Roxburgh Rohe
(Clutha/Mata-Au FMU)
River & Lake Water Quality
State and Trends

Benger Burn at Moa Flat Road

April 2021

Otago Regional Council

Contents

Executive Summary

This study analysed the available water quality data in rivers and lakes in section of the Clutha/Mata-Au catchment between the Clyde Dam and Beaumont. Four river sites and one lake sites are monitored in the Roxburgh Rohe and the state of water quality is reported relative to targets specified in the National Objectives Framework (NOF) of the National Policy Statement-Freshwater Management (NPSFM 2020). In addition, the study assessed water quality trends site by site. ORC engaged Land Water People (LWP) to evaluate water quality state and undertake trend analysis.

State analysis was based on water quality samples collected over a five-year period from 1 July 2015 to 30 June 2020 and compared to the five-year period 1 July 2012 to 30 June 2017, which is defined as the baseline state (NPSFM 2020).

This report describes only river and lake state and trends for the variables that specifically relate to the NPSFM (2020); chlorophyll-a, total nitrogen, total phosphorus, ammoniacal-nitrogen, nitrate, suspended fine sediment, macroinvertebrate community index (MCI), macroinvertebrate average score per metric (ASPM), dissolved reactive phosphorus and *E. coli*.

Sites were graded as a NOF Band (A, B, C, D, and for *E. coli*) (for NOF criteria) for each variable based on a comparison of the assessed state with the relevant criteria. Trend analysis was carried out for 10-year and 20-year periods ending on 1 September 2020 for all site and water quality variable combinations that met a minimum requirement for numbers of observations.

There is a lack of detailed information held by Otago Regional Council on local or catchment scale land use change or land management practice changes. This limits Council's ability to comment on drivers of trends evident across Otago. This will be addressed by requirements in the the NPSFM (2020), which requires that freshwater is managed in an integrated way that considers the effects of the use and development of land on a whole-of-catchment basis, including the effects on receiving environments.

1 Introduction

Otago Regional Council (ORC) operates a State of Environment (SoE) water quality monitoring network in lakes and rivers throughout the region for monitoring the state and trends in water quality and reporting on policy effectiveness. Prior to mid-2018, there were fewer monitoring sites in the Region, following a review (NIWA 2017), a more extensive monitoring programme commenced in mid-2018 to better represent environmental classes in the Otago region, based largely on the River Environment Classification (REC).

1.1 Freshwater management units

To give effect to the NPSFM (2020) and take a more localised approach to water and land management, ORC developed Freshwater Management Unit (FMU) boundaries incorporating the concept of ki uta ki tai (from the mountains to the sea).

The Clutha / Mata-Au FMU is one of five FMUs that were recognised, Figure 1; Clutha/Mata-Au, Taieri, North Otago, Dunedin and Coast, and Catlins. The Clutha/Mata Au FMU has been further divided in to five sub-areas, or ‘Rohe’, for a more tailored water management approach in these areas. These include the Upper Lakes Rohe, Dunstan Rohe, Manuherekia Rohe, Roxburgh Rohe and Lower Clutha Rohe.

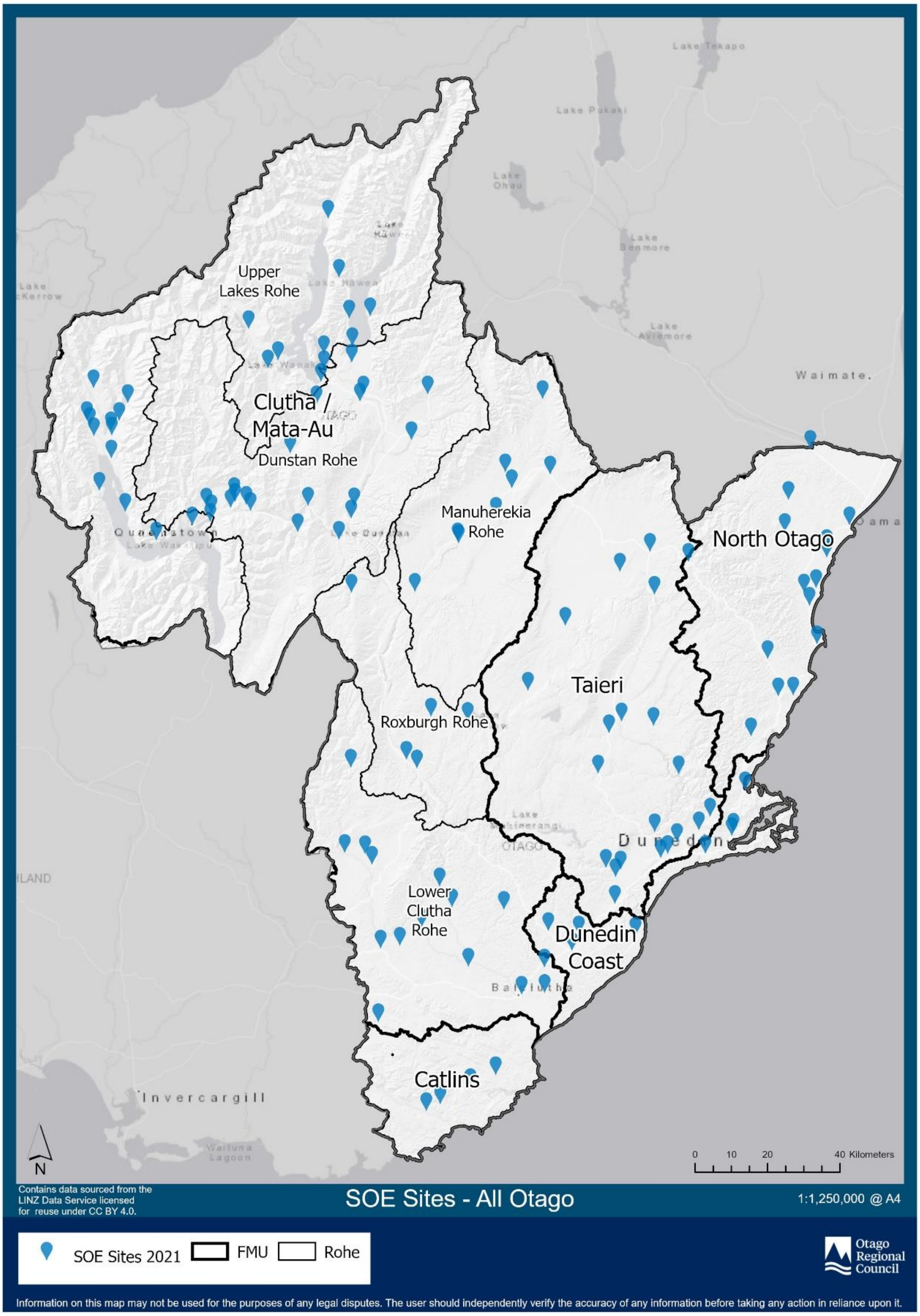

Figure 1 Map showing the FMU and Rohe boundaries, State of Environment monitoring site locations are also shown.

1.2 Roxburgh Rohe

The Roxburgh Rohe is bounded to the north by the Clyde Dam and to the South at Beamont. The Rohe covers just over 1,000 square kilometres and encompasses catchments of the Fraser River (327 km2), the Teviot River (332 km2) and the Benger Burn (131 km^2) as well as small tributaries entering the Clutha between Clyde and Beaumont. Lake Onslow is in the Rohe, it is a man-made 830ha lake at the head of the Teviot River, it lies 700m above sea level.

2 Water Quality

2.1 Water quality variables

Water quality was assessed using variables that characterise physical, chemical and microbiological conditions, and macroinvertebrate community composition. All variables included are attributes described in Appendix 2A or 2B of the NPSFM (2020).

2.1.1 Phytoplankton, Periphyton and Nutrients

Healthy freshwater ecosystems have low (oligotrophic) to intermediate (mesotrophic) levels of living material and primary production (growth of plants or algae). High levels of nutrients, primarily nitrogen (nitrate) and phosphorus (phosphate), can cause water bodies to become eutrophic. Eutrophic states are associated with periodic high biomass (blooms) of plants or algae, including suspended algae (phytoplankton) in lakes and algae on the beds of streams and rivers (periphyton).

Chlorophyll-*a* is a common method for estimating stream periphyton biomass (e.g., as used within Ministry for Environment, 2000) because all types of algae contain chlorophyll-*a*, this metric reflects the total amount of live algae in a sample. The trophic state of a water body is the amount of living material (biomass) that it supports. The NPSFM specifies attributes for tropic state based on phytoplankton biomass in lakes (Table 1, Appendix 2A, NPSFM) and periphyton biomass in rivers (Table 2, Appendix 2A, NPSFM), chlorophyll-*a* is the measure of biomass that the NOF phytoplankton and periphyton attributes are based on.

Nitrate (NO3N), ammoniacal-N (NH4N), dissolved reactive phosphorus (DRP), total nitrogen (TN) and total phosphorus (TP) influence the growth of benthic river algae (periphyton), lake planktonic algae (phytoplankton) and vascular plants (macrophytes). The NPSFM specifies additional attributes for TN and TP in lakes (Table 3 and Table 4, Appendix 2A, NPSFM).

The NPSFM does not specify nutrient concentration criteria to manage the trophic state of rivers, because the relationship between trophic state and nutrient concentrations varies between rivers even at the regional scale. The nutrient criteria to achieve periphyton biomass objectives in rivers are river specific and should be derived at the local level (MfE, 2018).

The Ministry for the Environment has produced guidance (MfE, 2020) for defining nutrient concentrations to manage the NPSFM periphyton attribute states in rivers. The guidance is centered around spatial exceedances for TN and DRP. Spatial exceedence is used because deriving nutrient targets to achieve a target periphyton growth cannot be 100% certain due to natural variability, complex interactions in the environment, and the complexity of the relationship between nutrients and periphyton abundance (MfE, 2020). Given the short record of chlorophyl-a observations in the region, these nutrient concentration criteria provide a useful alternative for estimating trophic state in the region’s rivers.

In this report TN and DRP median concentrations are compared to the spatial exceedence criteria of 20% (as opposed to 10% or 30%). At this level there is some risk (ie, 20%) that the chlorophyll *a* response at some sites will exceed the desired chlorophyll *a* threshold, even if the DRP or TN concentration targets are achieved.

In addition to the MfE guidance, the NPSFM provides an attribute table for DRP in rivers to protect ecosystem health. In combination with other conditions favouring eutrophication, DRP enrichment drives excessive primary production and significant changes in macroinvertebrate and fish communities, as taxa sensitive to hypoxia are lost. Table 20 (NPSFM, Appendix 2B) describes that at concentrations below the national bottom line it is expected that ecological communities are impacted by substantial DRP elevated above natural reference conditions.

2.1.2 Toxicants

When ammonia is present in water at high enough concentrations, it is difficult for aquatic organisms to sufficiently excrete the toxicant, leading to toxic build-up in internal tissues and blood, and potentially death. Environmental factors, such as pH and temperature, can affect ammonia toxicity to aquatic animals. The NPSFM has developed an ammonia toxicity risk framework (Table 5, Appendix 2A, NPSFM) when toxicity concentrations are below the national bottom line, toxicity starts impacting regularly on the 20% most sensitive species.

Nitrate generally impacts on trophic state at much lower concentrations than those that are toxic. Because of this, nitrate will generally be managed well within toxic levels by the requirement to manage trophic state (eg, periphyton). The NPSFM has developed a nitrate toxicity risk framework (Table 6, Appendix A, NPSFM) when toxicity concentrations are below the national bottom line, toxicity has growth effects on u[tp 20% of species

2.1.3 Suspended sediment

Suspended fine sediment can severely affect values around water, particularly around ecosystem health. High concentrations of suspended sediment have a '*high impact on instream biota and ecological communities are significantly altered* ***and sensitive fish and macroinvertebrate species are lost or at high risk of being lost'*** (NPSFM, 2020). Suspended fine sediment can be monitored by clarity or turbidity measurements.

Clarity is a measure of light attenuation due to absorption and scattering by dissolved and particulate material in the water column. Clarity is monitored because it affects primary production, plant distributions, animal behaviour, aesthetic quality and recreational values, and because it is correlated with suspended solids, which can impede fish feeding and cause riverbed sedimentation. Clarity is the metric used in the NPSFM suspended fine sediment attribute table (Table 8, Appendix A, NPSFM)

Turbidity which refers to light scattering by suspended particles. Nephelometric turbidity is generally inversely correlated with visual water clarity (Davies-Colley and Smith 2001), but unlike visual clarity, turbidity measurements do not account for the optical effects (i.e., absorption) of dissolved materials. The NPSFM allows for the conversion of turbidity to visual clarity, ORC does not measure visual clarity and applies this conversion.

2.1.4 Aquatic Life

Macroinvertebrates are an important component of streams and rivers because they aid ecosystem processes and provide food for fish and some birds. As macroinvertebrates have a relatively long-life span, they are good indicators of environmental conditions over a prolonged period.

Macroinvertebrates are included in the NPSFM as attributes requiring an action plan (Tables 14-15, NPSFM, Appendix 2B).

The main measure of macroinvertebrate communities, the MCI index, is designed specifically for stony-riffle substrates in flowing water. The MCI is responsive to multiple stressors, but not all stressors, and as such provides a good indicator of the overall condition of the macroinvertebrate component of stream ecosystem health.

MCI values can be affected by factors other than water quality, so it is more informative to consider changes in MCI values at the same site over a period, rather than among sites throughout the catchment. For example, a change in MCI value at a site may be due to human activities causing increased nitrogen or sedimentation with resulting ecological consequences (Clapcott et al. 2018). Sites with an MCI score of less than 80 are classified as poor, those scoring 80-100 as fair, those scoring 100-120 as good, and those scoring higher than 120 as excellent (Stark and Maxted 2007).

The NPSFM has attribute states for Macroinvertebrate Community Index (MCI) score; Quantitative Macroinvertebrate Community Index (QMCI) score and Macroinvertebrate Average Score Per Metric (ASPM). Historical monitoring by ORC has included the Semi-Quantitative Macroinvertebrate Community Index (SQMCI) score, rather than QMCI. As the two are not directly comparable the QMCI metric is not shown.

The Average Score Per Metric (ASPM) was introduced by Collier (2008), it is an aggregation method for assessing wadeable stream ecosystem health considering the relative responses of core metrics and is composed of three individual metrics, the MCI, EPT richness to the total taxa found and % EPT abundance. EPT Richness Index estimates water quality by the relative abundance of three major orders of stream insects that have low tolerance to water pollution. EPT can be expressed as a percentage of the sensitive orders (E= Ephemeroptera, P= Plecoptera, T= Tricoptera) and % EPT is the total number of EPT individuals divided by the total number of individuals in the sample).

2.1.5 *Escherichia coli* (*E. coli*)

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

‘Water contaminated by human or animal faeces may contain a range of pathogenic (disease-causing) micro-organisms. Viruses, bacteria, protozoa or intestinal worms can pose a health hazard when the water is used for drinking or recreational activities. It is difficult and impractical to routinely measure the level of all pathogens that may be present in fresh water. Instead, indicator bacteria are used to indicate the likely presence of untreated sewage and effluent contamination.

E. coli is a bacteria commonly found in the gut of warm-blooded organisms and is relatively easy to measure which makes it a useful indicator of faecal presence and therefore of disease-causing organisms that may be present. *E. coli* is the attribute for specifying human health for recreation objectives for fresh water because it is moderately well correlated with Campylobacter bacteria and numeric health risk levels can be calculated. Campylobacteriosis has the highest reporting rate of all New Zealand’s ‘notifiable’ diseases’ (MfE, 2018).

The NPSFM assesses river swimmability and the attribute states uses four statistical measures of *E. coli* concentrations, the overall state is determined by satisfying all numeric attribute states. (Table 9, Appendix 2A, NPSFM).

3 Methods

A detailed summary of water quality state and trend analysis presented in this report is provided in Appendix 1 and 2.

4 Results Roxburgh Rohe

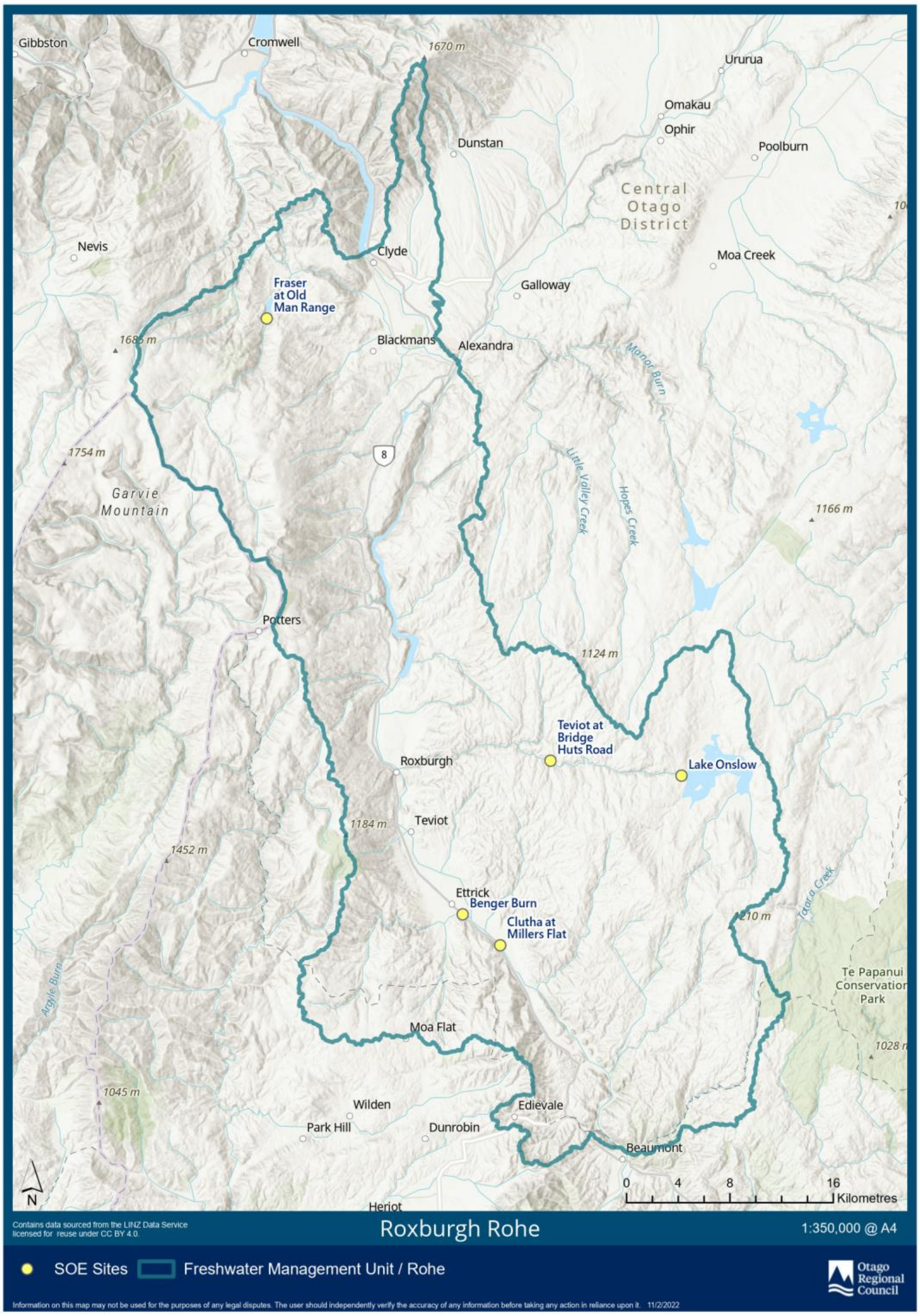

Figure 2 Location of water quality monitoring sites in the Roxburgh Rohe

4.1 State Analysis Results

The results of grading the SoE sites in the Roxburgh Rohe according to the NPSFM NOF criteria are summarised in Figure 3 and mapped Figure 4. Many sites in the Roxburgh Rohe did not meet the sample number requirements (shown in Table 5) and accordingly are shown as white cells with coloured circles

A small square in the upper left quadrant of the cells indicate the site grade for the baseline period (2012-2017) where the sample numbers for that period met the minimum sample number requirements.

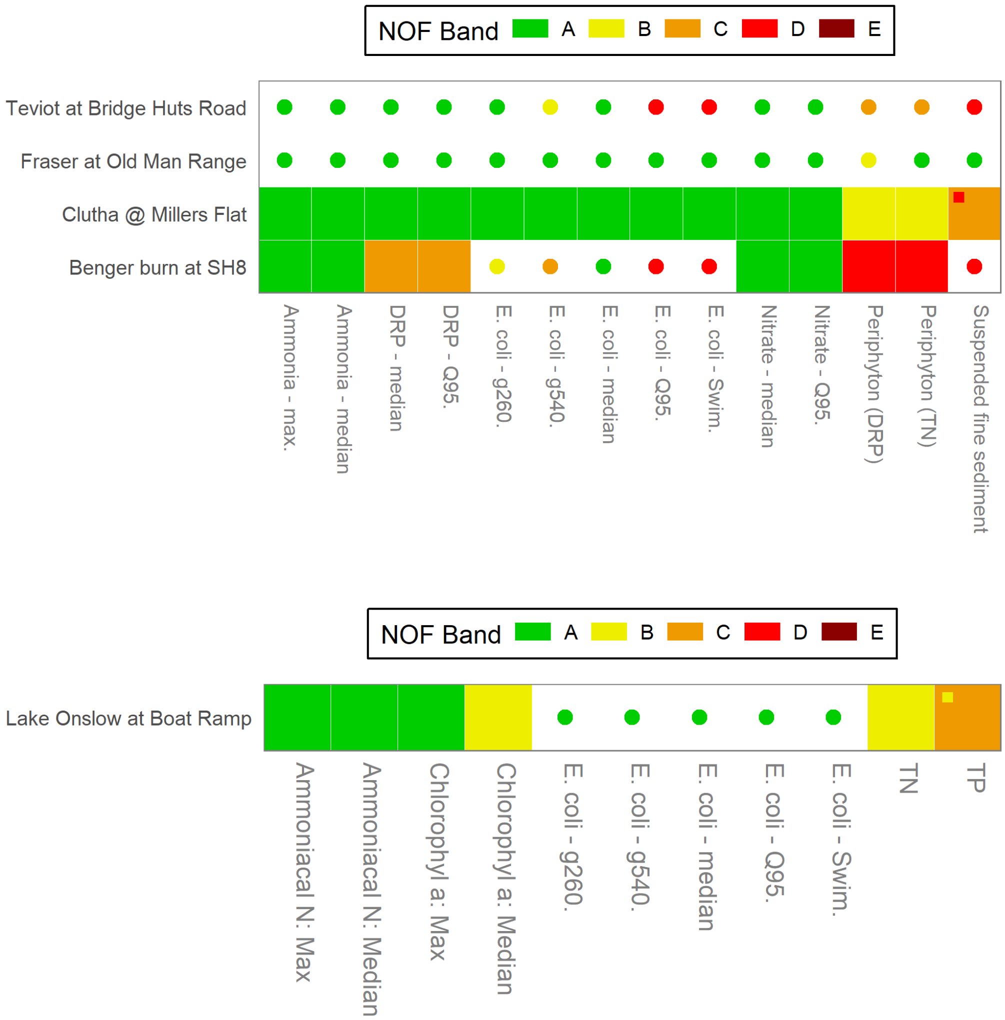

Figure 3 Grading of the river and lake sites in the Roxburgh based on the NOF criteria. Grades for sites that did not meet the sample number requirements in Table 5 are shown as white cells with coloured circles. The white cells indicate sites for which the variable was not monitored. Small square in the upper left quadrant of the cells indicate the site grade for the baseline

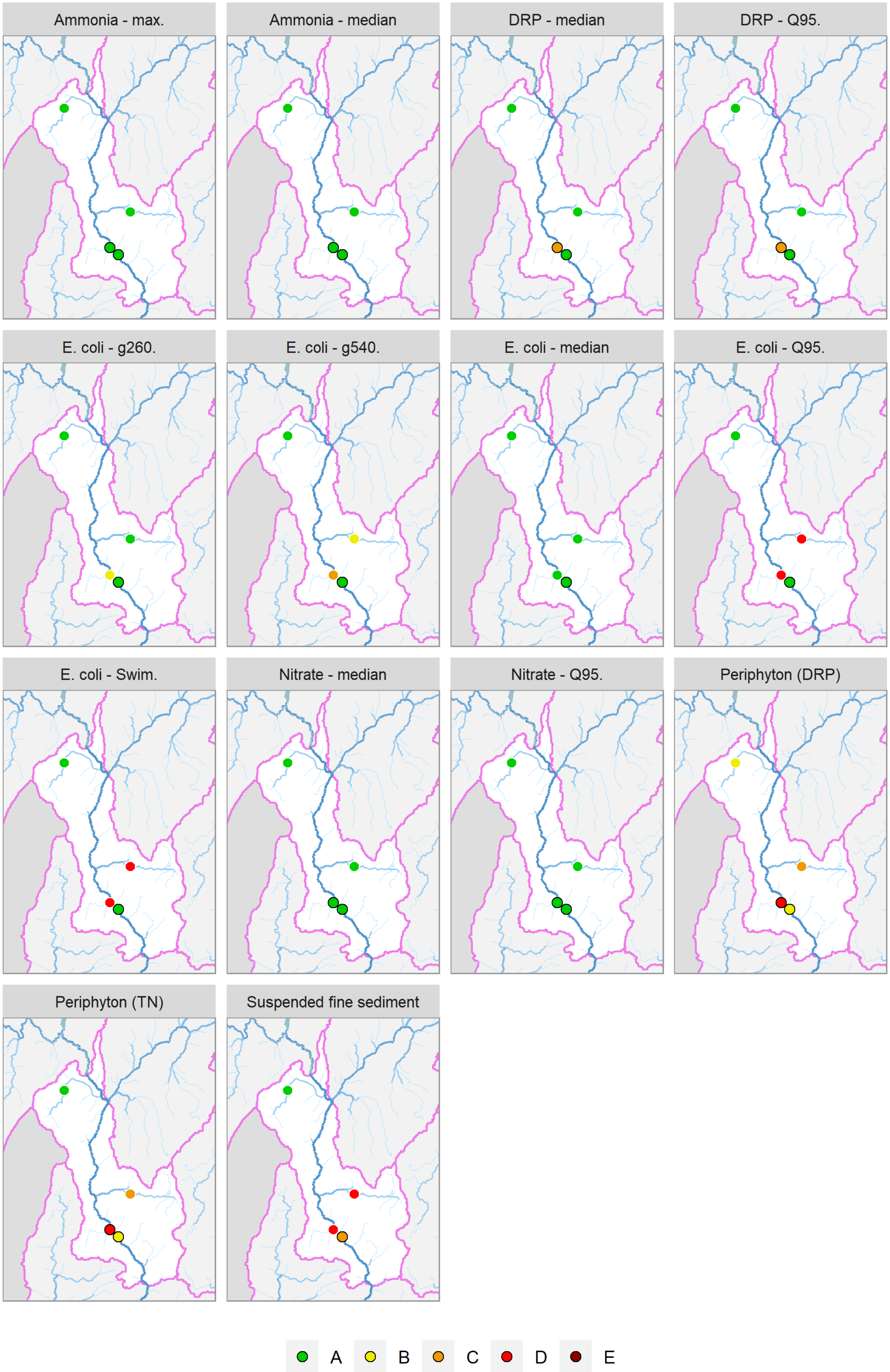

Figure 4 Maps showing Roxburgh Rohe sites coloured according to their state grading as indicated by NOF attribute bands. Bands for sites that did not meet the sample number requirements specified in Table 5 are shown without black outlines.

4.1.1 Periphyton and Nutrients

Results for the river periphyton trophic state results are shown in Figure 3 (periphyton). The Roxburgh Rohe does not have any sites that are monitored for chlorophyll a, but MfE (2020) TN concentrations to manage the NPSFM periphyton attribute states show that the Fraser River achieves a band 'A' as few results exceed 50 chl-*a*/m^2 meaning that blooms would be rare, reflecting negligible nutrient enrichment, the Clutha at Millers Flat a band 'B', the Teviot River a band 'C' and the Benger Burn achieves a band 'D'.

The MfE (2020) DRP concentrations to manage the NPSFM periphyton attribute states show the same pattern as the TN bands, other than the Fraser at Old Man Range which achieves a 'B' for DRP, when it achieved an 'A' band for TN.

Figure 4 also shows DRP attribute states for ecosystem health (DRP median and Q95). The results in the Roxburgh Rohe show that every site achieves a band 'A', other than the Benger burn which achieves band 'C'. The NPSFM (2020) describes band 'C' as *'Ecological communities are impacted by moderate DRP elevation above natural reference conditions. If other conditions also favour eutrophication, DRP enrichment may cause increased algal and plant growth, loss of sensitive macro-invertebrate and fish taxa, and high rates of respiration and decay'*

The NPSFM (2020) describes how phytoplankton affects lake ecological communities. If the chlorophyll a concentration is in the 'A' band, then '*Lake ecological communities are healthy and resilient, similar to natural reference conditions'*. Results for Lake Onslow are shown in Figure 4, the lake achieves an 'A' band for maximum chlorophyll a, but drops to a 'B' band for median chlorophyll a. Lake Onslow achieves a 'B' band for TN and a 'C' band for TP, which indicates that ecological communities are slightly-moderately impacted by additional algal and plant growth arising from nutrient levels above natural reference conditions.

4.1.2 Toxicants (Rivers)

In the Roxburgh Rohe the NOF attribute bands for NH4-N and nitrate (measured as NNN) toxicity) show excellent protection levels against toxicity risk as all monitoring sites return an 'A' band for NH4-N and NNN.

4.1.3 Suspended fine sediment

The clarity results for the Roxburgh Rohe are shown in Figure 3. The Fraser River returns a NOF band of 'A' which denotes '*minimal impact of suspended sediment on instream biota. Ecological communities are similar to those observed in natural reference conditions'* (NPSFM, 2020). The Clutha at Millers Flat returns a NOF band of 'B' and the Benger burn and Teviot return a NOF band of 'D' for suspended fine sediment, which is below the national bottom line.

4.1.4 Human contact

Figure 3 summarises compliance for *E. coli* against the four statistical tests of the NOF *E. coli* attribute. The overall attribute state is based on the worst grading with the national bottom line being a 'D' band.

Lake Onslow, the Fraser River and the Clutha at Millers Flat return 'A' bands across all four statistical tests, but the Teviot and Benger Burn achieve a 'D' band because their interim 95th percentile is >1200E.coli/100ml. Lake Onslow returns a NOF band of 'A'.

4.2 Trend Analysis

Results from trend analysis for the Roxburgh Rohe is shown in Figure 5.

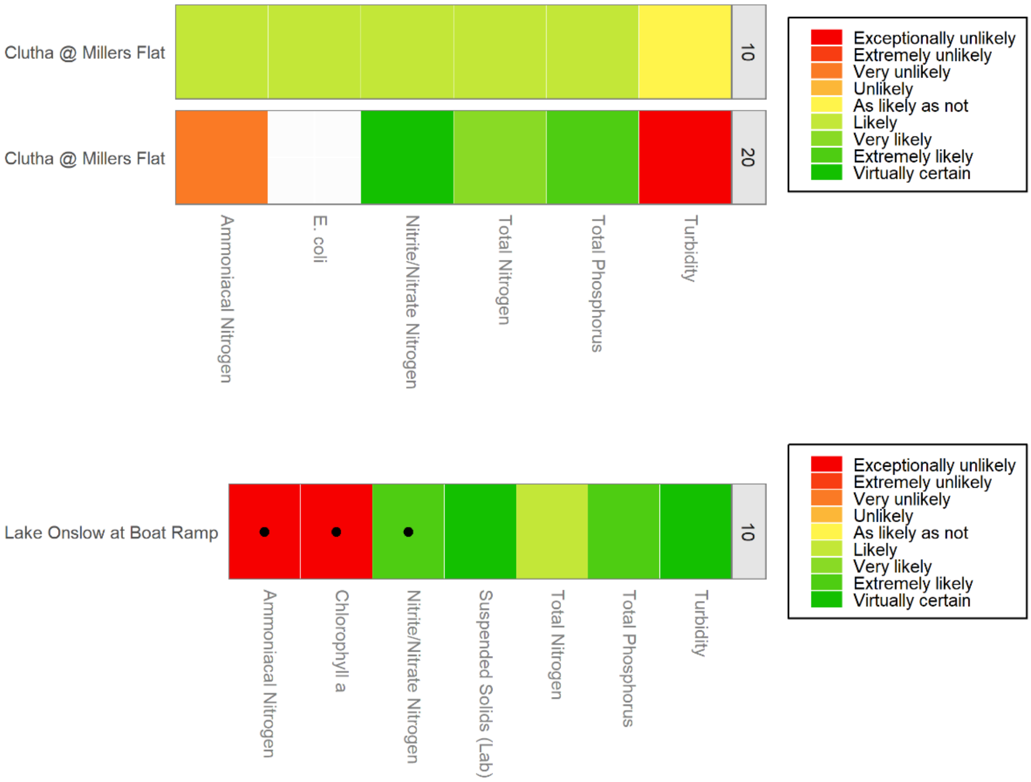

Figure 5 Summary of Roxburgh Rohe sites categorised according to the level of confidence that their 10 and 20-year raw water quality trends indicate improvement. Confidence that the trend indicates improvement is expressed using the categorical levels of confidence defined inTable 6. Cells containing a black dot indicate site/variable combinations where the Sen Slope was evaluated as zero (i.e., a trend rate that cannot be quantified given the precision of the monitoring). White cells indicate site/variables where there were insufficient data to assess the trend.

Trend analysis for both rivers and lakes are given in Figure 5. In the 10-year time frame, Lake Onslow shows 'exceptionally unlikely' improving trends for NH4-N and chlorophyll a, where most of the other variables show 'likely' to 'virtually certain' improving trends. For the Clutha River at Millers Flat, trend analysis shows a 20-year 'exceptionally unlikely' improvement in turbidity and an 'unlikely' improvement in NH4-N, however nutrient concentrations have improving trends, NNN is 'virtually certain' to have improved over 20 years and TP is 'extremely likely' to have improved.

4.3 Water quality summary Roxburgh Rohe

The tables in this section summarise:

1) River and lake sites where attributes where the national bottom line is not met (NPSFM, 2020)
2) Trends in river and lake sites when the trends are greater than 'likely' or 'unlikely'
3) All trends using raw data for rivers and continuous data for lakes over the two time-periods

Table 1 Summary of river state, red cells show where state does not meet the national bottom line in one or more variable.

sID	*NH4-N - max*	*NH4-N - median*	ASPM	DRP - median	DRP – Q95	E.coli	MCI	NNN - median	NNN – Q95	Periphyton	Periphyton (DRP)	Periphyton (TN)	Suspended fine sediment
Teviot at Bridge Huts													
Fraser at Old Mans													
Clutha at Millers Flat													
Benger burn at SH8													

Table 2 Summary of river sites where trends are greater than 'likely' or 'unlikely' (raw data). Confidence is expressed categorically based on the levels defined in Table 6.

npID	nObs	Freq	Period	AnnualSenSlope	DirectionConf	Descriptor
Clutha at Millers Flat						
Nitrite/Nitrate N	237	Month	20	-0.00034	Virtually certain	↑↑↑↑
Total Phosphorus	235	Month	20	-5.52417	Extremely likely	↑↑↑
Turbidity	237	Month	20	0.03662	Exceptionally unlikely	↓↓↓↓

Table 3 Summary of Lake Onslow trends when they are greater than 'likely' or 'unlikely'. Confidence is expressed categorically based on the levels defined in Table 6.

npID	nObs	Freq	Period	DirectionConf	AnnualSenSlope	Descriptor
Lake Onslow at Boat Ramp						
Ammoniacal N	49	BiMonth	10	Exceptionally unlikely	0	↓↓↓↓
Chlorophyll a	49	BiMonth	10	Exceptionally unlikely	0	↓↓↓↓
Nitrite/Nitrate N	49	BiMonth	10	Extremely likely	0	↑↑↑
Total Phosphorus	49	BiMonth	10	Extremely likely	-0.00096	↑↑↑
Turbidity	36	Qtr	10	Virtually certain	-0.36326	↑↑↑↑

Table 4 Overall summary of trends for Lake Onslow (continuous data). Confidence is expressed categorically based on the levels defined in Table 6..

	Virtually certain	Extremely likely	Very likely	Likely	As likely as not	Unlikely	Very unlikely	Extremely unlikely	Exceptionally unlikely
Descriptor	↑↑↑↑	↑↑↑	↑↑	↑	↔	↓	↓↓	↓↓↓	↓↓↓↓
River – 10 year trend				4	1				
River – 20 year trend	2	2	1						2
Lakes – 10 year trend									

The State analysis identified water quality in the Roxburgh Rohe rivers is generally good and the NPSFM band 'A' was achieved for most attributes. The only exceptions were for suspended fine sediment which was below the national bottom line in the Teviot and the Benger Burn. The suspended fine sediment in the Teviot is likely due to Lake Onslow, the main input to the river, as the lake is shallow and susceptible to sediment resuspension from wind-driven waves. *E.coli* was also below the national bottom line at these two sites.

In Table 2 and Table 3 only sites with 99%, 95% ,1% and 5% confidence levels are shown. These equate to the 'virtually certain', 'extremely likely', 'exceptionally unlikely' and extremely unlikely' categories. When sites have a zero sen slope alongside a reasonably high-level of confidence in trend direction the rate of the trend (i.e., the Sen slope) is at a level that is below the detection precision of the monitoring programme. In the Roxburgh Rohe, Lake Onslow had three parametes with a zero Sen slope; NH4-N, chlorophyll a and NNN. Lake Onslow, over 20 years had a 'virtually certain' improving trend for NNN and an 'extremely likely' improving trend for TP. In the same timeframe, turbidity showed an 'exceptionally unlikely' improving trend.

In summary:

- Water quality in the Clutha at Millers Flat and Fraser River generally achieve 'A' bands.
- The Teviot River does not meet the national bottom line for *E.coli* or suspended fine sediment.
- The Benger Burn does not meet the national bottom line for suspended fine sediment or *E.coli* and 'D' bands are achieved for periphyton DRP and periphyton TN
- The Clutha at Millers Flat has an 'exceptionally unlikely' improving trend for turbidity (over 20 years)

5 References

ANZECC, A., 2000. Australian and New Zealand Guidelines for Fresh and Marine Water Quality. Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra:1–103.

Burdon, F.J.; McIntosh, A.R.; Harding, J.S. Habitat loss drives threshold response of benthic invertebrate communities to deposited sediment in agricultural streams. Ecol. Appl. 2013, 23, 1036–1047

Clapcott J, Young R, Sinner J, Wilcox M, Storey R, Quinn J, Daughney C, Canning A, 2018. Freshwater biophysical ecosystem health framework. Prepared for Ministry for the Environment. Cawthron Report No. 3194. 89 p. plus appendices.

Choquette, A.F., R.M. Hirsch, J.C. Murphy, L.T. Johnson, and R.B. Confesor Jr, 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.

Fraser. C; Snelder, T. 2020. State of Lake and River Water Quality in the Otago Region. For records up to 30 June 2020

Fraser. C; Snelder, T. 2020b. ORC River, Groundwater and Lake water quality Trend analysis. For 10 and 20-year periods up to September 2020.

Franklin, P., D. Booker, and R. Stoffels, 2020. Contract 23184: Task 2 - Turbidity and Visual Clarity Threshold Conversion. NIWA.

Greenland, S., S.J. Senn, K.J. Rothman, J.B. Carlin, C. Poole, S.N. Goodman, and D.G. Altman, 2016. Statistical Tests, P Values, Confidence Intervals, and Power: A Guide to Misinterpretations. European Journal of Epidemiology 31:337–350.

Helsel, D.R., 2012. Reporting Limits. Statistics for Censored Environmental Data Using Minitab and R. John Wiley & Sons, pp. 22–36.

Helsel, D.R., R.M. Hirsch, K.R. Ryberg, S.A. Archfield, and E.J. Gilroy, 2020. Statistical Methods in Water Resources. Report, Reston, VA.

Hirsch, R.M., J.R. Slack, and R.A. Smith, 1982. Techniques of Trend Analysis for Monthly Water Quality Data. Water Resources Research 18:107–121.

Hickey, C., 2014. Derivation of Indicative Ammoniacal Nitrogen Guidelines for the National Objectives Framework. Memo prepared for Ms Vera Power, Ministry for the Environment, by NIWA.

Kienzle (S.W) and Schmidt (J). Hydrological impacts of irrigated agriculture in the Manuherekia catchment, Otago, New Zealand. Journal of Hydrology (NZ) 47 (2): 67-84 2008

Larned, S., A. Whitehead, C.E. Fraser, T. Snelder, and J. Yang, 2018. Water Quality State and Trends in New Zealand Rivers. Analyses of National-Scale Data Ending in 2017. prepared for Ministry for the Environment, NIWA.

Larned, S., T. Snelder, M. Unwin, G. McBride, P. Verburg, and H. McMillan, 2015. Analysis of Water Quality in New Zealand Lakes and Rivers. Prepared for the Ministry for the Environment. Wellington: Ministry for the Environment.

Larned, S., A. Whitehead, C.E. Fraser, T. Snelder, and J. Yang, 2018. Water Quality State and Trends in New Zealand Rivers. Analyses of National-Scale Data Ending in 2017. prepared for Ministry for the Environment, NIWA.

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

McBride, G.B., 2019. Has Water Quality Improved or Been Maintained? A Quantitative Assessment Procedure. Journal of Environmental Quality.

McBride, G., R.G. Cole, I. Westbrooke, and I. Jowett, 2014. Assessing Environmentally Significant Effects: A Better Strength-of-Evidence than a Single P Value? Environmental Monitoring and Assessment 186:2729–2740.

Ministry for the Environment. 2018. A Guide to Attributes in Appendix 2 of the National Policy Statement for Freshwater Management (as amended 2017). Wellington: Ministry for the Environment

Ministry for the Environment. 2020. Action for healthy waterways: Guidance on look-up tables for setting nutrient targets for periphyton. Wellington: Ministry for the Environment

Ministry for Environment and Ministry of Health, 2003. Microbiological Water Quality Guidelines for Marine and Freshwater Recreational Areas. Ministry for the Environment. https://www.mfe.govt.nz/sites/default/files/microbiological-quality-jun03.pdf.

NZ Government, 2020. National Policy Statement for Freshwater Management 2020.

Otago Regional Council, 2004. Regional Plan: Water for Otago

Otago Regional Council. (2004). Regional Plan: Water for Otago. Dunedin: Otago Regional Council.

Otago Regional Council, 2017. State of the environment: Surface water quality in Otago. 2006 to 2011. Otago Regional Council.

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

Snelder, T.H. and B.J.F. Biggs, 2002. Multi-Scale River Environment Classification for Water Resources Management. Journal of the American Water Resources Association 38:1225–1240.

Stark JD, Maxted JR, 2007. A user guide for the Macroinvertebrate Community Index. Prepared for the Ministry for the Environment. Cawthron Report No.1166. 58 p.

Stocker et al., 2014, IPCC Climate Change 2013: The Physical Science Basis - Findings and Lessons Learned

Whitehead, A.L., Booker, D.J. (2020). NZ River Maps: An interactive online tool for mapping predicted freshwater variables across New Zealand. NIWA, Christchurch. https://shiny.niwa.co.nz/nzrivermaps/
REC classes are CWM for M d/s Fork, CDM for Dunstan Creek and CDH for all other sites.

Wasserstein, R.L. and N.A. Lazar, 2016. The ASA Statement on P-Values: Context, Process, and Purpose. Taylor & Francis.

6 Appendix 1 – Water Quality State Analysis

ORC engaged Land Water People (LWP) to evaluate state at ORC's river and lake monitoring sites for nutrients and bacteria. This section details the methods LWP used for state analysis and is taken directly from Fraser (2020).

6.1.1 Grading of monitoring sites

The water quality state for river and lake monitoring sites is graded based on attributes and associated attribute state bands defined by the National Objectives Framework (NOF) of the NPSFM (2020) detailed in Table 5.

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

For each numeric attribute, the NOF defines categorical numeric attribute states as four (or five) attribute bands, which are designated A to D (or A to E, in the case of the *E. coli* attribute). The attribute bands represent a graduated range of support for environmental values from high (A band) to low (D or E band). The ranges for numeric attribute states that define each attribute band are defined in Appendix 2 of the NPSFM (2020). For most attributes, the D band represents a condition that is unacceptable (with the threshold between the C and the D band being referred to as "bottom line") in any waterbody nationally. In the case of the NO3N (toxicity) and NH4N (toxicity) attributes in the 2020 NPSFM, the C band is unacceptable, and for the DRP attribute, no bottom line is specified.

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

Table 5 River and lake water quality variables included in this report, including NPSFM reference and water body type

NPSFM Reference - NOF Attribute	Water body type	Minimum Sample Requirements	Numeric attribute state description	Units
A2A; Table 1 - Phytoplankton	Lakes		Median of phytoplankton chlorophyll-a	mg chl-a m^{-3}
			Annual maximum of phytoplankton chlorophyll-*a*	mg chl-a m^{-3}
A2A; Table 2 – Periphyton	Rivers	Minimum of 3 years of data	92nd percentile of periphyton chlorophyll-*a* for default river class[2]	mg chl-a m^{-3}
			83rd percentile of periphyton chlorophyll-*a* for productive river class[1]	mg chl-a m^{-3}
A2A; Table 3 – Total Nitrogen	Lakes		Median concentration of total nitrogen	mg m^{-3}
A2A; Table 4 – Total Phosphorus	Lakes		Median concentration of total phosphorus	mg m^{-3}
A2A; Table 5 - Ammonia	Lakes and Rivers		Median concentration of Ammoniacal-N	mg l^{-1}
			Maximum concentration of Ammoniacal-N	mg l^{-1}
A2A; Table 6 - Nitrate	Rivers		Median concentration of Nitrate	mg l^{-1}
			95th percentile concentration of Nitrate	mg l^{-1}
A2A.; Table 8 - Suspended fine sediment	Rivers	Median of 5 years of at least monthly samples (at least 60 samples)	Median visual clarity	m
A2A; Table 9 - *Escherichia coli*	Rivers and Lakes	Minimum of 60 samples over a maximum of 5 years	% exceedances over 260 cfu 100 mL^{-1}	%
			% exceedances over 540 cfu 100 mL^{-1}	%
			Median concentration of *E. coli*	cfu 100 ml^{-1}
			95th percentile concentration of *E. coli*	cfu 100 ml^{-1}
A2B; Table 14 - Macroinvertebrates	Rivers	State calculated as 5-year median	Median MCI score	-
A2B; Table 15 - Macroinvertebrates	Rivers		Median ASPM score	-
A2B; Table 20 - DRP	Rivers		Median concentration of DRP	mg l^{-1}
			95th percentile concentration of DRP	mg l^{-1}

A site can be graded for each attribute by assigning it to attribute bands (e.g., a site can be assigned to the A band for the NO3N toxicity attribute). A site grading is done by using the numeric attribute state (e.g., annual median nitrate-nitrogen) as a compliance statistic. The value of the compliance statistic for a site is calculated from a record of the relevant water quality variable (e.g., the median value is calculated from the observed monthly NO3N concentrations). The site's compliance statistic is then compared against the numeric ranges associated with each attribute band and a grade assigned for the site (e.g., an annual median NO3N concentration of 1.3 mg/l would be graded as 'B -band, because it lies in the range >1.0 to ≤2.4 mg/l). Note that for attributes with more than one numeric attribute state, a grade for each numeric attribute state has been provided (e.g., for the NO3N (toxicity) attribute, grades are defined for both the median and 95th percentile concentrations).

6.1.2 Time period for assessments

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

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

In this study, a period of 5 years represented a reasonable trade-off for most of the attributes because it yielded a sample size of 30 or more observations for many sites and attribute combinations. The five-year period for the state analyses is also consistent with national water-quality state analyses (e.g., Larned *et al.*, 2015, 2018), as well as guidance for a number of specific attributes within the NPSFM (2020) (4). Where no guidance was provided, a default filtering rule that required at least 30 observations in the 5-year time period was used. For annually sampled macroinvertebrate variables, which are generally less variable than physical or chemical water quality variables, the nominated minimum sample size requirement was reduced to 5.

For grading the suspended fine sediment and *E. coli* attributes, the NPSFM requires 60 observations over 5 years. For monthly monitoring, this requires collection of all monthly observations (i.e., no missing data). All ORC records have at least one missing observation associated with the national COVID-19 lockdown in April 2020, and so no sites met this requirement for the selected time periods. For this study, the rule to require observations for 90% of months over the 5-year period (54 observations) was relaxed. Both this relaxation and default sample number are subjective choices.

Therefore, within the supplementary files state assessments for all sites are provided regardless of whether they meet the filtering rules, as well as details about the number of observations and number of years with observations.

6.1.3 Calculation of water clarity

The NPSFM suspended fine sediment attribute is based on observations of visual clarity. ORC river monitoring programme does not include visual clarity but does routinely collect turbidity observations. Franklin et al. (2020) define a relationship between median clarity and median turbidity, based on a regression of 582 sites across New Zealand as:

$$\ln(CLAR) = 1.21 – 0.72 \ln(TURB)$$

where CLAR is site median visual clarity (m) and TURB is site median turbidity (NTU). In this study, median turbidity values over the 5-year time period was calculated first, and then calculated median clarity using the above relationship in order to grade the sites against the NPSFM suspended fine sediment attribute.

Sites operated by NIWA as part of the national monitoring network include observations of clarity, and therefore for these sites performance agains the NPSFM suspecnded fine sedinment attribute has been evaluated with the observed (rather than modelled) clarity values.

6.1.4 pH Adjustment of Ammonia

Ammonia is toxic to aquatic animals and is directly bioavailable. When in solution, ammonia occurs in two forms: the ammonium cation (NH_4^+) and unionised ammonia (NH_3); the relative proportions of the forms are strongly dependent on pH (and temperature). Unionised ammonia is significantly more toxic to fish than ammonium, hence the total ammonia toxicity increases with increasing pH (and/or temperature) (ANZECC, 2000). Standards related to ammoniacal-N concentrations in freshwater typically require a correction to account for pH and temperature. A pH correction to NH_4-N was applied to adjust values to equivalent pH 8 values, following the methodology outlined in Hickey (2014). For pH values outside the range of the correction relationship (pH 6-9), the maximum (pH<6) and minimum (pH>9) correction ratios were applied.

6.1.5 Evaluation of compliance statistics

For compliance statistics specified and “annual” (maximum, median, 95th percentile) in the NPSFM, have been calculated over the entire 5-year state period.

7 Appendix 2 – Water Quality Trend Analysis

ORC engaged Land Water People (LWP) to evaluate 10 and 20-year trends at ORC's river and lake monitoring sites for each measured variable (primarily nutrients and bacteria). This section details the methods LWP used for trend analysis and is taken directly from Fraser 2020b.

7.1.1 River water quality data

The river water quality data used in this analysis were supplied by ORC (110 sites) and NIWA (8 sites) and comprised 114,600 observations at 115 monitoring sites (3 sites overlapped between the ORC and NIWA data) of the variables at shown in Table 5.

7.1.2 Lake water quality data

The lake water quality data used in this study comprised 18,612 observations at 22 monitoring sites/depths of the 13 variables. Some sites had two depths associated with their water quality sampling. The different depths were treated as independent sampling sites.

The ORC lake monitoring programme underwent major changes over the period in 2016-2018. Several new sites were introduced, and older sites were phased out. Many of these older sites had long term records (starting in approximately 2000) but were ceased by mid-2018. Many of the water quality variables at the new sites were also monitored at these locations during an intensive investigation period between 2006-2009. These data were extracted from physical records for use in this study. The extracted data was not associated with censoring information. Observations were reinstated as censored values as part of the pre-processing based on the detection limits in operation for the same variables in other lakes over the same time period.

7.1.3 Flow data

Many of the river water quality monitoring sites were associated with flow records, which were also obtained from the ORC database. Flows associated with the NIWA sites were a combination of measured and modelled flows. Water quality observations can be strongly associated with flow, and the effect of flow on water quality can be accounted for in analysis of trends. Mean daily flows were associated with 51 of the 115 monitoring sites (and, of these sites, approximately 87% of all sample occasions had an associated flow).

7.1.4 Sampling dates, seasons and time periods for analyses

In trend assessments, there are several reasons why it is generally important to define the trend period and seasons and to assess whether the observations are adequately distributed over time. First, because variation in many water quality variables is associated with the time of the year or "season", the robustness of trend assessment is likely to be diminished if the observations are biased to certain times of the year. Second, a trend assessment will always represent a time period; essentially that defined by the first and last observations. The assessment's characterisation of the change in the observations over the time period is likely to be diminished if the observations are not reasonably evenly distributed across the time period. For these reasons, important steps in the data compilation process include specifying the seasons, the time period, and ensuring adequately distributed data.

Monitoring programs are generally designed to sample with a set frequency, (e.g., monthly, quarterly). The trend analysis 'season' is generally specified to match this sampling frequency (e.g., seasons are months, bi-months or quarters). There is therefore generally an observation for each sample interval (i.e., each season, such as month or quarter, within each year). Sampling frequency for some variables is annually. For example, annual sampling is common for biological sampling such as macro-invertebrates. In this case the 'season' is specified by the year.

Two common deviations from the prescribed sampling regime are (1) the collection of more than one observation in a sample interval (e.g., two observations within a month) and (2) a change in sampling interval within the time period. Both of these deviations occurred in the ORC datasets, particularly type (2), as there was a network wide change in sampling frequency in 2013, largely moving from bi-monthly to monthly monitoring for rivers, and from biannual to quarterly for groundwater in 2011. For type (1) deviations, the median within each sample interval was taken. For type (2) deviations, the coarser sampling interval to define seasons was used. For the part of the record with a higher frequency, the observations in each season were defined by taking the observation closest to the midpoint of the coarser season. The reason for not using the median value in this case is that it will induce a trend in variance, which will invalidate the null distribution of the test statistic (Helsel *et al.*, 2020).

The trend at all sites was characterised by the rate of change of the central tendency of the observations of each variable through time. Because water quality is constantly varying through time, the evaluated rate of change depends on the time-period over which it is assessed (e.g., Ballantine *et al.*, 2010; Larned *et al.*, 2016). Therefore, trend assessments are specific for a given period of analysis. Trend periods of 10 and 20 years were evaluated.

For a regional study that aims to allow robust comparison of trends between sites and to provide a synoptic assessment of trends across a whole region, such as the present study, it is important that trends are commensurate in terms of their statistical power and representativeness of the time period. In these types of studies, it is general practice to define consistent time periods (i.e., trend duration and start date) so that all sites are subjected to the same conditions (i.e., equivalent political, climate, economic conditions). It is also general practice to define the acceptable proportion of gaps and how these are distributed across sample intervals so that the reported trends are assessed from comparable data. The acceptable proportion of gaps and representation of sample intervals by observations within the time period are commonly referred to as site inclusion or filtering rules (e.g., Larned *et al.*, 2018) but this is also termed 'site screening criteria' and 'completeness criteria'.

There are no specific data requirements or filtering rules for trend assessments performed over many sites and variables such as the present study. The definition of filtering rules is complicated by a trade-off: more restrictive rules increase the robustness of the individual trend analyses but will generally exclude a larger number of sites thereby reducing spatial coverage. In general, this trade-off is also affected by the duration of trend period. Steadily increasing monitoring effort in New Zealand over the last two decades means that shorter and more recent trend periods will generally have a larger number of eligible sites.

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

In this study, filtering rules applied by Larned et al. (2019) were used, which set X and Y to 80%. Further, the definition of seasons was flexible in order to maximise the number of sites that were included. If the site failed to comply with filter rule (2) when seasons were set as months, a coarsening of the data to quarterly seasons was applied and the filter rule (2) was reassessed. If the data then complied with filter rule (2), the trend results based on the coarser (i.e., quarterly) seasons were retained for reporting. Bi-months were also included as an intermediate coarseness between months and quarters, as this sampling interval was historically used.

Using these filter rules, the number of site/variable combinations that would be included in the analysis under varying trend period end dates was explored. While the intention was to provide the most recent possible trend assessments (up to the end of the observations dataset, August 2020), the possibility of having an earlier end date was also considered, if that would significantly increase the number of sites that would comply with the filtering rules. End dates were considered at the end of months from December 2019 through to August 2020. The results of this analysis are not included in this report as generally, there was little variation in the number of sites that complied with the filtering rules for end dates between February 2020 and August 2020. The exception was for the macroinvertebrate metrics, which had a large reduction in the number of sites that complied with the filtering rules from the December 2019 cut-off point to all end dates in 2020 (generally a reduction from 26 to 13 sites). This arises due the cessation of several macroinvertebrate sites in 2018. In the interest of providing the most up to date trend assessments, the trends for rivers presented in this study were for 10- and 20-year periods ending at 31 August 2020.

A slightly different approach has been applied to the lake monitoring data in order to maximise the assessment of trends for these sites due the irregularity of the monitoring and changes in monitoring sites. The most recent end date to examine long term, fixed period, trends across all sites was identified. This date coincided with the termination of monitoring at a number of long-term sites at the end of June 2018. We evaluated trends for 10- and 18-year periods up to the end of June 2018. The 18-year period was selected as there were no lake data available prior to 2000. For these fixed period trend assessments, the data were subjected to the same filtering rules as used for the river and groundwater sites.

Another deviation for the trend analysis at lake sites was for a group of sites that were monitored for a period between 2006-2009 after which there was no monitoring until the program was re-established in 2018. These sites have been analysed using alternative trend assessment procedures that evaluate the change between the two time periods (see Section 7.1.11). However, it was important that the data still complied with the time period requirements relating to representativeness of the time periods, and that there was no bias toward any particular season in the records. Consequently, the two analysis time periods for these site/variable combinations to be three complete years: 1 May 2006 to 30 April 2009, and 1 June 2017 to 31 May 2020 were set. It was also required that at least 80% of observations were available in each time period.

7.1.5 Handling censored values

For several water-quality variables, true values are occasionally too low or too high to be measured with precision. These measurements are called censored values. The “detection limit” is the lowest value that can be measured by an analytical method (either a laboratory measurement or a measurement made in the field) and the “reporting limit” is the greatest value of a variable that can be measured. Water-quality datasets from New Zealand rivers and lakes often include DRP, TP and NH4N measurements that are censored because they are below detection limits, and ECOLI and CLAR measurements that are censored because they are above reporting limits.

Censored values are managed in a special way by the non-parametric trend assessment methods described in section 7.1.8. It is therefore important that censored values are correctly identified in the data. Detection limits or reporting limits that have changed through the trend time period (often due to analytical changes) can induce trends that are associated with the changing precision of the measurements rather than actual changes in the variable. This possibility needs to be accounted for in the trend analysis and this is another reason that it is important that censored values are correctly identified in the data.

A “hi-censor” filter was applied in the trend assessments to minimise biases that might be introduced due to changes in detection limits through the trend assessment period. The hi-censor filter identifies the highest detection limit for each water quality variable in the trend assessment period and replaces all observations below this level with the highest detection limit and identifies these as censored values. This procedure generally had limited impact on the trend assessment, with the exception of Ammoniacal Nitrogen, as there was a significant shift in the detection limit, and most of the observations were generally very small (of similar magnitude to the detection limit).

7.1.6 Flow adjustment

Where water quality observations are made in a river and are associated with a solute or particulate matter (e.g., a concentration or an optical measure such as clarity or turbidity) some of the variation can be associated with the river flow (i.e., discharge) at the time the observation was made. The observed values can vary systematically with flow rate due to two kinds of physical processes. The water quality observations may decrease systematically with increasing flow due to the effect of dilution of the contaminant, or increase with increasing flow due to wash-off of the contaminant (Smith *et al.*, 1996). Different mechanisms may dominate at different sites so that the same water quality variable can exhibit positive or negative relationships with flow. Some water quality variables can be associated with a combination of dilution and wash off with increasing flow. For example, a portion of the *E. coli* load may come from point sources discharges such as sewage treatment plants (dilution effect), but another portion may be derived from surface wash-off. Increasing flow in this situation may result is an initial dilution at the low end of the discharge range, followed by an increase with discharge at higher values of discharge.

Trend analysis seeks to quantify the relationship between the water quality observations and time. In this context, flow can be considered as a “covariate”; a variable that is also related to the water quality observations but whose influence is confounding the water quality – time relationship of interest. Statistical analysis can be used to remove the influence of the covariate on the water quality observations. For river data, this statistical analysis is called “flow adjustment”. The same principle can be applied to other types of environments (e.g., lakes, groundwater) and other covariates (e.g., wind, precipitation) and so the more general term is covariate adjustment.

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

Covariate adjustment involves fitting a model that describes the relationship between the water quality observation and the covariate, and then using the residuals of the model instead of the original water quality observations in the subsequent trend assessment step. In the description of the covariate adjustment method below, flow adjustment was the focus (i.e., removing the influence of flow at from water quality observations made in a river). However, in principle, the method is the same for any other type of covariate adjustment.

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

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

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

7.1.7 Seasonality assessment

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

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

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

7.1.8 Analysis of trends

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

Trends were evaluated using the LWPTrends functions in the R statistical computing software. A brief description of the theoretical basis for these functions is described below.

7.1.9 Trend direction assessment

The trend direction and the confidence in the trend direction were evaluated using either the Mann Kendall assessment or the Seasonal Kendall assessment. Although the non-parametric Sen slope regression also provides information about trend direction and its confidence, the Mann Kendall assessment is recommended, rather than Sen slope regression, because the former more robustly handles censored values.

The Mann Kendall assessment requires no *a priori* assumptions about the distribution of the data but does require that the observations are randomly sampled and independent (no serial correlation) and that there is a sample size of ≥ 8. Both the Mann Kendall and Seasonal Kendall assessments are based on calculating the Kendall *S* statistic, which is explained diagrammatically in Figure 6.

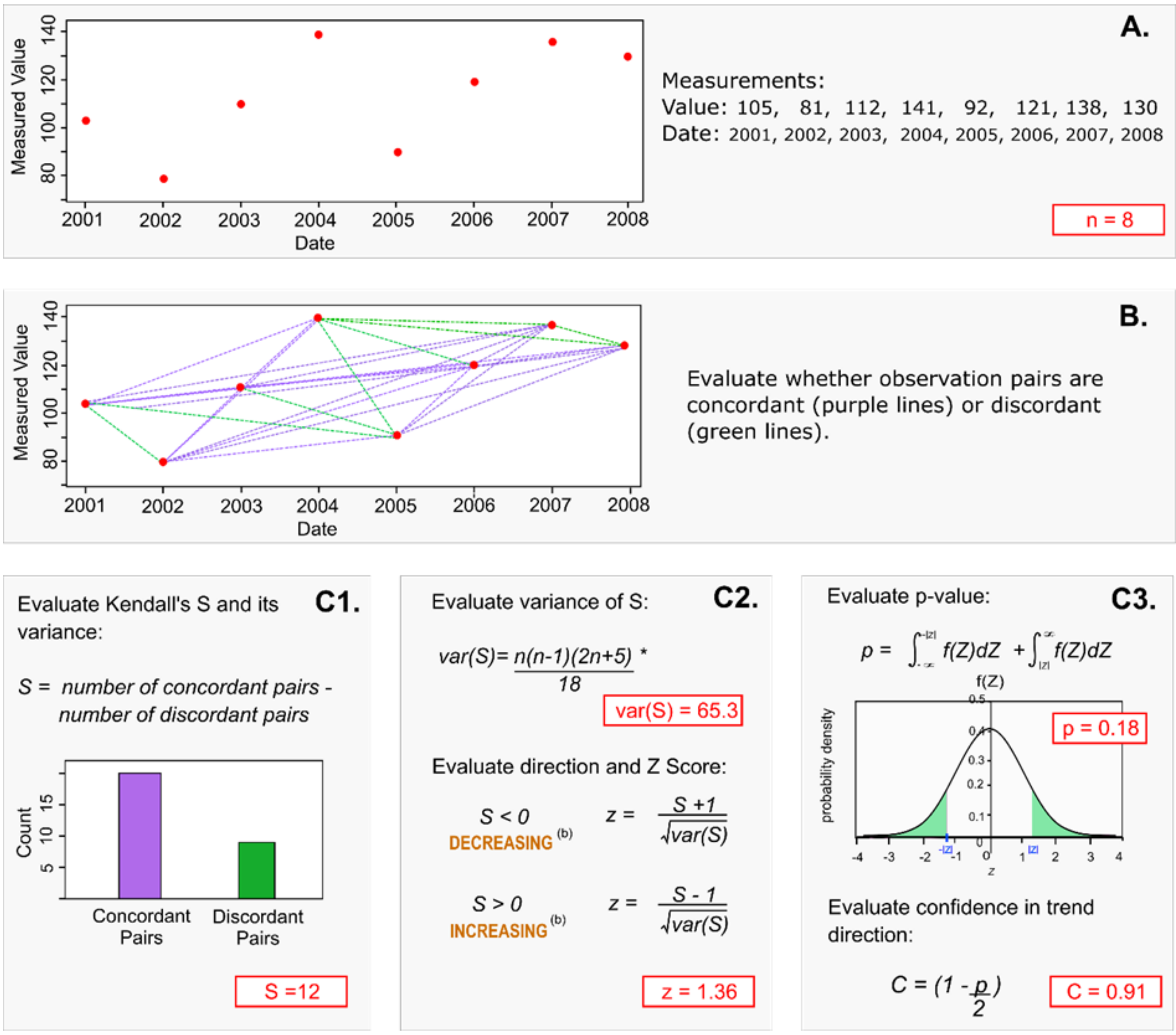

Figure 6 The Kendall S statistic is calculated by first evaluating the difference between all pairs of water quality observations (Figure 6, A and B). Positive differences are termed 'concordant' (i.e., the observations increased with increasing time) and negative differences are termed discordant (i.e., the observations decreased with increasing time). The Kendall S statistic is the number of concordant pairs minus the number of discordant pairs (Figure 6, C1). The sign of S

indicates the water quality trend direction with a positive or negative sign indicating that observations increased or decreased through time respectively (Figure 6, C2). In the special case that the z score is equal to zero, the trend would be pronounced "indeterminate", or equally likely to be increasing as decreasing.

7.1.10 Assessment of trend rate

The method used to assess trend rate is based on non-parametric Sen slope regressions of water quality observations against time. The Sen slope estimator (SSE; Hirsch *et al.,* 1982) is the slope parameter of a non-parametric regression. SSE is calculated as the median of all possible inter-observation slopes (i.e., the difference in the measured observations divided by the time between sample dates; Figure 7).

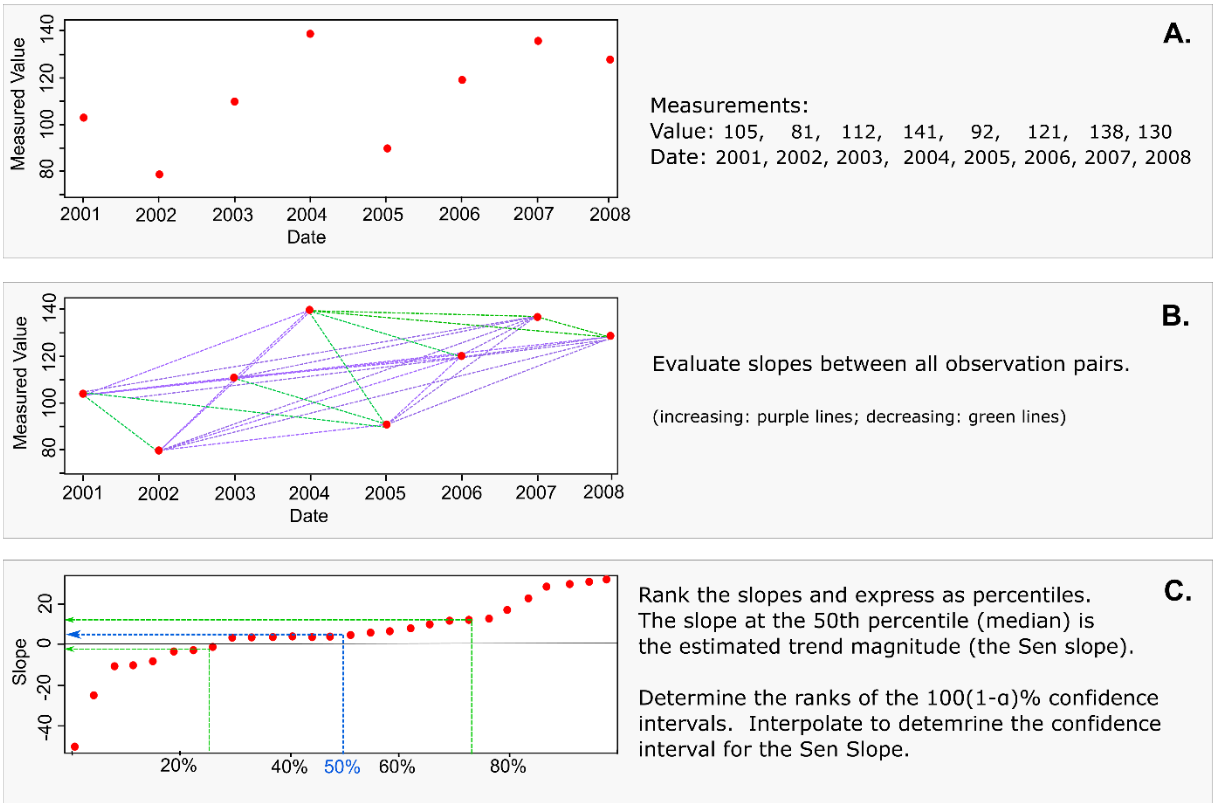

Figure 7 Pictogram of the calculation of the Sen slope, which is used to characterise trend rate.

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

7.1.11 Evaluating changes in discontinuous data

Some of the monitoring data for lake sites is broken into two distinct time periods, with a moderate gap (~ 4 years) between these periods. Following the USGS guidelines (Helsel et al. 2020), these types of datasets have been analysed using a step change approach. The analysis procedure uses a rank-sum test (and seasonal variant where appropriate) to test whether there is a change in the observations between the two periods, and the Hodges-Lehman (H-L) estimator to evaluate the magnitude, and direction of the change.

The H-L estimator is evaluated in a similar manner as the Sen Slope, with the exception that rather than evaluating the rate of change between all pairs of observations, only the differences are evaluated, and only between pairs from different periods. The H-L estimator is the median of all possible differences between the data in the before and after periods. A seasonal H-L estimator is evaluated when the observations are determined to be seasonal.

We also provide an estimate of the rate of change that the difference represents, by dividing the H-L estimator by the difference between the mid times of each time period. This measure is indicative only and should only be used as an approximation of the relative magnitude of the rate of change at these sites.

7.1.12 Interpretation of trends

The trend assessment procedure used here facilitates a more nuanced inference than the 'yes/no' output corresponding to the chosen acceptable misclassification error rate. The confidence in direction (C) can be transformed into a continuous scale of confidence the trend was decreasing (C_d). For all trends with $S < 0$, $C_d = C$, and for all $S > 0$ a transformation is applied so that $C_d = 1\text{-}C$. C_d ranges from 0 to 1.0. When C_d is very small, a decreasing trend is highly unlikely, which because the outcomes are binary, is the same as an increasing trend is highly likely.

The trend for each site/variable combination was assigned a categorical level of confidence that the trend was improving according to its evaluated confidence, direction and the categories shown in Table 6. Improvement is indicated by decreasing trends for all the water quality variables in this study except for MCI, SQMCI, ASPM and dissolved oxygen (for which increasing trends indicate improvement).

Table 6 Level of confidence categories used to convey the confidence that the trend (or step change) indicated improving water quality. The confidence categories are used by the Intergovernmental Panel on Climate Change (IPCC; Stocker et al., 2014).

Categorical level of confidence trend was decreasing	*Descriptor used in report*	*Value of C_d (%)*
Virtually certain	↑↑↑↑	0.99–1.00
Extremely likely	↑↑↑	0.95–0.99
Very likely	↑↑	0.90–0.95
Likely	↑	0.67–0.90
About as likely as not	↔	0.33–0.67
Unlikely	↓	0.10–0.33
Very unlikely	↓↓	0.05–0.10
Extremely unlikely	↓↓↓	0.01–0.05
Exceptionally unlikely	↓↓↓↓	0.0–0.01

Outputs from the trend analyses were also classified into four direction categories: improving, degrading, indeterminate, and not analysed. An increasing or decreasing trend category was assigned based on the sign of the S statistic from the Mann Kendall test. An indeterminate trend category was assigned when the Z score equalled zeros. Trends were classified as "not analysed" for two reasons:

1) When a large proportion of the values were censored (data has <5 non-censored values and/or <3 unique non-censored values). This arises because trend analysis is based on examining differences in the value of the variable under consideration between all pairs of sample occasions. When a value is censored, it cannot be compared with any other value and the comparison is treated as a "tie" (i.e., there is no change in the variable between the two sample occasions). When there are many ties there is little information content in the data and a meaningful statistic cannot be calculated.

2) When there is no, or very little, variation in the data because this also results in ties. This can occur because laboratory analysis of some variables has low precision (i.e., values have few or no significant figures). In this case, many samples have the same value, and this then results in ties.

Changes for discontinuous data were classified as "not analysed" when there were less than 3 unique observations in the entire record, or if seasonal, within any season.

7.1.13 River data availability

Following the application of the filtering rules, the total number of sites that were included in the analyses was reduced, a summary of the site numbers that were included in the final trend assessment is presented in Table 7. Confidence that the trend direction indicated improving water quality, was mapped for the raw (with high censor filter) for the 10 and 20 year trend periods.

Table 7 River water quality variables, measurement units and site numbers for which 10- and 20-year trends (Raw, and Flow Adjusted FA) were analysed by this study.

Variable	Number of sites	Number of sites that complied with filtering rules (10-years)		Number of sites that complied with filtering rules (20-years)	
		Raw	FA	Raw	FA
Ammoniacal Nitrogen	114	50	32	34	18
ASPM	51	10	6	0	0
Chlorophyll a	44	0	0	0	0
Dissolved Inorganic Nitrogen	108	0	0	0	0
Dissolved Reactive Phosphorus	108	50	32	33	18
E. coli	114	50	27	28	13
MCI	54	13	7	0	0
Nitrite/Nitrate Nitrogen	114	50	32	34	18
SQMCI Score	53	13	7	0	0
Total Nitrogen	114	50	32	33	18
Total Phosphorus	114	50	32	32	18
Turbidity	114	50	32	32	18