Using beta-content tolerance intervals to derive water quality guidelines.

By

Dr. David R. Fox
CSIRO, Perth Western Australia

Background

The new National Water Quality Monitoring Strategy and Water Quality Guidelines documents (ANZECC 2000) have attempted to identify robust statistical methods of assessing water quality status relative to a reference or trigger value. This search has been motivated by at least two factors: (i) a recognition that natural ecosystems are inherently 'noisy' and that traditional compliance driven approaches to water quality assessment have tended to overlook or at least underestimate the effect of this high variability; and (ii) statistical procedures based on 'classical' (ie. normal distribution theory linear models) approaches are not particularly robust to the nuances of water quality monitoring (eg. skewed distributions, spatial-temporal correlation etc.). The revised *ANZECC* water quality guidelines advocates the following approach for physical-chemical stressors:

> *A trigger for further investigation will be deemed to have occurred when the median concentration of n independent samples taken at a test site exceeds the eightieth percentile of the same compound at a suitably chosen reference site.*

This rule is statistically-based and accommodates natural background variation by comparison to a reference site. Its robustnest derives from the fact that it acknowledges site-specific anomalies and utilises robust statistical measures such as the median.

While the procedure has been shown to have acceptable performance characteristics for detecting a shift in location (Fox 2000a), it's utility for assessing 'compliance' based on an *individual* sample value may not be as great. For example, in the case where the distributions of some compound at the test and reference sites are identical, there is a 20% chance that a single observation from the test site will exceed the 80^{th}. percentile at the reference site. Thus, the Type I error rate when the sample size is n =1 is 20%. Fox (2000) shows that this can be reduced to the more conventional 5% level by increasing the test site sample size to 5 or 6. *ANZECC* has taken the view that the choice of sample size is a matter to be decided by those undertaking the analysis as it requires an individual assessment of the trade-offs between rate of false triggering and cost of additional sampling.

Although well-known to statisticians, *beta-content tolerance intervals* do not appear to have been routinely applied to water quality monitoring and assessment problems. Their use in this capacity is the subject of this note. The following development relies on classical, normal-distribution theory and is thus open to the criticism identified in (ii) above. An alternative, non-parametric procedure is described Fox (2000b) to be used in situations where the data are non-normal and cannot be suitably transformed.

One-Sided tolerance limits for the normal distribution

Let $X_i \sim f_{X_i}(x_i)$ where $f_X(\cdot)$ is the *pdf* of a normal distribution having mean μ and variance σ^2 and $0 < \beta < 1$ and $0 < \gamma < 1$.

The *random* interval $\left(-\infty, U\right)$ where $U = u\left(X_1, X_2, \ldots, X_n\right)$ is a **one-sided β-content tolerance interval** at level γ if

$$P\left[\int_{-\infty}^{U} f_X(x)\,dx \;\geq \beta\right] = \gamma \qquad (1)$$

Equation (1) says that the probability that at least 100β% of the distribution is 'captured' by the random interval $\left(-\infty, U\right)$ is γ.

The statistic U is assumed to be of the form:

$$U = \bar{X} + kS \qquad (2)$$

where $\bar{X}$ and S are respectively, the mean and standard deviation of n observations randomly selected from $f_X(\cdot)$.

Then

$$W = \int_{-\infty}^{U} f_X(x)\,dx \;\geq \beta \quad \Rightarrow \quad \frac{\bar{X} + kS - \mu}{\sigma} > z_\beta$$

where z_β is the ordinate of the standard normal distribution having right-tail area equal to β.

Equation (1) can be rewritten as

$$P\left[\frac{\bar{X} + kS - \mu}{\sigma} > z_\beta\right] = P\left[\frac{\sqrt{n}\,\frac{\left(\bar{X} - \mu\right)}{\sigma}}{S/\sigma} - \frac{\sqrt{n}\,z_\beta}{S/\sigma} > -k\sqrt{n}\right] = \gamma \qquad (3)$$

Equation (3) is equivalent to

$$P\left[T_{n-1}\left(-\sqrt{n}\,z_\beta\right) > -k\sqrt{n}\right] = \gamma \;\equiv\; P\left[T_{n-1}\left(\sqrt{n}\,z_\beta\right) < k\sqrt{n}\right] = \gamma$$

where $T_{n-1}\left(\sqrt{n}\ z_\beta\right)$ is the non-central T-distribution with *n-1* degrees of freedom and non-centrality parameter $\sqrt{n}\ z_\beta$.

Thus, the appropriate multiplier, *k* in equation 2 is found by finding the ordinate on the non-central T-distribution (call it t^*) such that

$$P\left[T_{n-1}\left(\sqrt{n}\ z_\beta < t^*\right)\right] = \gamma$$

k is then found by solving $t^* = \sqrt{n}\ k$.

Values for k have been computed for selected *n,* β, and γ are given in the table below.

n	**γ = 0.95**		**γ = 0.90**		**γ = 0.50**	
	β = 0.95	**β = 0.90**	**β = 0.95**	**β =0.90**	**β = 0.95**	**β =0.90**
2	22.261	20.583	13.090	10.253	2.339	1.784
3	7.656	6.1556	5.312	4.258	1.939	1.498
4	5.144	4.162	3.957	3.188	1.830	1.419
5	4.203	3.407	3.400	2.742	1.779	1.382
6	3.708	3.006	3.092	2.494	1.751	1.361
7	3.400	2.756	2.894	2.333	1.732	1.347
8	3.188	2.582	2.754	2.219	1.719	1.337
9	3.032	2.454	2.650	2.133	1.709	1.330
10	2.911	2.355	2.568	2.066	1.702	1.324
15	2.566	2.068	2.329	1.867	1.681	1.309
20	2.396	1.926	2.208	1.765	1.671	1.301
30	2.220	1.777	2.080	1.672	1.662	1.295

Comparison with a guideline or trigger value

Conceptually, to compare the data to a guideline or trigger value, the distribution at a test site is compared with a threshold value (the guideline concentration). For example, the guideline may be set so that some high proportion (e.g. 95%) of values at a reference site are less than the guideline or trigger value, G with some high probability (e.g. 90%). This gives rise to the notion of a 95:90 guideline or trigger value.

Assuming G has been established so that in an 'undisturbed' system, a proportion β of values will be less than G with probability γ, then compliance will be demonstrated provided the mean $\overline{X}$ of a sample of *n* readings satisfies

$$\overline{X} \le G - kS$$

Example

Suppose the trigger value for cadmium in marine waters is 5.5µg/l and that in an 'undisturbed' system, 95% of all Cd values should be less than this value with a high probability (eg. 0.95). Five water samples at a particular location had the following Cd values: 1.6; 1.4; 2.8; 1.7; and 1.1. For this data $\overline{X} = 1.72$ and $S = 0.646$.

With $\beta = \gamma = 0.95$ and n=5 we obtain k = 4.203 from the table. The comparison with the guideline is:

$$\overline{X} \leq 5.5 - (4.203)(0.646) = 2.78$$

In this case, the sample mean of 1.72 is less than 2.78 and no further investigations are indicated.

References

ANZECC/ARMCANZ (2000), National Water Quality Guidelines.

Fox, D.R. (2000a) An assessment of the proposed trigger method for physical-chemical stressors. Report EPO-TR/2000/2 Unpublished. CSIRO Environmental Projects Office, Perth Western Australia.

Fox, D.R. (2000b) A robust method for water quality monitoring and assessment: A discussion paper prepared for ANZECC review committee. Report EPO-TR/2000/1 Unpublished. CSIRO Environmental Projects Office, Perth Western Australia.