**Aquatic
Living
Resources**

A selection of nonparametric statistical methods for assessing trends in trawl survey indicators as part of an ecosystem approach to fisheries management (EAFM)

John Cotter[a]

Centre for Environment, Fisheries and Aquaculture Science Lowestoft, NR33 0HT, UK

Abstract – A selection of ten uni- and multivariate, nonparametric statistical methods thought to be useful for assessing trends in indicators estimated from trawl surveys is described. Nonparametric methods make minimal assumptions about the data and can therefore be suitable when parametric modelling methods, as typically preferred by fisheries scientists, are not suitable. The various methods described are sensitive to a variety of specific features of a trend thereby allowing many different forms of variation over time to be identified, e.g. changes in level, sloping straight or curved lines, and differing cyclicities. In contrast, model-based investigations can be constrained in this regard if the single model which is adopted and fitted to the trend is a simplification of its form.

Key words: Quantiles / Runs test / Mann-Kendall statistic / Spearman correlation / Jonckeere test / Nonparametric slope estimator / Cochran Q / Aligned rank test / Dietz-Killeen multivariate trend test

1 Introduction

These notes describe a selection of nonparametric statistical methods thought to be useful for assessing trends in indicators from trawl surveys or other fishery statistics, e.g. abundance-at-age, mean length, geostatistical indices. The trends referred to here are assumed to relate to time but they could also relate to a spatial transect. Several of the references cited come from the literature on monitoring of water pollution (Esterby 1993; Smith et al. 1993; Esterby 1996) where the sporadic and chaotic nature of variation combined with frequent gaps in the time-series has stimulated development of nonparametric methods requiring minimal assumptions. Fisheries scientists typically prefer parametric modelling for assessing trends in fish stocks and have exploited nonparametric methods relatively lightly. All the same, trends in fish stocks could sometimes be established with less reliance on assumptions if nonparametric methods were used. In addition, interest nowadays is shifting from estimation of quantities of fish in a single commercial stock to assessment of whole ecosystems, a task for which well established, structural models are not always available.

Neither nonparametric nor parametric methods can distinguish between trends in the underlying, true signal and trends

in the sampling or measurement errors associated with the variable of interest. Since trawl survey indicators typically have substantial coefficients of variation, +/– 50% is not exceptional, it is necessary to assume that measurement and sampling errors are well behaved statistically. The expected bias should be constant or zero over time so that it is not confused with signal, and the expected serial correlation should be near zero so that a trail of errors winding above and below the true signal does not contribute spuriously to an observed trend. Similarly, measurement and sampling errors should be independent of the level of the true signal so as to prevent any exaggeration or dampening of its variability. In practice, the truth of these assumptions is seldom verifiable and reliance must be placed on good sampling and measurement practices.

Process error can be another, distinct problem (Thomas et al. 2004). It arises when the signal is modelled, e.g. as a linear, or monotonic trend, but the true signal does not conform exactly with the model. So, a curved signal fitted with a straight line would be expected to show serially correlated process errors arising purely from the mismatching geometric forms. This is a particular disadvantage of straight lines fitted by least squares, because, when the signal is curved, the estimate of residual error variance and, consequently, the standard errors of fitted parameters are likely to be biased to give a spurious impression of accuracy. Nonparametric methods can also be adversely affected by process error, as when serially correlated data collected seasonally are being used to assess

---

a Corresponding author:
john.cotter@fishworldscience.com

Table 1. List of nonparametric methods described in this paper, their purposes, and opinions on ease of application.

Method	Purpose to:	Ease of application
1. Estimating quantiles	Estimate quantiles and confidence intervals for observed values or serial differences	Simple
2. Assessing trend relative to quantiles	Detect any kind of trend from one period to another	Simple
3. Runs test	Detect serial correlation around median or fitted trend	Simple
4. Mann-Kendall's K	Detect monotonic trend	Spreadsheet: onerous R: simple
5. Spearman's rho	Detect trend in ranks	Simple
6. Jonckheere's test	Detect monotonic trend; multiple observations at each time point	Onerous
7. Thiel's or Sen's slope estimator	Estimate linear slope	Onerous
8. Cochran's Q	Test whether multiple time series are responding similarly, but not a test of trend	Simple
9. Aligned rank test	Detect multivariate trend; different variables show same trend direction	Simple
10. Dietz-Killeen test	Detect multivariate monotonic trend; different variables may show different trend directions	Difficult; R program available

Table 2. International bottom trawl survey (IBTS) quarter 1: time series of abundance indices (numbers caught per hour) for cod in the North Sea in 6 age groups. Reporting of ages 3 to 6 did not begin until 1983. Medians relate to all available data (see also Fig. 1).

Year	Age1	Age2	Age3	Age4	Age5	Age6
1976	7.9	19.9				
1977	36.7	3.2				
1978	12.9	29.3				
1979	9.9	9.3				
1980	16.9	14.8				
1981	2.9	25.5				
1982	9.2	6.7				
1983	3.9	16.6	2.7	1.8	0.8	1.5
1984	15.2	8	3.9	0.9	1	0.9
1985	0.9	17.6	3.5	1.7	0.5	1
1986	17	3.6	6.8	2.3	1.3	1.1
1987	8.8	28.8	1.4	1.7	0.6	0.9
1988	3.6	6.1	5.8	0.6	0.9	1.1
1989	13.1	6.3	5	2.3	0.4	1
1990	3.4	15.2	2	1	1	0.8
1991	2.4	4.1	3.4	0.8	0.4	0.8
1992	13	4.5	1.2	1	0.3	0.5
1993	12.7	19.9	2	0.7	0.6	0.4
1994	14.8	4.4	3	0.8	0.5	0.5
1995	9.7	22.1	2.8	1.1	0.3	0.3
1996	3.5	8	6	0.7	0.6	0.4
1997	40	6.9	2.3	1.1	0.4	0.4
1998	2.7	26.4	2	0.9	0.5	0.4
1999	2.1	1.6	8.1	0.8	0.5	0.5
2000	6.6	3.8	0.7	2	0.4	0.5
2001	2.8	8.7	1.7	0.2	0.4	0.3
2002	7.8	3.4	4.3	0.5	0.1	0.2
2003	0.6	3	1	1.4	0.4	0.3
2004	7.5	1.3	1.2	0.30	0.4	0.01
Median	7.9	8.0	2.75	0.95	0.5	0.5

a trend over years (El-Shaarawi and Niculescu 1992). In the present paper, trawl surveys are assumed to be annual, as is typically the case, and the complications of seasonal variation are ignored.

The selection of nonparametric methods described here, as listed in Table 1, range from simple and univariate, to elaborate and multivariate. They do not include randomisation tests (Bell et al. 1981; Edgington 1995), procedures for finding step changes or "change points" (Pettitt 1979; Lanzante 1996), or nonparametric and semiparametric curve fitting methods which form a major topic by themselves (Hussian 2005; Wood 2006). The selected methods are applied to a single set of example fish survey data for comparative purposes. A practical and demonstrative approach is intended; theory behind most of the tests can be found from cited texts or other references. These sources also provide longer formulae for dealing with tied values in the data, a complication that is omitted here. Nonparametric methods can be implemented with numerous commercial software packages for statistical analysis, and special purpose programs are

also freely available, e.g. at http://www.ida.liu.se/divisions/ stat/research/engo/reports.html. All of the examples in this paper were analysed with Microsoft Excel using commands such as RANK (), MEDIAN(), BINOMDIST(), and with the open-source R programming system using functions such as rank(), median(), pbinom(), cor.test(). R is available from http://www.r-project.org.

Regarding notation, the following are standardised across the ten tests. Null and alternative hypotheses are labelled H and A, respectively, with the probability of a type 1 error being α, conventionally set at 0.05 though other values may sometimes be advisable. The observed test statistic is T except when traditionally labelled otherwise, e.g. K, J. Probability is denoted as p or as Pr () when used as a function. The standard normal deviate is z. An observed value is X, and the sample size, N, is either the number of annual observations, or more than that if more than one observation per year can be accepted by the test. Then k is used for the number of observation times. Most other notation is only relevant locally.

Example: North Sea cod abundance indices

Table 2 and Figure 1 show abundance indices (numbers of fish per hour) for North Sea cod as found by the ICES
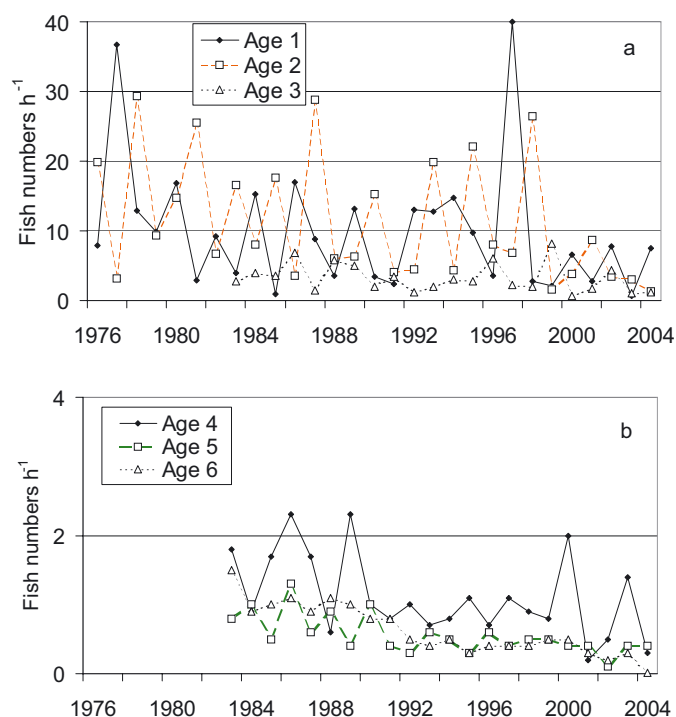
Fig. 1. International bottom trawl survey (IBTS) quarter 1: time series of abundance indices (numbers caught per hour) for cod in the North Sea by age groups: a) ages 1 to 3; b) ages 4 to 6 (with a different scale). Reporting of ages 3 to 6 did not begin until 1983 (see also Table 2).

International Bottom Trawl quarter 1 surveys from 1976 to 2004 at ages 1 to 6. Ages 3 to 6 were missing in the earlier years. These values were taken from a report of the ICES working group on fish stocks of the North Sea and Skagerrak. They will be referred to as "the cod indices" and used to illustrate application of various nonparametric statistical methods. Some use the full time series; others have to use only 1983 to 2004 when all age groups were observed. There are many tied values, particularly in the older age groups, because of the small number of significant figures for each index.

2 Ten nonparametric methods

2.1 Estimating quantiles for first differences

A time series can be characterised at a basic level using the median and its binomial confidence limits. A trend in a time series can be characterised similarly after transforming the observed values into a series of differences with an appropriate lag interval, e.g. at lag 1: $(X_2 - X_1)$, $(X_3 - X_2)$, etc. The median of the differences then estimates the median trend for that lag interval, and the confidence limits indicate whether zero difference, i.e. no trend, is likely. In the following, an "observed value" can be either the original value, or a lag 1 difference depending on which approach is being taken.

The sample median is defined here as the average of the observed values of rank $N/2$ and $N/2+1$ if N is even, and as the observed value of rank $(N + 1)/2$ if N is odd. Assuming that the

sample median coincides approximately with the true median of the signal over the observed time interval, the probability of a randomly timed observation being less than or equal to the sample median can be taken as 0.5. This allows confidence limits based on the binomial distribution to be estimated that are expected to have a known probability of enclosing the true median.

2.1.1 Method

1. Rank the observed values least to greatest and find the median value.
2. Using tables (e.g. Conover 1971) or software for the cumulative binomial distribution with probability of "success" $p = 0.5$ (for the median), find the ranks (a) and (b), with cumulative probabilities nearest to the required confidence limits, e.g. 2.5% and 97.5% for the case of limits of approximately 95%. Limits set exactly at some preset, rounded percentage are seldom possible with the binomial distribution.
3. Binomial confidence limits for the median, $\tilde{\mu}$, are obtained with

$$\Pr\{X_{(a)} \leqslant \tilde{\mu} < X_{(b)}\} = \sum_{i=0}^{b-1} B\{N, 0.5, i\} - \sum_{j=0}^{a-1} B\{N, 0.5, j\}$$

where $B\{N, p, n\}$ is the binomial probability of observing exactly n successes out of N trials, given p.

2.1.2 Example

From Table 2, the median index for the 29 indices for age 1 cod from 1976 to 2004 was 7.9 with 93.8% confidence limits of 3.6 and 12.7. These correspond to ranks $a = 10$ and $b = 20$ with cumulative binomial probabilities of 0.0307 and 0.969 respectively. The median of the 28 first differences (not shown) was –0.8 with 96.4% confidence limits of –5.3 and 6.3. This confidence interval includes zero quite centrally and so does not imply a trend between 1976 and 2004. Looking at Figure 1, the confidence interval might be narrowed by choosing a longer lag interval for calculating differences.

2.1.3 Comments

The median of X and its confidence limits are only applicable to the observed period and must be re-estimated as each new observed value arrives The presence of any trend over time would mean that the median is different from that of any other period, including those that partially overlap.

The observations need not be at equal time intervals, e.g. observation times might be chosen randomly in advance of the observed period in order to avoid the danger of bias from aliasing between a regular sampling interval and any cyclical variability present in the signal. In that case, lagged differences, if required, should be converted to differences per unit time. Investigations need not, of course, be limited to first differences. Second differences would estimate the rate of change

of a trend. For equal lag intervals, L, they are formulated as $\frac{(X_3-2X_2+X_1)}{L^2}$, $\frac{(X_4-2X_3+X_2)}{L^2}$, etc.

Quantiles other than the median can be estimated with confidence limits by adjusting (a), (b), and p. They could be useful for summarising peripheral values, e.g. $p = 0.25$ and 0.75 for quartiles, or $p = 0.1$ and 0.9 for deciles. However, the further the quantile to be estimated is from the median, the larger the sample size generally needed to give reasonable confidence intervals (Berthouex and Hau 1991), though multi-modal distributions may behave exceptionally in this regard. Binomial confidence limits can also be estimated for more than one quantile simultaneously, e.g. the 0.1, 0.5, and 0.9 quantiles (Cotter 1985).

The binomial test can be be made multivariate by using the spatial median and the angles between the observations (Brown 1983). The binomial test can also be re-arranged to estimate confidence limits for p, the proportion of "successes". A reservation, according to Brown et al. (2001), is that erratic coverage probabilities can then arise even for large sample sizes.

2.2 Assessing trend relative to quantiles

The binomial distribution can be used to provide a simple one-tailed test of H : "trend $\geqslant 0$" against A: "trend < 0" (or vice versa) by assuming, as for method (1) above, that the sample median is close to the true median for the observed period. Here "trend" refers to any change in level between the first and second halves of the observation period. It might be, for example, a change in level, a linear or curved trend, or possibly a differing cyclicity. Given enough observations, the test could also be applied subsequently to compare the first quarter of the observation period with the second, the third with the fourth, and so on in order to reveal something of the pattern of change over time, though there may then be a need to correct error rates for multiple hypothesis testing, e.g. using the Bonferroni inequality.

2.2.1 Method

1. Find the median value, $\tilde{\mu}$, and the median time, $\tilde{y}$, i.e. the halfway point for the observation period. This creates 4 quadrants containing all the data. Only the X's to the left of (or on) $\tilde{y}$ are free to be in either the upper or the lower quadrant; whatever the configuration, the X's to the right of $\tilde{y}$ are then fixed with respect to the upper and lower right quadrants by the two intersecting medians. (If not, they would no longer be medians.) Therefore probabilities for different configurations of points should only be based on one half of the data. The left side has one more point than the right for odd N and so is used here.
2. Count the number of observations in the top left quadrant, T_L, and the total number on the left side, N_L. Large T_L imply A, a downward trend.
3. Find the probability of observing T_L or more, given that H, meaning a level or upward trend, is true, i.e.

$$\Pr\{T_L \text{ or more} | N_L, H\} \quad \leqslant \quad \sum_{i=T_L}^{N_L} B\{N_L, 0.5, i\}.$$

A $\leqslant$ sign is needed, rather than an = sign, because $\Pr\{T_L \text{ or more} | trend > 0\} \quad < \quad \Pr\{T_L \text{ or more} | trend = 0\}$. With some tabulations of the binomial distribution, it may be more convenient to evaluate the equivalent $1 - \sum_{i=0}^{T_L-1} B\{N_L, 0.5, i\}$.

4. Accept A, meaning a downward trend, if $\max(\Pr\{T_L \text{ or more} | N_L, H\} \leqslant \alpha$.
5. The power of a statistical test is the probability of rejecting H when a specified A is true (Hollander and Wolfe 1973). The following method is suggested for finding the power of this test. First, specify an A by treating $\tilde{\mu}$ as the $p' < 0.5$ quantile for the first half of the total observation period and as the $(1 - p')$ quantile for the second half. This creates a putative step or trend downwards across the median time, with the median of X to the left being assumed above that to the right. The system is still constrained by $\tilde{\mu}$ and $\tilde{y}$ so only consider the left half of the time series. Now the probability of an observation in the left half falling *above* the p' quantile is $(1 - p')$, so

$$\Pr\{T_L \text{ or more} | N_L, A\} \quad = \quad \sum_{i=T_L}^{N_L} B\{N_L, (1 - p'), i\}.$$

An = sign is used here because A is specified exactly. This expression gives the power of the test when $T_L = T_\alpha$, the critical value chosen to test H.

2.2.2 Example

The locations relative to the upper and lower left quadrants for the cod indices of all ages are indicated (Table 3). Setting $\alpha = 0.05$, we accept downward trends in abundance for cod of ages 6 only for which $p = 0.033$. The other ages are not detectably different from being either level in abundance or trending upwards. For the power of the test, consider just age 2 for which $\Pr\{T_L \text{ or more} | N_L, H\} = 0.304$, and $T_L = 9$ and $N_L = 15$. Setting $p' = 0.25$, the power = 0.94 for $T_\alpha = T_L$.

2.2.3 Comments

The example includes 6 univariate tests, so use of a lower value of α might be preferred to allow for the increased possibilities of type I errors in multiple tests. A simple, if conservative, way to achieve this is with the Bonferroni inequality (Prins 2006). It instructs us to set $\alpha_g = \alpha/G$ when conducting $g = 1, ..., G$ tests. In this case, $\alpha_g = 0.0083$ implying that the downward trend for the 6 year-olds ($p = 0.033$, Table 3) cannot be accepted confidently as one of a group of equally important univariate tests. Of course, in practical terms, the age groups may not be equally important, and increasing α may be more sensible than neglecting a possible downward trend.

Binomial tests can also be applied to assess compliance with an ecological quality objective set as a quantile other than the median, e.g. as the 0.9 quantile in order to protect against the occurrence of extreme values of an indicator. Larger sample sizes may then be necessary to find statistical significance (Cotter 1985). Compliance testing with multiple objectives set as quantiles is further discussed by Cotter (1994).

Table 3. Binomial test of the null hypothesis, H = trend $\geqslant 0$ relative to the median value for the cod indices-at-age (Table 2). Only those from the first half of the observation period for each age are shown. 0 indicates index $\leqslant$ median; 1: index > median. T_L = number of observed values > median value and to left of median time; N_L = number of observed values to left of median time.

Year	Age1	Age2	Age3	Age4	Age5	Age6
1976	0	1				
1977	1	0				
1978	1	1				
1979	1	1				
1980	1	1				
1981	0	1				
1982	1	0				
1983	0	1	0	1	1	1
1984	1	0	1	0	1	1
1985	0	1	1	1	0	1
1986	1	0	1	1	1	1
1987	1	1	0	1	1	1
1988	0	0	1	0	1	1
1989	1	0	1	1	0	1
1990	0	1	0	1	1	1
1991			1	0	0	1
1992			0	1	0	0
1993			0	0	1	0
T_L	9	9	6	7	7	9
N_L	15	15	11	11	11	11
Pr(T_L or more|H)	0.304	0.304	0.500	0.274	0.274	0.033

2.3 Runs test

A "run" is defined as any sequence of 1 or more like elements drawn from two classes. In the present context, this could mean above or below a level line such as a median, or a given trend line. The runs test assesses whether the observed number of runs is likely from theory, assuming that individual values fall to either side of the line independently and with known probabilities. Non-randomness is usually represented by positive serial correlation, i.e. fewer than the expected number of runs, hence the test is usually one-sided. Runs relative to a fitted trend line would arise from serial correlation of the residuals and could indicate whether the fitted trend is a poor model of the signal. Brownlee (1965, Sect. 6.3) explains the formulae for the runs test. Those below refer to the simplest, most useful case when runs relative to the median are considered. Runs around a trend line can be analysed by treating it as the median at each observation time.

2.3.1 Method

1. Theory: Let the two classes of elements be a or b for "above" and "below" the sample median, respectively. Any values tied with the median should be omitted from consideration. Let m be the number of a's or b's, whichever is the least. The expectation of the number of runs, u, is

$$E(u) = 1 + m$$

and the variance is $V(u) = \frac{m(m-1)}{2m-1}$.

Tables for looking up the significance of the test statistic, u, are available (Swed and Eisenhart 1943; Conover 1971) but for series of reasonable length, say > 10 observations, it is easier, and justifiable under the Central Limit Theorem, to assume that the transformed statistic

$$\frac{u + 0.5 - (1 + m)}{\sqrt{m(m-1)/(2m-1)}}$$

is approximately a standard normal variate, z. The 0.5 in the numerator corrects for continuity.

2. Find the sample median value. Mark each observed value with a 1 if greater than the median, or with a 0 if less.
3. Count the runs of one or more similar marks. This is u.
4. For a one-sided test, compare u with tabulated values, or standardise it and compare with z_α. If $\Pr(u) < \alpha$, serial correlation around the median is detected.

2.3.2 Example

The runs test is illustrated using the cod indices for 1 and 2 year-olds from 1976 to 2004, and for 3 to 6 year-olds from 1983 to 2004. The numbers of runs for each age can be counted from the sequences of 0 and 1 much as they appear in Table 3 but using the whole series, not half of it, and after removing any zeros that represent values tied with the median. The results are shown in Table 4. The presence of tied values prevented the numbers of a's and b's from being equal in some age groups; m was therefore adjusted to be the minimum of the two numbers. The probabilities that the series were random, shown at the bottom of Table 4, indicates that only the 6 year-olds were non-random by this test. They showed a downward trend.

2.3.3 Comments

The runs test is sensitive specifically to sequences of observations on one side or the other of the overall median. Long runs of this type could appear as a trend in a short observation window. Conversely, all trends are expected to cycle around a median if the observation window is sufficiently long (since few natural variables rise or fall eternally). For this reason, runs are a type of trend for which the scale of viewing is important.

2.4 Mann-Kendall's K

Kendall's tau is used to measure concordance or correlation between two variables (Hollander and Wolfe 1973), one of which can be time so that trends may be assessed (Mann 1945). Each of the N observed values is paired with every value observed after it and the pair scored 1 if the first is less than the second, 0 if they are tied, and −1 if the first is greater than the second. The Mann-Kendall test statistic, K, is the sum of these values. As a result of this scoring, Kendall's method is considered to be sensitive to monotonic trends (Esterby 1993) in which every value is less (for a declining trend) than the previous one.

Table 4. Runs test applied to age groups 1 to 6 from 1983 to 2004 for the cod indices (Table 2) and scored as above or below the median; indices tied with the median were omitted. Please refer to text for terms and notation.

	Age1	Age2	Age3	Age4	Age5	Age6
no. runs (u)	14	18	13	16	8	2
no. > median (m)	14	13	11	11	8	9
no. $\leqslant$ median (n)	14	14	11	11	10	13
min(m,n)	14	13	11	11	8	9
$E(u)$	15	14	12	12	9	10
$V(u)$	6.74	6.24	5.24	5.24	3.73	4.24
z	–0.19	1.80	0.66	1.97	–0.26	–3.64
Probability	0.42	0.96	0.74	0.98	0.40	< 0.001

2.4.1 Method

1. Theory. Assuming that a negative trend is of interest, H is: "trend is not monotonic, negative" which is usually taken to mean the more general: "trend $\geqslant 0$". A is: "trend is monotonic, negative". K takes on large negative values when such a monotonic trend is present. One-tail probabilities of observing K under H are tabulated by Hollander and Wolfe (1973, Appendix A.21) but, for large N, $K_* = K / \sqrt{[N(N-1)(2N+5)/18]}$ is distributed as a standard normal variate if no data are tied. If there are ties, the square root denominator representing the standard error of K has to be inflated (Hollander and Wolfe 1973, p. 187). According to a citation in Yu et al. (1993), 10 observations are adequate for "large sample".

2. Form an $(N-1) \times (N-1)$ triangular matrix. Starting with the first observed value, fill the first column with 1s, 0s, or –1s to indicate whether each subsequent value is greater than, equal to, or less than the first. Repeat, starting with the second value and filling in the second column, and so on.

3. Sum all values in the triangular matrix. This is K. Compare with tabulated values, or compare $K*$ with z_α to carry out the one-tailed test.

2.4.2 Example

Scoring of observations for Kendall's K is illustrated in Table 5 for the age 1 cod indices from 1976 to 2004. The full triangular matrix has 28 columns, so only 3 years of scores are shown. The sum of all the scores for the period from 1976 to 2004 was $K = -106$ with $N = 29$, giving a large-sample standard normal approximation of $K* = -1.988$. The corresponding one-sided probability of observing this or a more extreme value given H is true is 0.023, less than the usual type I error probability, $\alpha = 0.05$, implying that a monotonic, downwardly trending signal was detected for the 1-year olds.

2.4.3 Comment

The Mann-Kendall test found a trend for age group 1 when tests (1) and (2) based on the binomial distribution did not. The Mann-Kendall test took account of the gradual lowering of values over time, whereas the binomial tests only look at variations around a median without regard to their sequence.

Table 5. Kendall's Tau test applied to age group 1 from 1976 to 2004 for the cod indices (Table 2) illustrating how the sequence of observed indices is scored for 1976 with subsequent years (Col. 1), and similarly for 1977 and 1978 (Cols. 2 and 3). X denotes observed value, u and v are years.

Year	Sign($x(v) - x(u)$), $u < k$		
	1976	1977	1978
1977	1		
1978	1	–1	
1979	1	–1	–1
1980	1	–1	1
1981	–1	–1	–1
1982	1	–1	–1
1983	–1	–1	–1
1984	1	–1	1
1985	–1	–1	–1
1986	1	–1	1
1987	1	–1	–1
1988	–1	–1	–1
1989	1	–1	1
1990	–1	–1	–1
1991	–1	–1	–1
1992	1	–1	1
1993	1	–1	–1
1994	1	–1	1
1995	1	–1	–1
1996	–1	–1	–1
1997	1	1	1
1998	–1	–1	–1
1999	–1	–1	–1
2000	–1	–1	–1
2001	–1	–1	–1
2002	–1	–1	–1
2003	–1	–1	–1
2004	–1	–1	–1

Looking at the original data in Table 2 and Figure 1a, a monotonic negative trend seems credible for age 1 provided that one can accept that the signal:noise ratio was very low.

Significant results with the Mann-Kendall test naturally prompt the fitting of a line for the slope. Special methods exist for monotonic trends (Barlow et al. 1972; Hussian 2005).

2.5 Spearman's ρ

Spearman's ρ is the product-moment correlation between the ranks of paired data. To test for trend, one member of the

pair is the time of observation, the other is the observed variable. This is also known as the Hotelling-Pabst test. It appears to be sensitive to monotonic trends, like the Mann-Kendall test.

2.5.1 Method

1. Order the N times, y, of observation and rank the observed values.
2. Find the test statistic, $T = \sum_{y=1}^{N} \left(R\left(X_y\right) - y \right)^2$, where $R\left(X_y\right)$ is the rank of the corresponding observation (Lettenmaier 1976). T is small when $R\left(X_y\right)$ and y are positively correlated, and large when they are negatively correlated.
3. Conover (1971, p. 389) gives quantiles for T for series of up to 30 observations. Alternatively, use

$$ w_p \approx \frac{1}{6}\left(N\left(N^2 - 1\right) + \frac{z_p\left(N\left(N^2 - 1\right)\right)}{\sqrt{N-1}} \right) $$

where z_p is the p'th quantile of a standard normal variate. T should be less than w_p (and z_p on the negative side of the standard normal distribution) for an upward trend, i.e. positive correlation with time, and T should be greater than w_p (and z_p on the positive side of the normal distribution) for a downward trend, i.e. negative correlation (Conover 1971).

2.5.2 Example

Calculation of T for Spearman's ρ is part-illustrated for the age 1 cod indices from 1976 to 2004 (Table 6). $T = 5586$ which is greater than $w_p = 5564$ with $z_{0.975} = 1.96$, implying that the downward trend for 1-year olds is significant at $\alpha = 0.025$.

2.5.3 Comments

Spearman's test gave a null probability of 0.025 which is comparable with 0.023 obtained with Mann-Kendall's K. Spearman's ρ works on ranks and therefore appears to provide a similar test to K which works with the relative magnitudes of succeeding values. Of the two, Spearman's rho is much easier to calculate with a spreadsheet. Though similar, they are thought not to be mathematically equivalent, e.g. see Hollander and Wolfe (1973, p. 191), and different authors use different tests so both have been included here for completeness.

2.6 Jonckheere's test

Jonckheere's test is a nonparametric version of a one-way analysis of variance with unequal sample sizes, except that it tests H: "no treatment effect" versus the special alternative, A: "the treatments are ordered in effect" (Hollander and Wolfe 1973, p. 120). This can be applied to trends by equating observation times to treatments and then arranging them in the order implied by the possible trend so as to make a one-sided

Table 6. Spearman's rho test applied to age group 1 cod indices from 1976 to 2004 (Table 2) illustrating the ranking (R) of years (y) and observed indices (X), and the computation of T.

Year, y	Age 1 index, X	$R(y)$	$R(X)$	$(R(X) - R(y))^2$
1976	7.9	1	15	196
1977	36.7	2	28	676
1978	12.9	3	21	324
1979	9.9	4	19	225
1980	16.9	5	26	441
1981	2.9	6	7	1
1982	9.2	7	17	100
1983	3.9	8	11	9
1984	15.2	9	25	256
1985	0.9	10	2	64
1986	17	11	27	256
1987	8.8	12	16	16
1988	3.6	13	10	9
1989	13.1	14	23	81
1990	3.4	15	8	49
1991	2.4	16	4	144
1992	13	17	22	25
1993	12.7	18	20	4
1994	14.8	19	24	25
1995	9.7	20	18	4
1996	3.5	21	9	144
1997	40	22	29	49
1998	2.7	23	5	324
1999	2.1	24	3	441
2000	6.6	25	12	169
2001	2.8	26	6	400
2002	7.8	27	14	169
2003	0.6	28	1	729
2004	7.54	29	13	256
Sum, T				5586

test. For monotonic trends, this would be the existing time ordering. Hypothesis A is then equivalent to a monotonic trend, as for Mann-Kendall's K. Jonckheere's test is applicable when observations are independently replicated at some time points.

2.6.1 Method

Find $k(k-1)/2$ Mann-Whitney counts U_{uv} where u and v index the years of observation and $1 \leqslant u < v \leqslant k$:

$$ U_{uv} = \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} \phi\left(X_{iu}, X_{jv}\right). $$

n_u, n_v are the numbers of observations at times u and v respectively, while $\phi(a,b) = 1$ if $a < b$, and 0 otherwise. The test statistic, $J = \sum_{u<v}^{k} U_{uv}$, can be compared with Table A.8 in Hollander and Wolfe (1973), or, when the minimum number of replicate observations at any point is "large", can be transformed to $J*$ which is approximately normally distributed:

$$ J* = \frac{J - \left\{\left(N^2 - \sum_{j=1}^{k} n_j^2\right)\big/4\right\}}{\sqrt{\left(N^2(2N+3) - \sum_{j=1}^{k} n_j^2\left(2n_j + 3\right)\right)\big/72}}. $$

Here $N = \sum_{j=1}^{k} n_j$. This is a one-tailed test. H is rejected if $J*$ exceeds z_α.

Table 7. Jonckheere's test applied to cod indices for age 1 from 1976 to 2004 (Table 2) and arbitrarily assigned to a shortened series of 10 years with variable replication annually for illustrative purposes. a) resulting test data and computation of terms for the test statistic; b) illustrating scoring for $u = 1976$ and $v = 1977, 1978$.

a) Resulting test data and computation of terms for the test statistic

Year	Dummy replicate observations					n_j	n_j^2	$n_j^2 * (2n_j + 3)$
1976	7.9	17	3.5			3	9	81
1977	36.7	8.8				2	4	28
1978	12.9	3.6	2.7			3	9	81
1979	9.9	13.1	2.1	40	7.8	5	25	325
1980	16.9	3.4	6.6			3	9	81
1981	2.9	2.4				2	4	28
1982	9.2	13				2	4	28
1983	3.9	12.7	0.6	2.8		4	16	176
1984	15.2	14.8	7.537			3	9	81
1985	0.9	9.7				2	4	28
					Sum=>	29	93	937

b)

$u = 1976$	Replicate 1	Replicate 2	Replicate 3
	$v = 1977$		
Replicate 1	1	1	1
Replicate 2	1	0	1
	$v = 1978$		
Replicate 1	1	0	1
Replicate 2	0	0	1
Replicate 3	0	0	0

2.6.2 Example

A fabricated set of data illustrates computations for Jonckheere's test (Table 7a). For this, the cod indices for 1-year olds were arbitrarily re-assigned to a shortened series of 10 years so that irregular numbers of replicate observations were present. Table 7b illustrates scoring for three arbitrarily chosen years, 1976-78. J was found to be 209 which together with the totals in the right 3 columns of Table 7a gave $J* = 0.8318$ which is less than 1.645, the standard normal deviate corresponding to 95% of the area under the normal curve. Not surprisingly, the arbitrarily re-arranged data did not show a significant monotonic trend.

2.6.3 Comment

Applying Jonckeere's test when k is large is very time-consuming when using a spreadsheet.

2.7 Thiel's or Sen's slope estimator

Thiel's slope estimator is used for a notional linear trend:

$$X_i = \alpha + \beta y_i + e_i, \quad i = 1, ..., N.$$

The e's must be mutually independent and from the same continuous population (Hollander and Wolfe 1973). Thiel's estimator for β is similar in construction to that for Mann-Kendall's K. Sen (1968a) generalised Thiel's estimator to allow for multiple observations at one time point, and examined statistical properties.

2.7.1 Method

1. Let u and v index the years of observation. Pair every value, X_i, observed at time u_i, with every value, X_j observed subsequently at time v_j, i.e. for all $u_i < v_j$. Calculate the slopes, $S_{ij} = (X_j - X_i)/(v_j - u_i)$. Thiel's estimator is the median of these values.
2. Small sample confidence limits are available using Hollander and Wolfe (1973, Chapter 9 and Table A.21). For large samples, find the rounded integer value of

$$C_\alpha \approx z_{\alpha/2} \left\{ \frac{N(N-1)(2N+5)}{18} \right\}^{0.5}.$$

3. The $1 - \alpha$ confidence interval is obtained from the $N' \leqslant \binom{N}{2}$ ranked slope values. Use $\{S_{(L)}, S_{(U)}\}$ where rank $L = (N' - C_\alpha)/2$ and rank $U = (N' + C_\alpha)/2$.

2.7.2 Example

Thiel's slope estimator and 95% confidence limits were calculated for the age 1 cod indices from 1976 to 2004. The median slope, -0.23, is drawn through the intersection of the median value of age 1 indices, 7.9 fish per hour, and the median observation time, 1990 (Fig. 2). Using $z_{\alpha/2} = 1.96$, $C_\alpha = 140$ was calculated. There were 29 observations, giving $N' = \frac{29!}{27!2!} = 406$ estimated slopes from which $L = 133$ and $U = 273$. From a list of ranked slopes, S_{ij}, lower and upper 95% confidence limits for the slope were -0.625 and 0.119. These limits include zero suggesting that a negative linear trend is not present.
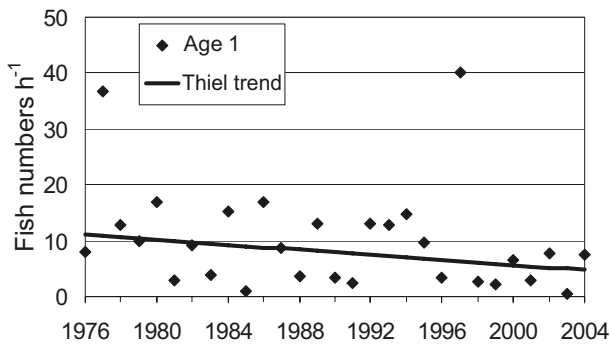
Fig. 2. Thiel's slope estimator for the age 1 cod indices, 1976 to 2004 (Table 2).

2.7.3 Comment

A truly linear trend is unlikely to exist for most environmental indicators but the fitting of a straight line may provide a useful summary of what is happening if the slope is clearly different from zero. Note, however, that the slope is likely to depend on the start and end times chosen for the fit, and that the assumption that all errors come from the same continuous distribution could be violated by process error. All the same, the Thiel slope is less likely to be affected by large outlying values than a least squares slope.

2.8 Cochran's Q test for similarity of multiple trends

Cochran's Q (Cochran 1950) tests H: "the probabilities of response are the same in different groups" against A: "they are detectably different". The test can be adapted to look at multiple time trends with the aim of avoiding the problem of multiple univariate tests mentioned in connection with binomial tests, above.

2.8.1 Method

1. Theory: let C_g be the column sums in the g'th time series, $g = 1, \ldots, G$, and $\bar{C}$ the mean of them. Let R_y be the y'th row sum, $y = 1, \ldots, N$. Then, Cochran's statistic is defined as

$$Q = \frac{G\,(G-1)\sum_g \left(C_g - \bar{C}\right)^2}{G\sum_y R_y - \sum_y R_y^2}.$$

2. Mark each observed value with a 1 if it is consistent with a downward trend, i.e. it is above the median value and located on or before the median time, or it is less than or equal to the median value and after the median time. Otherwise it is marked with a 0.
3. Find the column sums and means (for C), and the row sums and sums squared (for R). Calculate Q.
4. Q is distributed as χ^2 with $(G-1)$ degrees of freedom (Brownlee 1965, Sect. 7.10) under H.

Table 8. Cochran's Q test applied to cod indices, ages 1 to 6 from 1983 to 2004 (Table 2) illustrating how observed values are marked. 0 means index $\leqslant$ median and 1 means index > median if year $\leqslant$ 1993, and vice versa if year > 1993; 1993 is the median year.

Year	Age1	Age2	Age3	Age4	Age5	Age6
1983	0	1	0	1	1	1
1984	1	0	1	0	1	1
1985	0	1	1	1	0	1
1986	1	0	1	1	1	1
1987	1	1	0	1	1	1
1988	0	0	1	0	1	1
1989	1	0	1	1	0	1
1990	0	1	0	1	1	1
1991	0	0	1	0	0	1
1992	1	0	0	1	0	0
1993	1	1	0	0	1	0
1994	0	1	0	1	1	1
1995	0	0	0	0	1	1
1996	1	1	0	1	0	1
1997	0	1	1	0	1	1
1998	1	0	1	1	1	1
1999	1	1	0	1	1	1
2000	1	1	1	0	1	1
2001	1	0	1	1	1	1
2002	1	1	0	1	1	1
2003	1	1	1	0	1	1
2004	1	1	1	1	1	1

2.8.2 Example

Table 8 shows the cod indices for 1 to 6 year-olds from 1983 to 2004 scored 1 or 0 according to the consistency of each value with a downward trend relative to the overall median. From calculations, $Q = 3.56$ which is much less than $\chi^2\,(5) = 11.07$ implying that the different age groups are not showing detectably different trends, supposing (from some other test) that a general downward trend exists.

2.8.3 Comments

Note that Q is not sensitive to the total number of responses (since many of the R_y may be zero) hence, for our purposes, it does not, by itself, establish whether or not an overall trend is present in any of the groups. Cochran's Q appears to have similarities with Friedman's rank test for blocked data (Brownlee 1965), and with van Belle and Hughes' test for homogeneity of seasonal trend (Van Belle and Hughes 1984). El-Shaarawi (1993) suggests ways of extending the latter method to testing the notions that linear or quadratic patterns exist in the seasonal trends.

2.9 Aligned rank test

The aligned rank test is applicable when observations can be considered to be "blocked" and there are no missing values. As used in monthly water quality studies, *years* are considered as an ordered treatment replicated within months as

blocks (van Belle and Hughes 1984b). However, the test appears to be more generally applicable to the case when the *block* factor is a non-random treatment provided that it has a constant, linear effect from year to year, i.e. there is no *year × treatment* interaction. The treatments must all be expected to respond with the same direction of trend, up or down, over time. Van Belle and Hughes (1984b) provide an accessible account of the aligned rank test. It is not difficult to carry out as is summarised below. See also Sen (1968b) for the mathematical side, and Farrell (1980) who adapted the method for environmental work.

2.9.1 Method

1. Let $y = 1 \ldots N$ index years of observation, and $g = 1 \ldots G$ index observed variables, X. Transform them to $Y_{yg} = X_{yg} - \sum_y X_{yg} / N$. So if, for example, there is a joint trend downwards, these transformed values are expected to change from positive to negative over the time series for every variable. Next, all the Y_{yg} are ranked from 1 to GN giving the ranks, R_{yg}, with row and column averages $\bar{R}_y$ and $\bar{R}_g$, respectively.

2. Calculate

$$T = \left(\frac{12 G^2}{N(N+1) \sum_y \sum_g \left(R_{yg} - \bar{R}_g \right)^2} \right)^{1/2} \left(\sum_y \left(y - \frac{N+1}{2} \right) \right.$$
$$\left. \times \left(\bar{R}_y - \frac{NG+1}{2} \right) \right)$$

3. The probability of observing a magnitude $|T|$ or greater is found by direct comparison with the standard normal distribution. The normal approximation is reportedly good for all but the smallest number of observations (Van Belle and Hughes 1984a).

2.9.2 Example

The method was applied to the cod indices from 1983 to 2004 for 1 to 6 year-olds. Table 9 shows the ranks, R_{yg}, of the computed Y_{yg}, and their row and column averages, $\bar{R}_y$ and $\bar{R}_g$. Using these, $T = -4.34$ which can be compared with a standard normal deviate to give a one-tailed $p < 0.0001$. This implies that the downward trend in values of $\bar{R}_y$ represented a common trend for all year classes. A glance at Figure 1 indicates the feasibility of this result.

2.9.3 Comments

Other nonparametric tests can be applied to the 2-way blocked data structure suitable for the aligned rank test, e.g. Kendall's coefficient of concordance (related to Kendall's τ) or Friedman's test (Conover 1971). Van Belle and Hughes (1984) point out that such tests rely on separate rankings within each block. They therefore cannot take account of different relative sizes of *year* effects in different blocks which the aligned

Table 9. Aligned ranks found for the cod indices, ages 1 to 6 from 1983 to 2004 (Table 2), together with row and column averages needed for the test.

Year	Age1	Age2	Age3	Age4	Age5	Age6	$\bar{R}_y$
1983	19	125	43	106	91	108	82
1984	124	27	105	63	99	92	85
1985	4	126	94	103	78	97	84
1986	127	8	117	112	107	101	95
1987	87	131	30	103	84	92	88
1988	18	20	115	44	96	101	66
1989	120	21	114	112	65	97	88
1990	16	122	33	76	99	88	72
1991	9	13	90	53	65	88	53
1992	119	15	26	76	56	72	61
1993	118	128	33	48	84	58	78
1994	123	14	62	53	78	72	67
1995	110	129	47	82	56	50	79
1996	17	27	116	48	84	58	58
1997	132	22	38	82	65	58	66
1998	11	130	33	63	78	58	62
1999	6	2	121	53	78	72	55
2000	25	10	23	109	65	72	51
2001	12	32	31	39	65	50	38
2002	37	7	111	41	45	46	48
2003	3	5	24	95	65	50	40
2004	36	1	29	40	71	42	37
$\bar{R}_g$	58	51	65	73	76	74	

rank test, by ranking across blocks, does achieve. Judging from the fairly smooth and observable trend downwards in $\bar{R}_y$ in Table 9, the aligned rank test smooths out variability well and can effectively identify joint trends despite a high degree of year-to-year variability in the data.

A significant aligned rank test does not guarantee that all variables are responding in the same direction. If a visual examination of plotted time series leaves doubts about whether variables are changing together or not, a Cochran's Q test for similarity of trend, possibly on subsets of the variables showing contrasting behaviour, might cast light on what is happening.

2.10 Dietz and Killeen test for multivariate monotone trend

Dietz and Killeen (1981) derive a formula for the covariance matrix of Mann-Kendall statistics estimated from a multivariate monitoring programme, and a test statistic based on it that is asymptotically distributed as χ^2. H is: "the measured vectors of variables are randomly ordered in time" and A is: "a monotone trend exists for at least one of the variables". An advantage of the Dietz-Killeen method is that it is unaffected by the directions of the individual univariate trends, up or down. Use of multivariate methods "controls the overall significance level" and there were very few cases in a study of lake water quality "where the univariate methods perform better" (Loftis et al. 1991b). Dietz and Killeen (1981) also note that their multivariate test was able to reject H when separate univariate

Table 10. Dietz-Killeen test applied to age groups 1 to 6 from 1983 to 2004 for the cod indices (Table 2). Results are part of the output from Mvar.trend.r.

Age group	1	2	3	4	5	6
			Mann–Kendall statistics, K			
1	–45	–2	12	32	11	22
2	–2	–72	–31	27	34	27
3	12	–31	–60	–13	52	66
4	32	27	–13	–74	16	91
5	11	34	52	16	–97	95
6	22	27	66	91	95	–166
			Covariances for K			
1	1258	29	109	249	62	140
2	29	1257	–155	206	309	232
3	109	–155	1254	–90	394	475
4	249	206	–90	1248	107	662
5	62	309	394	107	1215	725
6	140	232	475	662	725	1233
Standard error [1]	35.5	35.4	35.4	35.3	34.9	35.1
Standardised K [2]	–1.27	–2.03	–1.69	–2.09	–2.78	–4.73
			Spearman's rank correlations			
1	1	0.025	0.089	0.202	0.049	0.112
2	0.025	0.999	–0.122	0.167	0.252	0.189
3	0.089	–0.122	0.998	–0.073	0.319	0.384
4	0.202	0.167	–0.073	0.994	0.086	0.535
5	0.049	0.252	0.319	0.086	0.971	0.587
6	0.112	0.189	0.384	0.535	0.587	0.984

(1) Square roots of diagonal elements of covariances V.
(2) K/Standard error.

Mann-Kendall tests, with significance levels adjusted using the Bonferroni approach, were not.

2.10.1 Method

1. Compute the matrix, K, of Mann-Kendall statistics. Those along the diagonal are computed as for the univariate Mann-Kendall statistics; see test (4) above. Those off the diagonal are

$$K_{gh} = \sum_{u<v} sign\left[\left(X_{gv} - X_{gu}\right)(X_{hv} - X_{hu})\right] \text{ for } g \neq h$$

where g and h index the G variables; u and v index the years of observation; $sign(x) = 1, 0, -1$ for positive, zero, and negative x respectively; and summation is over all N years of observation for v, and over all $u < v$. K_{gh} is high and positive when both the X_g and X_h variables are showing monotonic trends in the same direction, and high and negative when in opposite directions.

2. Compute the covariance matrix, V, of these statistics, i.e. (after Loftis et al. 1991b)

$$V_{gh} = \begin{cases} N(N-1)(2N+5)/18 & \text{for } g = h \\ \left(t_{gh} + r_{gh}\right)/3 & g \neq h \end{cases}$$

where

$$t_{gh} = \sum_{u<v} sign\left[\left(X_{gv} - X_{gu}\right)(X_{hv} - X_{hu})\right] \text{ and}$$

$$r_{gh} = \sum_{u,v,w} sign\left[\left(X_{gv} - X_{gu}\right)(X_{hv} - X_{hw})\right]$$

and the triple summation for r_{gh} is over all years of observation for u, v, and w.

3. Compute the test statistic, $K^T V^{-1} K$, and compare with χ^2 with degrees of freedom (d.f.) the same as the number of time series; all time series contribute to d.f. since no parameters are estimated.

4. Compute the Spearman rank correlations and the standardised K for each variable so as to assist with the interpretation of a significant multivariate result.

2.10.2 Example

Table 10 shows selected output from a computer program (in R) applied to the cod indices, ages 1 to 6, for 1983 to 2004: the Mann-Kendall statistics (K); their covariances (V), and the Spearman rank correlations of the observed values. The highest correlations occur among the older age groups. Note that although values along the diagonal of a correlation matrix are normally 1, some here are less than 1 due to tied values within the time-series. The test statistic, $K^T V^{-1} K = 25.09$ is compared with χ^2 with 6 d.f. and found to be highly significant,

$p < 0.005$; $\chi^2 = 18.548$. For 1-sided tests, standardised K greater in magnitude than 1.64 are likely to be contributing to the significance of the multivariate result. The results suggest that all except 1-year olds are contributing. Note that this finding comes from a shorter data series than that analysed as an example for the Mann-Kendall statistic above, when it was found that the trend for 1-year olds was significant at $p = 0.023$.

2.10.3 Comments

Interpretation of a significant multivariate result with the Dietz and Killeen test may not be straightforward, particularly if the series are showing trends in different directions and the series are noisy.

A related alternative to the Dietz and Killeen test when the aggregate trend of several variables is in the same direction is given by Hirsch and Slack (1984). Comparisons of these and other multivariate nonparametric methods in a water quality context are given by Lettenmaier (1988), Loftis et al. (1991a,b) and Smith et al. (1993). Libiseller and Grimvall (2002) give theory for a so-called 'partial Mann-Kendall test' in which the performance of a univariate Mann-Kendall test is improved by M-K statistics for a set of co-variates using the covariance matrix for all of them. Alvo and Park (2002) propose a new test statistic for dealing with missing values in multivariate nonparametric methods.

The computer code in R (Mvar.trend.r) for the Dietz and Killeen test is available from http://www.ifremer.fr/drvecohal/fisboat/index.htm; it was trialled with data given by Dietz and Killeen and found to give the same results. About 2 min of computer time were required for analysing the example data. Times increase roughly in proportion to the product of N^3G.

3 Conclusion

Use of nonparametric methods in fishery assessment work can provide statistical backing for the existence or otherwise of trends in the signal lying behind a time series of values measured with relatively large errors, as is typically the case in fish survey work. They can be more objective than model-based approaches, particularly the fitting of linear least squares models without regard for the process errors caused by any non-linearity in the true signal. Since a short series of purely random numbers can often give rise to convincing-looking trends (Kendall 1976), particularly if the start and end points for the trend are chosen flexibly and the shape of the trend is unimportant, careful nonparametric analysis based on frugal assumptions holds promise for distinguishing measurement and process errors from real signals characterising the ecosystem.

The nonparametric approach also opens up a different way of thinking about changes in a fishery or an aquatic environment. With the model-based approach, the selected model serves as a single simplification of what the true signal might look like. With the nonparametric approach on the other hand, a trend is seen as having many possible forms and each type of test may reveal a different, specific feature. For example, there may be a change of level, a straight or curved slope,

or there may be all manner of more complicated variability which, when viewed through the available observation window with the given observed values, are classifiable in various simple ways depending on which test results are significant or not.

Therefore, when checking trends in a research context, application of an assortment of nonparametric methods is worth considering, depending on which of the specific purposes listed (Table 1) are potentially relevant to the situation being investigated. Furthermore, trying to understand why one method gives a significant result when another does not is likely to be an instructive exercise both biologically and statistically. All the same, researchers should interpret results cautiously because of the dangers of finding that at least one test is statistically significant when many have been applied to one set of data. Use of the Bonferroni inequality to adjust significance levels for multiple tests would sometimes be advisable though over-acceptance of H, i.e. high probabilities of type II errors, may be a consequence.

A different situation prevails when nonparametric methods are being applied formally to test for compliance with some environmental quality criterion, perhaps as part of an ecosystem approach to fishery management. The most relevant test to use and the type I and II error probabilities must be agreed in advance of seeing the results if objective decisions are to be supported. A disadvantage of several nonparametric methods is that formulae for the probabilities of type II errors are not readily available, but their lack of strong assumptions is a notable advantage in a regulatory context, especially the freedom to abandon all faith in a mathematical model of trend.

References

Alvo M., Park J., 2002, Multivariate non-parametric tests of trend when the data are incomplete. Stat. Prob. Lett. 57, 281–290.

Barlow R.E., Bartholomew D.J., Bremner J.M., Brunk H.D., 1972, Statistical inference under order restrictions. J. Wiley & Sons, London.

Bell C.B., Conquest L.L., Pyke R., Smith E.P., 1981, Some nonparametric statistics for monitoring water quality using benthic species counts. Environmetrics, Soc. Indust. Appl. Math. (SIAM) Philadelphia, 8, 100–120.

Berthouex P.M., Hau I., 1991, Difficulties related to using extreme percentiles for water quality regulations. Res. J. Water Pollut. Control Fed. 63, 873–879.

Brown B.M., 1983, Statistical uses of the spatial median. J. R. Stat. Soc. Ser. B, 45, 25–30.

Brown L.D., Cai T.T., DasGupta A., 2001, Interval estimation for a binomial proportion. Stat. Sci. 16, 101–133.

Brownlee K.A., 1965, Statistical theory and methodology in science and engineering. John Wiley & Sons, New York.

Burnham K.P., Anderson D.R., 2002, Model selection and multimodel inference. Springer, New York.

Cochran W.G., 1950, The comparison of percentages in matched samples. Biometrika 37, 256–266.

Conover W.J., 1971, Practical nonparametric statistics. John Wiley & Sons, New York.

Cotter A.J.R., 1985, Water quality surveys: a statistical method based on determinism, quantiles and the binomial distribution. Water Res. 19, 1179–1189.

Cotter A.J.R., 1994, Compliance testing of two or more water quality determinands using quantiles. Environmetrics 5, 29–45.

Dietz E.J., Killeen T.J., 1981, A nonparametric multivariate test for monotone trend with pharmaceutical applications. J. Am. Stat. Assoc. 76, 169–174.

Edgington E.S., 1995, Randomization tests. Marcel Dekker, Inc., New York.

El-Shaarawi A.H., 1993, Environmental monitoring, assessment and prediction of change. Environmetrics 4, 381–398.

El-Shaarawi A.H., Niculescu S.P., 1992, On Kendall's Tau as a test of trend in time series data. Environmetrics 3, 385–411.

Esterby S.R., 1993, Trend analysis methods for environmental data. Environmetrics 4, 459–481.

Esterby S.R., 1996, Review of methods for the detection and estimation of trends with emphasis on water quality applications. Hydrol. Process. 10, 127–149.

Farrell R., 1980, Methods for classifying changes in environmental conditions. Vector Research Inc., Ann Arbor, Michigan, Techn. Rep. VRF-EPA7.4-FR80-1.

Hirsch R.M., Slack J.R., 1984, A nonparametric test for seasonal data with serial dependence. Water Resour. Res. 20, 727–732.

Hollander M., Wolfe D.A., 1973, Nonparametric statistical methods. John Wiley & Sons, New York.

Hussian M., 2005, Monotonic and semiparametric regression for the detection of trends in environmental quality data. Department of Mathematics, Division of Statistics, Linköpings Universitet, SE-581 85 Linköping, Sweden. http://www.ida.liu.se/divisions/stat/research/engo/Hussianthesis.pdf

Kendall M.A., 1976, Time-series. Charles Griffin and Co. Ltd., London.

Lanzante J.R., 1996, Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde data. Internat. J. Climatol. 16, 1197–1226.

Lettenmaier D.P., 1976, Detection of trends in water quality data from records with dependent observations. Water Resour. Res. 12, 1037–1046.

Lettenmaier D.P., 1988, Multivariate nonparametric tests for trend in water quality. Water Resour. Bull. Am. Water Resour. Assoc. 24, 505–512.

Libiseller C., Grimvall A., 2002, Performance of partial Mann-Kendall tests for trend detection in the presence of covariates. Environmetrics 13, 71–84.

Loftis J.C., Taylor C.H., Chapman P.L., 1991a, Multivariate tests for trend in water quality. Water Resour. Res. 27, 1419–1429.

Loftis J.C., Taylor C.H., Newell A.D., Chapman P.L., 1991b, Multivariate trend testing of lake water quality. Water Resour. Bull. Am. Water Resour. Assoc. 27, 461–473.

Mann H.B., 1945, Nonparametric tests against trend. Econometrica 13, 245–259.

Pettitt A.N., 1979, A non-parametric approach to the change-point problem. Appl. Stat. 28, 126–135.

Prins J. (2006). "NIST/SEMATECH e-Handbook of Statistical Methods" from http://www.itl.nist.gov/div898/handbook/prc/section4/prc473.htm.

Sen P.K., 1968a, Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat. Assoc. 63, 1379–1389.

Sen P.K., 1968b, On a class of aligned rank order tests in two-way layouts. Ann. Math. Stat. 39, 1115–1124.

Smith E.P., Rheem S., Holtzman G.I., 1993, Multivariate assessment of trend in environmental variables. In: Patil G.P., Rao C.R. (Eds.) Multivariate environmental statistics. Amsterdam, Elsevier, pp. 491–507.

Swed F.S., Eisenhart C., 1943, Tables for testing randomness of grouping in a sequence of alternatives. Ann. Math. Stat. 14, 66–87.

Thomas L., Burnham K.P., Buckland S.T., 2004, Temporal inferences from distance sampling surveys. In: Buckland S.T., Andersen D.R., Burnham K.P., Laake J.L., Borchers D.L., Thomas L. (Eds.) Advanced distance sampling. Oxford University Press, pp. 71–107.

Van Belle G., Hughes J.P., 1984, Nonparametric tests for trend in water quality. Water Resour. Res. 20, 127–136.

Wood S.N., 2006, Generalized additive models; an introduction with R. Chapman & Hall/CRC, Boca Raton.

Yu Y.-S., Zou S., Whittemore D., 1993, Non-parametric trend analysis of water quality data of rivers in Kansas. J. Hydrol. 150, 61–80.