DIN and DRP (Figure 26 and Figure 27).

The second feature of the data highlighted by the WRTDS models is that the predicted median values oscillated through the period. For example, predicted median NH4N at the Hoteo site was lower in the period from 2001 to 2003 compared to the period 2004 to 2006 and was again lower in the period from 2009 to 2011. The oscillations in state indicated by the WRTDS models are consistent with the trend assessments for the 5-years duration reported in Section 5.5.2. For example, for the 5-year trend assessment periods, trend rates and directions for all four variables oscillated with an approximate duration of a full cycle being six to seven years (Figure 22 and Figure 23).

The third feature of the data highlighted by the WRTDS models is that flow normalised median values (blue lines on Figure 31) exhibit appreciably less oscillation than the medians derived from the predicted values and the median values calculated from the water quality observations. For example, while there was a secular decreasing trend over the whole period for DRP at the Hoteo site, values were higher in succeeding years on several occasions through the period. For the Rangitopuni site, predicted DRP and the calculated DRP median face values also oscillated appreciably so that state changed between the NOF D band and C band several times over the period of record. In contrast, at both sites and for all variables the flow normalised median values exhibited much less oscillation and tended to indicate consistent trends through the whole record. This indicates that the oscillations are explained by flow regime variation and that the WRTDS flow normalisation procedure is effective in removing their effect. It is noted that oscillations in the rolling trends are shown in the preceding analysis (e.g., Figure 22, Figure 24) even for the flow-adjusted trends. This indicates that that flow normalisation by WRTDS is reasonably able to account for the impact of flow regime variation whereas flow-adjustment of instantaneous flow does not.

Finally, the fourth feature of the data highlighted by the WRTDS models is that the rolling 5-year median values calcuated from the flow normalised WRTDS predictions were sometimes outside of the precision of the medians calculated from the monitoring data (Figure 31). This indicates water quality state, as represented by the monitoring data, is appreciably associated with flow regime variation. Further, this indicates that water quality state assessments produced from different assessment periods may differ in association with flow regime variation even in the absence of changes in anthropogenic pressure on water quality. In other words, assessments produced from monitoring data for a specific assessment period may indicate poorer or better or water quality for the site than that which would be obtained if the assessment period had represented the “average” flow regime.

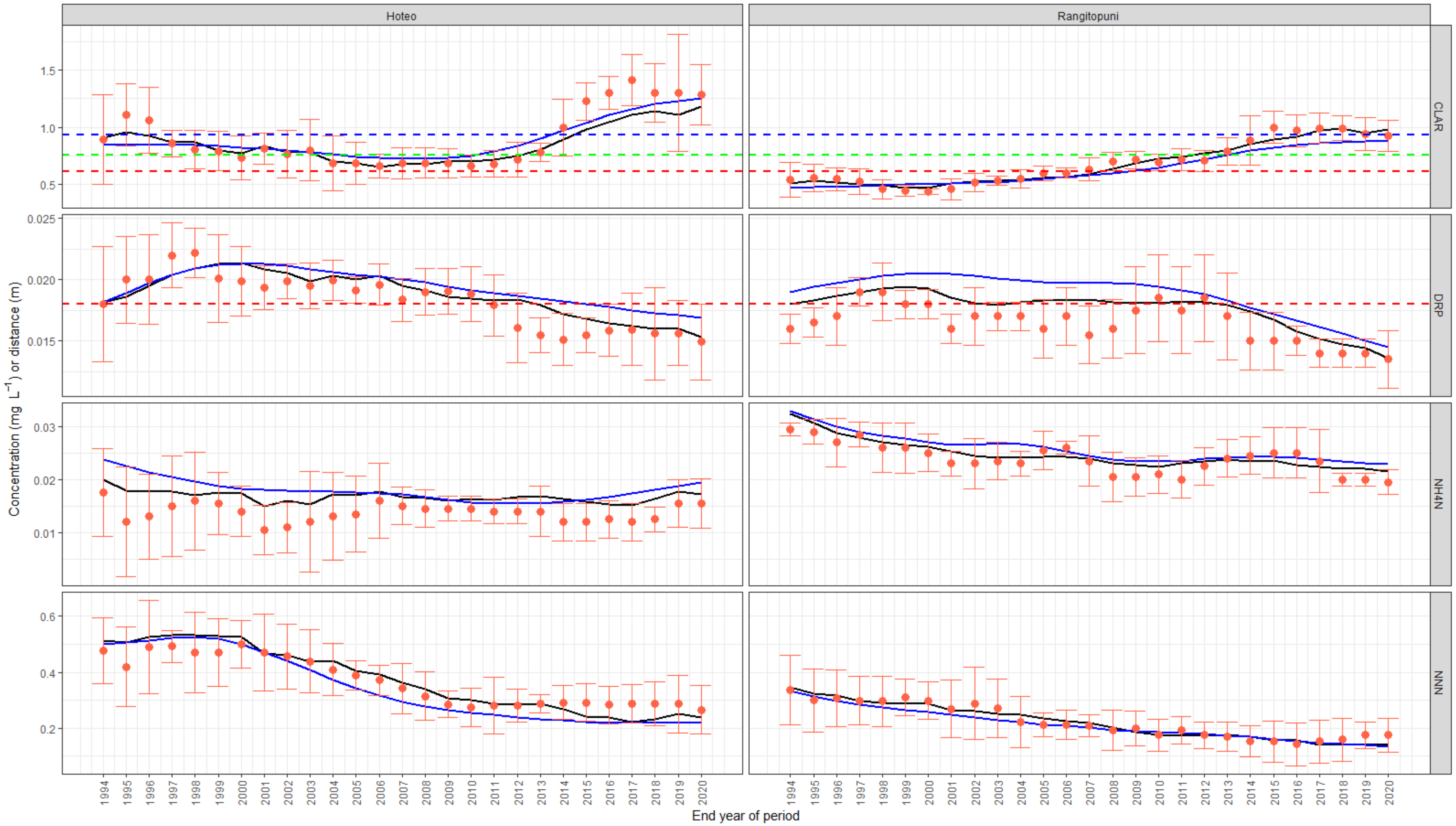


Figure 31. Rolling five-year median values for the preceding five-year assessment period. The black lines indicate the WRTDS predicted median values and the blue lines are the WRTDS predicted flow normalised medians. The red dots indicate the median values of the observations in the preceding five-year period and the error bars indicate the associated precision as 95% confidence intervals. Note that y-axis indicates units of concentration for NNN, DRP and NH4N and distance for CLAR. The dashed horizontal blue, green and red lines indicate lower thresholds for the A, B and C NOF bands. Note that both sites belong to NOF Suspended Sediment Class 2 and concentrations of NH4N and NNN at both sites were consistently within the A band.

6 Discussion

The National Policy Statement on Freshwater Management (NPS-FM) and section 35 of the RMA requires AC to monitor water quality, regularly assess and report on water quality state, including the state of compulsory attributes identified as part of the National Objectives Framework (NOF), assess water quality trends, and take appropriate action in the case that those trends indicate degradation. In streams and rivers, these tasks are complicated because water quality observations are variable and are influenced by both the instantaneous flow at the time of observation and the preceding hydrological conditions (i.e., flow regime) at timescales of weeks, months and even years. Both the influence of instantaneous flow and the flow regime at longer time scales on water quality have implications for the assessment of water quality state, trends, and the attribution of trends to causes.

In this study, we undertook a series of analyses of water quality and associated flow data at 15 long term monitoring locations across the Auckland region. Based on the findings of those analyses and our experience and expertise in state and trend assessment, we make the following observations.

6.1 Uncertainty in assessment of attribute state

6.1.1 Precision of state assessments

Assessments of water quality state are uncertain. The most obvious component of uncertainty is associated with sample error. Sample error can be understood as the uncertainty associated with an estimate of the true water quality state that is made from monitoring data (a “sample”). This uncertainty arises because water quality is variable over time and observations are only a snapshot of what actually occurred (the “true” water quality state) over the assessment period. The uncertainty associated with sample error can be quantified and we refer to it as the precision of the state assessment (Section 5.1). Limited precision means that there is uncertainty in assessments of state, and therefore assignment of sites to NOF bands. In the context of assessing attribute state, precision can be understood as the range over which we could expect the assessed state to vary if there had been multiple independent sets of samples taken.

6.1.2 Instantaneous flow rate contributes to water quality variability

Instantaneous flow rate was represented in this study by mean daily flows and was shown to be generally associated with river water quality variability and therefore the precision of state assessments. We showed that water quality variables are correlated with instantaneous flow rate (Section 5.4). If water quality assessments, including attribute states, are to describe characteristics of the true water quality state, in the assessment period, water quality sampling needs to be unbiased with respect to instantaneous flow rate. If samples are unbiased with respect to instantaneous flow, then over an assessment period, the range of flows sampled should approximate the distribution of all flows in the assessment period. For the 15 monitoring sites that were the focus of this study, we found that this was always true. In any five-year assessment period, the distribution of instantaneous flows on sample occasion was never significantly different to the distribution of all flows (Section 5.3.2). This means that the sampling was unbiased with respect to instantaneous flow.

6.1.3 Flow regime variability contributes to water quality variability

Flow regime variability refers to variation in flows at longer than instantaneous (i.e., daily) timescales. Flow regime variability produces another aspect of uncertainty of state assessments that is not quantified. This uncertainty is associated with the fact that flows for any five-year assessment period are not a perfect representation of the long-term “average” flow regime (Section 5.2). From this we can infer that the flow regime can also be expected to vary significantly between state assessment periods (i.e., 5-year periods). State assessments can therefore also be expected to vary between assessment periods because of differences in the flow regime between those periods. We refer to this second component of uncertainty associated with water quality state assessments as unquantified uncertainty of type A. With respect to state assessment, unquantified uncertainty of type A can be understood as the difference in state assessments between five-year assessment periods that can be expected due to hydrological differences (manifested as flow regime variation) between the periods. This uncertainty is not readily quantified, although its existence can retrospectively be seen as oscillations in attribute state over time (Sections 5.5.2and 5.6).

6.2 Unquantified uncertainty confounds comparisons between baseline, current and target attribute states

The unquantified uncertainty of type A associated with state assessments is relevant to the NPS-FM requirement to publish comparisons of current and target attribute states (S3.30(2)(b)) and to assess whether target attribute states are being, or are likely to be, achieved (S3.30(2)(c)), and to assessments of trends and their causes (S3.30(2)(d)). The intent of these requirements is to identify changes in water quality in a timely fashion and understand their causes so that appropriate action can be taken if they are due to effects of human resource use rather than natural or unmanageable causes. However, in this study, we showed that water quality is strongly influenced by flow regime variation (Section 5.6). Therefore, both trends and state assessments pertaining to different time-periods will differ due to differences in flow regimes between the periods. Because flow regime variability is a strong drivers of water quality variability, they are a confounding factor with respect to the intent of the above NPS-FM requirements. We note that the influence of flow regime variability on water quality may be due to variation in climatically-driven processes that control the mobilisation, storage and transport of contaminants in catchments. However, additionally the influence may be exerted by anthropogenic responses to climatic variation. For example, land management practices might be different in wet and dry years and these differences may be reflected in water quality differences.

Unquantified uncertainty of type A is also relevant to setting target attribute states because these are based on assessments of baseline and current state. Because flow regime variability is associated with oscillations in water quality state (Sections 5.5.2 and 5.6), a baseline or current state assessment may represent an unusually good or poor water quality for a site over the long run. The risk associated with this uncertainty is that baseline state assessments may estimate better water quality than that associated with the “average” flow regime, and this may result in targets that are not ambitious enough to effect change. Alternatively, the assessed baseline may represent better water quality than that associated with the “average” flow regime and targets may be set that are not achievable.

6.3 Uncertainty in assessment of water quality trends

6.3.1 Flow-adjustment addresses instantaneous flow but cannot control for flow regime variability

Flow conditions can be considered a "covariate" when we are interested in how water quality observations are changing over time. This means that flow conditions are related to the water quality observations, but their influence confounds identifying whether anthropogenic factors are involved and whether action needs to be taken. Because flow is a covariate, an accepted practice in trend assessment is to undertake "flow-adjustment". Flow-adjustment attempts to remove the influence of instantaneous flow on the water quality observations prior to trend analysis. There is an important distinction between instantaneous flow rate and flow regime variability at longer time scales. Flow-adjustment uses a statistical process to control for instantaneous flow rate on water quality observations (Section 5.4.1). However, this does not account for all the influence of flow regime variability on water quality because hydrological processes vary over a range of temporal scales.

In this study we show that trend magnitude (indicated by Sen slopes) and direction (indicated by confidence the trend was decreasing; C_d) produced by rolling trend assessments oscillate over time (see Section 5.5.2). The magnitude and frequency of the oscillations tend to decrease with increasing assessment period duration due to temporal smoothing. The most plausible explanation for these oscillations is the influence of climatic processes. Various studies have shown that the El Niño Southern Oscillation climate pattern (ENSO) translates into a predictable variation in water quality trends (Scarsbrook et al. 2003; Snelder et al. 2021c, b). In this study we also showed that flow-adjustment of water quality data prior to trend analysis does not reduce the magnitude and frequency of the oscillations seen in results of rolling trend assessments (Section 5.5.2). This indicates that the influence of flow regime variation on water quality is not removed by flow-adjustment. Flow regime variation therefore remains a confounding factor when considering how to respond to degrading trends as required by NPS-FM (S3.30(2)(d)).

6.3.2 Trend assessments are associated with unquantified uncertainty

There are commonly accepted methods for fitting trend models that we employed in this study. A key determination from trend models is trend rate and confidence in the trend direction. Confidence in assessments of trend rates and direction can be understood as analogous to the precision of state assessments; limitations to confidence is equivalent to limitations to precision and is due to sample error. However, because trend analysis is based on statistical models, it involves simplifications, assumptions and procedural decisions. The influence of these simplifications, assumptions and procedural decisions are not reflected in the evaluation of the confidence of a trend assessment (i.e., are not quantified) but are a source of uncertainty that should not be overlooked.

In the context of trend analysis, there are two sources of unquantified uncertainty. First, there will be differences in trend assessments between different time-periods that are associated with differences in the flow regimes over those periods that we refer to as unquantified uncertainty of type A. Type A uncertainty means that trend rate and direction indicated by an "up-to-date" trend assessment, for example an assessment of a short duration trend (e.g., 5 or 10-years duration) that ends with the most recent observations, may be a response to flow regime variation and may be reversed in a subsequent time step. In this study, we demonstrated that rolling trend assessments vary considerably in trend magnitude and direction, even between assessment periods that are close in time. For example, rolling trend assessments periods of 5 and 10 years duration that increment by one year, can go from "Likely" or "Very likely" decreasing to "Likely" or "Very likely" increasing when the assessment period is incremented by only one or two years (Section 5.5.2). The likelihood of reversal of short duration "up-to-date" trend means they cannot be regarded as the sole basis for making decisions to act. The second source of unquantified uncertainty in trend analysis arises due to there being other potential models that describe the trend, these may result in differences in the assessment, and these alternative models may be equally credible. In this study, we refer to this type of unquantified uncertainty as type B and demonstrated this by showing that there are alternative credible flow-adjustments and these produce differences in trend assessments (Section 5.5.1).

The first step in flow-adjustment is to model the relationship between instantaneous flow and the water quality observations. In this study, we showed that this process is subjective and there is generally more than one plausible model of the association (Section 5.4). Furthermore, we demonstrated that these different flow-adjustments result in differences in trend assessments (see Section 5.4.1). Consequently, flow-adjustment adds to the uncertainty associated with trend assessment and this uncertainty is generally not quantified in the application of commonly accepted trend assessment methods.

6.3.3 Trend assessments alone are not a reliable early warning mechanism

Management of water quality would be facilitated by early warning of degradation and whether specific anthropogenic activities are causing the degradation. This suggests that trend assessments based on short assessment periods (e.g., 5-years duration) ending with the most recent observations are desirable because these describe changes that have occurred in the recent past. However, rolling trend assessments show that the shorter the time-period, the greater the likelihood of reversal of the assessed trend at the subsequent trend assessment (see Section 5.5.2). We also showed that the tendency for water quality state to oscillate is strongly decreased when flow regime variation is accounted for (flow normalised WRTDS predictions, Section 5.6). This indicates that short term trends are likely to be strongly influenced by flow regime variation, even when these trends are flow-adjusted. Therefore, on their own, short term trend assessments are not a reliable early warning mechanism.

More generally, a trend assessment produces no information regarding the causes of the observed trend. The effects of hydrological variation may amplify or counteract the effects of other drivers of water quality trends. Therefore, there is a risk that reporting water quality trends without robust attempts to identify the causes may lead to speculative attribution of the trends to anthropogenic drivers. This may then lead to management actions to mitigate anthropogenic drivers that are ineffective in reversing degrading trends, or complacency that water quality is being protected when in fact anthropogenic degradation has been counteracted by effects of hydrological variation.

6.4 Models are needed to understand water quality variation

Temporal variation in water quality is complex. Our ability to understand and describe this variation is increasing as data records increase in duration, and our ability to model the complex relationships improves. In this study we demonstrated WRTDS as a new approach to modelling the evolution of water quality over time. The WRTDS models indicated that interannual oscillations in water quality are strongly associated with flow regime variation because flow normalised predictions produced by WRTDS were considerably smoother than the non-normalised counterparts (Section 5.6). These results are consistent with the established link between the ENSO climate pattern and variation in water quality trends

(Snelder et al. 2021c, b). Importantly, because WRTDS can remove the influence of flow regime variation on water quality, it is likely to be more useful than traditional flow-adjusted trend analysis when assessments aim to determine whether anthropogenic pressure or actions are changing water quality (e.g., Choquette et al. 2019; Murphy 2020).

The analyses undertaken by this study are examples of the need to use models, of increasing complexity, to make sense of water quality data and to carry out the requirements of the NPS-FM effectively and robustly. Therefore, we consider that Section 1.6 of the NPS-FM needs to be interpreted carefully and broadly. In our opinion, raw data is not useful information and the "best information available" is obtained from a combination of data and modelling. In addition, it should always be acknowledged that estimates of state and trends are representations of reality with associated uncertainties; both quantified and unquantified. Unquantified uncertainties should be understood as differences in assessments of state and trends that arise because:

- no two assessment periods are alike, from, at least, a flow regime perspective
- modelling involves simplifications of reality and different models and modellers are likely to produce different but equally plausible assessments using the same data.

6.5 Attribution of cause(es) is a very significant challenge

The requirement under NPS-FM S3.30(2)(d) to assess causes of trends is referred to by Snelder et al. (2021b, c) as "attribution". The present study was not explicitly concerned with attribution. However, flow-adjustment and flow normalisation, as undertaken in this study, can be regarded as statistical approaches to removing the influence of instantaneous flow and longer-term flow regime variability, respectively. The purpose of these procedures is to allow attribution of trends to factors other than instantaneous flow and flow regime variability. We define "rigorous attribution" (of cause) to mean quantitative analyses of relationships between water quality trends and drivers, and consideration of multiple alternative drivers (Ryberg 2017; Ryberg et al. 2018; Murphy 2020). Rigorous attribution of cause will generally be based on statistical models that include multiple alternative drivers, consideration of the physical plausibility of the associations, and quantification of the confidence in the inferred causes. We identify weaker alternatives to "rigorous attribution" to include qualitative reasoning, references to previous studies, and simple speculation.

In our opinion, AC should strive to undertake robust attribution of cause(s) in seeking to carry out the requirements of NPS-FM S3.30(2)(d). However, this is extremely challenging for two reasons. First, suitable data characterising spatio-temporal variation in environmental drivers of water quality are scarce and fragmented. Suitable data consist of time-series of measurements of land use and management and point sources of contaminants, with durations and frequencies that correspond to water quality time-series, and which are spatially congruent with water quality monitoring sites (Snelder et al. 2021b). Second, water quality is generally influenced by multiple environmental drivers, including anthropogenic drivers such as land use and natural drivers such as climate variability and its impact on flow regimes. There may be additive, compensatory or synergistic interactions among these drivers, making it difficult to reliably attribute water quality responses to specific water quality pressures. The influences can only be elucidated by modelling and models are dependent on there being sufficient sites for the signals (i.e., causes) to rise above the noise.

7 Recommendations

In this section we provide recommendations for dealing with the complications that arise in carrying out the requirements of the NPS-FM due to the relationship between water quality variables, including NPS-FM attributes, and flow. These recommendations are narrowly focussed on technical issues and are based on our technical interpretation of the relevant NPS-FM sections and the limitations to scientific quantification of water quality state and trends. Our recommendations are distinguished below by bold italic text.

We recognise that the wording of the NPS-FM can be interpreted less narrowly than our interpretation for the purposes explored in this report. Broader interpretation of these policy requirements may provide greater discretion for responding to the policy intent. Therefore, we consider that our recommendations need to be considered by people with expertise in NPS-FM implementation. Furthermore, we consider the details of AC's implementation should be a blend of our suggestions with those of policy experts and those charged with NPS-FM implementation. The aim should be to strike an appropriate balance between the intent of the NPS-FM and what is technically possible and defensible. In our opinion, the need to strike this balance means that there is no perfect solution to the problems exposed in this study with respect to the uncertainty of state and trend assessments. Therefore, in carrying out its functions it is important that AC acknowledges the limitations and is transparent about how uncertainty has been dealt with in implementation of the NPS-FM. In addition, we suggest that this report and our recommendations should not be regarded as comprehensive or final. This is a complex topic that involves multiple disciplines and best practice is therefore likely to evolve over time.

7.1 Recommendations for monitoring and assessment of attribute states

Our study has shown that water quality variables observed in rivers are correlated to varying degrees to instantaneous flow. So that assessments of attribute state describe the true water quality state, water quality sampling needs to be unbiased with respect to instantaneous flow rate. This ensures that the range of variability in the population of a water quality variable is represented by the observations (i.e., the sample). ***We therefore recommend that water quality sampling continues to be carried out so that it is unbiased with respect to instantaneous flow.***

Because assessments of attribute state are made from a sample, they should be considered as model outputs that contain unavoidable uncertainty. However, the "face value" of an assessed attribute state (i.e., the evaluated numeric attribute state or NOF band) is the best estimate of the state in the assessment period, given the available data. ***We therefore recommend that:***

- ***an assessed baseline, or current attribute state is regarded as the "best information at the time" as defined by NPS-FM Section 1.6(1) and the uncertainty of the assessment is not an adequate reason to delay giving effect to the NPS-FM.***
- ***the precision of the current state estimates (as implemented in this study or similar) is included when considering and publishing data describing attributes and the associated uncertainty (NPS-FM S3.30(1c)).***
- ***when appropriate, unquantified uncertainty of type A associated with baseline/current attribute state is broadly described as arising from the***

influence of flow regime variability on water quality at timescales of weeks to years.

- ***when appropriate, it is transparently stated that unquantified uncertainty of type A will impact on future water quality state assessments and that when appropriate, it is clarified that these fluctuations confound the identification of anthropogenic causes of water quality change and the formulation of appropriate actions.***

We note that the current MFE guidance regarding S3.18 of the NPS-FM (MFE 2022) does not acknowledge that there are uncertainties associated with information derived from monitoring data. The MFE guidance does indicate that the monitoring method must be fit for purpose and there is therefore a broader question about whether the uncertainties associated with monitoring are acceptable. In our opinion, there is insufficient research on the potential impact of unquantified uncertainty of type A to answer this question robustly.

An obvious question is whether water quality monitoring sample frequency is sufficient to satisfy the fit for purpose criteria suggested by the MFE guidance. We don't have any recommendations regarding changing water quality monitoring frequency for the purposes of state and trend assessment. All other things being equal, increasing monitoring frequency will increase the precision of state assessments and confidence in trend assessments. However, it is not clear how helpful this will be because it does not address the issue of unquantified uncertainty of type A. ***We recommend that more research is needed into increasing the certainty of state and trend assessments.***

7.2 Recommendations for target attribute states

We recommend that the impact of unquantified uncertainty of type A on assessment of baseline, and current state is considered when setting target attribute states and developing actions to improve water quality. This could take the form of sensitivity analyses that test the extent to which planned actions may fail to achieve target attribute states in future assessment periods due to foreseeable influence of flow regime variability on water quality.

We recommend that analysis of water quality time series is used to attempt to quantify the potential magnitude of foreseeable fluctuations in water quality due to flow regime variation. We suggest that the WRTDS model is a promising tool for this type of investigation. We also suggest that AC's process-based Freshwater Management Tool (FWMT) is potentially useful for this type of analysis.

7.3 Recommendations for analysing and reporting on trends in water quality over time

We recommend that water quality trend assessments are always represented as model outputs that are unavoidably uncertain. To manage the uncertainty, ***we recommend that AC consider that there are two types of application of trend analysis; "regional application" and "local application" as described by the current guidance on trend assessment*** (Snelder et al. 2021a). A "regional application" of trend analysis should be regarded as assessing and reporting trends across many sites and variables in the context of regional SOE monitoring programmes, and to satisfy requirements to publish information about state and trends set out in the NPS-FM. ***We recommend that AC utilises "regional application" of trend analysis to fulfil its NPS-FM requirements to assess progress towards target attribute states under S3.30(2)(c).*** A "regional application" should use a consistent methodology over sites and variables and should be regarded as a screening exercise that seeks to identify problem locations for closer inspection. The currently accepted approach to “regional application” of trend assessment is the use of non-parametric correlation and regression as set out in (Snelder et al. 2021a) and as used to make trend assessments in this study. In relation to “regional application” of trend analysis, we recommend that:

- ***flow-adjustment is not undertaken and only raw (un-adjusted) trends are reported under S3.30(2)(c).***
- ***it is made clear in reporting that trend assessments only describe changes in water quality that were observed and not what they were caused by.***
- ***consideration is given to reporting trend assessment periods of at least 10-year periods, and possibly longer to reduce the likelihood that abrupt changes in these assessments occur if reporting occurs frequently (e.g., annually).***

We recommend that AC regards “local application” of trend analysis to be associated with the requirements to assess trends and their causes under NPS-FM S3.30(2)(d). We recommend that “local application” of trend analysis is “triggered” where (i.e., for those sites and variables) “regional application” undertaken to fulfil S3.30(2)(c) requirements provides evidence that deterioration was observed. The objective of a local application is to extract as much information as possible about the trend direction and rate of change from the available data (Snelder et al. 2021a). A local application may therefore utilise more than one statistical method and may produce assessments that are inconsistent with assessments made using the approach recommended for a “regional application”. NPS-FM S3.30(2)(c) and S3.30(2)(d) require assessment of trends and their causes. Therefore, ***we recommend that it is appropriate for local application of trend analysis to incorporate flow-adjustment.*** However, we recommend caution with inferences made from flow-adjusted trends and that it is kept in mind that flow-adjustment adds unquantified uncertainty of type B to the assessment and does not remove the influence of flow regime variation to trend assessments. It is important to acknowledge that there is no definitive numeric data driven assessment that can, with complete accuracy, assess progress towards target attribute states or attribute a trend to a cause. A combination of analysis and modelling and use of different lines of evidence (e.g., water quality observations, measured actions and/or physical changes in the catchment) will need to be used to arrive at a “reasoned judgement” about progress in river health and causes of trends.

We recommend that the NPS-FM requirements to assess trends and their causes under NPS-FM S3.30(2)(d) should be applied by ensuring, whenever possible, that water quality monitoring is associated with flow data. In our opinion, the interpretability and therefore, value of water quality data, is dependent on associated flow measurement. The manner in which the WRTDS model combines daily flow records with less frequent water quality observations indicates that flow data needs to be, at least, at the daily timescale. Flow data at least at the daily timescale is also required for estimating contaminant loads, which is also likely to be needed to support ongoing NPS-FM implementation. We acknowledge that it will not always be possible to have river water quality monitoring sites co-located with flow sites and that water quality monitoring sites that do not have measured flows are nevertheless useful. ***We therefore recommend that AC considers collecting data over the long term to improve the ability to use models to provide synthetic flow records for water quality monitoring sites that do not have measured flow.***

We recommend that, to the extent that it is possible, trend cause attribution incorporates assessment of hydrological drivers. We don't have specific recommendations for how to do this and consider that the study of attribution of causes of water quality trends should be an active area of research. We suggest that new water quality modelling tools such as WRTDS and AC's FWMT model are potentially useful for this type of analysis.

7.4 Recommendations for acting on degrading trends

We recommend a cautious and staged approach with respect to taking action when degrading trends are detected. We recommend that degrading trends indicated by "regional application" of trend analysis to fulfil NPS-FM S3.30(2)(c) requirements are treated as "triggers" for closer analysis by "local application" of trend analysis (see the recommendations under Section 7.3 above). The evidence provided by "local application" of trend analysis can then be used to make judgements about taking proportionate action at stage 1. Stage 1 may include taking cautious action on the ground and/or potentially increased monitoring effort and ongoing surveillance of possible water quality pressures. In situations where there is an absence of information about trend drivers, such as changes in land use and management, and contaminant discharges, an appropriate response might be to retrospectively gather this information while continuing to monitor water quality. If deteriorating trends continue and/or confidence in the causes of these trends is judged sufficiently high, then stage 2 would be triggered that would involve significant intervention in the catchment to halt and reverse degradation. This staged approach is consistent with current MFE guidance regarding S3.20 of the NPS-FM (MFE 2022). The guidance points out that S3.19 "allows councils discretion based on risk, and on whether it is possible to determine unnatural cause, before declaring an attribute is 'degrading'". The guidance also indicates that the intent of the policy is that "the response should be proportionate to the likelihood of degradation, the magnitude and the risk to the environment, and the risk of not achieving the target attribute state".

We recommend gathering data describing possible causes of trends, such as changes in land use practices and intensity, changes in point source discharge loads, and adoption of actions in the catchments of monitoring sites and across the Region in general. These data will be useful explanatory variables in any future attempt to robustly attribute water quality changes to anthropogenic causes and in our experience are frequently lacking. We note that this recommendation is consistent with the current MFE guidance regarding S3.18 of the NPS-FM (MFE 2022). That guidance suggests that "monitoring should not be limited to the state of the water body" and that it should include drivers such as land use and land use intensity as well as the implementation of actions such as "the rules and actions aim to halt expansion or intensity".

We recommend that thought is given to how physical changes in relevant water quality drivers are measured and recorded. We suggest that to some extent at least, the relevant measurements are going to be indicated by existing water quality models such as AC's FWMT. In our experience, limitations around accurate representation of water quality drivers includes issues such as insufficient data describing concentrations and flow of point source discharges, insufficient data describing land use (particularly agricultural land use) and insufficient data describing land management practices.

It is unclear whether more frequent monitoring would improve our confidence to act, because attributing the change to causes so that appropriate action is taken will remain a significant issue even if the precision of trend assessments is increased. ***We recommend that more research is needed into attributing water quality changes to causes.***

Robust attribution of cause(s) is difficult and will always be facilitated by increasing the number of monitoring sites. Increasing the site coverage will improve attribution of cause(s) in the long term. This study has also shown the value of having continuous daily flow data at monitoring sites. ***Therefore, we recommend that priority is given to adding sites to the water quality monitoring network that can be associated with flow data. We acknowledge that synthetic (i.e., modelled) flow data may be suitable for this purpose.*** If a flow modelling approach is to be taken at state of environment monitoring sites, AC should consider the ability to predict flows at those sites and the ability to model flows at new sites should be considered along with any other new site selection criteria.

Finally, because attribution of causes to trends is dependent on both the collection of data and analysis (i.e., modelling), ***we recommend monitoring and modelling are treated as equal and mutually informative processes that must work together to fulfil AC's functions and duties under the RMA and NPS-FM.***

8 Acknowledgements

We thank Auckland Council (RIMU) for provision of data and conversations during the course of this study. We thank Caroline Fraser and Ned Norton (LWP Ltd) for help with analysis and review comments on early drafts of this report.

9 References

Ballantine DJ, Davies-Colley RJ (2014) Water quality trends in New Zealand rivers: 1989–2009. Environmental Monitoring and Assessment 186:1939–1950

Best J (2019) Anthropogenic stresses on the world's big rivers. Nature Geoscience 12:7–21

Brown LD, Cai TT, DasGupta A (2001) Interval estimation for a binomial proportion. Statistical science 16:101–133

Chen T, Bao A, Jiapaer G, et al (2019) Disentangling the relative impacts of climate change and human activities on arid and semiarid grasslands in Central Asia during 1982–2015. Science of the Total Environment 653:1311–1325

Choquette AF, Hirsch RM, Murphy JC, et al (2019) Tracking changes in nutrient delivery to western Lake Erie: Approaches to compensate for variability and trends in streamflow. Journal of Great Lakes Research 45:21–39

Franklin P, Stoffels R, Clapcott J, et al (2019) Deriving potential fine sediment attribute thresholds for the National Objectives Framework. NIWA, Hamilton, New Zealand

Gadd J, Williamson B, Mills GN, et al (2019) Developing Auckland-specific ecosystem health attributes for copper and zinc: summary of work to date and identification of future tasks. Auckland Council

Gascuel-Odoux C, Aurousseau P, Durand P, et al (2010) The role of climate on inter-annual variation in stream nitrate fluxes and concentrations. Science of the total environment 408:5657–5666

Helsel DR, Hirsch RM, Ryberg KR, et al (2020) Statistical methods in water resources. Reston, VA

Hirsch RM, Archfield SA, De Cicco LA (2015) A bootstrap method for estimating uncertainty of water quality trends. Environmental Modelling & Software 73:148–166. https://doi.org/10.1016/j.envsoft.2015.07.017

Hirsch RM, De Cicco LA (2015) User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data. US Geological Survey

Hirsch RM, Moyer DL, Archfield SA (2010) Weighted regressions on time, discharge, and season (WRTDS), with an application to Chesapeake Bay river inputs 1. JAWRA Journal of the American Water Resources Association 46:857–880

Ingley R (2021a) River Water Quality State and Trends in Tāmaki Makaurau / Auckland 2010-2019. State of the Environment Reporting. Auckland Council, Auckland, New Zealand

Ingley R (2021b) River Water Quality State and Trends in Tāmaki Makaurau / Auckland 2010-2019. State of the Environment Reporting. Auckland Council, Auckland

Ingley R, Groom J (2022) River Water Quality in Tāmaki Makaurau / Auckland 2020 Annual Reporting and National Policy Statement for Freshwater Management Current State Assessment. Auckland Council, Auckland, New Zealand

Ingley R, Groom J (2021) River Water Quality in Tāmaki Makaurau / Auckland 2019 Annual Reporting and National Policy Statement for Freshwater Management Current State Assessment. Auckland Council, Auckland

Kundzewicz Z, Robson A (2000) Detecting trend and other changes in hydrological data. World Meteorological Organization

Lanzante JR (2021) Testing for differences between two distributions in the presence of serial correlation using the Kolmogorov–Smirnov and Kuiper's tests. International Journal of Climatology 41:6314–6323

Larned S, Whitehead A, Fraser CE, et al (2018) Water quality state and trends in New Zealand rivers. Analyses of national-scale data ending in 2017. NIWA

Larned ST, Snelder T, Unwin MJ, McBride GB (2016) Water quality in New Zealand rivers: current state and trends. New Zealand Journal of Marine and Freshwater Research 50:389–417

Lee H, Lau S-L, Kayhanian M, Stenstrom MK (2004) Seasonal first flush phenomenon of urban stormwater discharges. Water research 38:4153–4163

Margariti J, Rangecroft S, Parry S, et al (2019) Anthropogenic activities alter drought termination. Elementa: Science of the Anthropocene 7:

Massey FJ (1951) The Kolmogorov-Smirnov test for goodness of fit. Journal of the American statistical Association 46:68–78

McBride GB (2016) National Objectives Framework. Statistical considerations for design and assessment. NIWA, Hamilton, New Zealand

McBride GB (2005) Using statistical methods for water quality management: issues, problems and solutions. John Wiley & Sons, New Jersey

MFE (2020) National Policy Statement for Freshwater Management 2020

MFE (2022) Guidance on the National Objectives Framework of the National Policy Statement for Freshwater Management. Ministry for the Environment., Wellington, New Zealand

Moriasi DN, Gitau MW, Pai N, Daggupati P (2015) Hydrologic and water quality models: Performance measures and evaluation criteria. Transactions of the ASABE 58:1763–1785

Mullan B (1996) Effects of ENSO on New Zealand and the South pacific. The Royal Society of New Zealand, Wellington, New Zealand

Murphy JC (2020) Changing suspended sediment in United States rivers and streams: linking sediment trends to changes in land use/cover, hydrology and climate. Hydrology & Earth System Sciences 24:991–1010

NZ Government (1991) Resource Management Act

Ryberg KR (2017) Structural equation model of total phosphorus loads in the Red River of the North Basin, USA and Canada. Journal of environmental quality 46:1072–1080

Ryberg KR, Blomquist JD, Sprague LA, et al (2018) Modeling drivers of phosphorus loads in Chesapeake Bay tributaries and inferences about long-term change. Science of the Total Environment 616:1423–1430

Salinger MJ, Mullan AB (1999) New Zealand climate: temperature and precipitation variations and their links with atmospheric circulation. International Journal of Climatology: A Journal of the Royal Meteorological Society 19:1049–1071

Scarsbrook MR, McBride CG, McBride GB, Bryers GG (2003) Effects of Climate Variability on Rivers: Consequences for Long Term Water Quality Analysis. JAWRA Journal of the American Water Resources Association 39:1435–1447. https://doi.org/10.1111/j.1752-1688.2003.tb04429.x

Smith DG, McBride GB, Bryers GG, et al (1996) Trends in New Zealand's national river water quality network. New Zealand Journal of Marine and Freshwater Research 30:485–500

Snelder T, Fraser C, Larned S, Whitehead A (2021a) Guidance for the analysis of temporal trends in environmental data. NIWA, Christchurch

Snelder T, Fraser C, Whitehead A (2022) Continuous measures of confidence in direction of environmental trends at site and other spatial scales. Environmental Challenges In press:1–11

Snelder TH, Fraser C, Larned ST, et al (2021b) Attribution of river water-quality trends to agricultural land use and climate variability in New Zealand. Marine and Freshwater Research

Snelder TH, J. Booker D (2013) Natural flow regime classifications are sensitive to definition procedures. River Research and Applications 29:822–838

Snelder TH, Larned ST, Fraser C, De Malmanche S (2021c) Effect of climate variability on water quality trends in New Zealand rivers. Marine and Freshwater Research

Snelder TH, McDowell RW, Fraser CE (2017) Estimation of catchment nutrient loads in New Zealand using monthly water quality monitoring data. JAWRA Journal of the American Water Resources Association 53:158–178

Sofi MS, Bhat SU, Rashid I, Kuniyal JC (2020) The natural flow regime: A master variable for maintaining river ecosystem health. Ecohydrology 13:e2247

Whitehead A, Fraser CE, Snelder TH, et al (2021) Water quality state and trends in New Zealand Rivers. Analyses of national data ending in 2020. NIWA, Christchurch

Wilson E (1927) Calculating a confidence interval of a proportion. J Am Stat Assoc 22:209–212

Wilusz DC, Harman CJ, Ball WP (2017) Sensitivity of catchment transit times to rainfall variability under present and future climates. Water Resources Research 53:10231–10256

Zar JH (1999) Biostatistical Analysis. Prentice-Hall, New Jersey

Appendix A Site Details

Table 8. Details of the 15 water quality sites and related flow sites that were used in the analyses presented in this report.

Water Quality Site Name	Water Quality Site ID	Water quality record length (years) and number of observations	Flow Site Name	Flow Site ID	Flow Record length (years)	Comment on water quality and flow site locations
Hoteo	45703	32 (386)	Hoteo River	45703	43.6	Paired
Mahurangi River (Warkworth)	6804	28 (321)	College	6806	36.1	Water quality site approx. 750 m downstream of flow site
West Hoe Stream	7206	19 (216)	West Hoe Stream	7206	18.2	Paired
Kaukapakapa River	45415	12 (143)	Kaukapakapa @Taylors	45415	22.6	Paired
Vaughan Stream	7506	20 (223)	Vaughn Stream	7506	18	Paired
Lucas Creek	7830	28 (327)	Lucas Creek	7830	14.3	Paired
Oteha River	7811	35 (413)	Oteha Stream	7811	40.4	Paired
Rangitopuni River	7805	35 (325)	Rangitopuni River	7805	45.8	Water quality site approx. 900 m downstream of flow site
Opanuku Stream	7904	35 (325)	Opanuku Stream	7904	14.6	Paired
Otara Creek (East)	8205	35 (415)	Hills Road Bridge	8208	29.1	Water quality site approx. 1.2 km upstream of flow site
Puhinui Stream	43807	29 (347)	Puhinui Stream	43807	41.6	Paired
Wairoa River	8516	27 (321)	Wairoa River	8516	42	Paired
Papakura Stream (Lower)	43856	35 (414)	Gt Sth Rd	43803	48.6	Water quality site approx. 2km upstream of flow site
Ngakoroa Stream	43829	28 (329)	Ngakoroa Stream	43829	40.8	Paired
Waitangi Stream	43601	28 (329)	Waitangi @ Weir	43602	55.9	Water quality site approx. 1.2 km downstream of flow site

Appendix B Hydrographs

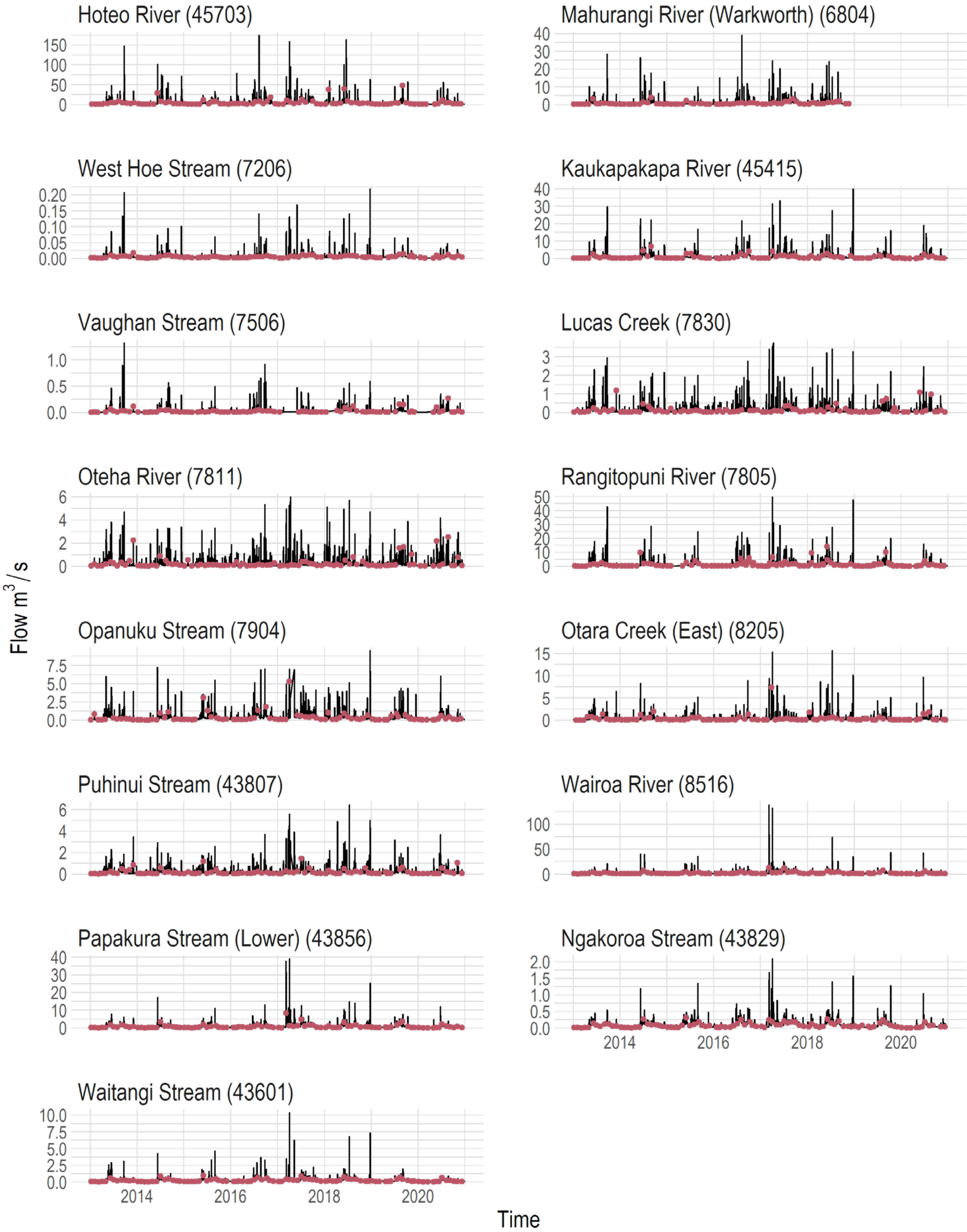


Figure 32. Hydrographs and water quality observation day flows for the 15 water quality sites used in the analyses presented in this report. The hydrographs are indicated as black lines and the observation day flows as maroon dots. Larger versions of these plots are available in the supplementary file "Appendix B Hydrographs.pdf".

Appendix C Flow distributions

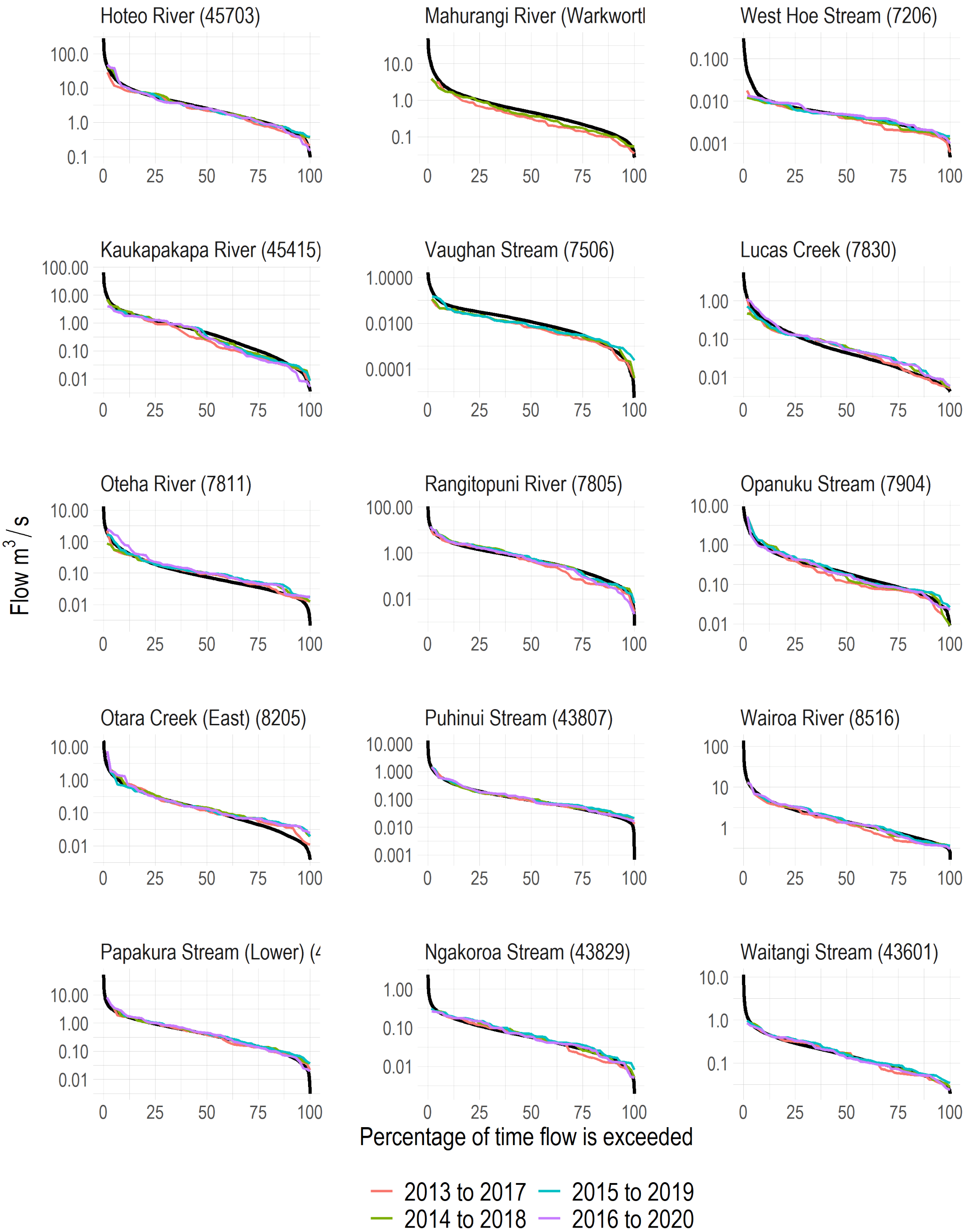


Figure 33. Flow duration curves for 5-year periods and for the full flow record for the 15 water quality sites used in the analyses presented in this report. Note the y-axes are log transformed. Larger versions of these plots are available in the supplementary file “Appendix C Flow distributions.pdf”.

Appendix D Site-Variable flow-adjustment models

Table 9. Site-variable combinations with plausible flow-adjustment models. The criteria for plausibility is described in Section 4.4 and includes r^2 ≥ 20% and p<0.01. All subjectively selected models had p values less than 0.001

Site	Variable	Model	r^2	Number of observations
Hoteo River (45703)	CLAR	GAM	0.76	381
Hoteo River (45703)	TN	LinLog	0.61	367
Hoteo River (45703)	DIN	LinLog	0.49	536
Hoteo River (45703)	NNN	LinLog	0.46	563
Hoteo River (45703)	TP	GAM	0.4	647
Hoteo River (45703)	ECOLI	GAM	0.33	217
Mahurangi River (Warkworth) (6804)	CLAR	GAM	0.73	175
Mahurangi River (Warkworth) (6804)	CU	LinLog	0.35	100
Mahurangi River (Warkworth) (6804)	TN	LinLog	0.35	118
Mahurangi River (Warkworth) (6804)	DIN	LinLog	0.33	290
Mahurangi River (Warkworth) (6804)	NNN	LinLog	0.33	295
Mahurangi River (Warkworth) (6804)	TP	GAM	0.32	295
West Hoe Stream (7206)	CLAR	GAM	0.22	195
Kaukapakapa River (45415)	TN	LinLog	0.61	142
Kaukapakapa River (45415)	DIN	LinLog	0.59	141
Kaukapakapa River (45415)	NNN	LinLog	0.59	141
Kaukapakapa River (45415)	CLAR	LinLog	0.42	143
Kaukapakapa River (45415)	ECOLI	GAM	0.25	143
Kaukapakapa River (45415)	TP	GAM	0.22	143
Vaughan Stream (7506)	CLAR	LinLog	0.32	186
Vaughan Stream (7506)	CU	GAM	0.32	168
Vaughan Stream (7506)	TN	GAM	0.27	126
Vaughan Stream (7506)	TP	GAM	0.24	191
Vaughan Stream (7506)	NNN	LinLog	0.19	190
Vaughan Stream (7506)	DIN	LinLog	0.17	188
Lucas Creek (7830)	CLAR	LinLog	0.54	157
Lucas Creek (7830)	CU	LinLog	0.43	157

Site	Variable	Model	r^2	Number of observations
Lucas Creek (7830)	ECOLI	LinLog	0.18	157
Oteha River (7811)	CLAR	GAM	0.64	201
Oteha River (7811)	ECOLI	LinLog	0.21	173
Oteha River (7811)	ZN	LinLog	0.21	295
Rangitopuni River (7805)	CLAR	LinLog	0.62	465
Rangitopuni River (7805)	NNN	LinLog	0.49	586
Rangitopuni River (7805)	DIN	LinLog	0.48	568
Rangitopuni River (7805)	TN	LinLog	0.45	394
Rangitopuni River (7805)	TP	GAM	0.31	668
Rangitopuni River (7805)	ECOLI	GAM	0.28	247
Opanuku Stream (7904)	CLAR	LinLog	0.57	167
Opanuku Stream (7904)	TN	LinLog	0.22	140
Otara Creek (East) (8205)	CLAR	LinLog	0.49	201
Otara Creek (East) (8205)	NNN	LinLog	0.38	339
Otara Creek (East) (8205)	TN	LinLog	0.33	141
Otara Creek (East) (8205)	DIN	LinLog	0.29	337
Puhinui Stream (43807)	CLAR	LinLog	0.45	201
Puhinui Stream (43807)	TN	LinLog	0.38	143
Puhinui Stream (43807)	DIN	LinLog	0.14	311
Puhinui Stream (43807)	NNN	LinLog	0.13	313
Wairoa River (8516)	CLAR	GAM	0.64	202
Wairoa River (8516)	TN	LinLog	0.64	143
Wairoa River (8516)	DIN	LinLog	0.58	341
Wairoa River (8516)	NNN	LinLog	0.57	343
Wairoa River (8516)	ECOLI	GAM	0.34	172
Wairoa River (8516)	CU	LinLog	0.16	125
Papakura Stream (Lower) (43856)	CU	LinLog	0.59	126
Papakura Stream (Lower) (43856)	CLAR	LinLog	0.58	203
Papakura Stream (Lower) (43856)	TN	GAM	0.58	144
Papakura Stream (Lower) (43856)	NNN	LinLog	0.52	325
Papakura Stream (Lower) (43856)	DIN	LinLog	0.5	323

Site	Variable	Model	r^2	Number of observations
Papakura Stream (Lower) (43856)	TP	GAM	0.42	325
Papakura Stream (Lower) (43856)	ZN	LinLog	0.4	125
Ngakoroa Stream (43829)	TN	GAM	0.41	143
Ngakoroa Stream (43829)	CLAR	LinLog	0.35	202
Ngakoroa Stream (43829)	DIN	GAM	0.13	320
Ngakoroa Stream (43829)	NNN	GAM	0.13	322
Waitangi Stream (43601)	TN	LinLog	0.71	144
Waitangi Stream (43601)	CLAR	LinLog	0.67	143
Waitangi Stream (43601)	DIN	LinLog	0.64	142
Waitangi Stream (43601)	NNN	LinLog	0.63	142
Waitangi Stream (43601)	TP	GAM	0.51	142
Waitangi Stream (43601)	NH4N	GAM	0.41	140
Waitangi Stream (43601)	DRP	LinLog	0.26	141
Waitangi Stream (43601)	ECOLI	GAM	0.23	144

Appendix E Trend Assessments

Table 10. Trend likelihood categories and their related Mann-Kendall statistics.

Trend likelihood category	Trend symbol	Mann-Kendall S statistic	Confidence the trend was decreasing
Highly likely decreasing	+++	Negative	0.95 – 1.0
Very likely decreasing	++	Negative	0.9 – 0.95
Likely decreasing	+	Negative	0.67 – 0.9
As likely increasing as decreasing	±	Negative	0.50 – 0.67
As likely increasing as decreasing	±	Positive	0.50 – 0.67
Likely increasing	-	Positive	0.67 – 0.90
Very likely increasing	--	Positive	0.90 – 0.95
Highly likely increasing	---	Positive	0.95 – 1.0

Table 11. Trend categories for site-variable combinations before and after flow-adjustment.

Site	Variable	2007-2017		2008-2018		2009-2019		2010-2020	
		Raw	Adj	Raw	Adj	Raw	Adj	Raw	Adj
Hoteo River (45703)	TP	-	--	+	±	+	+	--	±
Hoteo River (45703)	TN	-	±	+	±	+	±	±	++
Hoteo River (45703)	NNN	---	--	±	±	±	±	±	NA
Hoteo River (45703)	ECOLI	-	-	+	±	+	+	++	+++
Hoteo River (45703)	DIN	---	--	±	±	±	±	±	+
Hoteo River (45703)	CLAR	+++	+++	±	+	±	+	++	±
Mahurangi River (Warkworth) (6804)	TP	++	+++	++	+	±	++	-	+
Mahurangi River (Warkworth) (6804)	TN	--	--	-	--	±	±	+	±
Mahurangi River (Warkworth) (6804)	NNN	--	±	±	±	+++	±	+	±
Mahurangi River (Warkworth) (6804)	DIN	---	±	-	-	+	±	±	-
Mahurangi River (Warkworth) (6804)	CU	-	±	+	±	+	-	+	-
Mahurangi River (Warkworth) (6804)	CLAR	++	++	±	±	±	-	+++	±
West Hoe Stream (7206)	CLAR	+++	+++	+++	+++	+++	+++	+++	+++
Kaukapakapa River (45415)	TP	---	---	---	---	---	---	---	---
Kaukapakapa River (45415)	TN	---	--	--	---	-	-	--	+
Kaukapakapa River (45415)	NNN	-	±	-	±	-	-	---	±
Kaukapakapa River (45415)	ECOLI	±	±	±	±	-	±	-	±
Kaukapakapa River (45415)	DIN	-	-	±	--	-	-	--	±
Kaukapakapa River (45415)	CLAR	+++	+++	+++	+++	+	±	++	±
Vaughan Stream (7506)	TP	-	±	-	±	---	-	--	--
Vaughan Stream (7506)	TN	---	-	---	---	---	--	-	±
Vaughan Stream (7506)	NNN	+	+++	+	+	+	±	+	±
Vaughan Stream (7506)	DIN	+	+++	±	+	-	±	-	+
Vaughan Stream (7506)	CU	+	+++	+	+++	++	++	+++	+++
Vaughan Stream (7506)	CLAR	-	---	---	---	---	---	-	---

Site	Variable	2007-2017		2008-2018		2009-2019		2010-2020	
		Raw	Adj	Raw	Adj	Raw	Adj	Raw	Adj
Lucas Creek (7830)	ECOLI	+	+	+++	+	+	±	±	±
Lucas Creek (7830)	CU	-	--	++	+	+++	+++	+++	+++
Lucas Creek (7830)	CLAR	+++	+++	±	+++	+	+++	++	+++
Oteha River (7811)	ZN	±	+	±	±	+	±	+	++
Oteha River (7811)	ECOLI	++	+++	+++	+	+++	+	+	±
Oteha River (7811)	CLAR	±	±	---	+	--	±	±	+
Rangitopuni River (7805)	TP	---	---	--	---	---	---	---	---
Rangitopuni River (7805)	TN	±	+	+	±	±	-	---	-
Rangitopuni River (7805)	NNN	-	-	-	NA	--	NA	---	NA
Rangitopuni River (7805)	ECOLI	-	-	±	±	±	±	-	±
Rangitopuni River (7805)	DIN	--	-	-	---	---	--	---	NA
Rangitopuni River (7805)	CLAR	+	+	-	±	-	-	+++	+++
Opanuku Stream (7904)	TN	--	-	-	-	±	±	+	++
Opanuku Stream (7904)	CLAR	+	+++	+	+	+	±	+	+
Otara Creek (East) (8205)	TN	---	---	---	---	---	---	-	-
Otara Creek (East) (8205)	NNN	---	---	-	---	-	±	±	++
Otara Creek (East) (8205)	DIN	---	---	---	---	---	-	-	+
Otara Creek (East) (8205)	CLAR	±	+++	-	±	±	±	+	+
Puhinui Stream (43807)	TN	---	---	--	---	-	--	---	-
Puhinui Stream (43807)	NNN	---	---	-	±	±	±	-	±
Puhinui Stream (43807)	DIN	---	---	-	-	±	-	-	±
Puhinui Stream (43807)	CLAR	+++	+++	+++	+++	+++	+++	+++	+++
Wairoa River (8516)	TN	---	---	--	---	±	-	±	+
Wairoa River (8516)	NNN	---	---	-	---	+	±	+	+++
Wairoa River (8516)	ECOLI	---	---	---	---	---	---	--	---
Wairoa River (8516)	DIN	---	---	-	---	++	±	+	+++
Wairoa River (8516)	CU	-	--	±	±	++	+	+	++
Wairoa River (8516)	CLAR	+++	+++	++	+++	±	+	+++	+
Papakura Stream (Lower) (43856)	ZN	±	-	-	---	-	---	-	-

Site	Variable	2007-2017		2008-2018		2009-2019		2010-2020	
		Raw	Adj	Raw	Adj	Raw	Adj	Raw	Adj
Papakura Stream (Lower) (43856)	TP	---	---	---	---	±	--	±	--
Papakura Stream (Lower) (43856)	TN	-	---	--	---	-	--	±	±
Papakura Stream (Lower) (43856)	NNN	-	---	±	---	±	-	±	+++
Papakura Stream (Lower) (43856)	DIN	---	---	-	---	-	-	±	+++
Papakura Stream (Lower) (43856)	CU	+	±	++	+	+++	+++	+	++
Papakura Stream (Lower) (43856)	CLAR	+	+++	--	±	---	---	±	---
Ngakoroa Stream (43829)	TN	+	±	±	±	-	-	-	-
Ngakoroa Stream (43829)	NNN	+++	+++	+++	+++	+++	++	++	+++
Ngakoroa Stream (43829)	DIN	+++	+++	+++	+++	+++	++	++	+++
Ngakoroa Stream (43829)	CLAR	±	+	--	-	±	+	+++	+++
Waitangi Stream (43601)	TP	+	-	+	-	-	---	---	---
Waitangi Stream (43601)	TN	++	+++	+++	+++	+++	+++	±	+++
Waitangi Stream (43601)	NNN	+++	+++	+++	+++	+++	+++	+++	+++
Waitangi Stream (43601)	NH4N	---	---	---	---	--	---	-	-
Waitangi Stream (43601)	ECOLI	---	---	---	---	---	---	--	---
Waitangi Stream (43601)	DRP	+++	+++	±	±	--	--	---	---
Waitangi Stream (43601)	DIN	+++	+++	+++	+++	+++	+++	+++	+++
Waitangi Stream (43601)	CLAR	+	++	±	+	++	+	+++	+++