Environmental Challenges 9 (2022) 100601

Contents lists available at ScienceDirect

Environmental Challenges

journal homepage: www.elsevier.com/locate/envc

Continuous measures of confidence in direction of environmental trends at site and other spatial scales

T, H Snelder[a,*], C Fraser[a], A.L. Whitehead[b]

[a] LWP Ltd, Unit 13, 212 Antigua Street, Christchurch, Select State 8023, New Zealand
[b] NIWA, 10 Kyle Street, Riccarton, Christchurch 8011, New Zealand

ARTICLE INFO

Keywords:
Environmental monitoring
Trend analysis
Spatial correlation
Water quality

ABSTRACT

Managers and decision makers need to know if variables measured by environmental monitoring programs are increasing or decreasing, both at individual sites and at larger spatial scales, and the degree of statistical support for these assessments. Traditionally, null hypothesis significance testing (NHST) has been used to evaluate whether an assessed trend is a reliable estimate of the true (i.e., population) trend but has two shortcomings. First, failure to achieve "statistical significance" is often falsely interpreted as evidence that there was no trend. Second, the acceptable Type 1 error risk tends to be chosen arbitrarily and without consideration of the risk of failing to identify important trends. As an alternative to NHST, we propose a continuous measure of confidence in the direction of an individual site trend based on the posterior probability distribution. Confidence that the trend direction is correctly inferred (i.e., that the assessed trend direction has the same sign as the population value) is expressed as a probability. The approach is extended to assessing confidence in the direction of aggregate trends (i.e., trends observed over multiple sites representing a spatial domain such as a geographic region). The aggregate trend assessment accounts for the confidence in the individual site trends and spatial correlation in the observations, which reduces the effective size of the dataset. The approach is demonstrated for site and aggregate river water quality trends for 352 sites in New Zealand. Compact graphical reporting of the results indicated appreciable variation in trend direction between sites for all variables, as well as patterns in trend direction at larger spatial scales. The new method provides decision makers with a more complete description of the statistical support for the assessment of trend direction than an arbitrary "significant/not significant" designation associated with NHST.

1. Introduction

Environmental management is generally associated with regular monitoring (e.g., weekly, monthly, annually) and reporting of characteristic indicators of environmental quality such as air quality (Sicard et al. 2021), water quality (Behmel et al. 2016) and biological measures (Cairns and Pratt 1993). Typically, monitoring data is subject to trend analysis to assess how the measures have changed at a site over time (Davies-Colley et al. 2011; Oelsner et al. 2017). Generally, managers and decision makers are interested in the trend direction, strength (i.e., consistency of increases or decreases in successive observations) and rate (change in the observed quantity per unit time) at both individual sites and at larger spatial scales. Assessment of trends at scales larger than individual sites is also relevant because environmental changes are often driven by pressures that act over broad areas, such as diffuse source emissions of contaminants or climatic changes (Isbell et al. 2017), and management actions are generally implemented over similarly broad areas, thereby potentially influencing many monitoring sites simultaneously (Cash et al. 2006; Cumming et al. 2006).

The most commonly used approaches to assessing the direction and strength of a temporal trend in an environmental variable and the trend's rate at the site-scale are Kendall's tau (τ, Mann, 1945; Hirsch and Slack, 1984) and Sen-Theil regression (Theil 1950; Hirsch et al. 1982), respectively. These methods quantify monotone trends (i.e., increasing or decreasing) and are used because they are robust to non-normal data, missing values, and censored values, which are common in environmental monitoring data (Hirsch and Slack 1984; Helsel et al. 2020).

Because trend analysis is based on building a statistical model from a limited number of observations (i.e., the sample), quantifications of direction and strength are accompanied by assessments of their reliability as estimates of the real (i.e., population) trend. The traditional approach to assessing the reliability of the estimated τ is the Mann–Kendall test, which tests the null hypothesis that $\tau = 0$. However, null hypothesis significance testing (NHST) has been subject to considerable criticism (e.g., Rozeboom 1960; Cohen 1994; Wasserstein and Lazar 2016)

* Corresponding author.
E-mail address: ton@lwp.nz (T.H. Snelder).

https://doi.org/10.1016/j.envc.2022.100601
Received 9 July 2022; Received in revised form 14 August 2022; Accepted 15 August 2022

and its use in trend assessment has been challenged for two key reasons (Vogel et al. 2013; Hirsch et al. 2015). First, the failure to achieve statistical significance is often falsely interpreted as evidence that there was no trend. This is an incorrect conclusion; a 'large' p-value (i.e., $p > 0.05$) indicates only that the data are not unusual if the null hypothesis were true, and none of the other assumptions were violated (Greenland et al. 2016). The null hypothesis is also "unrealistic and misleading" because there will always be a trend, which "may be trivially small but will always be positive or negative" (Cohen 1994; Jones and Tukey 2000). The problem with testing the null hypothesis that the trend is zero is that it suggests that the null hypothesis might be true and therefore encourages the incorrect interpretation of a non-significant result as indicating no trend.

The second reason for challenging the use of NHST in trend analysis is that it involves the use of an arbitrary value alpha (α) to designate a trend as significant. Generally, α is set at a low value (e.g., 0.05) to minimize the risk of incorrectly rejecting the null hypothesis (i.e., making Type I error). However, from a management perspective, the acceptable Type 1 error risk should not be defined by an arbitrary statistical rule. The acceptable risk of a Type 1 error is a normative decision that needs to be balanced against the risk of Type 2 errors (failure to recognize an important trend that should have provided the basis for acting; Vogel *et al.* 2013).

Alternatives to NHST for making inferences about trend direction have been suggested. McBride (2019) proposed a method to assess the risk of misclassifying trend direction that is based on calculating Bayesian credible intervals from the posterior probability distribution. The approach does not require postulating an unrealistic null hypothesis and assumes that there is always a trend no matter how small. McBride (2019) demonstrated this approach based on estimating the slope of the Sen-Theil regression line. McBride (2019) calculated the risk of misclassification of the trend direction, using the $100(1-2\alpha)$ credible interval. If this interval does not contain zero, the trend direction misclassification risk is $< \alpha$. The use of the $100(1-2\alpha)$ credible interval to control the error risk to $< \alpha$ arises because a trend can only be in one direction. Hirsch *et al.* (2015) and Murphy (2020) also used the posterior distribution to estimate the likelihood that the assessed trend direction was correct, however the underlying statistical model in this case was the Weighted Regressions on Time, Discharge, and Season method. Choquette *et al.* (2019) used the Mann–Kendall test p-value to define a continuous measure of likelihood that the trend was positive or negative. They estimated the likelihood that the assessed direction of the trend was correct (i.e., was the same as the true trend) from the Mann–Kendall test p-value as $[1 - p\text{-value}/2]$.

There is often interest in making statements about the direction of a trend in an environmental variable over scales larger than that represented by individual sites (Helsel and Frans 2006). Accordingly, results of trend assessments for multiple individual sites are often aggregated in tables or graphs. For example, Larned *et al.* (2016) presented tabulations of numbers of degrading and improving 10-year trends in a variety of water quality variables for sites across New Zealand grouped by environmentally defined classes and presented box plots showing the distributions of the Theil-Sen slopes. Similarly, Monteith *et al.* (2014) plotted and tabulated rates of change of anions and cations at 22 lakes and streams of the UK Acid Waters Monitoring Network. This method of presenting results loses some of the information that is available because it does not account for confidence in the assessed individual site trends. An alternative approach for quantifying aggregate trends is the regional Kendall test (Helsel and Frans 2006).

Both the simple tabulation of site-scale trend assessments and the regional Kendall test have the limitation that they do not account for spatial correlation that may exist among sites within a monitoring network. Spatial correlation implies a degree of redundancy in the available data because a trend direction at a site is likely to be consistent with nearby sites. From a statistical perspective, this redundancy reduces the effective size of the dataset and results in an overly liberal assessment of confidence in the aggregate trend (Douglas et al. 2000; Yue and Wang 2002).

In this article we show how the strength of evidence in assessments of trend direction made using the Mann–Kendall statistic can be quantified by a continuous measure of confidence in trend direction. Our method is novel in that it does not need to refer to the p-value and does not require the analyst to make decisions concerning acceptable error risk. We also describe a novel method for assessing trend direction and strength at scales larger than individual sites that is based on aggregation of individual site trend assessments. Confidence in the aggregate trend direction is assessed analogously to our method for assessment of confidence in the direction of individual site trends, and accounts for the confidence associated with each individual site trend and spatial autocorrelation between sites. To demonstrate our methods for trend direction assessment, we analyse 20-year river water quality trends of six water quality variables at 352 monitoring sites across New Zealand. We quantify confidence in trend direction at the scale of sites, the whole country and selected regions and suggest some compact graphical methods to present these assessments.

2. Material and methods

2.1. Assessment of confidence in trend direction for individual sites

Kendall's tau (τ) evaluates the direction and strength of a monotonic trend in a variable X, representing observations at a site of an environmental characteristic (e.g., a water quality variable measured each month in a river, Helsel et al. 2020). To calculate τ for data that is systematically collected with a set frequency, the observations of a variable X are first ordered by increasing time and then each observation is compared to all its subsequent observations. From these comparisons, the Mann–Kendall statistic ($\hat{S}$) is calculated as:

$$\hat{S} = \sum_{i=1}^{N-1} \sum_{j=1+i}^{N} sign\left(x_i - x_j\right) \tag{1}$$

where N is the number of observations, x_i and x_j are the observations at times i and j and $sign()$ is equal to +1 if x_i is greater than x_j and -1 if x_i is less than x_j. The statistic $\hat{S}$ can be understood as the number of concordant pairs minus the number of discordant pairs so that positive values indicate a tendency for values to increase over time, an increasing trend, and vice versa. We include the hat operator when denoting $\hat{S}$ to indicate that is an estimate, based on the observations, of the population value of S. Finally, $\hat{S}$ is standardised by the number of pairs of compared observations so that τ is a standardised measure of the strength of the monotonic trend in X, which varies between -1 and +1:

$$\tau = \frac{\hat{S}}{N(N-1)/2} \tag{2}$$

and:

$$D^{\tau} = \begin{cases} Decreasing\ if\ \tau < 0 \\ Increasing\ if\ \tau > 0 \end{cases} \tag{3}$$

where D^{τ} is the assessed site trend direction.

Traditionally, the strength of the inference associated with τ is based on the Mann–Kendall test, which is a test of the null hypothesis (H_0) that $S = 0$ (i.e., there has been no change in the central tendency of the observations). A trend is declared to be statistically significant if that hypothesis is rejected, in which case the p-value associated with the hypothesis is less than the prescribed significance level (α). The p-value is calculated by first noting that, when $N > 10$, S is asymptotically normal (Mann 1945; Kendall 1975). In the absence of tied values (i.e., pairs for which $x_j = x_i$), the variance of S is:

$$Var(S) = (N/18)(N-1)(2N+5) \tag{4}$$

(Hirsch et al. 1982). The p-value is determined by transforming the observed value of $\hat{S}$ to its equivalent standard normal deviate (Z score)

as follows:

$$Z_S = \begin{cases} \frac{\hat{S}-1}{\sqrt{Var(S)}} \; if \; \hat{S} > 0 \\ 0 \; if \; \hat{S} = 0 \\ \frac{\hat{S}+1}{\sqrt{Var(S)}} \; if \; \hat{S} < 0 \end{cases} \tag{5}$$

The resulting Z_S is evaluated against the standard normal cumulative distribution function and H_0 is rejected if $|Z_S| > Z_{\alpha/2}$, where α is the confidence level, and $Z_{\alpha/2}$, is the value needed to generate an area of $\alpha/2$ in each tail of the normal distribution. When observations include ties and censored values there are adjustments to the calculation of $\hat{S}$ and $Var(S)$ (Helsel 2011). When the observations are seasonal, seasonal versions of τ and $\hat{S}$ are calculated in two steps. First, for each season, the calculations of $\hat{S}$ and $Var(S)$ are made from data pertaining to observations in that season. Second, the seasonal values of $\hat{S}$ and $Var(S)$ are summed over all seasons and are used to calculate τ, the seasonal Kendall test statistic and its variance (Hirsch et al. 1982).

We propose calculating a continuous measure of confidence in the assessed trend direction based on the posterior probability distribution of S, the true (i.e., population) difference in concordant and discordant pairs. The posterior probability distribution of S is given by a normal distribution with mean of $\hat{S}$ and variance of $Var(S)$ (as described in Eq. (4)). From this information, confidence in assessed trend direction can be evaluated as the proportion of the probability distribution that has the same sign as $\hat{S}$. A graphical demonstration of the procedure is shown in Fig. 1, where the posterior probability distribution for $\hat{S}$ is shown by the blue line.

In Fig. 1(a), the trend is assessed to be positive because $\hat{S}$ lies to the right of the zero (red) line. The shaded area to the right of zero represents the probability that the true (i.e., population) trend has the direction indicated by $\hat{S}$, which represents the confidence in the assessed direction. In Fig. 1(b), the trend is assessed to be negative because $\hat{S}$ lies to the left of the zero (red) line. The shaded area to the left of zero represents the confidence in the assessed direction.

The probability the assessed site trend direction is the same as the true (i.e., population) trend is described mathematically as:

$$C^\tau = \begin{cases} \int_0^\infty N\left(\hat{S}, Var(S)\right) \; if \; \hat{S} > 0 \\ 0.5 \; if \; \hat{S} = 0 \\ \int_{-\infty}^0 N\left(\hat{S}, Var(S)\right) \; if \; \hat{S} < 0 \end{cases} \tag{6}$$

Where we refer to C^τ as the confidence in the assessed trend direction, and N represents the normal distribution function. In practice the integrals described in Eq. (6) can be calculated by first transforming the value of $S = 0$ on the posterior probability distribution into a standard normal deviate Z as follows:

$$Z_0 = \begin{cases} \frac{\hat{S}-1}{\sqrt{Var(S)}} \; if \; \hat{S} > 0 \\ 0 \; if \; \hat{S} = 0 \\ \frac{-(\hat{S}+1)}{\sqrt{Var(S)}} \; if \; \hat{S} < 0 \end{cases} \tag{7}$$

C^τ is then calculated as area under the standard normal distribution to the left of Z_0 using the quantile function for the normal distribution. The value C^τ can be interpreted as the probability that D^τ (i.e., the sign of the calculated τ statistic) indicates the same direction as the true trend (i.e., that the assessed trend direction is correct). The value C^τ ranges between 0.5, indicating the true trend direction is equally likely to be in the opposite direction to that indicated by D^τ, to 1, indicating complete confidence that D^τ is the same as the true trend.

2.2. Aggregate trend direction and strength assessment

For a domain of interest with several monitoring sites, we define a standardized measure of aggregate trend strength to be the proportion of site trends that are in the modal direction. We denote this statistic as capital tau with a hat operator to signify that it is an estimate of the population value ($\hat{\mathrm{T}}$). We also define the aggregate trend direction to be the modal (i.e., most frequently occurring) direction of the individual site trends (D^T).

The statistic $\hat{\mathrm{T}}$, and confidence in the aggregate trend direction (C^T), are calculated by letting the sites within the domain be indexed by m, so that $m \in \{1, \ldots, M\}$. The aggregate trend direction is calculated as:

$$D^\mathrm{T} = sign\left(\sum_{\mathrm{m}=1}^{\mathrm{M}} sign\left(\hat{S}_m\right)\right) \tag{8}$$

Let I be a random Bernoulli distributed variable for which the value 1 indicates the assessed site trend direction is the same as the aggregate trend direction (D^T) with probability p given by:

$$p = \begin{cases} C^\tau \; if \; sign\left(\hat{S}\right) = \mathrm{D}^\mathrm{T} \\ 1 - C^\tau \; if \; sign\left(\hat{S}\right) \neq \mathrm{D}^\mathrm{T} \end{cases} \tag{9}$$

and for which the value 0 indicates the direction of a site trend is opposite to the modal direction with probability $1 - p$. Then, the estimated proportion of sites with trends in the modal direction is:

$$\hat{\mathrm{T}} = \frac{1}{M}\sum_{m=1}^{M} I_m \tag{10}$$

where $\hat{\mathrm{T}}$ can vary between 0.5 and 1. Given the variance of a random Bernoulli distributed variable is $Var(I) = p(1-p)$, and assuming the site trends are independent, the variance of T is:

$$Var(\mathrm{T}) = \frac{1}{M^2}\sum_{m=1}^{M} Var\left(I_m\right) = \frac{1}{M^2}\sum_{m=1}^{M} p_m\left(1 - p_m\right) \tag{11}$$

A continuous measure of confidence in the aggregate trend direction (D^T) is found analogously to the confidence in the direction of a site trend by calculating the proportion of the posterior probability distribution for which $\mathrm{T} > 0.5$.

$$C^\mathrm{T} = \int_{0.5}^{\infty} N\left(\hat{\mathrm{T}}, Var(\mathrm{T})\right) \tag{12}$$

where C^T is the confidence in the aggregate trend direction. In practice, C^T can be calculated by first transforming the value of $\mathrm{T} = 0.5$ on the posterior probability distribution into a standard normal deviate as follows:

$$Z_{0.5} = \frac{\hat{\mathrm{T}} - 0.5}{\sqrt{Var(\mathrm{T})}} \tag{13}$$

C^T is then calculated as area under the standard normal distribution to the left of $Z_{0.5}$ using the quantile function for the normal distribution.

If there is spatial correlation in the assessed trend directions at the individual sites, the assumption of independence when calculating the variance of T above is violated (Douglas *et al.* 2000). Spatial correlation means that the effective sample size of the dataset is less than the number of sites and this results in under-estimation of the variance and therefore over-estimation of C^T. The method of Douglas *et al.* (2000) can be used to calculate the variance of T that is "corrected" for spatial correlation:

$$Var(\mathrm{T}) = \frac{1}{M^2}\left[\sum_{k=1}^{M} Var\left(I_k\right) + 2\sum_{k=1}^{M-1}\sum_{l=1}^{M-k} Cov\left(I_k, I_{k+l}\right)\right] \tag{14}$$

where the covariance between monitoring sites is calculated as:

$$Cov\left(I_k, I_{k+l}\right) = \sqrt{Var\left(I_k\right)Var\left(I_{k+l}\right)\rho^c_{k,\,k+l}} \tag{15}$$

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

$$r_{k,k+l} = \frac{\frac{1}{W}\sum_{i=1}^{W}\left(x_{k,i} - \overline{x_k}\right)\left(x_{k+l,i} - \overline{x_{k+l}}\right)}{\sqrt{Var\left(x_k\right)Var\left(x_{k+l}\right)}} \tag{16}$$

where x^i_k and x^i_{k+l} represent one set of W concurrent observations at the sites. The alternative estimate of $Var(\mathrm{T})$ given by Eq. 14 can be used to calculate confidence in the aggregate trend direction C^T using Eq. (13).

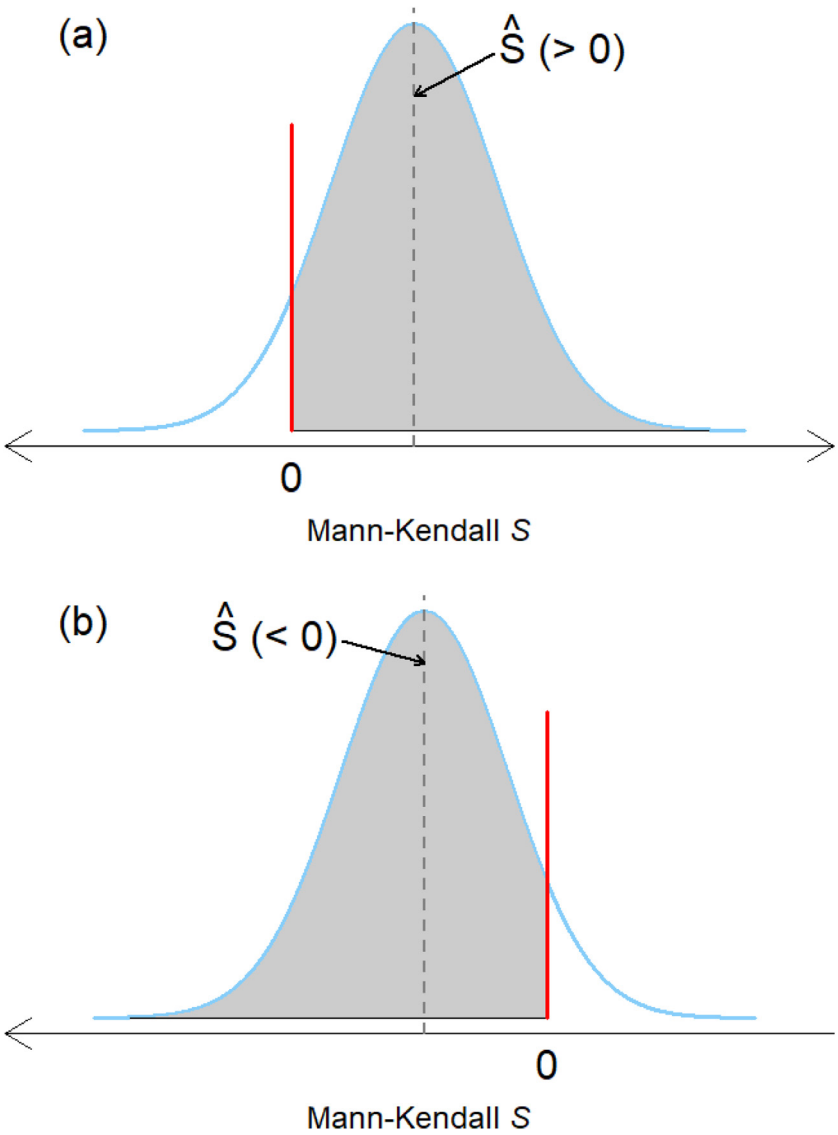

Fig. 1. Using the posterior probability distribution to assess confidence in the trend direction. The blue curve is the posterior probability distribution that encompasses an area of unity and is centered on the estimated value of the Mann–Kendall statistic ($\hat{S}$) with variance $\boldsymbol{Var}(\boldsymbol{S})$. For (a), the estimated trend is positive and the area under the curve with $\hat{S} > 0$ indicates the probability that the true trend is positive and that the estimated trend has the same sign as the true trend. For (b), the estimated trend is negative and the area under the curve with $\hat{S} < 0$ indicates the probability the trend has the same sign as the true trend.

2.3. Example application to water quality data in New Zealand

River water quality is monitored at sites distributed across New Zealand by 15 regional councils and the National Institute of Water and Atmospheric Research (NIWA) (Fig. 2). These monitoring programmes are well-established and timeseries of observations have increased over time so that there are now more than 1000 sites at which a range of water quality variables have been regularly observed for up to 30 years (Larned et al. 2016). These river water quality data are regularly used to produce numerous reports on water quality state and trends (e.g., Larned *et al.* 2004, 2016).

In this study we assessed river water quality trends for the period 2001 to 2020 (i.e., 20 years) for five physico-chemical variables: dissolved reactive phosphorus (DRP mg m^{-3}), nitrate-nitrite nitrogen (NNN mg m^{-3}), total nitrogen (TN mg m^{-3}), total phosphorus (TP mg m^{-3}), visual clarity (CLAR m), and the microbiological variable *Escherichia coli* (ECOLI MPN 100 ml^{-1}). A period of 20 years was chosen because previous studies have shown that trends for shorter timescales are strongly influenced by interannual climate variability (Snelder et al. 2021a, b).

We acquired and collated the water quality data for the 20-year period from each council and NIWA. As part of collating the datasets we undertook data checks and corrections described by Larned *et al.* (2016). For most variables, there was variation in analytical methods across the data collecting agencies. For each variable, only data corresponding to the most widely used and comparable procedures were retained and the remaining data omitted as described by Larned *et al.* (2016). The individual datasets contained censored values indicating that reported values were either below the analytical detection limit or above the reporting limit. Censored values were identified in all datasets and these entries were consistently indicated by the combination of the reported values and a flag indicating the type of censoring.

All variables were observed at monthly, bi-monthly or quarterly sampling intervals, which we treated as 'seasons' for the trend analysis. There were two common deviations from a fixed sample interval: (1) the collection of more than one observation in a sample interval (e.g., two observations within a month) and (2) a change in the sample interval within the time period. The second type of deviation was most common because many sites had changed from lower frequency (e.g., bi-monthly or quarterly) to monthly sampling during the time period. For the first type of deviation, we took the observation closest to the midpoint of the sample interval to represent the season's observation. For the second type of deviation, we derived a time series that was consistently of the lower frequency by taking, from the higher frequency part of the record, those observations that were closest to the midpoint of the coarser sampling interval. Taking the observation closest to the midpoint of the sample interval, rather than averaging over samples

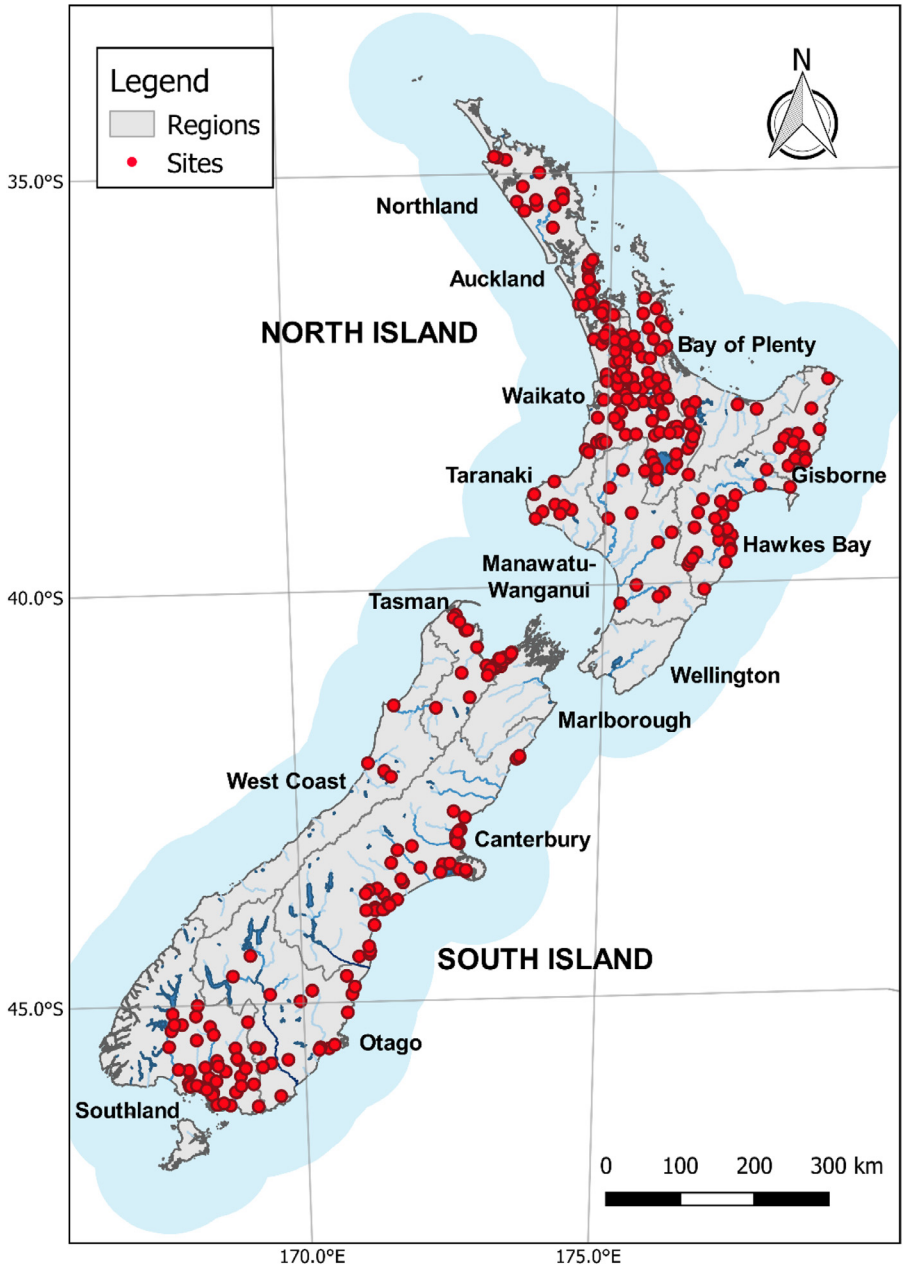

Fig. 2. Map of New Zealand showing the location of the 352 river monitoring sites.

within the sample interval, was applied to avoid inducing a trend in the variance (Helsel et al. 2020).

To provide for robust representation of the 20-year time period for each site and variable combination and comparison of trends between sites, we filtered the sites to have an acceptable proportion of gaps (i.e., missing values in sampling intervals) and distribution of gaps between the start and end date of the time period. Our choice of filtering rules represented a trade-off between highly restrictive rules that would ensure a high level of robustness of the individual trend analyses but excluded numerous sites thereby reducing spatial coverage, and highly lenient rules that would retain more sites but decrease the robustness of the individual trend assessments. Our filtering rules required that sites had observations for at least 90% of the years in the period and 90% of the sample intervals. If a site failed to comply with the filtering rules when sampling intervals were months, the data were coarsened to bi-monthly and then quarterly and the highest frequency times series that complied with the filter rules was retained for analysis. It is noted that the retained data implies that the analysis has variable levels of statistical power and temporal representativeness across the sites. After application of the filtering rules, the dataset comprised 308,300 observed values of the variables at 352 sites across New Zealand (Table 1).

We assessed the seasonality of each of the filtered site and variable combinations. Where there was a statistically significant difference in the observations grouped by season (Kruskal Wallis test $\alpha \leq 0.05$), we used the Seasonal Kendall statistic to calculate $\hat{S}$ and C^{τ} otherwise we used the Mann–Kendall statistic. For each site and variable combination, we expressed confidence in trend direction (C^{τ}) using four categories used by Choquette et al. (2019; Table 2). For each variable, we assessed the direction of trends at national and regional scales and confidence in those assessments, by aggregation of all sites (i.e., national scale) and by the regions shown in Fig. 2, respectively, based on D^{T} and C^{T}.

All analyses were performed in the R Statistical Computing Environment (R Core Team 2021). The NADA package (Lopaka Lee 2020) was used to calculate the Mann Kendall statistic in the presence of tied and censored values using methods described by (Helsel 2011). All data and code are provided as supplementary material.

Table 1
Sites and data retained for trend assessment of six water quality variables after application of the filtering rules.

Variable	Abbreviation	Number of sites	Number of observations	Proportion of sites with monthly, bimonthly and quarterly observations (%)	Proportion of sites that were seasonal (%)
Visual clarity	CLAR	207	42393	59, 15, 26	69
Dissolved reactive phosphorus	DRP	313	68221	61, 12, 27	61
Nitrate-nitrite nitrogen	NNN	303	62744	58, 13, 29	94
Total nitrogen	TN	137	27455	52, 19, 29	82
Total phosphorus	TP	272	58454	66, 12, 22	58
Escherichia coli	ECOLI	262	49033	37, 11, 52	69

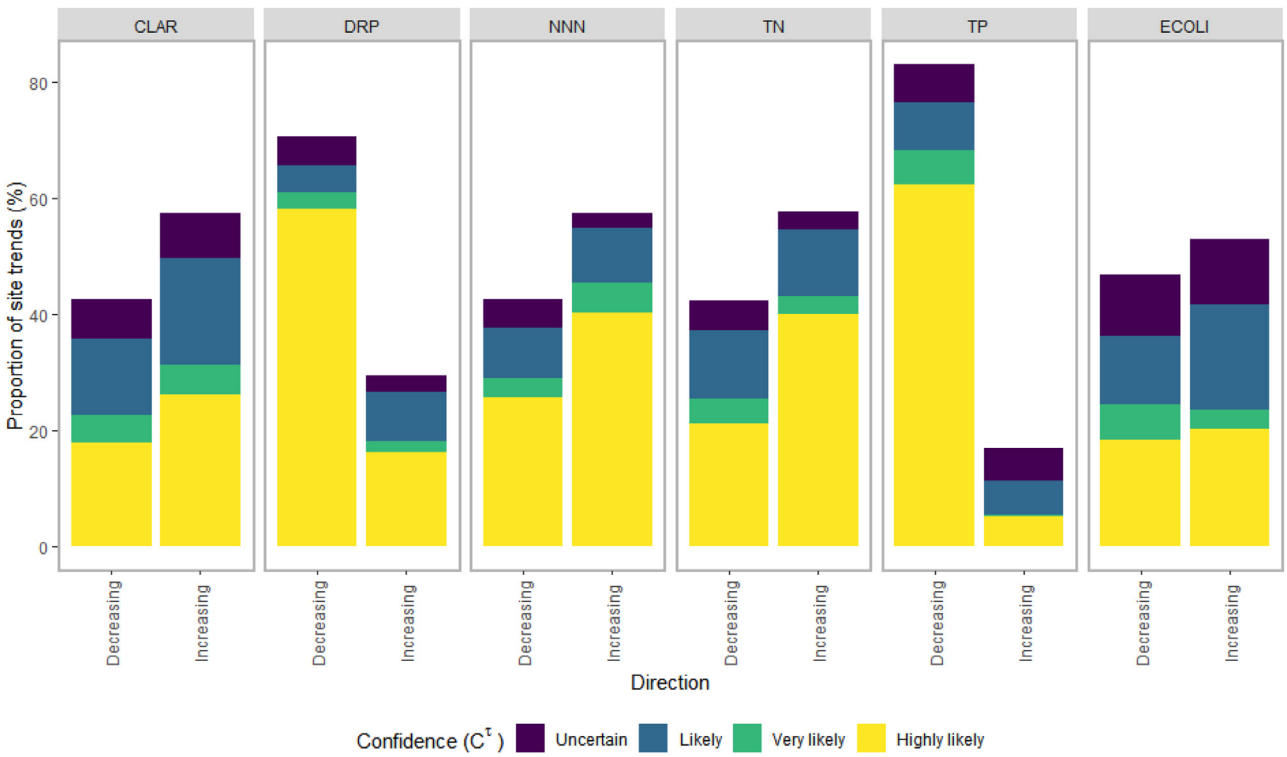

Fig. 3. Summary plot representing the proportion of sites with increasing and decreasing 20-year trends for six water quality variables at each categorical level of confidence for $\mathbf{C}^{\tau}$ defined in Table 2. Decreasing trends indicate water quality improvement and all variables except CLAR, for which increasing trends indicate improvement.

Table 2
Level of confidence categories used to convey the confidence in the assessed trend direction.

Categorical level of confidence in assessed trend direction	Value of C
Highly likely	0.95 to 1.00
Very likely	0.90 – 0.95
Likely	0.67 – 0.90
Uncertain	0.50 – 0.67

3. Results

The proportions of individual site trends belonging to each categorical level of confidence in the assessed direction of the 20-year trends are shown in Fig. 3. Decreasing trends indicate water quality improvement and all variables except CLAR, for which increasing trends indicate improvement. At the national scale, there was a dominance of sites with decreasing trends with at least "likely" levels of confidence for DRP, and TP (Fig. 3). For DRP and TP, 58% and 63% of decreasing site trends were in the "Highly likely" category, respectively (the strongest evidence of decrease). Correspondingly for DRP and TP, a minority of increasing site trends (27% and 11% respectively) were assigned to at least the "Likely" category. For NNN and TN, a majority of sites had increasing trends (indicating degradation) with at least "Likely" levels of confidence (55% and 55%, respectively, Fig. 3). For CLAR and ECOLI there was a more even split in the proportions of increasing and decreasing trends. For CLAR and ECOLI decreasing trends, 36% and 36% of sites were assigned to at least the "Likely" category, respectively, and for increasing trends, 49% and 42% of sites were assigned to at least the "Likely" category, respectively.

Observations made in each sample interval exhibited a degree of cross-correlation between all pairs of sites for all water quality variables (Fig. 4). Between site cross-correlation could be both negative and positive but there was a tendency for correlation to be positive. The mean of the cross-correlations was largest (0.3) for NNN and least for ECOLI (0.03).

A compact graphical representation of the results of the trend direction and strength assessment for sites aggregated nationally is shown in Fig. 5 and in more detail in Table 3. The aggregate trend direc-

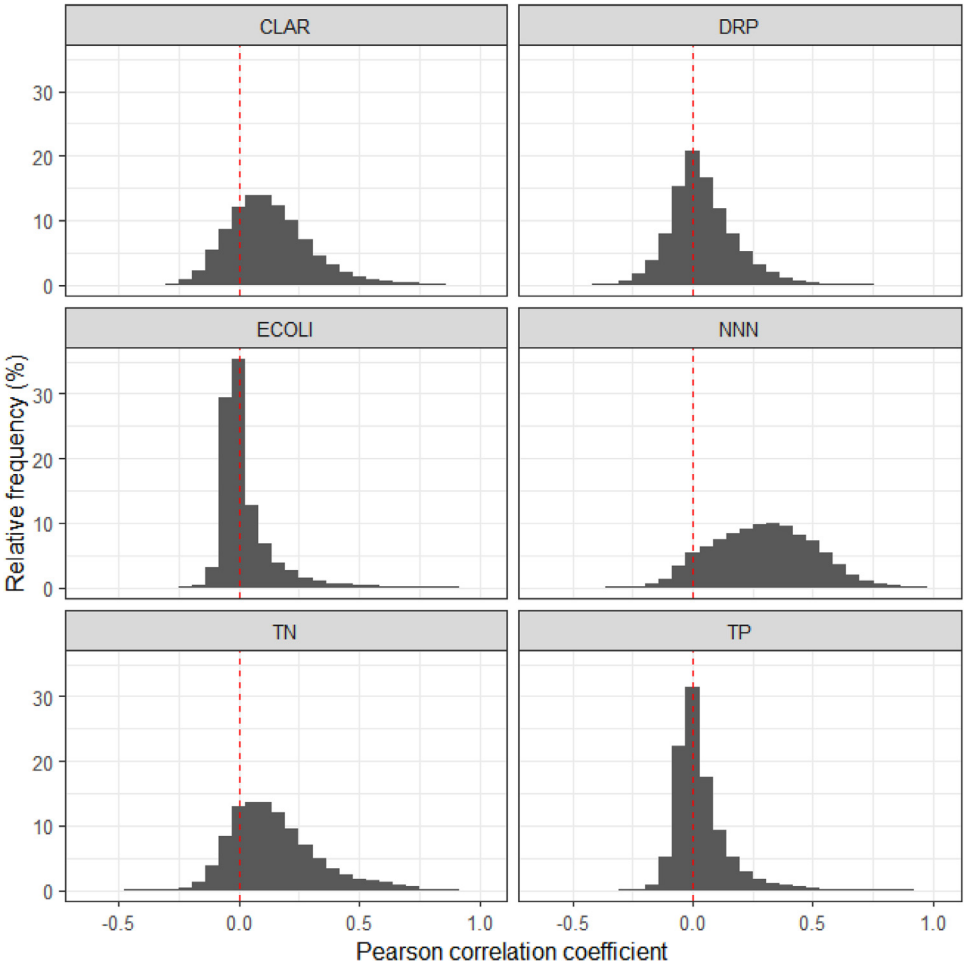

Fig. 4. Distributions of cross-correlations between observations between all pairs of sites for each of the six water quality variables. The red dashed line indicates correlation of zero.

Table 3
Aggregate trend strength ($\hat{\mathbf{T}}$), direction ($\boldsymbol{D}^{\mathbf{T}}$) and confidence in assessed aggregate direction ($\boldsymbol{C}^{\mathbf{T}}$) for six variables and sites aggregated nationally and by four regions (for DRP and NNN only). Confidence in assessed aggregate direction ($\boldsymbol{C}^{\mathbf{T}}$) is shown for calculations that have been corrected for spatial correlation and calculations that have been left uncorrected. Confidence categories express $\boldsymbol{C}^{\mathbf{T}}$ as defined by Table 2.

Variable	Domain	Number of sites	$\hat{T}$	D^T	C^T (Corrected)	Confidence category (Corrected)	C^T (Uncorrected)	Confidence category (Uncorrected)
CLAR	National	207	0.57	Increasing	0.78	Likely	1.00	Highly likely
DRP	National	313	0.71	Decreasing	1.00	Highly likely	1.00	Highly likely
NNN	National	303	0.57	Increasing	0.79	Likely	1.00	Highly likely
TN	National	137	0.58	Increasing	0.84	Likely	1.00	Highly likely
TP	National	272	0.83	Decreasing	1.00	Highly likely	1.00	Highly likely
ECOLI	National	262	0.53	Increasing	0.74	Likely	0.95	Very likely
DRP	Waikato	103	0.93	Decreasing	1.00	Highly likely	1.00	Highly likely
	Hawkes Bay	25	0.60	Increasing	0.78	Likely	0.95	Very likely
	Canterbury	42	0.79	Decreasing	1.00	Highly likely	1.00	Highly likely
	Southland	36	0.83	Decreasing	1.00	Highly likely	1.00	Highly likely
NNN	Waikato	103	0.61	Increasing	0.89	Likely	1.00	Highly likely
	Hawkes Bay	26	0.88	Decreasing	1.00	Highly likely	1.00	Highly likely
	Canterbury	45	0.62	Increasing	0.83	Likely	1.00	Highly likely
	Southland	38	0.84	Increasing	1.00	Highly likely	1.00	Highly likely

tion (D^T) was decreasing for DRP and TP and aggregate trend strength ($\hat{T}$) was >0.7 (Table 3). This is consistent with the dominance of decreasing individual sites trends for these variables (Fig. 3). Confidence in these assessed directions were "Highly likely" (Fig. 5) irrespective whether or not the assignment had been corrected for spatial correlation (Table 3). The aggregate trend direction (D^T) was increasing for variables CLAR, NNN, TN and ECOLI. The aggregate trend strength ($\hat{T}$) for these variables was between 0.53 (ECOLI) and 0.58 (TN). The uncorrected confidence in the assessed aggregate trend directions was "Highly likely" or "Very likely" for these variables but reduced to "Likely" for all four variables when confidence was corrected for spatial correlation (Table 3).

There were four regions with > 25 monitoring sites with DRP and NNN data that was consistent with the filtering rules. The aggregate trend direction (D^T) for three regions for DRP and NNN was consistent with the national scale direction (i.e., decreasing and increasing, respectively; Fig. 6). However, for the Hawkes Bay region, these directions were reversed. After correction for spatial correlation, confidence in the

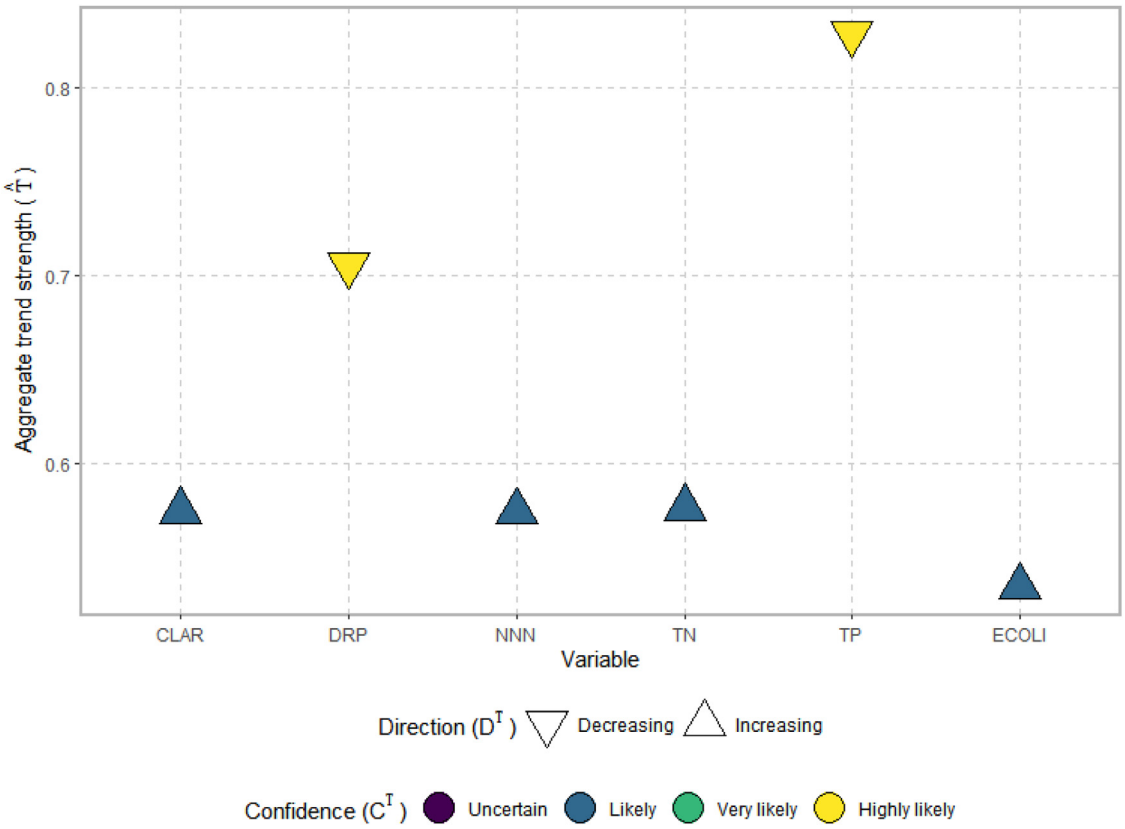

Fig. 5. Aggregate trend strength ($\hat{T}$) and direction ($\boldsymbol{D}^{\mathbf{T}}$) for 20-year trends for six water quality variables over all sites. Confidence in the aggregate direction (CT) is indicated by the four confidence categories (see Table 2 for details). Confidence is shown for calculations that have been corrected for spatial correlation (Eq. (14)).

Fig. 6. Aggregate trend strength ($\hat{T}$) and direction ($\boldsymbol{D}^{\mathbf{T}}$) for 20-year trends for DRP and NNN for sites in four regions with >25 sites. See Fig. 5 for explanation of plot key.

aggregate trend direction (C^T) was "Highly Likely" for both variables in only the Southland region (Table 3).

4. Discussion

4.1. Assessing confidence in trend direction rather than hypothesis testing

The continuous measure of confidence in trend direction (C^τ) is an alternative to the use of NHST for assessing the reliability of a site trend. The difference between approaches is demonstrated by considering two trend assessments, A and B, with positive Mann–Kendall $\hat{S}$ values and p-values of 0.04 and 0.14, respectively. A significance test with $\alpha = 0.05$, would reject the null-hypothesis for A at the 95% confidence level and would not reject the null-hypothesis for B. Using our trend direction assessment, positive trends for A and B would be inferred with 98% and 93% confidence in the direction, respectively. Because C^τ quantifies the evidence that the trend is in the assessed direction (i.e., positive, or negative), it is consistent with the principle that there is always a trend and does not allow the conclusion that there was "no trend" to be drawn. Consequently, our trend direction assessment provides decision makers with a more complete description of the evidence than an arbitrary "significant/not significant" designation.

The continuous measure of confidence in trend direction (C^τ) links McBride's (2019) use of the Bayesian credible interval to Choquette *et al.'s* (2019) use of the Mann–Kendall test p-value to estimate the likelihood that the assessed trend direction is correct. Confidence in trend direction (C^τ) retains the simple and robust frameworks for trend assessment provided by the Mann–Kendall correlation statistic but has the advantage that it removes the need to choose an arbitrary α value to define a credible interval and does not invoke NHST. However, the quantity C^τ is equal to Choquette *et al.'s* (2019) measure of likelihood [1 − p/2]. C^τ is also numerically equivalent to a Bayesian index of effect existence called Probability of Direction (Greenland and Poole 2013; Makowski et al. 2019). The Probability of Direction index does not require a prior distribution, nor does it rely on a null hypothesis, and is mathematically defined as the proportion of the posterior distribution that is of the median's sign (Makowski et al. 2019). An important point that arises from this is that there is nothing intrinsically wrong with p-values, it is their misinterpretation that is the problem (Makowski et al. 2019).

Chen et al. (2021) refer to confidence and likelihood as qualitative and quantitative representations of uncertainty, respectively, with confidence including a degree of expert judgement. Hirsch *et al.* (2015), Choquette *et al.* (2019) and Murphy (2020) used likelihood to refer to their probabilistic (i.e., based on quantitative statistical analysis) estimates that the assessed direction of the trend was correct. We use 'confidence' because it is a more commonly used and understood term by non-scientists, however. we acknowledge that it has the same meaning a likelihood as defined by Chen et al. (2021).

We note that as the size of the sample (i.e., the number of observations) increases, confidence in trend direction increases. When the sample size is very large, C^τ can be high, even if the trend rate is very low. It is important, therefore, that C^τ is interpreted correctly as the confidence in direction and not as the importance of the trend.

4.2. Aggregate trend strength and direction

A further problem with using NHST arises when trends across multiple sites are aggregated to assess the direction of trends at scales larger than that represented by individual sites. If only significant trends are counted, there is a loss of information about the general trend direction. However, if all trends are counted, it is important that the aggregation accounts for the confidence in each trend's direction. The C^T statistic provides a continuous measure of confidence in the aggregate trend direction that is derived from all available site trends, and which accounts for the confidence in the individual trend direction assessments. An advantage of the C^Tstatistic over existing regional trend assessment methods is that it utilises individual site trend assessments as the input data thereby creating a direct link between site-scale and aggregate trends. This also means the C^Tstatistic can be calculated from mixtures of individual site-scale trends including those that are assessed from data of different monitoring frequencies and for site trends that are judged to be seasonal and non-seasonal. As a continuous measure of confidence in aggregate trend direction, C^T has the same advantages as confidence in the direction of individual trends (C^τ) (i.e., it does not allow the conclusion that there was "no trend" to be drawn and it provides decision makers a complete description of the evidence).

The C^T statistic also accounts for spatial correlation that may exist among sites within a monitoring network, which is not accounted for by simple tabulation of site trends and is not always explicitly accounted for by 'regional' trend assessment methods (e.g., Helsel and Frans 2006). The presence of spatial correlation results in under-estimation of the variance and over-estimation of confidence in trend assessment and prevents the analytical derivation of the posterior probability distribution of T (Douglas et al. 2000; Yue and Wang 2002). The approach presented here develops an approximate distribution for T based on the cross-correlations of the observations (Eq. 16). The implications of accounting for spatial correlation are shown by our study where, when C^T was corrected for spatial correlation, confidence in the aggregate trend directions often decreased (Table 3).

We refer to 'aggregate' rather than 'regional' trends because long-term monitoring sites are generally selected for a range of purposes and therefore unlikely to be spatially representative of a region, however that is defined. For example, the national network of river water quality monitoring sites in New Zealand over-represents low elevation sites that are more likely to be degraded than upland sites whose catchments are less dominated by agricultural and urban land use (Larned et al. 2014). Therefore, the national and regional $\hat{T}$ and C^T statistics presented here represent the aggregate trend direction as indicated by the monitoring network but cannot be considered to represent all national or regional rivers. In addition, site aggregation can be undertaken based on other than contiguous geographic regions (Helsel and Frans 2006). For example, sites could be aggregated by classes based on environmental characteristics such as catchment climate and land cover (Larned et al. 2016).

4.3. Twenty-year water quality trends

At the national scale, the 20-year river water quality trends in New Zealand indicate appreciable variation in trend direction between sites for all variables but also some general patterns in aggregate trend direction, which differed between variables. A striking aggregate pattern at the national scale is increasing nitrogen but decreasing phosphorus. These patterns have been identified in New Zealand's river water quality data by previous studies (McDowell et al. 2019; Snelder et al. 2021a). In the past 40 years, pastoral agriculture in New Zealand has experienced significant intensification and diversification (Smith and Montgomery 2004; MacLeod and Moller 2006). The changes include increased fertilizer and supplementary feed input, expansion of irrigation, and increased dairy farming and contraction of sheep and beef farming (MacLeod and Moller 2006). Concerns about water quality impacts have triggered requirements for the agriculture sector to improve management of fertilizer, livestock effluent and irrigation water and to reduce stock access to streams, increase riparian protection and increase tree planting on erodible hill country (Monaghan et al. 2021).

McDowell *et al.* (2019) suggest that decreasing trends in river DRP and TP concentrations over the last 20 years may be attributable to the growing use of mitigation measures to reduce the loss of phosphorus from agricultural land (e.g., shifting from high to low solubility fertilizers). In contrast, increasing trends in NNN and TN are consistent with limitations in the ability to mitigate nitrogen loss from farms (Monaghan et al. 2021) and increasing nitrogen fertilizer use on agri-

cultural land, which is driven by replacement of the formerly dominant sheep industry by dairy farming (Dymond et al. 2013; MFE and StatsNZ 2019).

Our analysis indicates that aggregate trend direction varies between spatial domains. For example, the national-scale pattern of increasing NNN and decreasing DRP was repeated in the Waikato, Canterbury and Southland regions but was reversed in the Hawkes Bay region (Fig. 6). Snelder et al. (2021a) showed that regional variation in changes in the intensity of agriculture, as indicated by changes in the density of pastoral animals, explains some of the variation in water quality trends within New Zealand over the last 20 years. The finding by this study that aggregate trend direction varies between regions is consistent with inter-regional variation in changes in agricultural land use of the last 20 years.

5. Conclusions

The minimum level of information that land and water managers require regarding changes in monitored environmental variables over time is whether they are increasing or decreasing, both at individual sites and at larger spatial scales, and the degree of statistical support for these assessments. In this article we provide approaches for assessing confidence in trend direction at site and larger scales in continuous, rather than binary (i.e., significant/non-significant) terms. These continuous measures of confidence can be expressed categorically, and statements can be made such as "it is likely that the national trend in nitrate between 2001 and 2020 was increasing". This means that we assess there is at least a 67 out of 100 chance that the direction was increasing. This allows for a natural approach to the presentation of information about the certainty associated with a trend assessment (Hirsch et al. 2015). Expressing confidence in this way also helps to clarify to decision makers that they must consider trade-offs involving the risks, costs and benefits of taking action or not.

Our example analysis reveals that there are patterns in the direction of trends in New Zealand's river water quality at various spatial scales. Assessment of patterns in trend direction at various spatial scales, and robust evaluation of confidence in these assessments, is relevant to managers and decision makers who need to consider the pressures or management actions that are driving these trends and to formulate appropriate responses.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Snelder, Fraser, Whitehead reports financial support was provided by LWP Ltd.

CRediT authorship contribution statement

TS and CF conceived and developed the theory. TS developed and performed the computations. AW assembled the data. TS drafted the manuscript. CF and AW provided critical feedback and helped with the analysis and manuscript.

Data Availability

Data will be made available on request.

Acknowledgments

We thank the staff of regional and district councils and NIWA who provided water quality data and details about monitoring programmes and methods. We thank Vic Duoba of Statistics New Zealand for statistical advice. We thank two anonymous reviewers whose comments greatly improved our original manuscript.

Funding

This work was funded as an output from the Suitability programme of the Our Land and Water National Science Challenge [Ministry of Business, Innovation and Employment contract C10 × 1507].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.envc.2022.100601.

References

Behmel, S, Damour, M, Ludwig, R, Rodriguez, MJ, 2016. Water quality monitoring strategies—a review and future perspectives. Sci. Total Environ. 571, 1312–1329.
Cairns, J, Pratt, JR, 1993. A history of biological monitoring using benthic macroinvertebrates. Freshwater biomonitoring and benthic macroinvertebrates 10, 27.
Cash, DW, Adger, WN, Berkes, F, et al., 2006. Scale and cross-scale dynamics: governance and information in a multilevel world. Ecol. Soc. 11.
Chen, D, Rojas, M, Samset, BH, et al., 2021. Framing, context, and methods. In: Climate Change 2021: the Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, pp. 147–286.
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. J. Great Lakes Res. 45, 21–39.
Cohen, J, 1994. The earth is round (p<. 05). Am. Psychol. 49, 997.
Cumming, G, Cumming, DHM, Redman, C, 2006. Scale mismatches in social-ecological systems: causes, consequences, and solutions. Ecol. Soc. 11. doi:10.5751/ES-01569-110114.
Davies-Colley, RJ, Smith, DG, Ward, RC, et al., 2011. Twenty years of New Zealand's national rivers water quality network: benefits of careful design and consistent operation 1. JAWRA J. Am. Water Resour. Assoc. 47, 750–771.
Douglas, EM, Vogel, RM, Kroll, CN, 2000. Trends in floods and low flows in the United States: impact of spatial correlation. J. Hydrol. 240, 90–105.
Dymond, J, Ausseil, A-G, Parfitt, R, et al., 2013. Nitrate and phosphorus leaching in New Zealand: a national perspective. N. Z. J. Agric. Res. 56, 49–59. doi:10.1080/00288233.2012.747185.
Greenland, S, Poole, C, 2013. Living with P values: resurrecting a Bayesian perspective on Frequentisi statistics. Epidemiology 62–68.
Greenland, S, Senn, SJ, Rothman, KJ, Carlin, JB, Poole, C, Goodman, SN, Altman, DG, 2016. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European Journal of Epidemiology 31, 337–350.
Helsel, D, Frans, L, 2006. Regional Kendall test for trend. Environ. Sci. Technol. 40, 4066–4073.
Helsel, DR, 2011. Statistics for Censored Environmental Data using Minitab and R. John Wiley & Sons.
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. Environ. Model. Softw. 73, 148–166. doi:10.1016/j.envsoft.2015.07.017.
Hirsch, RM, Slack, JR, 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resour. Res. 20, 727–732.
Hirsch, RM, Slack, JR, Smith, RA, 1982. Techniques of trend analysis for monthly water quality data. Water Resour. Res. 18, 107–121.
Isbell, F, Gonzalez, A, Loreau, M, et al., 2017. Linking the influence and dependence of people on biodiversity across scales. Nature 546, 65–72.
Jones, LV, Tukey, JW, 2000. A sensible formulation of the significance test. Psychol. Methods 5, 411.
Kendall, MG, 1975. Rank Correlation Methods. Charles Griffin, London.
Larned S, Unwin M, Sykes J (2014) Recommendations for new sites to improve representativeness in the New Zealand river environmental monitoring network. NIWA, Christchurch, New Zealand
Larned, ST, Scarsbrook, MR, Snelder, T, et al., 2004. Water quality in low-elevation streams and rivers of New Zealand. N. Z. J. Mar. Freshw. Res. 38, 347–366.
Larned, ST, Snelder, T, Unwin, MJ, McBride, GB, 2016. Water quality in New Zealand rivers: current state and trends. N.Z. J. Mar. Freshw. Res. 50, 389–417.
Lopaka Lee (2020) NADA: nondetects and data analysis for environmental data. Version R package version 1.6-1.1URL https://CRAN.R-project.org/package=NADA
MacLeod, CJ, Moller, H, 2006. Intensification and diversification of New Zealand agriculture since 1960: an evaluation of current indicators of land use change. Agric, Ecosyst. Environ. 115, 201–218.
Makowski, D, Ben-Shachar, MS, Chen, SH, Lüdecke, D, 2019. Indices of effect existence and significance in the Bayesian framework. Front. Psychol. 10, 2767.
Mann, HB, 1945. Nonparametric tests against trend. Econometrica 245–259.
McBride, GB, 2019. Has water quality improved or been maintained? A quantitative assessment procedure. J. Environ. Qual. 48, 412–420.
McDowell, RW, Hedley, MJ, Pletnyakov, P, et al., 2019. Why are median phosphorus concentrations improving in New Zealand streams and rivers? J. R. Soc. N. Z. 1–28.
MFE & StatsNZ (2019) Environment Aotearoa 2019. Ministry for Environment and Statistics New Zealand, Wellington, New Zealand

Monaghan, RM, Basher, L, Spiekermann, R, et al., 2021. Quantifying contaminant losses to water from pastoral landuses in New Zealand II. The effects of some farm mitigation actions over the past two decades. N.Z. J. Agric. Res. doi:10.1080/00288233.2021.1876741.

Monteith, DT, Evans, CD, Henrys, PA, et al., 2014. Trends in the hydrochemistry of acid-sensitive surface waters in the UK 1988–2008. Ecol. Indic. 37, 287–303.

Murphy, JC, 2020. Changing suspended sediment in United States rivers and streams: linking sediment trends to changes in land use/cover, hydrology and climate. Hydrol. Earth Syst. Sci. 24, 991–1010.

Oelsner GP, Sprague LA, Murphy JC, et al (2017) Water-quality trends in the nation's rivers and streams, 1972–2012—data preparation, statistical methods, and trend results. US Geological Survey

R Core Team, 2021. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rozeboom, WW, 1960. The fallacy of the null-hypothesis significance test. Psychol. Bull. 57, 416.

Sicard, P, Agathokleous, E, De Marco, A, et al., 2021. Urban population exposure to air pollution in Europe over the last decades. Environ. Sci. Eur. 33, 1–12.

Smith, W, Montgomery, H, 2004. Revolution or evolution? New Zealand agriculture since 1984. GeoJournal 59, 107–118.

Snelder, TH, Fraser, C, Larned, ST, et al., 2021a. Attribution of river water-quality trends to agricultural land use and climate variability in New Zealand. Mar. Freshw. Res..

Snelder, TH, Larned, ST, Fraser, C, De Malmanche, S, 2021b. Effect of climate variability on water quality trends in New Zealand rivers. Mar. Freshw. Res..

Theil, H, 1950. A rank-invariant method of linear and polynomial regression analysis. Indag. Math. 12, 173.

Vogel, RM, Rosner, A, Kirshen, PH, 2013. Brief Communication: likelihood of societal preparedness for global change: trend detection. Nat. Hazards Earth Syst. Sci. 13, 1773–1778.

Wasserstein, RL, Lazar, NA, 2016. The ASA statement on p-values: context, process, and purpose. Am. Stat. 70, 129–133. doi:10.1080/00031305.2016.1154108.

Yue, S, Wang, CY, 2002. Regional streamflow trend detection with consideration of both temporal and spatial correlation. Int. J. Climatol. 22, 933–946.