Water quality trend detection in the presence of changes in analytical laboratory protocols

David G Smith and P Bryce McCann

David G. Smith: Director of Aquatic Studies (Limnology, Hydrology, Pathogens, and Reservoir Modeling Programs) for New York City's Department of Environmental Protection. Prior to this was Senior Scientist at the National Institute of Water and Atmospheric Research, New Zealand: led the design of, and managed, the NZ National Rivers Monitoring Network; conducted research in optical water quality and co-authored a book on the subject. Current research interests: design of water quality monitoring networks, water quality trend analysis, and optical water quality management.

P. Bryce McCann: Scientist/Water Ecologist with the City of New York's Department of Environmental Protection, Division of Drinking Water Quality Control. Primary focus is the study of the pathogens *Cryptosporidium*, *Giardia* and enteric viruses. Work also includes field methods development and data analysis.

ABSTRACT: Changes in analytical procedures, and detection limits with sub-detection limit data censoring, can have profound effects on trend analysis and add tedium and uncertainty to the process.

Less accurate methods early in the record may produce not only high-biased data but also greater variance. The consequence of this is an artificial induction of a down-trend. In datasets with multiple procedure changes embracing relatively short time periods, there can be considerable difficulty in judging whether the data from adjacent time blocks can be coalesced to increase the length of the time period being analyzed and hence increase trend detection power. In the absence of paired comparisons between 'old' and 'new' procedures, combining adjacent data blocks adds uncertainty to any conclusions drawn.

The consequence of changing (usually lowering) detection limits with data censoring at the detection limits may induce artificial (usually down) trends. Data censoring may be of concern when using non-parametric statistics because multiple tied values may result, the consequence of which may be the paradoxical situation of significant trend but a zero slope.

For trend detection, it is preferable that there should be no analytical procedure changes without compelling reason, and networks should be designed with this in mind.

If a thorough assessment of the consequences of multiple analytical procedure changes and multiple detection limits with sub-detection limit data censoring is not made during data analysis, the consequence may be that the data analyst is merely *measuring changes in the monitoring program not the environment*.

INTRODUCTION

During the course of long-term data collection there can be many analytical method changes and improvements, and changing detection limits. What are the consequences of these changes for trend detection? This question is not new but there is little in the literature which addresses it using a long-term environmental dataset. Much of the literature uses simulated data.

However, there are some notable exceptions. Shapiro and Swain (1983) found that a much-reported decline in silica content in Lake Michigan was, in fact, due to changes in analytical method and as a consequence it was not safe to conclude anything about long-term trends. Alexander et al. (1993) corrected stream water quality trends for laboratory measurement bias using US Geological Survey (USGS) data and parametric statistics. They found an improvement in the accuracy of trend slope estimates but a significantly lower precision as reflected by a reduced number of detected trends after correction.

If there is a change in analytical procedure it may be possible to calibrate the old data with respect to the new method using both methods employed over a period of time. Newell et al. (1993) examined the consequences of using such overlapping analytical data, and compared the technique with that of blocking data (using data in discreet time blocks) before and after an analytical method change. They concluded that the power of the blocked test exceeded the power of the calibration approach only when the calibration error was extremely large. Newell and Morrison (1993) investigated method changes and found statistically significant differences for different filter types, instruments, and sampling techniques.

Aside from method changes, there are problems associated with analytical detection limits. And data censoring (i.e., not reporting values determined to be less than a "detection limit") is commonly carried out. The effects of data censoring have been discussed by, for example, Gilliom et al. (1984), Hughes and Millard (1988), Millard and Deverel (1988), Bell (1990), Helsel and Cohn (1988) and Porter et al. (1988). Bell (1990) showed that data censoring masked an upward trend in site pollution.

Data censoring can be a major concern in trend detection analysis, which is exacerbated if the detection limit changes over time (normally downwards). If, for a particular determinand, the detection limit changes over time from say, 10 gm^{-3} to 2 gm^{-3} part way through the data record, and the data are censored, a concentration of, say, 3 gm^{-3} initially reported as < 10 gm^{-3} would now be reported as 3 gm^{-3}. More importantly, if many of the data over the whole record are actually <2 gm^{-3}, the data will be reported as <10 and then <2 gm^{-3}, respectively. If the default is to set less-than values as half the detection limit (e.g., Ellis and Gilbert, 1980), we now have a record with a substantial number of equivalent values set at 5 gm^{-3} and then at 1 gm^{-3}. The result of this is the introduction of a step trend in the record producing a monotonic down-trend overall. Small step-trends may be very difficult to observe visually on the time plot of the data.

To avoid problems such as these, which may compromise trend detectability, Smith et al. (1996) cautioned that, once set up, the network (and this includes analytical methods) should remain unaltered. Of course, this necessitates that appropriate analytical methods are incorporated into the network at the outset. These will, in part, depend upon the objectives of the network, which must be very carefully considered in the design phase (e.g., Ward et al., 1990). Of course, over enough time, there will of necessity have to be some changes in, for instance, equipment, but it is essential to allow for such changes by, for example, paired method comparison so that historic data can be adjusted if required (see, for example Newell et al., 1993)

Here, we examine the consequences for trend detection of changing analytical procedures for three determinands (turbidity, ammoniacal-nitrogen (NH_x–N), and total phosphorus (TP)) at one river site in New York City's wide-ranging river and reservoir monitoring network, which started well over a decade ago. Trend detection was not the prime objective at that time and therefore not specifically designed for (and so, for example, the requirement incorporating unchanging analytical methods was not made). Several changes have been made to some of the methods and to detection limits (with data censoring) over the study period. We have used data obtained by the City's Department of Environmental Protection (DEP) to assess the consequences for trend detection of these changes over the period of record. Because many other agencies appear to be in a similar position, these findings are probably of much wider interest given the fact that there are many long-term records in existence and the perceived growing interest in trend analysis. It is the main purpose of this paper to highlight the issues referred to above; it is not the intent to solve some of the problems encountered.

The determinands chosen for this study are a selection from a larger study by McCann (1999) whose objective was the development of a relatively simple, easy-to-use, "boiler plate" protocol for trend analysis for the whole DEP river and reservoir network. This incorporates 130 stream sites and 86 reservoir main sites with various sampling depths, three laboratories, and a multiplicity of determinands.

APPROACH

The Neversink River in New York State's Catskill Park was selected for study (McCann, 1999). This site (coded NK7A —USGS site S29) has good flow and extensive water quality records, and the water is relatively pristine and perceived to be relatively unchanging over the record period.

Turbidity, NH_x–N, and TP data for this site were obtained from the New York City Department of Environmental Protection's extensive and well-maintained database. This database also records analytical methodology, with changes appended to each data request. However changes made to detection limits are not included in the record but can be obtained by examination of the raw data; values below detection limit are reported as "-X", where X is the detection limit. Such values were replaced by X/2 for the purposes of this study. There are no paired sets of overlapping data for any procedure changes.

The initial analytical thrust consisted of producing temporal plots of the data from which potential outliers were observed. In the event, all possible outliers were <u>in</u>cluded in the dataset subsequently used because there were no reasons to <u>ex</u>clude them. All outliers (i.e., 'high' values) were deemed the consequence of high flow.

Following acceptance of a usable record, the data were analyzed as indicated in Figure 1. Trend analysis over time using ordinary linear regression is not usually appropriate for water quality data because the required assumption of normally distributed residuals is so often violated. (Smith and Maasdam, 1994). Also, it fails to account for seasonal components of variability, so that statistical power to detect trends can be greatly diminished. Consequently, the non-parametric Seasonal Kendall and Mann-Kendall tests for trends were used, together with their associated Seasonal Kendall (SKSE) and Sen Slope Estimators (SSE), respectively, to determine trend slopes. However, for comparison purposes and because it is readily available in software and therefore commonly used, linear regression was included where appropriate. Flow-adjustment was also carried out and trend analysis performed for both raw and flow-adjusted data. The presence of seasonality was assessed using the Kruskal-Wallis test ($\alpha = 0.05$).

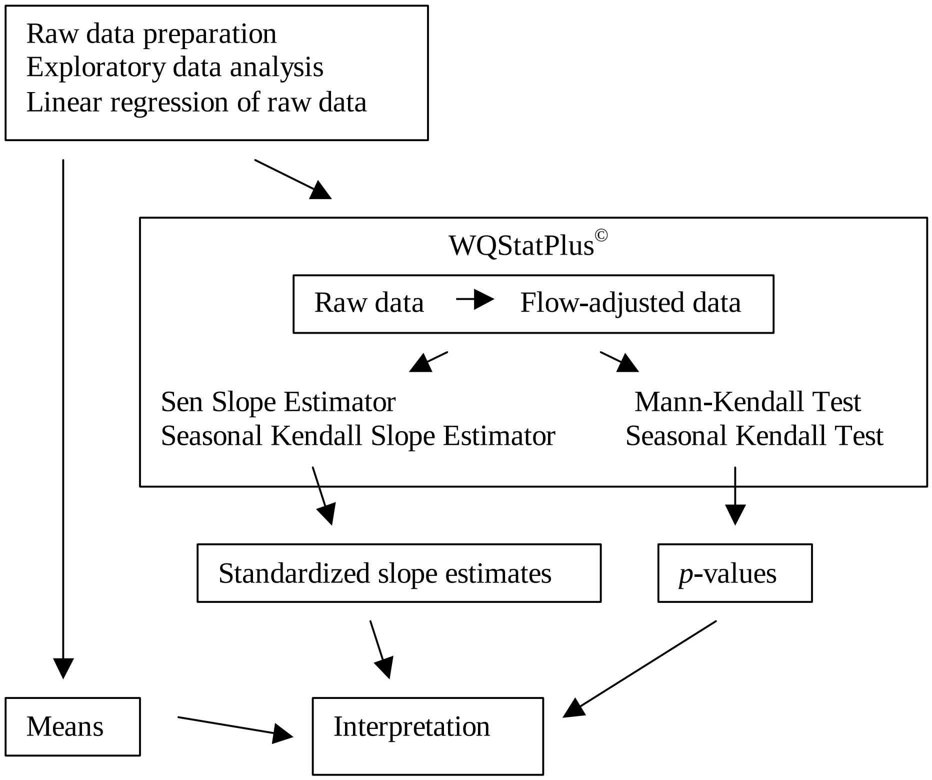

Figure 1. Flow diagram depicting temporal trend analysis protocol. "Standardized slope estimates" means simply the annual slope divided by the median of the raw data.

The software WQStatPlus© (obtainable from Intelligent Decision Technologies, Ltd., Longmont, Colorado) was used for the study. Its flow-adjustment technique is limited to linear regression of log/log transformed data, which is not always the most appropriate for water quality data (Smith et al, 1996); the preferred LOWESS smoother technique (Smith et al. 1996) is not included in the software. Using this software, it was not possible to conduct linear regression on the flow-adjusted data.

To put the trends into context, slope estimates were standardized by dividing the slope by the period median and expressing the result as an annual percentage change.

Flow data (mean daily flows) were obtained from the USGS.

Following Smith et al. (1996) we ignore the effects of serial correlation which we justify by confining our attention solely to the period of record and not considering changes in the underlying processes causing the trend (Loftis et al., 1991). We have used conventional hypothesis testing with associated *p* values whilst being mindful of the potential problems of this procedure (e.g., Goodman, 1993).

RESULTS AND DISCUSSION

Firstly, the flow data were plotted against time and the temporal trend assessed. There were no changes in flow measurement procedure over the study period which could potentially interfere with this assessment. There was no statistical significance (see Table 1) using all the trend assessment techniques although the slopes were of a magnitude (e.g., 1.2% per annum by the SKSE method–13% over the record) to suggest that a component of any trend in the determinands examined could be a consequence of flow.

Results for each determinand are presented as three graphs: (a) the raw data (including a notation when there were analytical procedure and detection limit changes); (b) the raw data plus trend lines for the whole record; (c) the raw data plus non-parametric trend lines for the dataset partitioned into appropriate time blocks. The term 'analytical procedure' changes includes method, instrument, filter type, autoanalyzer detector, and laboratory location changes.

WQStatPlus does not allow for export of flow-adjusted data so only raw data are plotted here. All plots were made using the Kaleidagraph™ software, and trend lines were translated from the WQStatPlus plots. For clarity, the SKSE slopes are the only non-parametric slopes presented in the figures; the SSE slopes were generally very similar. The SKSE slopes are placed on the graphs such that their mid points coincide with the period median value.

Turbidity (Figure 2)
Seasonality is apparent from the graph, supported by the Kruskall-Wallis test (p < 0.05), with lower values tending to appear in the summer period coincident with lower flows. There are six analytical procedure changes over the record period. In mid-1992 there is an apparent data 'discontinuity' which does not seem linked to reported procedural changes. This 'discontinuity' is also apparent in the flow-adjusted data. Because the turbidities reported are low, and the procedure itself is somewhat subjective, it is possible that this 'discontinuity' is linked to personnel changes in the laboratory at that time. But we cannot be certain. A visual assessment of Figure 2a suggested that the dataset should be divided into two approximately equal time blocks coinciding with this 'discontinuity'. There was no visual evidence to suggest that the data be further partitioned. Further statistical analysis to assist here is beyond the scope of this study. We acknowledge that the inclusion of many method changes may, to some extent, invalidate our analysis because we have no readily available assessment of the effects of these changes on data integrity.

Figure 2b shows the least squares fit plus the flow-adjusted (FA) SKSE slope for the full record. It is of interest to note that the although the least squares fit indicates a down-slope approximately twice that of the two non-parametric methods, it is classed as not significant (p > 0.1) whereas, in this instance where there is seasonality and a non-normal data distribution, the non-parametric slopes are highly significant (Table 1). This is because much of the variance has been removed by the deseasonalizing procedure and the higher values at the start of the record has had a much bigger effect on the least squares procedure than on the non-parametric test which uses ranks.

Figure 2c depicts the FA SKSE slopes for the partitioned dataset (with an end 6/92 partitioning point). A different picture now emerges although it should be borne in mind that the trend detection power is now smaller than for the full dataset and, on average, p-values will be higher. Whereas using the whole dataset revealed strong negative slopes over the full period of record, we now see two positive slopes with the FA SKSE slope for the 4/87-6/92 block being non-significant, whereas the slope for the 7/92-12/97 block is highly significant. In this instance, the two up-slopes have been masked by the effect of the down-stepping 'discontinuity' in the middle of the full time block. Further examination of the 7/92-12/97 dataset shows an additional potential 'discontinuity' in mid-1995 with no 0.1NTU values being obtained afterwards. At this point, we cannot say if this is due to some unknown laboratory effect or an environmental change. This latter block was further partitioned into two approximately equal, but very short, time blocks which showed non-significant down- and

Table 1. Summary statistics for trend analysis for raw (R) and flow-adjusted (FA) data. For each slope type, the slope (un % change/annum[1], and statistical significance[2] are presented.

Determinand (units)	Period	*n*	Data type	Linear regression slope			Sen Slope			Seasonal Kend	
				Annual	% change	Sig.	Annual	% change	Sig.[3]	Annual	% chang
Discharge (l/s)	4/87-12/97	278	R	+73	+2.4	ns	+32	+1.0	ns	+36	+1.2
Turbidity (NTU)	4/87-12/97	278	R	-0.073	-14.6	ns	-0.028	-5.6	***	-0.032	-6.8
			FA				-0.035	-6.9	***	-0.034	-6.4
	4/87-6/92	137	R	+0.46	+77	ns	0.000	0.0	ns	0.000	0.0
			FA				+0.020	+3.3	ns	+0.019	+3.2
	7/92-12/97	141	R	+0.042	+14.0	**	0.000	0.0	**	+0.025	+8.3
			FA				+0.024	+8.0	***	+0.027	+8.7
	7/92-5/95	76	R	+0.007	+2.6	ns	0.000	0.0	ns	0.000	0.0
			FA				-0.019	-7.6	ns	-0.011	-4.4
	6/95-12/97	65	R	-0.008	-2.1	ns	0.000	0.0	ns	0.000	0.0
			FA				-0.004	-1.0	ns	+0.009	+2.3
NH_x-N (g/m^3)	4/87-12/97	263	R	-0.003	-33.4	***	-0.001	-11.1	***	-0.001	-13.8
			FA				-0.001	-11.1	***	-0.001	-13.8
	6/88-12/97	233	R	-0.001	-16.6	***	-0.001	-16.6	***	-0.0006	-10.2
			FA				-0.001	-16.6	***	-0.0007	-11.3
	6/88-12/93	136	R	+0.0007	+11.7	ns	0.000	0.0	ns	0.000	0.0
			FA				0.000	0.0	ns	0.000	0.0
	2/95-12/97	73	R	+0.0004	+16.0	ns	0.000	0.0	ns	0.000	0.0
			FA				0.000	0.0	ns	0.000	0.2
TP (mg/m^3)	4/87-12/97	276	R	-0.58	-14.6	***	-0.296	-7.4	***	-0.326	-8.1
			FA				-0.299	-7.5	***	-0.322	-8.2
	4/87-10/89	66	R	-0.73	-11.8	ns	-0.652	-10.5	*	-0.984	-15.9
			FA				-0.658	-10.6	*	-1.019	-16.4
	11/89-12/97	210	R	-0.365	-9.7	ns	0.000	0.0	*	-0.144	-3.8
			FA				-0.147	-3.9	*	-0.162	-4.3
	11/89-8/93	100	R	+0.365	+9.1	ns	0.000	0.0	ns	0.000	0.0
			FA				+0.060	+1.5	ns	-0.093	-2.3
	9/93-12/97	110	R	-0.256	-8.5	ns	0.000	0.0	**	0.000	0.0
			FA				-0.099	-3.3	**	-0.117	-3.9

[1] % change/annum = Annual change x 100/period median.

[2] Statistical significances: ns = non significant , p∃0.1; * = significant, $p<0.1$; ** = significant, $p<0.05$; *** = significant, $p<0.01$. WQStatPlus tests at confidence levels up to 99% for the Mann-Kendall test, whereas it tests at confidence levels up to 95% for the

[3] Mann-Kendall test. [4] Seasonal Kendall test. [5] Kruskall Wallis test (α = 0.05).

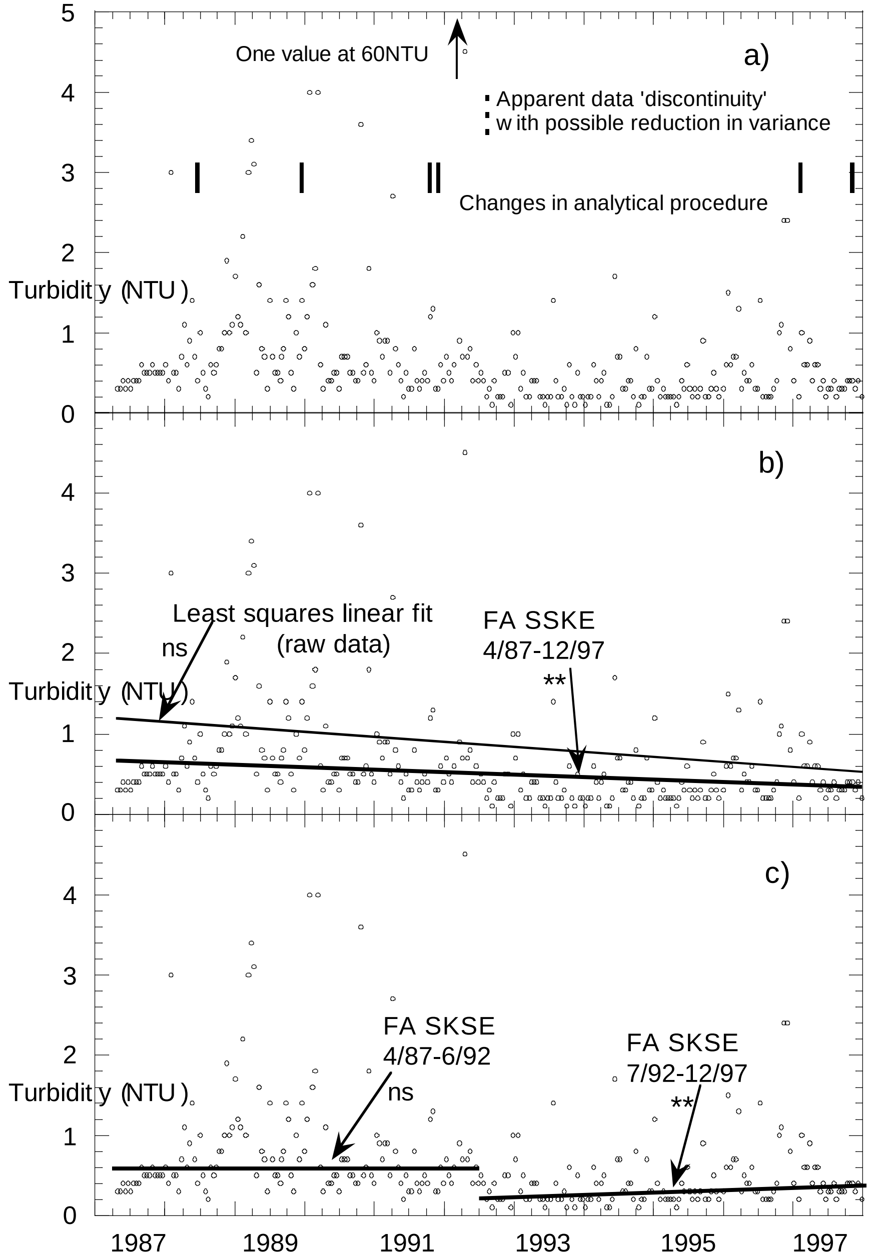

Figure 2. Trends in turbidity: a) raw data with analytical changes noted; b) raw data with overall trend lines; c) raw data for partitioned datasets. The non-parametric slopes drawn are the flow-adjusted Seasonal Kendall Slope Estimate (FA SKSE) with each slope's mid-point located at the period median. The trend test significances are marked adjacent to the arrow locating the slope. It was not possible to export the flow-adjusted data from the software.

up-slopes, respectively with the SKSE (Table 1). It is too early to judge the potential effects of the two procedural changes in 1997.

It should be noted that the log-log flow-adjustment produced a lower r^2 value than did the simple raw data linear regression, 0.17 vs 0.31, however it was not possible to perform a flow-adjustment in WQStatPlus using raw data.

The Mann-Kendall trend test yielded a highly significant trend for the raw data for the period 7/92-12/97 but the associated SSE produced a zero slope. This apparent paradox (a significant trend test but zero slope) was noted by Smith et al (1996) when examining trend statistics from the first five years of data from the New Zealand National Water Quality Monitoring Network. The explanation was addressed by McBride and Loftis (1994) who pointed out that the slope estimator is not very appropriate in the presence of many tied values (as can happen in the case of many "less than" values) because of the mechanics of the test procedures. Flow adjustment has presumably removed most of the tied values by making small and different adjustments to these values because of differing flows at sampling time.

Ammoniacal N (NH_x- N) (Figure 3)
The concentrations during the first 18 months are higher than the rest of the dataset (Figure 3a). This was not caused by flow issues (the flows in this early period were not unusual) but appear to be the consequence of the analytical method used at that time. No seasonality is apparent in the plot, confirmed by the Kruskall-Wallis test. There are seven analytical procedure changes over the record period and two changes in detection limit. In the period 6/88 to 12/93 there are many data having the same value. The lower value (0.005 g/m^3) is half the detection limit; the others are reported to 2 decimal places because of the requirements of the database during this period.

A visual assessment of Figure 3a suggested that the dataset should be divided into three periods with the first period (up until the end of 5/88) being discarded because the data appear unreliable when compared with the rest of the dataset. In 06/88 there was a major methodological change in ammonia analysis. Apart from the consequences of the change in detection limit, there is no visual evidence to suggest that the data be further partitioned. Again, further statistical analysis to assist here is beyond the scope of this study. At the beginning of 1994 and 1995 the changes in detection limit (from 0.01 to 0.002 and then to 0.005 g/m^3) are apparent from the record. These are noticeable in Figure 3a but more so in Figure 3c which has an expanded ordinate scale. The concentrations from early 1994 to the end of the record (12/97) are consistently lower than those preceding 1994 suggesting an improved laboratory technique.

Figure 3b depicts a trend analysis for the full record. In both cases there are highly significant and substantial down-trends. Flow-adjustment (for the non-parametric tests only) showed no substantial effect. This is not surprising given that the log-log plot is virtually horizontal (r^2 = 0.0006); for the raw data, r^2 = 0.01. The effects on slope of the 4/87-5/88 data are marked for the linear regression but less so for the non-parametric slopes (Table 1).

For the dataset minus the early probably unreliable data (6/88-12/97), there is an apparent (highly significant) downtrend (Figure 3c and Table 1). This is all accounted for by the detection limit changes noted; when the dataset is further partitioned (but omitting a period 1/94-1/95) we see two horizontal slopes. Both non-parametric tests (and least squares regression) indicate non-significance ($p > 0.1$).

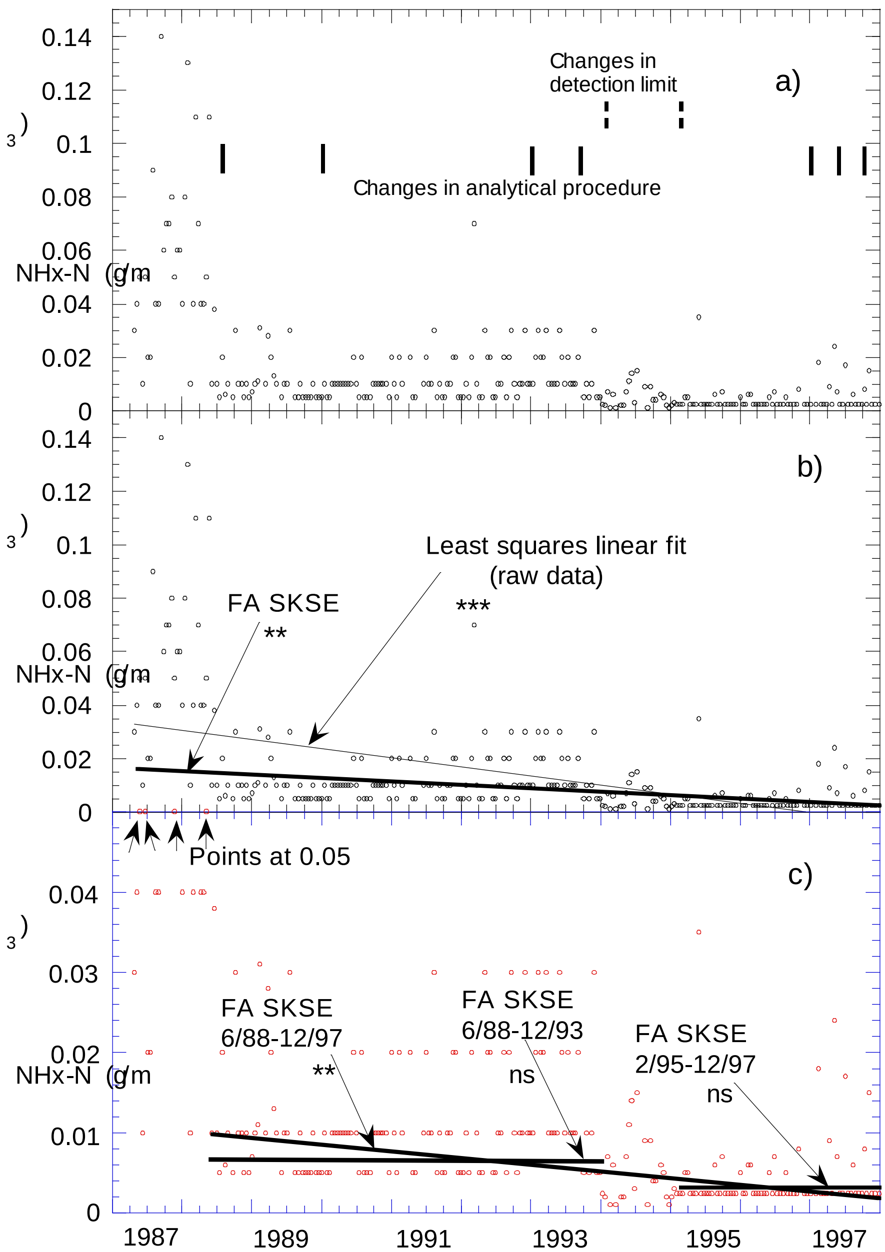

Figure 3. Trends in ammoniacal nitrogen (NH_x-N): a) raw data with analytical changes noted; b) raw data with overall trend lines; c) raw data for partitioned datasets (note scale change). The non-parametric slopes drawn are the flow-adjusted Seasonal Kendall Slope Estimate (FA SKSE) with each slope's mid-point located at the period median. The trend test significances are marked adjacent to the arrow locating the slope. The 2/95-12/97 FA SKSE line has been upwardly displaced from its correct position (virtually horizontal at 0.0025 g/m^3), for clarity. It was not possible to export the flow-adjusted data from the software.

Total Phosphorus (TP) (Figure 4)
Seasonality is apparent in Figure 4a, confirmed by the Kruskall-Wallis test. There are five analytical procedure changes over the record period and one change in detection limit. Major analytical equipment changes in late 1989 and 1993 appear to produce 'discontinuities' in the data. The change in detection limit is apparent in early 1988 (from 10 to 2 mg/m^3). A visual assessment of Figure 4a suggested that the dataset should be divided into four blocks coincident with the change in detection limit and the two analytical changes (at end 10/89 and beginning 9/93) which appear to cause discontinuities in the data. This would create two unacceptably short blocks in the early, already short, time period (4/87-11/89). As a consequence, the dataset was divided into three blocks, the first block incorporating the detection limit change to show a change in the very short time as a probable consequence. Apart from the obvious consequences of the change in detection limit, there was no evidence to suggest that the data be further partitioned. Further statistical analysis to assist here is beyond the scope of this study.

Figure 4b shows the analysis for the whole record. For both trend lines shown, there are highly significant and substantial down-trends. Flow-adjustment (for the parametric tests only) had no effect on the outcomes. As with NH_x-N, this is not surprising given that the log-log plots are virtually horizontally (r^2 = 0.0004). For the raw data, r^2 = 0.18; this would have been a preferable flow adjustment method, were it available within the software used.

In Figure 4c the dataset is partitioned into three blocks. The first block (4/87-10/89) also incorporated a major drop in detection limit and so, not surprisingly, a substantial down-trend was found. Test significance is only apparent for the non-parametric tests (Table 1).

For the two remaining blocks (11/89-8/93 and 9/93-12/97) combined, only the non-parametric tests show significance but with the SSE showing a paradoxical zero slope for the raw data. For the 11/89-8/93 block, there are no trends shown by any of the trend test techniques although the least squares analysis gives a substantial slope; the non-parametric slopes are zero for the raw data but small and different for the SSE (positive) and SKSE (negative) for the flow-adjusted data. For the second block (9/93-12/97), both non-parametric tests show significant trends with the paradoxical zero slope. The zero slopes disappear with flow-adjustment as was found for turbidity, as a probable consequence of the removal of multiple tied values. We now see slight, significant, negative slopes (3-4% per annum) in both cases. We note that for the FA SKSE there is a down-slope over the 11/89-12/97 block and down-slopes over the two partitioned blocks (11/89-8/93 and 9/93-12/97) lending support to a down-trend in stream TP since 11/89, although we cannot rule out a potential confounding influence because of the two procedural changes in the last year of the record.

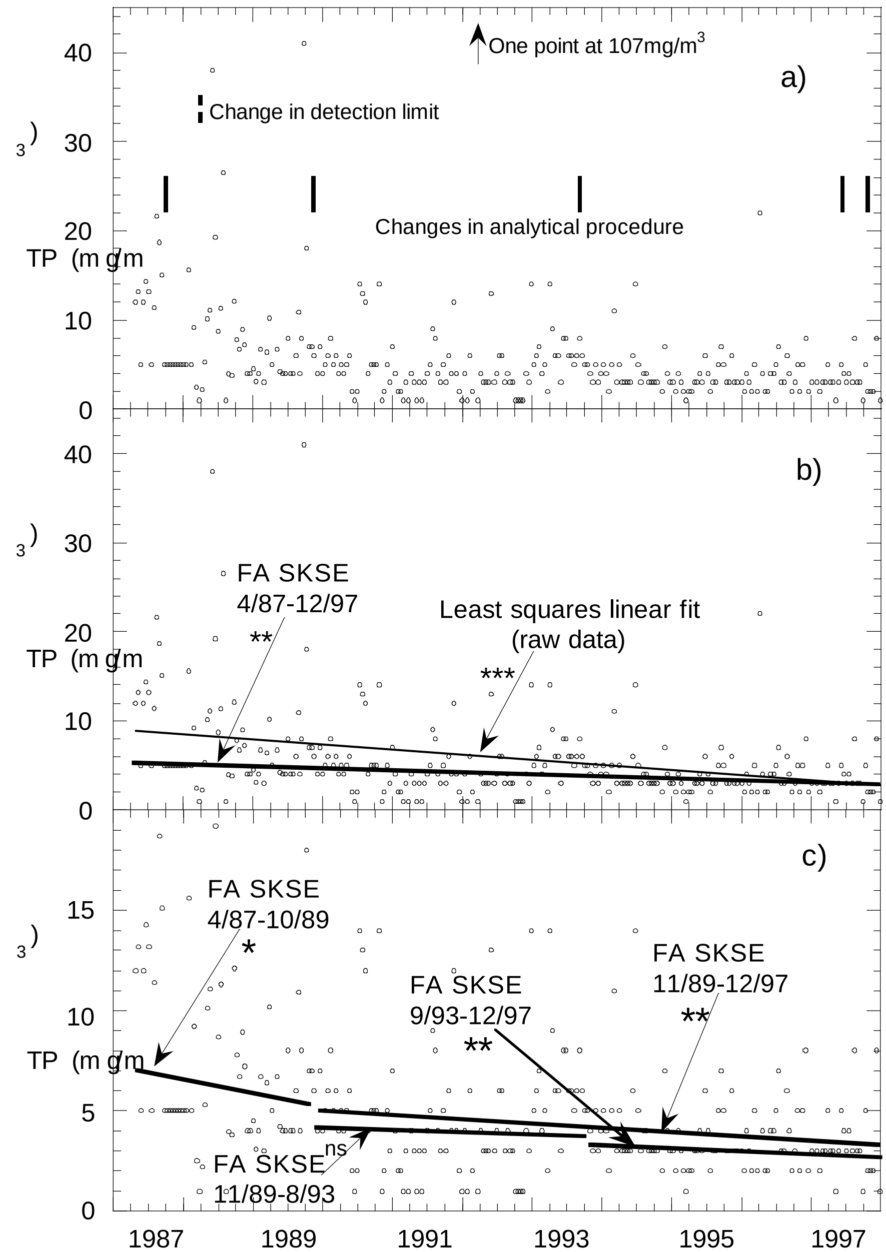

Figure 4. Trends in total phosphorus (TP): a) raw data with analytical changes noted; b) raw data with overall trend lines; c) raw data for partitioned datasets (note scale change). The non-parametric slopes drawn are the flow-adjusted Seasonal Kendall Slope Estimate (FA SKSE) with each slope's mid-point located at the period median. The trend test significances are marked adjacent to the arrow locating the slope. It was not possible to export the flow-adjusted data from the software.

CONCLUSIONS

We have used three actual water quality datasets to examine the effects on trend detection of changing analytical procedures and detection limits with sub-detection limit data censoring. Several points have emerged from this study which are of importance for water authorities, and others, in possession of a long data time series and who are contemplating trend analysis.

1. Such changes can have strong effects on trend analysis and add tedium and uncertainty to the process. As a consequence, a simple, easy-to-use protocol is not possible and much interpretation of the data is required.

2. Less accurate methods early in the record may produce not only high-biased data but also greater variance. The consequence of this is an artificial induction of a down-trend. Such data are therefore of little value in trend detection.

3. In datasets with multiple procedure changes embracing relatively short time periods, say of the order of a few years, there is considerable difficulty in judging whether the data from adjacent time blocks can be coalesced to increase the length of the time period being analyzed, thus potentially increasing trend power. In the absence of paired comparisons between 'old' and 'new' procedures, combining adjacent data blocks adds uncertainty to any conclusions drawn. For trend detection, the default assumption must be that there should be no analytical procedure changes to the network, and networks must be designed with this in mind.

4. The consequence of changing (usually lowering) detection limits with data censoring at the detection limits induces artificial (usually down) trends. Censoring of data is therefore a real issue. A detected value below detection limit is still a better estimate of "truth" than the statement "< detection limit". For trend detection, reported below detection limit concentrations notwithstanding their larger error, are still of much greater value than using a substitute value, e.g., half the detection limit. Data censoring is of concern when using non-parametric statistics because of the production of multiple tied values.

REFERENCES

Alexander, R. B., R. A. Smith, and G. E. Schwarz, 1993. Correction of stream quality trends for the effects of laboratory measurement bias. Water Resources Research 29: 3821-3833.

Analytical Methods Committee 1987. Recommendations for the definition, estimation and use of the detection limit. The Analyst 112: 199-204.

Bell, H. F., 1990. IBM groundwater quality monitoring program at East Fishkill New York. In: Ward et al. (1990), op cit.

Ellis, J. C., and C. F. Gilbert, 1980. How to handle 'less-than' data when forming summaries. Water Research Centre Enquiry Report ER 764. Water Research Centre, Medmenham, England.

Gilliom, R. J., R. M. Hirsch, and E. J. Gilroy, 1984. Effect of censoring trace-level water-quality data on trend-detection capability. Environmental Science & Technology 18: 530-535.

Goodman, S. N., 1993. p values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. American Journal of Epidemiology 137: 485-496.

Helsel, D. R., and T.A. Cohn, 1988. Estimation of descriptive statistics for multiply censored water quality data. Water Resources Research 24: 1997-2004.

Hughes, J, and S. P. Millard, 1988. A tau-like test for trend in the presence of multiple censoring. Water Resources Bulletin 24: 521-532.

Loftis, J. C., G. B. McBride, and J. C. Ellis, 1991. Considerations of scale in water quality monitoring and data analysis. Water Resources Bulletin 27: 255-264.

Millard, S. P., and S. J. Deverel, 1988. Nonparametric statistical methods for comparing two sites based on data with multiple nondetect limits. Water Resources Research 24: 2087-2098.

McBride, G. B., and J. C. Loftis, 1994. The most important statistical aspects. In: Adriaanse, M., J. van der Kraats, P. G. Stoks, and R. C. Ward, ed. Monitoring tailor-made. RIZA–The Netherlands Institute for Inland Water Management and Waste Water Treatment, Leystad, The Netherlands, pp. 153-161.

McCann, P. B., 1999. An analysis of water quality trends in the Neversink River, Catskill Region, New York. A thesis presented to the Graduate Faculty of Western Connecticut State University in partial fulfillment of the requirements for the degree Master of Arts in Oceanography and Limnology.

Newell, A. D., and M. L. Morrison, 1993. Use of overlap studies to evaluate method changes in water chemistry protocols. Water, Air, and Soil Pollution 67: 433-456.

Newell, A. D., D. J. Blick, and R. C. Hjort, 1993. Testing for trends when there are changes in methods. Water, Air, and Soil Pollution 67: 457-468

Porter, P. S., R. C. Ward, and H. F. Bell, 1988. The detection limit. Environmental Science & Technology 22: 856-861.

Shapiro, J., and E. B. Swain, 1983. Lessons from the silica "decline" in Lake Michigan. Science 221: 457-459.

Smith, D. G., and R. Maasdam, 1994. New Zealand's national water quality network. 1. Design and physico-chemical characterisation. New Zealand Journal of Marine and Freshwater Research 28: 19-35.

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

Ward, R. C., J. C. Loftis, and G. B. McBride, 1990. Design of water quality monitoring systems. Van Nostrand Reinhold, New York. 231pp.