SOME STATISTICAL CONSIDERATIONS IN THE ASSESSMENT OF COMPLIANCE

S.R. Esterby

NWRI Contribution No. 89-26

Lakes Research Branch
National Water Research Institute
Canada Centre for Inland Waters
Burlington, Ontario L7R 4A6
Canada

May 1989

CONSIDÉRATIONS STATISTIQUES DANS L'ÉVALUATION DE LA CONFORMITÉ

S.R. Esterby

Direction de la recherche sur les lacs

Institut national de recherche sur les eaux

Burlington (Ontario) Canada

RÉSUMÉ

On a évalué par surveillance la conformité aux critères limitant le déversement de polluants ou aux normes de qualité de l'eau dans le milieu récepteur. Afin d'établir des limites ou des normes, il faut déterminer les caractéristiques du déversement ou du plan d'eau à surveiller. Cette démarche comporte généralement l'ajustement d'une distribution des probabilités aux données chronologiques ou à celles relevées lors d'un échantillonnage préliminaire, ainsi que la sélection d'une valeur statistique pour la limite ou la norme. On suppose que la distribution des données de surveillance recueillies pour évaluer la conformité est semblable à celle des

données chronologiques ou préliminaires. Une caractérisation correcte de ces dernières contribuera à assurer que l'hypothèse est confirmée. Les méthodes statistiques qui supposent une distribution pour la variable de qualité, ou comportent l'emploi, après transformation, d'une variable binaire, sont comparées. On examine la validité des hypothèses sous-jacentes dans l'application des méthodes aux données sur la qualité de l'eau ou des effluents.

SOME STATISTICAL CONSIDERATIONS IN THE ASSESSMENT OF COMPLIANCE

S.R. Esterby
Lakes Research Branch
National Water Research Institute
Burlington, Ontario, Canada

ABSTRACT

Compliance with criteria limiting the discharge of pollutants or standards of water quality in the receiving water body are assessed by monitoring. In order to set limits or standards, the features of the discharge or the water body to be monitored must be characterized. This generally involves the fitting of a probability distribution to historical data or data from preliminary sampling and the choice of a statistic for the limit or standard. Monitoring data collected to assess compliance are assumed to follow the same distribution as that of the historical or preliminary data. Proper characterization of this latter data will help to ensure that the assumption is met. Statistical methods which either assume a distribution for the quality variable or involve a transformation to a binary variable are compared. The validity of the underlying assumptions in the application of the methods to water or effluent quality data is discussed.

INTRODUCTION

Consider the situation where samples are taken to assess whether an acceptable level of quality is being maintained. This involves the setting of a standard or a limit against which the results of the sampling are compared. When the samples do not provide evidence that the source being sampled is at variance with the standard, the source is considered to be in compliance. There are a number of ways in which the standard or limit may be defined and these include external considerations, such as the level which makes the water acceptable for a specific use, and values determined from either prior sampling of the source to be assessed or another source, for example, one considered to provide background levels. It is clear that the definition of the standard or limit involves a measure of location. Due to the inherent variability of effluent or water quality parameters, a measure of the variability must also be incorporated into the method of assessment. A single probability distribution may adequately characterize this variability. However, there will often be structure within a data set or a concomitant variable that will account for some of the variability, and then a model involving both deterministic and random components will be needed.

Methods based on the number of times a quality parameter exceeds a limit and methods which use the distribution of a measure of location of the quality parameter directly have been applied in the assessment of compliance. These will be discussed using examples from the literature. For the methods using the quality parameter distribution directly, one example takes the standard as given, while the other accounts for the variability in both background data and samples taken for the assessment of compliance. The assumptions underlying the methods of analysis will be discussed.

BINOMIAL VARIABLES

An application of the theory of hypothesis testing and confidence limits, known as sampling inspection or acceptance sampling (e.g., Brownlee, 1965 or Mandel, 1967), has become standard methodology in the control of the quality of manufactured items. This is based on the binomial distribution. The number of times a limit is exceeded in water or effluent quality assessment has also been treated as a binomial random variable (Warn and Matthews, 1984; Ellis, 1985; Crabtree et al., 1987). The use of the binomial distribution in sampling inspection is briefly reviewed and then the applicability of these methods to the assessment of the compliance of effluent or water quality parameters is considered. This comparison permits clarification of a difficulty expressed by the above mentioned authors. The role of the distribution of the quality variable is shown and estimation of a percentile is considered.

Simple Sampling Inspection for Attributes

These methods have been applied in industry, where items in a manufacturing process can be sampled and classified as defective or nondefective. A random sample of n items is inspected and X, the number of defective items, is compared with a specified standard x_0 (e.g., Brownlee, 1965). If the sample size, n, is small relative to the total number of items, then X follows a binomial distribution with parameters n and p, where p is the fraction of defective items in the batch from which the sample was drawn. The probability that the process is found to be in compliance is

$$P(X \leq x_0) = \sum_{x=0}^{x_0} \binom{n}{x} p^x (1-p)^{n-x} \qquad (1)$$

By specifying the acceptable fraction defective, p_0, and a fraction defective, p_1, which is considered to be relatively bad, a sampling plan determining n and x_0 in (1), can be found for given type 1 error, α, and type 2 error, β. That is, n and x_0 are chosen so that the probabilities α, the probability of finding the process out of compliance when the fraction defective is acceptable, i.e., $p = p_0$, and β, the probability of finding the process in compliance when too many items are defective, i.e., $p = p_1$, are as specified. Alternatively, the operating characteristic curve, a plot of A(p) versus p, where

$$A(p) = P\ (X \leq x_0;\ n,\ p) \tag{2}$$

can be used to determine how a specific combination of n and x_0 perform for different values of p.

Binomial Methods for Quality Variables

Let Y be the concentration of the quality parameter in a particular effluent or water body, and L, the upper limit for acceptable concentrations. For example, L might be the 95th percentile determined from a large number of prior samples from the source being assessed. In general, for L defined so that p in the expression

$$P\ (Y > L) = p \tag{3}$$

is known, the continuous variable Y, can be transformed to a binary variable, Z, where

$$P\ (Z = 1) = P\ (Y > L) = p$$

and (4)

$$P\ (Z = 0) = P\ (Y \leq L) = 1 - p$$

Thus the probability, that x out of n samples exceed L, is given by the binomial distribution

$$P\,(X = x) = \binom{n}{x}\, p^{x}\, (1 - p)^{n-x} \tag{5}$$

provided that the assumptions of independence and constant p are met.

Regardless of the distribution of Y, the distribution of X is binomial under the assumptions given above. However, the distribution of Y determines the value of L. If L is taken to be the 95th percentile of the distribution of Y, then L is determined from (3) with $p = 0.05$. For p fixed and small, differences between the values of L for different distributions depends on the characteristics of the distributions in the right tail. The symmetric normal and asymmetric lognormal distributions are compared for several values of the means and variances in Table 1. Let L_N be the 95th percentile if Y is normally distributed with mean μ and variance σ^2 and L_ℓ, the 95th percentile if Y is lognormally distributed with mean μ_ℓ and variance σ_ℓ^2. The relationship between μ, σ^2 and μ_ℓ, σ_ℓ^2 is given, for example, by Aitchison and Brown (1981). The lognormal distributions, with means and variances as given in rows two and three of Table 1, are highly skewed and that, in row one, nearly symmetric, yet the 95th percentiles differ by at most two. The value of L differs by 6.5 between a lognormal distribution with $\mu_\ell = 10$, $\sigma_\ell = 15$ and a normal distribution with $\mu = 10$ and $\sigma = 18$. This latter case shows how the estimate of L is affected by assuming a normal distribution when a lognormal distribution is appropriate. To use a higher variance for the normal distribution is a realistic case and the value $\sigma_\ell = 18$ used here equals the sample variance of a random sample from a lognormal distribution with $\mu_\ell = 10$ and $\sigma_\ell = 15$. As another example, the effect of the assumed form of the distribution on the probability of compliance with bacterial water quality guidelines has been shown by Esterby (1982).

The binomial distribution with L defined as the 95th percentile has been used by some Water Authorities in the United Kingdom to assess the compliance of effluents (Warn and Matthews, 1984 and Ellis, 1985). Warn and Matthews cited practical problems with this method, including published yearly failure rates (i.e., the fraction of samples in a year which exceed L) which are higher than the actual 5 percent rate, p, and opted for a method which uses the quality variable directly. Ellis discussed the balance between type 1 and type 2 errors and proposed lowering the type 1 error to keep the failure rate down. The inconsistency between these rates was shown numerically by both Warn and Matthews (1984) and Ellis (1985), and stems from trying to equate the probability of an individual sample exceeding L,

$$p = P\ (Y > L) = 0.05 \tag{6}$$

to the probability of more than 0.05 n of the samples exceeding L,

$$P\ (X > 0.05\ n) = \sum_{x=x_0}^{n} \binom{n}{x} (0.05)^x\ (0.95)^{n-x} \qquad (7)$$

where x_0 is the smallest integer greater than 0.05 n. This results in P(X > 0.05 n) = 0.05, which is satisfied when n = 1, that is, only one sample is collected during the year. Ellis notes this in stating that the legal definition of compliance relates to an individual sample.

The difficulty arises from taking the objective, that fewer than 5 percent of the samples exceed the limit L, to mean that L should be the 95th percentile of the distribution of the quality variable Y. If the objective is to have the probability, of more than 0.05 n samples exceeding L, equal to 0.05, then instead of taking L to be the 95th percentile, the value of L which satisfies

$$0.05 = P\ (X > 0.05\ n) \qquad (8)$$

should be determined. Writing equation (7) in terms of p gives

$$P(X > 0.05\ n) = \sum_{x=x_0}^{n} \binom{n}{x}\ p^x\ (1-p)^{n-x} \qquad (9)$$

From equation (9) and, for given n, the value of p satisfying (8) can be determined, and, using this value of p in (3), the value of

L, the (1-p)th percentile of the distribution of Y, can be obtained. For example, if n = 26, equation (8) becomes

$$0.05 = P\{X > 1.3\} = P\{X \geq 2\}$$

which is satisfied if p = 0.0138. Assuming Y is distributed $N(10,10^2)$, the 1-p = 0.9862 percentile is 32. Compare this with the 0.95 percentile of 26.45.

In the preceeding paragraphs of this section, the probability of events and the values of percentiles have been calculated assuming the probability distribution is known. This is useful in designing a program but once data is collected, quantities must be estimated. Estimation of the 95th percentile from water quality data sets was considered by Crabtree _et al_. (1987) and the discussion applies equally to other percentiles. The authors compared estimates obtained using parametric and nonparametric estimates on water quality data sets. The nonparametric estimates consist of sample quantiles but the convention used for their determination was not stated. Freund and Perles (1987) give three alternatives that can be used for determining the position of quantiles by interpolation and these are applicable for any quantile including percentiles. Their second alternative would determine the position of the 95th percentile at 0.5 + 0.95 n and thus would be the 48th observation, for observations in ascending order, in a sample of size n = 50.

METHODS FOR CONTINUOUS VARIABLES

Confidence Interval for the 95th Percentile

Warn and Matthews (1984) use confidence limits for the 95th percentile determined from the samples taken for the assessment of compliance and compare these limits with a predetermined limit L. The discharge is considered out of compliance if L is less than the lower confidence limit, in compliance if L is greater than the upper confidence limit, and unresolved if the confidence interval includes L. The water or effluent quality variable Y, as observed or suitably transformed, is assumed to be normally distributed.

Let $\bar{y}$ and s_y^2 be the mean and variance calculated from the sample of size n, taken to assess compliance. The confidence interval for, L, the (1-p)th percentile of the distribution of Y, is obtained by noting that $\sqrt{n}k$, where

$$k = (L - \bar{y})/s_y \tag{10}$$

follows the non-central t distribution, denoted by t', with degrees of freedom $\nu = n - 1$ and non-centrality parameter $\Delta = \sqrt{n}\, K$, for K given by the (1-p)th percentile of the N(0,1) distribution (Pearson and Hartley, 1976).

Comparison with the Binomial Method

Consider the case where it is of interest to know only whether the sample provides evidence that the source is out of compliance. The test is then one of the null hypothesis H_0, that the true 95th percentile equals L_0 against the alternative, H, that it equals L_1 where $L_1 > L_0$. The performance of the two methods can be compared by calculating the power of the test, that is, the probability of finding the source out of compliance when in fact its true 95th percentile is larger than the limit L_0. The distribution of Y is assumed to be $N(\mu_0, \sigma_0^2)$ and $N(\mu_1, \sigma_1^2)$ under H_0 and H_1, respectively.

For significance level α, the test of the null hypothesis based on Y provides t'_0 such that

$$P\ (t' < t'_0;\ \nu_0,\ \Delta_0) = \alpha \qquad (11)$$

where $t' = \sqrt{n}k$ has $\nu_0 = n\text{-}1$ and $\Delta_0 = \sqrt{n}K$. The significance level, α, is an area in the left tail of the distribution because values of k, smaller than expected assuming the limit equals L_0, provide evidence against H_0. Under H_1, it can be shown that

$$\sqrt{n}k = \{-\sqrt{n}\ (\frac{\bar{y} - \mu_1}{\sigma_1}) + \sqrt{n}\ (K + (L_0 - L_1)/\sigma_1)\} + s_y/\sigma_1 \qquad (12)$$

and thus $\sqrt{n}k$ follows a non-central t distribution with $\nu_1 = n\text{-}1$ and $\Delta_1 = \sqrt{n}\ (K + (L_0 - L_1)/\sigma_1)$. The power of the test is given by

$$\pi = P\ (t' < t'_0;\ \nu_1,\ \Delta_1) \tag{13}$$

Tables for the non-central t are available (Pearson and Hartley, 1976, Table 26) for 8 levels of the tail probability. An analytical solution of a quantity ℓ, required for use of these tables, can be obtained by solving a quartic equation. For values of power calculated here a very simple computer search calculating t' as a function of ℓ was used instead.

The test of the null hypothesis, based on the binomial variable X, provides x_0 such that

$$P\ (X > x_0;\ n,\ p_0) = \alpha \tag{14}$$

where $p_0 = P\ (Y > L_0)$. Since X is discrete, there is generally no x_0 satisfying (13) for a specified α. It will often be appropriate to use the inequality

$$P\ (X > x_0;\ n,\ p_0) \leq \alpha \tag{15}$$

instead since this reduces that the risk of finding a source out of compliance. Thus x_0 will be one less than the smallest x for which (14) exceeds α, if an exact solution of (14) is unavailable. Having determined x_0, the power is

$$\pi = P\ (X > x_0;\ n,\ p_1) \tag{16}$$

where $p_1 = P\ (Y > L_0)$ for Y assumed $N(\mu_1, \sigma_1^2)$.

A loss of power would be expected in the transformation from a continuous to binary variable in view of results known for other tests. For example, 50 pairs of observations are required for the sign test to have the same power as the paired t-test with 32 pairs, assuming normality (Snedecor and Cochran, 1980, p. 140). The power of the tests as described above have been calculated for three examples (Table 2), chosen to be reasonable for the assessment of compliance in effluents. The values of the means and variances are consistent with those of effluent BOD reported by Adams and Gemmell (1973) and the number of samples n = 26 and 91 correspond to samples taken every two weeks and every fourth day over one year, intervals long enough to expect no serial correlation (Berthouex and Hunter, 1975). Power calculated using the binomial distribution and the χ^2 approximation (Brownlee, 1965) are in reasonable agreement. The results using the non-central t are approximate because linear interpolation within the table for one particular tail probability and graphical interpolation between these tables were used. Considerable loss in power occurs by transforming to a binary variable. However, in any application this must be assessed against the suitability of a normal assumption.

Prediction Intervals

In the previous section, a confidence interval for a percentile was obtained from a sample collected to assess compliance and this interval was compared with a preassigned value L. In practice, L is an estimate of the percentile obtained from prior sampling when the sample size is considered large enough to treat L as known. If the limit was determined from sampling the same source as that being monitored for compliance, then the change to be detected is a shift in level. Further, all samples collected over a year were used for a single test of compliance.

In the case of a prediction interval, the background data is used to form an interval which will be compared with future samples collected to assess compliance. Davis and McNichols (1987) describe a situation in the monitoring of groundwater in the vicinity of hazardous waste management facilities and describe a method for constructing an interval $(-\infty, \bar{Y} + KS_y)$, which will contain at least q of m observations on each of r future occasions with probability $(1 - \alpha)$. The mean, $\bar{Y}$, and standard deviation, S_y are calculated from the background sample of size n. The variance of an individual sample is assumed to be the same for both background and future samples. The purpose of the monitoring is to detect a shift in level, or equivalently, the condition that the source is out of compliance relative to the background level. Their procedure has the feature that compliance can be assessed at each sampling occasion, where there are r sampling occasions and m samples on each occasion for a total of rxm samples, and the overall risk of a type I error for all these tests is controlled at the α level. There is a correspondence between methods based on percentiles and based on at least q of m observations, since the qth order statistic $y_{(q)}$ is the 100 q/m percentile, for q/m an integer.

Davis and McNichols give limited tables of the value of K and outline the algorithm for the computation of K for specified α, n, r, q and m. The sample sizes are small compared with those used in effluent monitoring with n = 10, 15, 20, m $\leq$ 6 and r = 1,2,4,8,16. Under semiannual sampling, as used in the groundwater monitoring situation which motivated their paper, r = 16 covers 8 years. However, the procedure could be applied with the total number of samples in a year being divided into r sets of m samples, and this would allow for the detection of non-compliance within the year.

ADEQUACY OF A SINGLE PROBABILITY DISTRIBUTION

The methods discussed above are based on the assumption that the observations made on a quality variable, while the process is in control or the water body is receiving a constant load of pollutant, can be characterized as a sample of independent observations from a single probability distribution. Further some methods also required the assumption that the quality variable follows a normal distribution.

The assumption of normality can often be satisfied by an appropriate transformation such as the logarithmic transformation. It may be more appropriate, however,to model the quality variable, Y, as the sum of a structural component, which would account for variability due, for example, to seasonality, and an error component which might then be adequately described by a normal distribution, although a transformation will sometimes still be required. That is

$$Y_i = \mu + f(t_i) + \epsilon_i \tag{17}$$

where $f(t_i)$ is an appropriate function of time.

Berthouex and Hunter (1975), in discussing treatment plant monitoring programs, note that lack of normality may be the result of including observations from periods when the plant was out of control. If a normal distribution is appropriate when the plant is in control, then it can still be used for control charts since the latter are not used when the plant is out of control. A transformation or alternate distributional assumption will be necessary if nonnormality is present while the plant is in control. The point made here is that a statistical distribution is being used to characterize the data when the plant is operating under a particular set of conditions. If data is included from a period when the plant is operating under another set of conditions, then the combined data does not represent the first condition only. A similar argument could be applied to water quality data. Berthouex and Hunter make the further point that because of the central limit theorem, statistics such as the mean tend to satisfy the normality assumption when the individual observations do not.

Equally as important as the adequacy of the assumptions about the probability distribution, is the elimination of effects which obscure the difference that the procedure is trying to detect. Davis and McNichols (1987) give careful consideration to the assumptions underlying the method of analysis and not only transform the original data, but also use differences between approximately simultaneous upgradient and downgradient samples to eliminate effects of seasonality, temperature and sample-handling methods. Examples from water quality studies are the fitting of a seasonal cycle (El-Shaarawi et al., 1983), the pairing of stations to eliminate seasonal and other time-varying effects (El-Shaarawi et al., 1985) and the spatial zonation of a lake to account for heterogeneity (Esterby and El-Shaarawi, 1984). Crabtree et al. (1987) found that only half of the 334 sets of water quality data analyzed, where each set consisted of either daily or monthly samples for an entire year for periods up to 3 years, could be fitted by one of three distributions (normal, lognormal or Pearson type 3). Although the authors did not comment on structural features which may have made distribution fitting difficult, this is a possibility.

One practical approach to the lack of independence is to avoid it by spacing observations far enough apart. Berthouex and Hunter (1975) suggest that sampling wastewater treatment plants every third or fourth day should avoid the problem and every fourth day (Berthouex et al., 1981) has other advantages. Similarly, Davis and McNichols (1987) use adequate spacing as a method of avoiding serial correlation of observations. Aggregation is an alternative and van Belle and Hughes (1984) discuss the use of means or medians in the analysis of water quality data. As noted in an earlier section, the assumption of independence is made when the binomial distribution is used. When closely spaced observations are required, models may have to be modified to include serial correlation.

ACKNOWLEDGEMENTS

The author would like to thank D. Chapman for providing some useful references.

REFERENCES

Adams, B.J., and Gemmell, R.S.: 1973, 'Performance of Regionally Related Wastewater Treatment Plants', Jour. Water Poll. Control Fed. 45, 2088-2103.

Aitchison, J., and Brown, J.A.C.: 1981, The Lognormal Distribution, Cambridge University Press, Cambridge, 176 pp.

Berthouex, P.M., and Hunter, W.G.: 1975, 'Treatment Plant Monitoring Programs: A Preliminary Analysis', Jour. Water Poll. Control Fed. 47, 2143-2156.

Berthouex, P.M., Hunter, W.G., and Pallesen, L.: 1981, 'Wastewater Treatment: A Review of Statistical Applications', Environmetrics 81: Selected Papers, SIAM, Philadelphia, pp.77-99.

Brownlee, K.A.: 1965, Statistical Theory and Methodology, Wiley, New York, 590 pp.

Crabtree, R.W., Cluckie, I.D., and Forster, C.F.: 1987, 'Percentile Estimation for Water Quality Data', Wat. Res. 21, 583-590.

Davis, C.B., and McNichols, R.J.: 1987, 'One-Sided Intervals for at Least p of m Observations from a Normal Population on Each of r Future Occasions', Technometrics 29, 359-370.

Ellis, J.C.: 1985, Determination of Pollutants in Effluents, Environment TR230, Water Research Centre, Marlow, UK, 46 pp.

El-Shaarawi, A.H., Esterby, S.R., and Kuntz, K.W.: 1983, 'A Statistical Evaluation of Trends in the Water Quality of the Niagara River', J. Great Lakes Res. 9, 234-240.

El-Shaarawi, A.H., Esterby, S.R., Warry, N.D., and Kuntz, K.W.: 1985, 'Evidence of Contaminant Loading to Lake Ontario from the Niagara River', Can. J. Fish. Aquat. Sci. 42, 1278-1289.

Esterby, S.R.: 1982, 'Fitting Probability Distributions to Bacteriological Data: Considerations for Regulations and Guidelines', Journal Francais d'Hydrologie 13, 189-203.

Esterby, S.R., and El-Shaarawi, A.H.: 1984, 'Coliform Concentrations in Lake Erie - 1966 to 1970', Hydrobiologia 111, 133-146.

Freund, J.E. and Perles, B.M.: 1987, 'A New Look at Quartiles of Ungrouped Data', The American Statistician 41, 200-202.

Mandel, J.: 1967, The Statistical Analysis of Experimental Data, Interscience Publishers, Wiley, New York, 410 pp.

Pearson, E.S. and Hartley, H.O.: 1976, Biometrika Tables for Statisticians, Vol. 2, Charles Griffin and Company Ltd., Buckinghamshire, 385 pp.

Snedecor, G.W., and Cochran, W.G.: 1980, Statistical Methods, The Iowa State University Press, Ames, Iowa, 507 pp.

Warn, A.E., and Matthews, P.J.: 1984, 'Calculation of Compliance of Discharges with Emission Standards', Wat. Sci. Tech. 16, 183-196.

van Belle, G., and Hughes, J.P.: 1984, 'Nonparametric Tests for Trend in Water Quality', Water Resour. Res. 20, 127-136.

Table 1. Comparison of the 95th percentile for some normal and lognormal distributions.

Normal			Lognormal		
μ	σ	L_N	μ_ℓ	σ_ℓ	L_ℓ
35	10	51.5	35	10	53.4
10	10	26.5	10	10	27.8
10	15	34.7	10	15	33.1
10	18	39.6			

Table 2. Comparison of the power of the test of the hypothesis that the 95th percentile is L_0 against the alternative that it is L_1 using Y and the binomial variable X.

H_1			Non-central t				Binomial					
										Exact		χ^2
n	μ_1	σ^2_1	Δ_0	Δ_1	α	π	p_0	p_1	x_0	α	π	π
26	12.5	10^2	8.39	7.11	0.05	0.23	0.05	0.082	3	0.039	0.16	0.17
91	12.5	10^2	15.69	13.31	0.05	0.51	0.05	0.082	8	0.038	0.32	0.33
91	15.0	10^2	15.69	10.92	0.05	0.96	0.05	0.126	8	0.038	0.83	0.80

Under the null hypothesis $Y \sim N(10, 10^2)$ and $L_0 = 26.45$.
$L_1 = 28.95$ and 31.45 for $\mu_1 = 12.5$ and 15.0, respectively.