Calculating compliance with targets: modified from Swan River Trust 2004

To assess whether or not the water quality at a site is passing a target, sample data are compared with the target concentrations. Developing the sampling program – known as a compliance-monitoring scheme – and assessing the achievement of targets require the definition of three key things:

- When, where and how often samples are collected. In the catchments, compliance sampling is carried out while the rivers are flowing, between June and October.
- The statistic used to characterise quality may be any percentile value. For compliance with the nutrient targets in the catchment the 50th percentile is used. Note that compliance is not based on a single sample value, but on a set of data points obtained throughout monitoring.
- The 'compliance/breach' criterion that will be applied to the target. Taking into account sampling error, how many samples are permitted to exceed the target before it is decided that the target is breached (or met)?

In all catchments data from three consecutive years is pooled to compare both TN and TP values to the targets. Pooling three years of data minimises the impact of unusually wet or dry years and gives greater confidence that any changes in compliance are real.

The test statistic

Because of natural variation in water quality, compliance is not assessed against a single sample value, but against a key statistic derived from the set of monitoring data.

Assessing compliance with percentile water quality targets is actually about examining the rate of excursion from target levels. The excursion rate is the period in which water quality is worse than the limit established by the target, and the maximum allowable excursion rate is specified by the population statistic used.

The compliance/breach criteria

From the pooled samples collected in each catchment (about 30 for the Swan-Canning catchment) it is easy to calculate an excursion rate, which is simply the percentage of samples that exceed the target value. However, because the sampling data is collected at intervals (as opposed to continuous monitoring) the true rate of excursion above the target value cannot be known. The excursion rate determined from the available data is an estimate of the actual excursion rate in the stream. However, statistics can be used to calculate a range around the sample excursion rate within which the actual excursion rate is likely to lie. This range of values is known as a confidence interval. For catchment compliance, decisions of compliance or breach are taken using a 95 per cent confidence interval around the sample excursion rate (see Figure 1). There is a 95 per cent probability that the true population rate of excursion lies within this range.

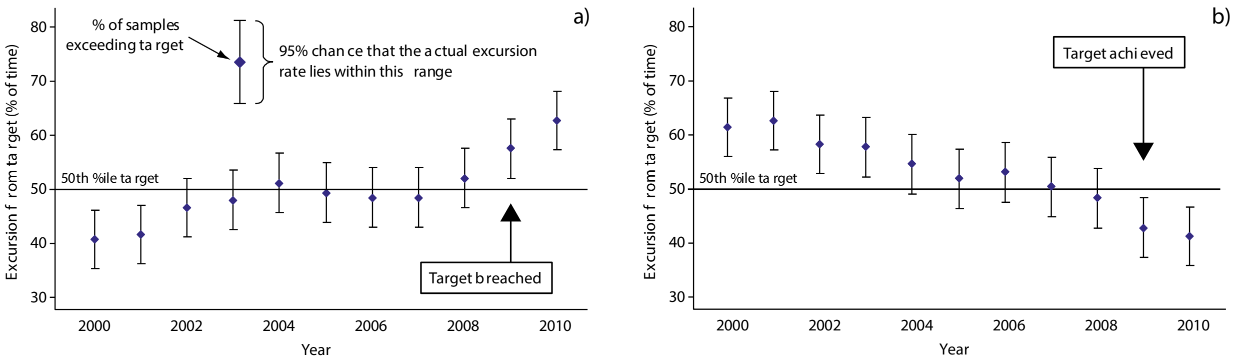


Figure 1: The confidence interval method for measuring a) breach and b) achievement (or compliance) of a catchment nutrient target.

For those catchment tributaries where water quality is currently acceptable the target is breached when the entire confidence interval lies above the target value (Figure 1a). in this case it can be safely concluded (with 95 per cent probability) that water quality has exceeded the target level. For those tributaries where water quality is currently worse than target levels the benefit of the doubt goes the other way. That is, the target is only achieved where there is a 95 per cent probability that water quality has actually improved to better than target levels (Figure 1b).

Using the confidence interval method, decision rules have been developed which specify exactly how many samples are allowed to exceed each target before the target has been breached, or met. Examples of the decision rules are given in Table 1 below. The confidence interval method results in two separate decision rules depending on whether or not current water quality is acceptable. Decision rules can be developed in exactly the same way for any number of collected samples which is necessary in practice because the number of samples collected over a monitoring period varies from year to year and from catchment to catchment (as there may be different sampling frequencies).

Table 1: Examples of decision rules for catchment nutrient targets. The rules cover compliance decisions for 50th percentile targets using 30 samples.

Prior condition	Decision rule
Water quality target met (passing target)	20 or more samples > target = water quality target no longer met (tributary fails target)
Water quality target not met (failing target)	11 or fewer samples > target = water quality target met (tributary passes target).

References

Swan River Trust 2004, *Developing targets for the Swan-Canning Cleanup Program, River Science 7*, Swan River Trust, Perth.

Calculating trends in nutrient data

Detecting trends in a water quality data series is not as simple as drawing a 'line of best fit' and measuring the slope. There are likely to be multiple factors contributing to variation in water quality over time - many of which can hide or exaggerate trend components in the data. The most likely sources of variation include: flow variation, seasonal variation, trend and random components. Changes in water quality brought about by human activity will usually be superimposed on natural sources of variation. Therefore, the influence of flow and seasonal variation needs to be examined prior to analysing for trending periods in a water quality data series. Although the primary objective is to detect trends over time, sources of natural variation must be known and adjusted for prior to analysis. This will provide management with an improved perception of changes in water quality that are (more than likely) linked to human intervention in the catchment.

Assumptions of the trend tests

Non-parametric significance tests were used to identify statistically significant trending periods in a water quality data series. Non-parametric tests were used because they are not affected when the distribution of data is not normal, insensitive to outliers and are not affected by missing or censured data (Loftis *et al*., 1991).

An assumption of trend tests is that trends are consistently increasing or decreasing, otherwise known as monotonic changes (Helshel and Hirsch, 1992). If concentrations vary non-monotonically over the period being analysed the results of linear tests for trend may be misleading (Robson and Neal, 1996). For this analysis, the assumption of monotonic change was verified by a visual examination of a LOWESS (Locally Weighted Scatterplot Smooth) line fitted to the data over the period of interest (Helshel and Hersh 1992, Aulenbach *et al.*, 1996). If the water quality data series was non-monotonic over the whole monitoring period, only the most recent period of monotonic change was examined for trend.

Another assumption of the trend tests is that samples in a data series must be independent. If the data series are not independent (that is, exhibits auto-correlation) the risk of falsely detecting a trend is increased (Esterby 1996, Ward *et al*., 1990). A correlated data series contains surplus data and ultimately results in the little or no nett information gain. As a rule, the level of serial correlation in a data series increases as the frequency of sampling increases. The maximum sampling frequency possible without encountering serial correlation can be thought of as the point of information saturation (Ward *et al.*, 1990).

Five years of data is required to test for trends in this assessment. If five consecutive years of data was not available then trend analysis could not be undertaken for that site.

Testing for statistically significant changes

The Mann Kendall test was used to determine the statistical significance of the trending periods (Gilbert, 1987). It is an example of a non-parametric test and was only used when the data series exhibited independence (ie. no correlation in the data series).

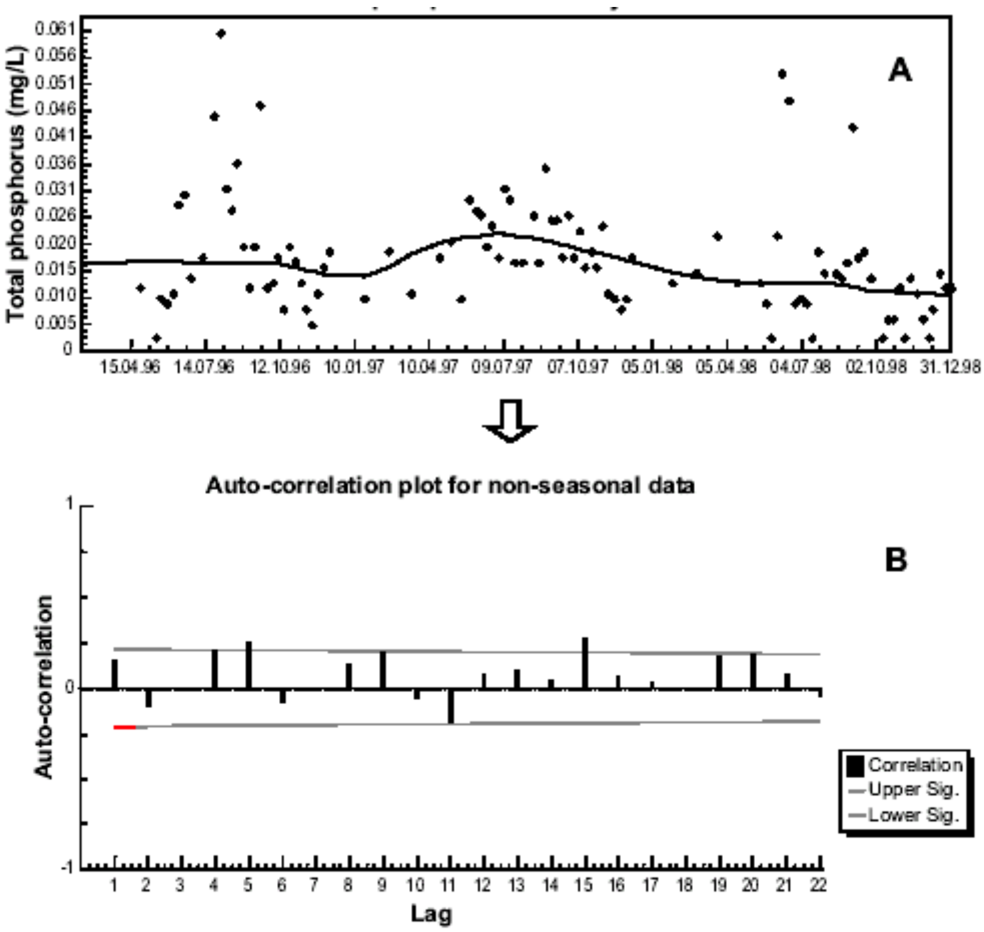


Figure A above shows an example of a time series for TP with little evidence of seasonal variation (LOWESS smooth line). Figure B above shows an autocorrelation plot which indicates that the data points in the time series are mostly independent of each other.

When seasonal cycles were evident in a data series the Seasonal Kendall test was used to test for trend. The Seasonal Kendall test is a variant of the Mann Kendall test that accounts for the presence of seasonal cycles in the data series (Gilbert, 1987). Seasonal cycles in water quality are common in waterways and can be introduced by natural cycles in rainfall, runoff, tributary hydrology and seasonal variation in groundwater. The presence of seasonal cycles in a data series can introduce correlation to the data series which will complicate the detection of trends. The detection of seasonal variation in the data series was tested for by using an auto-correlation analysis.

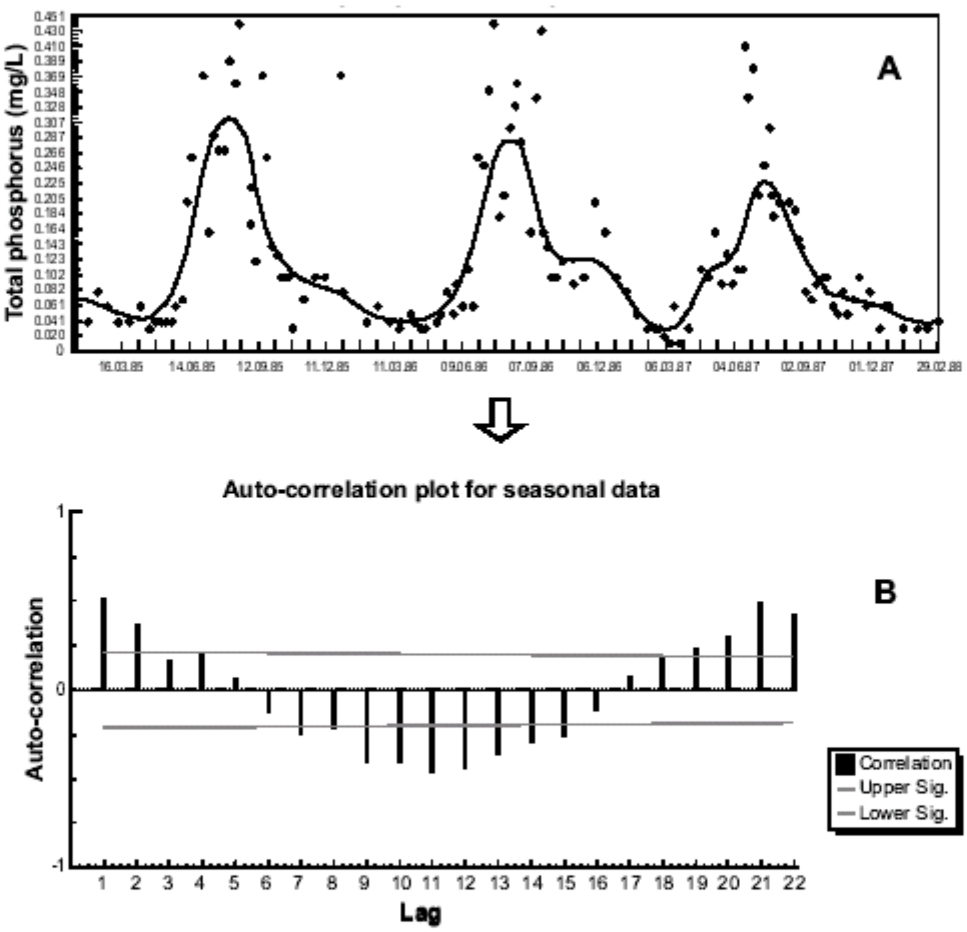


Figure A above shows an example of a time series for TP with seasonal variation (LOWESS smooth line). Figure B above shows an autocorrelation plot which indicates that the data points in the time series are dependent on each other.

A trend will be found to be statistically significant when the magnitude of the change is large relative to the variation of the data around the trend line. Unfortunately, when analysing long periods with large sample sizes any trend no matter how small will be statistically significant (Loftis, 1996; McBride *et al.*, 1993; Loftis *et al.*, 1991). The identification of a statistically significant trend should be seen as filter that removes small drifts in concentration from further consideration. Further analysis using sample size estimates are required to determine whether a sufficient number of ‘independent’ samples were collected to detect a trend.

Removing variation due to flow

Water quality in waterways can also be affected by changes in discharge that may create or hide trends in a fixed-interval data series. For this reason, trend analysis was also carried out on the data after it was adjusted for the effects of variation due to flow. The relationship between nutrient concentration and flow was modelled using a LOWESS fit on the flow / concentration response (Esterby, 1996, Robson and Neil, 1996, Lettenmaier *et al* 1991). The difference or ‘residuals’ between the observed concentration and the LOWESS modelled concentration is known as a flow-adjusted concentration (Hipel and McLeod, 1994). Subsequently, the flow-adjusted concentrations were reordered in time and then analysed for trend (Gilbert, 1987, Helshel and Hersh, 1992, Harned *et al* 1981, Hipel and McLeod 1994, Lettenmaier *et al* 1991). The flow-adjustment process often helped to remove seasonal variation, although some evidence of seasonal variation often remained in the flow-adjusted data series.

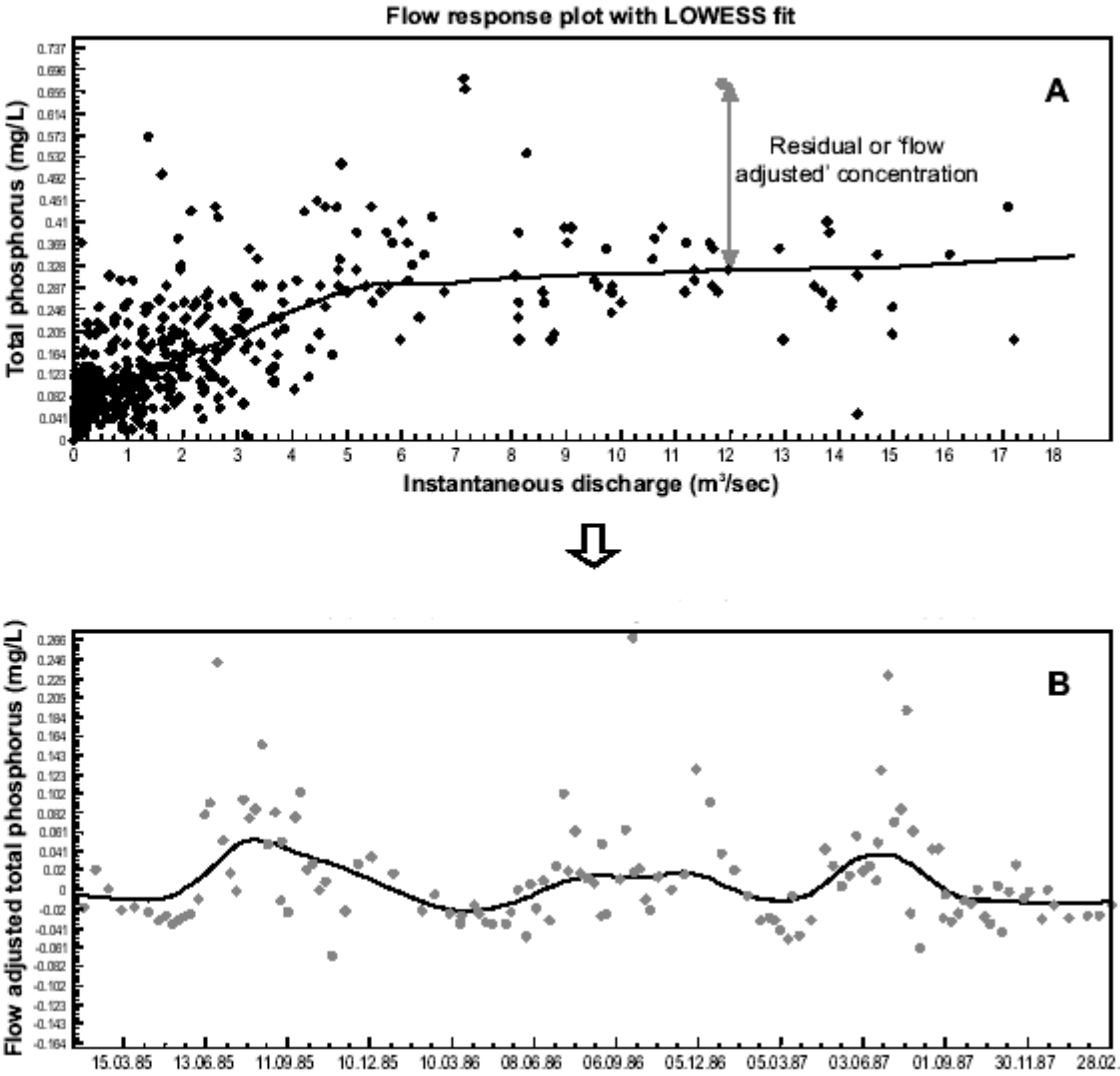
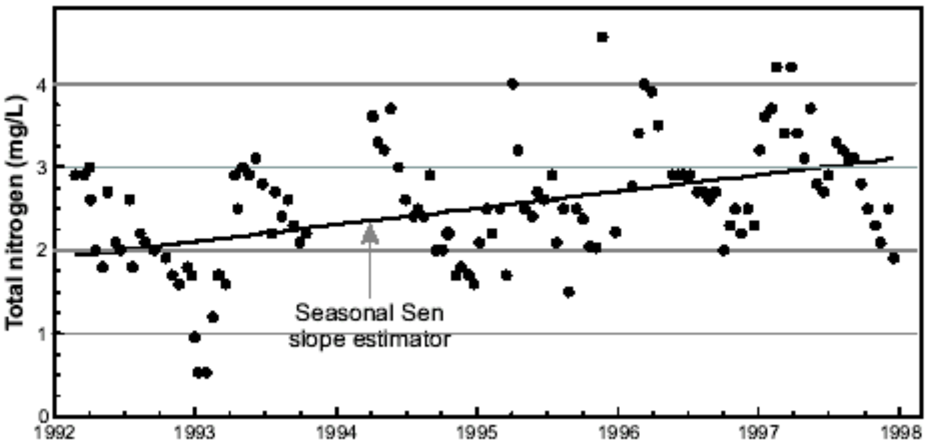


Figure A above shows a flow / TP concentration relationship curve (otherwise known as a flow response) with a LOWESS smooth line. Modelled concentrations are derived for every flow level and compared to the observed data. The difference between the modelled and observed data are known as residuals. Figure B above shows the residuals reordered in time with a LOWESS smooth line and are considered to have flow effects removed from the data.

The Sen slope estimator was used to estimate the slope of the trend line (Gilbert, 1987). The Sen estimate is the median slope of all slopes calculated using all inter-annual pairs of observations. In the presence of seasonal cycles the Seasonal-Kendall slope estimator was used (Gilbert, 1987), which is the median slope of all slopes calculated using pairs of observations collected at the same time each year.



An example of the Seasonal Sen slope estimator being applied to TN monitoring data. This line is used to estimate the slope of the trend in the data series.

Sample size estimates

A period of change being analysed was found to be statistically significant when the Kendall Test had a p-value less than or equal to 0.05. This was not enough evidence to conclude a trend was present. '*A-posteriori*' calculations were subsequently carried out to assess whether enough independent samples had been collected and used in the trend test to meet the criteria specified by the nominated statistical error risks (a = 0.05 and b = 0.10). This was achieved by comparing the effective information content in the collected data series with the number of independent samples required to detect a trend.

The effective information content in the data series, that is the effective number of independent values, was estimated for each of the data series analysed for trend using the formula provided by Bayly and Hammersley (1946) (*op cit* Lettenmaier, 1976, Lachance, 1992, Close, 1989, Zhou, 1996):

$$n^* = [1/n + 2/n^2 \sum_{j=1}^{n-1} (n - j)\, \rho\, (jt)]^{-1}$$

where:

n^* = effective number of independent observations

n = number of samples

j = lag number

t = sampling interval

ρ = coefficient of correlation

Where seasonal cycles were found the data series was de-trended and de-seasonalised (using seasonal medians) prior to calculating the number of independent samples (n*).

The estimated number of independent samples needed to detect a linear trend (in a variable distributed normally about the trend line) was estimated using the function (Lettermaier, 1976; Ward *et al.,* 1990):

$$n^{\#} = 12\, \sigma^2\, [\, t_{\alpha/2,(n-2)} + t_{\beta\,(n-2)}\,]^2/\Delta^2$$

where:

$n^{\#}$ = estimated number of samples needed to detect a trend

σ = the standard deviation of the de-trended series

Δ = the magnitude of the trend

t = the critical values of the t-distribution, using $\alpha = 0.05$ and $\beta = 0.1$

This function relies on probabilities predicted by the t-distribution and is therefore from the parametric family of statistical procedures. Data requirements for parametric and the equivalent non-parametric tests are similar, so this equation will approximate the sample size needed for non-parametric tests of significance (Ward *et al.,* 1990).

Detecting the trend

A trend in the data series was considered to be detected only when two criteria were met. Firstly, the Kendall test for trend on the data series must be statistically significant (ie. $p<0.05$). Secondly, the number of independent samples collected (n*) had to approximately equal or exceed the ‘estimated’ number of independent samples (n#) required to detect a trend. The direction of a detected trend either increases (representing a deterioration of water quality) or decreases (representing an improvement in water quality). If any of the above two criteria were breached then the result was 'no trend'. If $p<0.05$

and the number of independent samples collected was less than the 'estimated' number of independent samples required to detect a trend, the trend was described as "emerging" (either increasing or decreasing). Sites with a 'no trend' result may be a consequence of poor monitoring program design and implementation over time and, if it is widely believed there should have been a detected trend over the monitoring period, then the monitoring program needs to be re-evaluated.

References

Aulenbach, BT, Hooper, RT & Bricker, EP 1996, 'Trends in the chemistry of precipitation and surface water in a national network of small watersheds', *Hydrological Processes* 10(2): 151-181.

Close, ME 1989, 'Effect of serial correlation on groundwater water quality sampling frequency', *Water Resources Bulletin* **25**: 507-515.

Esterby, SR 1996, 'Review of methods for the detection and estimation of trends with emphasis on water quality applications'*, Hydrological Processes*, 10(2): 127-149.

Gilbert, RO 1987, *Statistical methods for environmental pollution monitoring*. Van Nostrand Reinhold: New York.

Harned, DA, Daniel, CC III & Crawford, JJ 1981, 'Methods of discharge compensation as an aid to the evaluation of water quality trends', *Water Resources Research*, 17: 1389-1400.

Helshel, DR & Hirsch, RM 1992, *Statistical methods in water resources*, Elsevier, Amsterdam, p. 288

Hipel, KW and McLeod, AI. (1994), *Time Series Modelling of Environmental and Water Resources Systems.* Elsevier, Amsterdam.

Lachance, M 1992, Monitoring lakes in Quebec. Case study in: *Design of water quality monitoring systems,* R. Ward, J. Loftis & G. McBride. Van Nostrand Reinhold, New York.

Lettenmaier, DP 1976, 'Detection of trends in water quality from records with independent observations', *Water Resources Research* 12(5): 1037-1046.

Lettenmaier, DP, Hooper, ER, Wagoner, C and Faris, K 1991, 'Trends in stream quality in the continental United States, 1978-1987', *Water Resources Research*, 273: 327-339.

Loftis, JC, McBride, GB & Ellis, JC 1991, 'Considerations of scale in water quality monitoring and data analysis', *Water Resources Bulletin*, 27(2): 255-264

Loftis, J 1996, 'Trends in groundwater quality', *Hydrological Processes* **10** (2): 335-355.

McBride, GB, Loftis, JC & Adkins, NC 1993, 'What do significance tests really tell us about the environment?' *Environmental Management* 17: 423-432.

Robson, AJ & Neal, C 1996, ‘Water quality trends at an upland site in Wales’, *Hydrological Processes*, 140(2): 183-203.

Ward, R, Loftis, J & McBride, G 1990, *Design of water quality monitoring systems*. Van Nostrand Reinhold: New York.

Zhou, Yangxiao 1996, Sampling frequency for monitoring the actual state of groundwater systems, *Journal of Hydrology* **180**: 301-318.

Calculating status

TN and TP concentrations are described in terms of the nutrient classification shown in Table 1.

Table 1: *Classifications used to assess the status of TN and TP concentrations in monitored waterways*

Status	Total nitrogen (mg/L)	Total phosphorus (mg/L)
Very high	> 2.0	> 0.20
High	> 1.2 – 2.0	> 0.08 – 0.20
Moderate	> 0.75 – 1.2	> 0.02 – 0.08
Low	< 0.75	< 0.02

Depending on trends, chance sampling and sources of natural variation, the nutrient concentrations sampled from a monitored site will change. The nutrient status for a waterway is initially assigned using the median nutrient concentration for the first year of sampling. Subsequent status periods are assessed using the median and 90 per cent confidence interval. If the median or all or part of the confidence interval remains in the earlier classification band, then there is no change in status. Status only changes once both the median and entire 90 per cent confidence interval move to a different classification band.

Figure 1 shows how TP status at Mayfields Main Drain (in the Peel-Harvey catchment) was originally classified as high (the median was between 1.2 and 2.0 mg/L). By the 1992–94 period, the median had decreased and fallen within the moderate classification band (0.75–1.2 mg/L); however, part of the 90% confidence interval was still in the high classification band and so the status remained high. In the 1994–96 period, both the median and 90% confidence interval fell below the high classification and hence the status changed to moderate. During the 1996–98 period the median once again dropped to a lower classification band (<0.75 mg/L); however, it wasn't until the 1998–2000 period that the actual classification status changed to low.

In summary, the nutrient status for a waterway is assigned by using the median of nutrient concentration over a three-year period. The three-year period is used to diminish the influence of natural variation between years.

Change in status requires the median and whole 90% confidence interval to be within the new status concentration range.

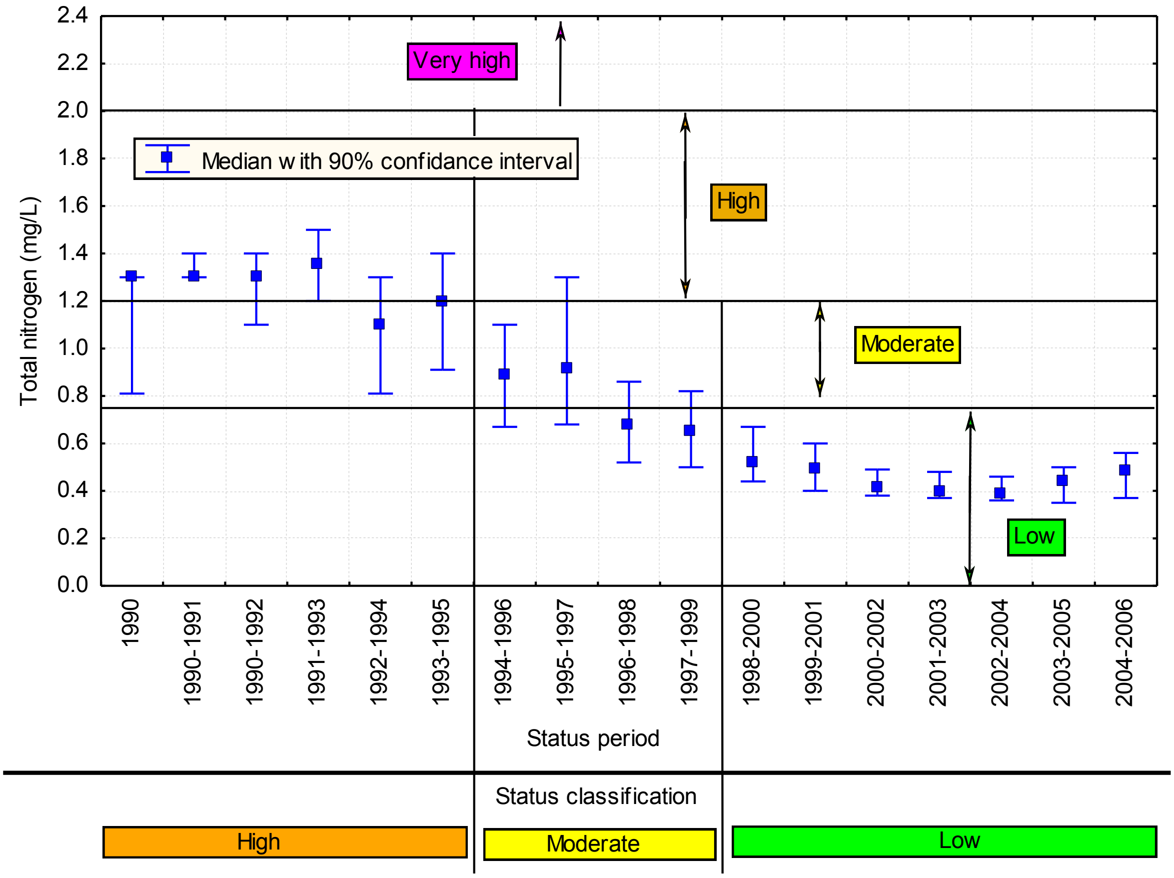


Figure 1: *Total phosphorus status classification for Mayfields Main Drain (AWRC 613031)*

References

Hall, J 2010, *Water quality management in urban catchments of the Swan Coastal Plain: analysis of the Bartram Road catchment*, Water Science Technical Series, report no. 22, Department of Water, Western Australia.

Calculating loads

Loads were calculated for those sites where there was sufficient flow and nutrient data available.

Annual loads are calculated by multiplying daily flow with daily nutrient concentration and aggregating over the year. Daily concentration measurements are not available as samples were taken weekly at best, so daily concentration data needs to be in-filled to calculate loads. To calculate the in-filled nutrient data the locally estimated scatterplot smoothing (LOESS) algorithm (Cleveland 1979) was used.

LOESS creates a flow-concentration curve by fitting a low-degree polynomial to a subset of the flow-concentration data to estimate the concentration for the flow at the centre point of the data subset. This is done for each flow value in the dataset. For days on which nutrient data were collected daily loads are calculated from observed concentrations and flows. For days with no data, daily loads are calculated from the daily flow and the estimated concentration from the LOESS flow-concentration curve. The assumption of the LOESS algorithm is that there is a relationship between flow and concentration.

References

Cleveland, WS 1979, 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association,* Vol 74, pp. 829-836.

Hall, J 2010, *Water quality management in urban catchments of the Swan Coastal Plain: analysis of the Bartram Road catchment*, Water Science Technical Series, report no. 22, Department of Water, Western Australia.

Calculating loads using modelled data

The Streamflow Quality Affecting Rivers and Estuaries (SQUARE) model reports estimated flow, nitrogen and phosphorus loads from a catchment. It can also estimate the source of the nutrient loads within the catchment.

The following catchments have been modelled:

- Swan-Canning
- Peel-Harvey
- Leschenault
- Geographe Bay
- Scott River

Nutrient sources

Land use and associated nutrient export rates were used to determine the sources of nutrients exported from a catchment.

Estimated loads

The loads presented in this section differ to those calculated from water quality and flow data at one site using a LOESS algorithm as they have been modelled utilising land use statistics and calibrated with water quality and flow data.

Nitrogen (kg/ha/year)	Remediation priority	Phosphorus (kg/ha/year)
> 6.0	High	> 1.0
3.0 – 6.0	Medium	0.5 – 1.0
< 3.0	Low	< 0.5

Remediation priority

Using the nutrient load per cleared area a simple remediation prioritisation ranking was determined.

March 2015