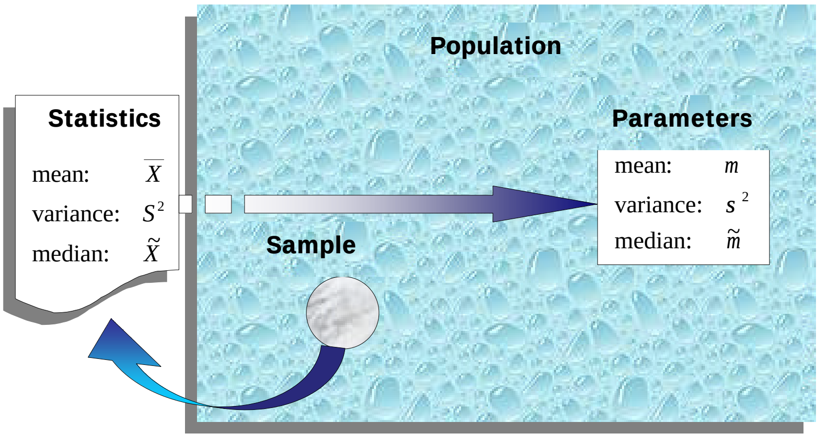
UNDERSTANDING THE NEW *ANZECC* WATER QUALITY GUIDELINES

David R Fox, CSIRO Land and Water, Perth



Introduction

The new *ANZECC/ARMCANZ* Water Quality Guidelines (ANZECC & ARMCANZ 2001a) and the companion document, the Monitoring and Reporting Guidelines (ANZECC & ARMCANZ 2001b) pave the way for a quantum leap forward in the way in which the environmental condition of water bodies in Australia and New Zealand will be monitored and assessed. The development of new 'trigger' values and associated monitoring has been underpinned by two overriding objectives: 1. Risk equity; and 2. Simplicity. The risk-based framework represents a significant and welcome departure from the old 'command and control' mode whereby 'compliance' was assessed by a prosaic comparison of sample data with a fixed criterion. Both the industry and its regulators have been aware of deficiencies in this approach for some time, the most significant perhaps being the inherent lack of recognition of natural variation in water quality parameters over time and space – even for relatively undisturbed systems.

Thus the challenge for the *ANZECC / ARMCANZ* team was to devise statistical methods that were:

- Fit for purpose
- Easy to implement and understand
- Robust under a wide range of operating conditions
- Easy to interpret

With respect to the derivation of 'trigger values' for toxicants, the second dot point has been the most difficult to satisfy. This is largely a consequence of *ANZECC / ARMCANZ* adopting the now well-established statistical method of Aldenberg and Slob (1993).

The companion Monitoring and Reporting Guidelines document (ANZECC & ARMCANZ 2001b), provides a review of statistical inference and related concepts such as power, level of significance, and sample size calculations. While the tools of statistical inference provide us with a consistent and scientifically credible way of making decisions under uncertainty, they are not foolproof and errors invariably occur. In the context of statistical hypothesis testing, two types of error are possible and statisticians refer to these (somewhat unimaginatively) as *Type I* and *Type II* errors. A *Type I* error arises when a true hypothesis is incorrectly rejected while a *Type II* error will have been committed if the statistical test leads us to incorrectly accepting a false hypothesis. The ability to 'get it right' in this situation is an important feature of any statistical test procedure and it is called the test's *power*. It is intuitively obvious that statistical power increases as the sample size *n* increases. Curves that depict this relationship for a specific statistical test are often consulted at the planning stage of an investigation to help balance effort (cost) and Type II error. The calculations underpinning these curves are complex and best consigned to computer software. The Monitoring and Reporting Guidelines document (ANZECC & ARMCANZ 2001b) gives more information about power and sample size calculations as well as references to freeware. *CSIRO's* own product, *PowerPlant*® and accompanying documentation can be freely downloaded from the following site:

ftp://ftp.per.its.csiro.au/csiro-wa/biometrics

Establishing trigger values for toxicants

Considerable effort has been expended in refining the manner in which Australian guidelines for toxic chemicals are established. The aim of the previous *ANZECC / ARMCANZ* (1992) guidelines was to protect all forms of aquatic life and all aspects of the aquatic life cycle and this was done using the 'assessment factor' method. The sequence of steps is as follows:

1. Define an appropriate *end-point* (eg. mortality) for the (chronic) toxicity test;
2. Establish the 'no observable effects concentration' or NOEC for a number of test species. The sample NOEC is the highest concentration tested which is not significantly different to the control;

3. Scale the *NOEC* by an arbitrary factor (the 'assessment factor') in an attempt to introduce an added level of protection so as to protect the most sensitive species.

There are difficulties with each of the steps outlined above. The identification of an appropriate end-point has been problematical, although there is agreement that these ought relate to functions of life (mortality, reproduction, growth) rather than behavioural or biochemical characteristics (Holdway 1996, McCarty and Munkittrick 1996). Determination of *NOECs* is usually statistically based (eg. regression or analysis of variance methods) and is also fraught with difficulties (Fox 1999, Chapman et al. 1996) while the arbitrariness of the magnitude of the assessment factor itself (Hart 1974, Nicholson 1984, OECD 1992, 1995, Rand et al. 1995) and lack of scientific underpinning (Warne 1998) has attracted criticism. A number of alternative methods have been developed over the last decade (Stephan et al. 1985, Kooijman 1987, van Straalen and Denneman 1989, Wagner and Lokke 1991, Aldenberg and Slob 1993) in an attempt to alleviate these concerns and provide a more rational and scientifically defensible method of developing water quality criteria. The extent to which these objectives have been met is difficult, if not impossible to assess. What is certain is that these new methods have relied on increasingly more sophisticated statistical methods, although as I have argued elsewhere (Fox 1999), this does not necessarily guarantee a superior outcome. The method by which the new *ANZECC / ARMCANZ* guidelines for toxicants have been derived is an adaptation of the Aldenberg and Slob (1993) approach. The objective is to determine the concentration in either fresh or marine waters for a toxicant such that some high percentage (typically 95%) of *all* species in that environment will be protected. An extra statistical dimension is introduced by attaching a level of confidence to the stated concentration. This is intended to acknowledge the inherent uncertainty in the estimated concentration that arises from using only a small sample of *NOECs* for the analysis. The end result is the somewhat awkward concept of, for example a *95:95* trigger value. The interpretation of this number is that it is a concentration for which it is claimed that 95% of all species will be protected - with 95% confidence. Clearly there are an infinite number of possibilities – *95:50, 50:50 etc.* although the *ANZECC / ARMCANZ* guidelines have used *95:50* as the basis for setting triggers for 'slightly' to 'moderately' disturbed ecosystems. The justification for these choices rests on the following arguments: (i) a 95% level of protection is thought to be sufficient to protect the ecosystem; and (ii) high levels of confidence applied to high protection levels are difficult to defend and tend to produce nonsensical results (eg. extremely low metal concentrations which were often below background concentrations).

In the remainder of this section I shall attempt to provide a simple explanation of the A&S technique. First, consider the derivation of *NOECs* for the test species. There are a number of ways in which this can be done, although perhaps the most common is a statistical analysis of the data obtained from a series of dilution experiments. For example, organisms might be exposed to five concentrations of a chemical toxicant as well as a 'control' in which the toxicant is absent. The *NOEC* is then the highest concentration for which the result at that concentration is statistically indistinguishable from the result for the control. A more rigorous, although more time consuming approach is to characterise the *dose-response* relationship for each species. This is simply a plot of %mortality (say) against concentration. The *NOEC* can then be estimated by extrapolating these curves back to the horizontal axis – that is, the concentration at which zero mortality is observed (Eg Mayer et al., 1994; Lee et al., 1995). By repeating either approach for a (usually small) number of species, a sample of *estimated NOECs* is obtained. A critical assumption of the A&S methodology is that this sample of *NOECs* is from a larger population (see the diagram at the beginning of this article) whose distribution is characterised by a theoretical model called the log-logistic distribution.

The next step in the A&S methodology is to use the imputed log-logistic distribution to estimate the concentration that is exceeded by 95% of *all NOECs.* This concentration is the basis of the *ANZECC / ARMCANZ* trigger value.

The imposition of a confidence level on the estimated trigger value is more complex and will not be described here except to say that it has the effect of reducing the initial 95% value.

In an attempt to overcome some of the limitations of the A&S methodology identified in Fox (1999), Shao (2000) used a more flexible family of probability distributions of which the log-logistic is a member. This modification has been adopted by *ANZECC / ARMCANZ* (Warne, 2001) and embodied in the companion software package *BurrliOz*® which is available for download from CSIRO at:

http://www.cmis.csiro.au/products.html

Monitoring physical-chemical stressors

A significant shortcoming of any omnibus guideline or environmental criterion is the lack of recognition of site-specific conditions and/or anomalies. The revised *ANZECC / ARMCANZ* guidelines have a number of features built into them that attempt to acknowledge regional differences, disturbance category, and ecosystem type. Ecosystems are classified as: upland and lowland rivers; lakes, reservoirs, and wetlands; estuaries and marine. Regional groupings are: south-east Australia (VIC, NSW, ACT, south-east QLD, and TAS); south-west Australia (southern WA); tropical Australia (northern WA, NT, northern QLD); south central Australia — low rainfall area (SA) and New Zealand. Disturbance categories are: high conservation/ecological value (condition 1 ecosystems); slightly or moderately disturbed (condition 2 ecosystems), and highly disturbed (condition 3 ecosystems), each having an associated level of protection. Natural resource managers are further encouraged to adopt site-specific monitoring with acceptable water quality criteria judged relative to local reference site conditions. The use of reference sites is a key feature of the new guidelines. Before outlining the approach it will be instructive to make explicit the competing objectives of any monitoring program. Monitoring is a key requirement of environmental protection, although the objectives are somewhat different for regulators and operators. This subtle, although nonetheless important difference is exemplified by the different emphasis placed on the components of the 'triple bottom line'. For the regulatory agencies, the primary concern is the environment. The economics of environmental protection is to some extent a secondary issue. Their interest is in minimising the *Type II* error, that is that a detrimental impact goes undetected. For industry, environmental performance cannot be de-coupled from economic considerations. An operator seeks to have low *Type I* error, that is that a low/no impact situation is declared detrimental. Statistical inference is inextricably linked with environmental monitoring since this is an application of decision-making under uncertainty. The new *ANZECC / ARMCANZ* guidelines have avoided being prescriptive on this issue and instead have provided an opportunity for the water quality manager to decide on a level of sampling that balances risk with cost.

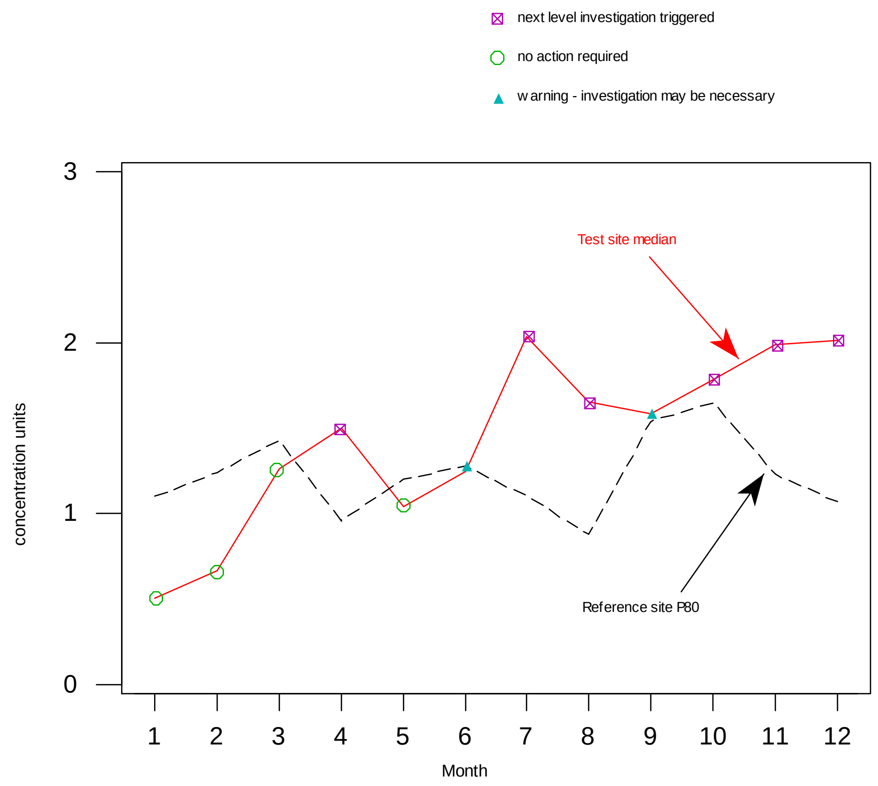
Mindful of the difficulties encountered with the application of 'classical' statistical tests to environmental data (Fox 2001), the *ANZECC / ARMCANZ* team developed a more robust monitoring tool known as the 'P_{80}:P_{50}' comparison procedure. Though its genesis is a little sketchy, the team concluded that the approach

- Is intuitively appealing
- Is simple to implement
- Requires no assumptions to be made about the underlying statistical distributions
- Is robust under a wide range of conditions and environments
- Has acceptable statistical performance characteristics
- Is flexible

Assuming a suitable reference location has been identified (the guidelines provide advice on this issue) it is suggested that at least 24 readings from regular sampling (eg. monthly sampling for two years) be obtained from the reference location. The 80th percentile of the sample of 24 readings (P_{80}) from the reference location becomes the trigger value for the *current* comparison at the test location. The sample *median* (i.e. the 50th percentile, P_{50}) from a series of readings from the test location is used for this comparison. The sample size (i.e. number of readings) at the test location used to compute this median is determined by the analyst and could be as small as n=1, a single reading. This is where the risk trade-offs occur. When n=1, the *Type I* error (probability of a false positive) is 20% although this can be halved (for example) if three samples are used rather than one. Thus the issue for the water quality manager is to balance the consequences of incorrectly triggering further action with the extra cost of sampling. This is a matter of individual (or corporate) utility and cannot be mandated by any regulatory agency. Although the use of the 80th percentile is somewhat arbitrary, the *ANZECC / ARCMANZ* team felt that a median at the test location that was numerically equal to the 80th percentile at the reference location represented a shift worthy of further investigation. This links in with the notion of ecological significance but avoids quantification of this problematic concept. The advantage of the percentile comparison is that it is based on a *relative* change and not an absolute. We have referred to the magnitude of the P_{80}-P_{50} shift as a 'measurable perturbation'. Whether this shift is ecologically significant is another matter entirely, which requires an informed judgement of the consequences of such a shift to the ecosystem. Judging ecological importance is a vexed issue, which is explored further in background material provided in the *ANZECC / ARCMANZ* guidelines.

The control chart – a management tool for water quality managers

It is evident that these quantitative approaches to water quality monitoring and assessment demand a higher level of technical sophistication. However, uptake of these methods can only be assured if managers are provided with easy to use tools that facilitate the decision-making process. To this end, the new guidelines advocate the use of control charts wherever appropriate. Control charts are not new, they were developed under the umbrella of Statistical Process Control (*SPC*) for the manufacturing industries back in the 1930's. Unfortunately their migration from the industrial setting to the environmental setting has been slow to occur and it is only more recently that the utility of these tools for natural resource management has been recognised. Chapter Six of the new Monitoring and Reporting Guidelines (ANZECC & ARMCANZ 2001b) provides more detail and examples of the use of control charts. One important development has been the linking of the reference site – test site comparisons with the control chart. Assuming sampling is conducted on a monthly basis, the initial reference value is obtained as the 80th percentile from the first two years data (as described in the previous section). A *rolling* 80th percentile is used for all subsequent comparisons by dropping the oldest reading and adding the reading for the current month and recomputing the 80th percentile from this set of twenty-four observations. Clearly, this revised estimate will incorporate a significant amount of 'history' which will make it resilient to rapid fluctuations caused by intermittent 'spikes'. It will nevertheless respond to a constant upwards or downwards trend over time. These were seen to be desirable features of a local reference value. The control chart which reflects the monthly comparisons is nothing more than a plot of the rolling 80th



percentile as the reference value and the monthly median at the test site. An example of such a plot appears in the figure.

Concluding Remarks

The new *ANZECC / ARMCANZ* guidelines represent a significant shift in thinking about how water quality in Australia and New Zealand is monitored and assessed. While the National Water Quality Guidelines have undergone substantial development and review it is recognised that the real test of the strategy's efficacy will only occur through implementation and experience. This is about to happen. Practitioners are encouraged to adopt the new methods and report their experiences (both positive and negative) back to *ANZECC/ARMCANZ.* This will help ensure that the guidelines remain relevant, robust, and achieve their aims of balancing the competing risks identified in this article.

References

ANZECC 1992 National water quality management strategy: Australian water quality guidelines for fresh and marine waters. Australia and New Zealand Environment and Conservation Council, Canberra.

ANZECC and ARMCANZ 2001a. Australian and New Zealand guidelines for fresh and marine waters. National Water Quality Management Strategy Paper No 4, Australian and New Zealand Environment and Conservation Council & Agriculture Resource Management Council of Australia and New Zealand, Canberra.

ANZECC and ARMCANZ 2001b. Australian guidelines for water quality monitoring and reporting. National Water Quality Management Strategy Paper No 7, Australian and New Zealand Environment and Conservation Council & Agriculture Resource Management Council of Australia and New Zealand, Canberra.

Aldenberg, T. and Slob, W. 1992 Confidence limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicology and Environmental Safety*, **25**, 48-63.

Chapman, P.M. Cardwell, R.S., and Chapman, P.F. 1996 A warning: NOECS are inappropriate for regulatory use. *Environ. Toxicol. Chem.*, **15**, 77-79.

Fox, D.R. 1999 Setting water quality guidelines: a statician's perspective. *SETAC News*, May 1999, 17-18.

Hart, B.T. 1974 A compilation of Australian water quality criteria. AWRC Technical Paper 77, Australian Government Publishing Service, Canberra.

Holdway, D.A. 1996 The role of biomarkers in risk assessment. *Human Ecol. Risk Assess.*, **2**, 236-267.

Koojiman, S.A.L.M. 1987 A safety factor for LC_{50} value allowing for differences in sensitivity among species. *Water Research*, **21**, 269-276.

OECD (Organisation for Economic Cooperation and Development) 1992. Report of the OECD workshop on extrapolation of laboratory aquatic toxicity data to the real environment. OECD Environmental Monographs 59, OECD, Paris.

OECD (Organisation for Economic Cooperation and Development)

1995. Guideline document for aquatic effects assessment. OECD Environmental Monographs 92, OECD, Paris.

Mayer, F.L., Krause, G.F., Buckler, D.R., Ellersieck, M.R., and Lee, G. 1994 Predicting chronic lethality of chemicals to fishes from acute toxicity tests data: concepts and linear regression analysis. *Environ. Toxicol. Chem.*, **13**, 671-678.

Lee, G., Ellersieck, M.R., Mayer, F.L., and Krause, G.F. 1995 Predicting chronic lethality of chemicals to fishes from acute toxicity tests data: Multifactor probit analysis. *Environ. Toxicol. Chem.*, **14**, 345-349.

McCarty, L.S. and Munkittrick 1996 Environmental biomarkers in aquatic toxicology: Fiction, fantasy or fictional? *Human Ecol. Risk Assess.*, **2**, 268-274.

Rand, G.M., Wells, P.G., and McCarty, L.S. 1995 Introduction to aquatic toxicology. In Fundamentals of Aquatic Toxicology, Effects, Environmental Fate and Risk Assessment. 2Editn. Rand, G.M. Ed. Taylor and Francis, Washington D.C. USA pp 3-67.

Van Straalan, N.M. and Denneman, C.A.J. 1989 Ecotoxicological evaluation of soil quality criteria. *Ecotoxicol. Environ. Saf.* **18**, 241-251.

Shao, Q. 2000 Estimation for hazardous concentrations based on NOEC toxicity data: an alternative approach. *Environmetrics*, **11(5)**, 583-595.

Stephan, C.E., Mount, D.I., Hansen, D.J., Gentile, J.H., Chapman, G.A., and Brungs, W.A. 19895 Guideliens for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. USEPA PB 85-227049.

Wagner, C. and Lϕkke, H. 1991 Estimation of ecotoxicological protection levels from NOEC toxicity data. *Water Research*, **25**, 1237-1242.

Warne, M.StJ. 1998 Critical review of methods to derive water quality guidelines for toxicants and a proposal for a new framework. Supervising Scientist Report 135, Supervising Scientist, Canberra.

Warne, M.StJ. 2001 Derivation of the ANZECC and ARMCANZ water quality guidelines for toxicants. *Australas. J. Ecotoxicol.*, In press.

Acknowledgements

The author is grateful for the helpful advice and feedback from a number of colleagues in the preparation of this article. In articular, Dr. Graeme Batley (CSIRO Energy Technology), Dr. Chris Humphrey (ERISS), Dr. Barry Hart (Monash University), Dr. John Chapman (NSW EPA), Dr. Bill Maher (Canberra University), Mr. Kevin McAlpine (WA DEP), Dr. Leon Barmuta (University of Tasmania), and Dr. Michael Warne (NSW EPA). Thanks are also due to Bob Swinton for his editorial assistance.

About the Author

Dr. David Fox is a Program Leader in CSIRO's Division of Land and Water. He received his PhD. in Statistics from the University of Wyoming and was recipient of the *Deming Award* for distinguished performance in his PhD studies. He is a Chartered Statistician and Fellow of the Royal Statistical Society, London as well as a member of the American Statistical Association. He was joint recipient of CSIRO's Chairman's Gold Medal in 1996 for his contributions to the Port Philip Bay Environmental Study. He has held research and teaching positions in statistics departments in Australia, United Kingdom, and the United States. David is also Director of the new Adelaide Coastal Waters Study and Director of the Boags Rocks Outfall Monitoring Project.

Contact: (08) 9333-6625
david.fox@per.clw.csiro.au