Environmental and Ecological Statistics 9, 379–392, 2002

A Bayesian approach to evaluating site impairment

KEYING YE and ERIC P. SMITH

Department of Statistics, Virginia Tech, Blacksburg, VA 24061
E-mail: epsmith@vt.edu

Received December 1999; Revised October 2001

In the United States, each state is required to list water resources that are declared to be impaired under guidelines set by the Clean Water Act. Measurements are typically collected on a number of chemical constituents and compared with a standard. If there are too many measurements exceeding the standard, then the site is declared impaired. The approach is non-statistical but similar to a Binomial test. The Binomial approach would convert the measurements to binary data then test if the proportion exceeding the standard is excessive. Both methods convert measurements to binary values hence exclude potentially important information in the data. We present a statistical approach using a Bayesian model that uses the raw data instead of the binary transformed data. The population distribution of a family of location-scale parameter models is studied under the model. Posterior distributions from the Bayesian analysis are used in the decision-making process and error probabilities for the Bayesian and the Binomial approaches are compared for a normal population.

Keywords: hypotheses testing, location-scale parameter model, mean squared errors, standards, reference priors, posterior distribution, Type I error probability, Type II error probability

1. Introduction

To assess violations of water quality standards, Section 303 (d) of the Clean Water Act mandates that states undertake an assessment of, and then report on, the condition of specific segments (a lake, bay or river) of water within the state. For river and stream systems, reports are based on information collected at a large number of monitoring locations, each associated with a different site. Reports describe the condition of stream segments with each segment typically associated with a single sampling station or site. The monitoring typically occurs on a quarterly (or more frequent) basis and the assessment is based on two or more years of data. Each sample from a water segment is assumed to represent a background population of water quality conditions. The assessment challenge is to interpret a limited amount of sample data to determine whether stream conditions are violating standards more than 10% of the time at each site, recognizing that the sample measurements are taken from a population of stream conditions affected by variability in human activity and natural (or background) conditions.

For the evaluation of stream condition, Virginia's Department of Environmental Quality

(DEQ) has proposed that data interpretation is best addressed as a classical problem of statistical inference and has employed a well recognized statistical approach as part of the data interpretation (see Commonwealth of Virginia, 1997, Appendix D or http://www.deq.state.va.us/water/305b.html). Specifically, the DEQ has chosen the "Binomial" approach as the appropriate tool for assessment. The approach is based on estimating p, the true probability of the chemical measurement exceeding the standard. The Binomial distribution is used to conduct a statistical test of whether this p is more than 10% or not. A site is declared impaired if there is sufficient evidence to declare p is bigger than 10%, based on the sample proportion.

On the other hand, the U.S. Environmental Protection Agency (EPA) recommends a simpler method we refer to as the "raw score" analysis. If 10% or more of the sample measurements exceed the standard, the site would be declared impaired. A statistical representation of the raw score approach can be made with reference to Fig. 1. Suppose that a standard of 2.5 units is given for certain water quality variable. When the site is not impaired, a possible distribution of the water quality variable may be drawn as shown with the 90th percentile of the population distribution less than or equal to 2.5. If this is the case, the water quality standard requires that a concentration of 2.5 should not be exceeded more than 10% of the time. On the other hand, if a site is in violation of the standard there will be some number of samples that will exceed the 2.5 standard and the true 90th percentile of the distribution from the data will exceed this standard. Note that this conceptual model is reasonable for measurements that fluctuate naturally (e.g., dissolved oxygen). For other quantities (e.g., synthetic chemicals such as PCB's) more stringent views are required.

Both of the aforementioned methods use truncated data information, namely binary data. A data point is recorded as either exceeding the standard or not. One concern is the information loss due to the transformation. In this article, we try to use the raw data directly in the assessment process and will focus on the problem of the evaluation of the quality of water in a stream segment based on information from a single site. A Bayesian approach is used to develop a decision process. In Sections 2 and 3, we will set up the Bayesian approach and discuss certain properties of the method when the underlying

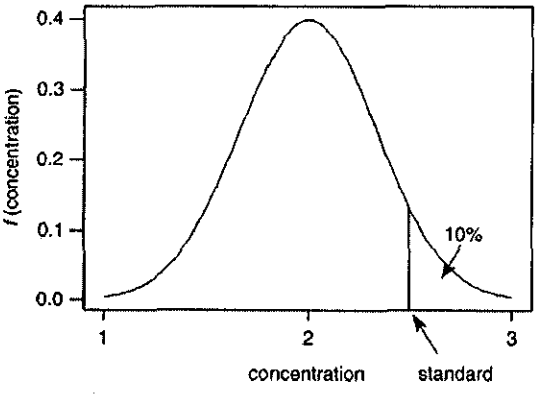

Figure 1. Plot of the distribution of a hypothetical chemical concentration. The standard value is set so 90% or more of the concentrations are less than the standard if the site is to meet the use designations required by law.

distribution comes from a location-scale parameter model. In Section 4, we discuss and compare the methods using the normal distribution.

2. Location-scale models

2.1 *The hypotheses and Bayesian test*

Suppose that in a particular spatial and temporal region, the response variable (e.g., dissolved oxygen, fecal coliform, etc.) follows a certain probability distribution. Denote by θ the desired percentile of the distribution (e.g., 90th percentile) that is of interest. Further assume that the standard is known and denoted by θ_0. This value represents the known percentile for the distribution assuming the site is healthy. We assume this to be an upper percentile although similar results follow for lower percentiles. In the following we consider a family of location-scale parameter models that includes popular distributions such as the normal and student's T distributions.

Let $X_1, \ldots, X_n$ be a random sample of measurements on a single constituent which are independently and identically distributed with the form of density function

$$f(x|\mu,\sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right), \tag{1}$$

where μ and σ are the population mean and standard deviation, respectively. Since we are interested in the ξth percentile, θ, rather than the mean, we need to derive the relation between θ and (μ,σ). Define $Z = (X-\mu)/\sigma$. The density of Z has the form $f(z)$ which is free from any parameters. Hence, denoting by p_ξ the ξth percentile of Z, we obtain

$$\theta = \mu + p_\xi \sigma, \tag{2}$$

since $\xi = P(X \le \theta) = P(Z \le (\theta-\mu)/\sigma)$. To check the impairment of the monitoring site, rather than using EPAs approach that declares impairment if at least 10% of the observations exceed the standard θ_0, we use a statistical tool to declare impairment if the true percentage of the population distribution exceeds the standard. In other words, we test if $\theta > \theta_0$. This is the approach in acceptance sampling by variables (Duncan, 1974). Here in the form (2), ξ in the parameter θ is usually predetermined. To match the EPAs 10% rule, we use $\xi = 0.9$. Therefore, we need to use the data to test

$$H_0 : \theta \le \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0. \tag{3}$$

To test the hypotheses in (3), one may consider rejecting the null hypothesis when the posterior probability of the alternate hypothesis given the data is large, i.e.

$$P^{\pi(\theta,\sigma|\underline{x})}(H_1) > q, \tag{4}$$

where q is a predetermined number. In form (4), $P^{\pi(\theta,\sigma|\underline{x})}(H_1)$ stands for the posterior probability of the alternative hypothesis H_1, given that the data is observed.

2.2 *An example with a normal population*

The analysis will consider data as arising from a normal distribution. Although the distribution of many measurements is not normal, it is quite common that under certain transformation of the variables, the underlying distribution of the measurements may approximately follow a normal distribution (see for example Wild *et al.* (1996)). The properties of the statistical tests using the Bayesian approach under a normal population will be discussed in detail in Sections 3 and 4. In this section, we describe the test in (3) where the data is a sample from normal population. Further, we will apply the method to a real data set.

Suppose that $X_1, \ldots, X_n$ is a random sample from a normal distribution and we want to test the hypotheses in (3). Using the Bayesian approach described in Section 3, we evaluate the null hypothesis using the statistic

$$T = \sqrt{n}\frac{\overline{X} - \theta_0}{S_X},$$

where $\overline{X}$ is the sample mean and S_X the sample standard deviation. The hypothesis is rejected if $T > t_0$ where t_0 is the critical value for standards at the upper tail of the distribution and $T < -t_0$ for standards based on the lower tail. Critical values for T based on small sample sizes are provided in Table 1.

Table 2 provides an example of dissolved oxygen data collected at one site over a two-year period. Using the raw score approach (EPA's method) with a standard of 5.0 (low dissolved oxygen is not good), we find that three observations are below the standard. Since $N = 24$ the site would be declared as impacted by the EPAs raw score approach. If a Binomial model with $p = 0.10$ is assumed, a value of 3 is actually quite likely (the probability of observing 3 or fewer values below the standard of 5.0 is 0.786). Hence, declaration of non-impairment would be made using this approach. On the other hand, using the above test with sample mean = 7.03 and standard deviation of 1.787, the value

Table 1. Cutoff values for the non-central t statistic T when the value q is set at 0.9 and the standard is above the mean. In the table, n is sample size and t_0, which is described in Result 6, is the critical value associated with the given sample size.

n	2	3	4	5	6	7	8	9
t_0	−0.569	−0.926	−1.234	−1.510	−1.762	−1.996	−2.216	−2.423
n	10	11	12	13	14	15	16	17
t_0	−2.620	−2.808	−2.988	−3.161	−3.328	−3.490	−3.647	−3.799
n	18	19	20	21	22	23	24	25
t_0	−3.947	−4.091	−4.232	−4.369	−4.503	−4.634	−4.763	−4.889
n	26	27	28	29	30	31	32	33
t_0	−5.013	−5.134	−5.253	−5.371	−5.486	−5.600	−5.711	−5.821
n	34	35	36	37	38	39	40	41
t_0	−5.929	−6.036	−6.142	−6.246	−6.349	−6.450	−6.550	−6.649
n	42	43	44	45	46	47	48	49
t_0	−6.747	−6.843	−6.939	−7.034	−7.127	−7.220	−7.311	−7.402

Table 2. Dissolved oxygen data for 1995 and 1996. Values smaller than the standard are indicated by an *.

Year	*Month*	*Dissolved Oxygen*
95	1	8.70
95	2	8.70
95	3	8.43
95	4	9.21
95	5	5.57
95	6	5.95
95	7	5.88
95	8	3.61*
95	9	5.23
95	10	6.50
95	11	7.89
95	12	8.66
96	1	8.07
96	2	8.57
96	3	8.92
96	4	10.2
96	5	6.63
96	6	6.33
96	7	5.07
96	8	4.07*
96	9	4.92*
96	10	5.79
96	11	7.89
96	12	7.80

of the statistic is $T = 5.56$. The decision is to reject the hypothesis of no impairment if the statistic is smaller than the critical value. With a critical value of $t_0 = 4.763$ for $n = 24$, the no impairment decision would also be made. Evaluation of the assumptions of the model indicated that the data were autocorrelated. We used the adjustment procedure in Darken *et al.* (2000) and Wilks (1997) to account for variance inflation due to autocorrelation assuming an AR(1) model and found the decision did not change.

3. Bayesian analysis

3.1 *Reference priors*

To conduct a Bayesian analysis, prior information needs to be elicited. If there is *a priori* information about the hypotheses, clearly we need to incorporate the information in the analysis. Otherwise, noninformative priors such as reference priors (Bernardo, 1979) may be used. Reference priors are considered as "default" priors or "objective Bayes priors" because they usually maintain certain frequentist properties while Bayesian flexibility is still available. They are quite commonly used in many statistical models. Although several different approaches for obtaining non-informative priors have been proposed in literature

(for details, see Kass and Wasserman, 1996 and the references therein), the reference prior algorithm will be used in this study. This algorithm has an advantage in multi-dimensional problems if parameters of interest and nuisance parameters can be distinguished (see also Berger and Bernardo, 1992). Since the reference prior obtained in this section is uniform through the location parameter, it may be viewed as avoiding bias by using informative priors.

The parameter setting for the problem in Section 2 is no longer (μ, σ). Instead, we are interested in (θ, σ) where θ is the parameter of interest and σ is a nuisance parameter. The reference prior for this setting is provided by the following result whose proof is shown in the Appendix.

Result 1 *Suppose that a location-scale parameter model as in Equation (1) is considered with the parameters (θ, σ) where θ is defined in Equation (2). Under certain regularity conditions, the reference prior of (θ, σ) when θ is the parameter of interest and σ is a nuisance parameter can be expressed as*

$$\pi(\theta, \sigma) \propto \frac{1}{\sigma}, \tag{5}$$

where " $\propto$ " means "proportional to."

The posterior distribution of (θ, σ) thus can be expressed as follows,

$$\pi(\theta, \sigma \mid \underline{x}) \propto \text{ prior} \times \text{posterior } \propto \frac{1}{\sigma} \times \prod_{i=1}^{n} f\left(\frac{x_i - \mu}{\sigma}\right)$$

$$\propto \frac{1}{\sigma^{n+1}} \prod_{i=1}^{n} f\left(\frac{x_i - \theta}{\sigma} + p_\xi\right). \tag{6}$$

It is noted that for a one-sided hypothesis testing problem such as Equation (3), one can use non-informative priors to yield results similar to a frequentist analysis. In particular, the posterior probability of the null hypothesis is similar to the p-value (Berger, 1985, page 147). For instance, if one is willing to reject the null hypothesis at the level of significance 0.1, one can simply set the q value in Equation (4) to 0.9.

3.2 *Error Probabilities*

To further study the posterior probability of the alternative hypothesis in Equation (4) and the two types of error probabilities for a location-scale parameter model, we have the following result whose proof is located in the Appendix.

Result 2 *For the model in Equation (1) with the hypotheses in Equation (3), suppose the rejection region is determined by $P(H_1|\underline{x}) > q$. Define $z_i = (x_i - \theta_0)/\sigma_0 + p_\xi$ and $w_i = (x_i - \theta_1)/\sigma_1 + p_\xi$. Then the Type I error probability of rejecting H_0 is*

$$\alpha = P_{\theta_0, \sigma_0}(P(H_1|\underline{x}) > q) = \int_{P_\alpha(H_1|\underline{z}) > q} \prod_{i=1}^{n} f(z_i) d\underline{z}.$$

where

$$P_\alpha(H_1|\underline{z}) = \frac{\int_{u>0}\int_0^\infty v^{n-2}\prod_{i=1}^n f(p_\xi + v(z_i - p_\xi) - u)dv\,du}{\int_{-\infty}^\infty\int_0^\infty v^{n-2}\prod_{i=1}^n f(p_\xi + v(z_i - p_\xi) - u)dv\,du}. \tag{7}$$

which is free from the scale parameter σ_0. The Type II error probability when the true θ is $\theta_1 = \theta_0 + (p_\xi - p_{\xi_1})\sigma_1$ is

$$\beta = P_{\theta_1,\sigma_1}(P(H_1|\underline{x}) \le q) = \int_{P_\beta(H_1|\underline{w}) \le q} \prod_{i=1}^n f(w_i)d\underline{w}.$$

where

$$P_\beta(H_1|\underline{w}) = \frac{\int_{u>v(p_{\xi_1} - p_\xi)}\int_0^\infty v^{n-2}\prod_{i=1}^n f(p_\xi + v(w_i - p_\xi) - u)dv\,du}{\int_{-\infty}^\infty\int_0^\infty v^{n-2}\prod_{i=1}^n f(p_\xi + v(w_i - p_\xi) - u)dv\,du}. \tag{8}$$

which is also free of the scale parameter σ_1.

Note that the scale parameter does not affect calculation of the Type I error probability. Furthermore, if we try to distinguish the quantiles p_ξ and p_{ξ_1} in calculating the Type II error, the scale parameter also has no effect. The argument for this is similar to the argument in calculating the Type II error probability for the Binomial model (see Smith *et al.*, 2001).

3.3 *Bayesian analysis with certain information*

Although people may not be able to specify *a priori* information on the parameters involved in the model, it is quite common that the experimenters or researchers can assign certain prior probability over the assumed hypotheses. If this is the case, we may assume that $P(H_1) = v = 1 - P(H_0)$, where H_0 and H_1 are defined in Equation (3). If the same priors as in Section 3.2 are used here, i.e., uniform priors for the quantile parameter and the logarithm of the standard deviation, we need to recalculate the posterior probabilities of the two hypotheses. The following result can be easily obtained using the simple posterior distribution calculation and thus its proof is omitted.

Result 3 *Denote by $P^v(H_1|\underline{x})$ the posterior probability of the alternative hypothesis H_1 when the prior probability of this alternative hypothesis is v. Also, denote by $P(H_1|\underline{x})$ the posterior probability of the alternative hypothesis H_1 when there is no a priori information for the hypotheses. Then we have the relation*

$$P^v(H_1|\underline{x}) = \frac{vP(H_1|\underline{x})}{(2v-1)P(H_1|\underline{x}) + (1-v)}.$$

Note that from the above relation, the criteria for rejecting the null hypothesis are similar between this setup and the one in Section 3.2. The only difference is the cutoff value under the same level of significance. It is interesting to notice that when $v = 0.5$, the situation where there is no preference between the null and alternative hypotheses, both posterior probabilities are the same. Hence, the same cutoff value should be used. On the other hand,

if ν is more than 0.5 which implies that H_1 is more likely, then $P^{\nu}(H_1|\underline{x}) > P(H_1|\underline{x})$ which shows that the null hypothesis is more likely to be rejected. Suppose for instance $\nu = 0.75$ and the rejection rule using P^{ν} is $P^{\nu}(H_1|\underline{x}) > 0.9$ (which is essentially using a significance level of 0.1). This rejection rule is similar to using the criterion in Section 3.2 $(P(H_1|\underline{x}))$ with a significance level 0.25. Thus it is easier to reject the null hypothesis.

The decision rules as well as error probability calculations from Section 3.2 can also be applied here since both posterior probabilities for the alternative hypothesis are simple functions of each other.

3.4 *Transformations*

In environmental applications, data may be transformed using operations such as logarithm, squared root, etc. to make the data approximately normal. Some distributions such as the lognormal are generic transformed distributions of location-scale models. In this section, we investigate the impact of a transformation on our problem. The result is presented in the following proposition.

Result 4 *Suppose that a location-scale parameter model as in Equation (1) is considered with the parameter setting (θ, σ) where θ is defined as in Equation (2). Define a transformation $Y = g(X)$ such that g is a strictly increasing function in X with finite 1st order derivative. Then $\eta = g(\theta)$ is the ξth quantile of Y. The reference prior for (η, σ) is $\pi(\eta, \sigma) \propto |h'(\eta)|/\sigma$ where h is the inverse function of g. Furthermore, Results 2 and 3 above as well as the results in the next sections are all valid.*

The proof of this result is trivial since one only needs to do the straightforward transformation calculations. To obtain the reference prior of (η, σ), the invariance result for change of parameters in Datta and Ghosh (1995) can be used directly since the transformation of the parameter only involves the parameters of interest (θ to η).

It should be noted that if the transformation function g is a strictly decreasing function in X then we could simply make another transformation $-g$ and the above result still holds.

4. Normal population

In this section, the procedure we described in Section 3 will be applied to the normal populations.

4.1 *Error probability comparisons*

In this subsection, we give the expressions of the error probabilities in Result 2 for a normal population. Furthermore, they are used to compare the error probabilities of the tests using *EPA*'s "raw score" method and Virginia *DEQ*'s binomial distribution method.

Suppose $X_1, \ldots, X_n$ are i.i.d. normal random variables with mean μ and standard deviation σ. Result 2 in Section 3 can be simplified to the following result whose proof is in the Appendix.

Result 5 *For the normal case, the Type I and Type II error probabilities can be expressed as*

$$\alpha = E^{\bar{z},S_z^2}\left[I_{\{P_\alpha(H_1|\bar{z},S_z^2)>q\}}\right], \quad \text{and} \quad \beta = E^{\bar{z},S_z^2}\left[I_{\{P_\beta(H_1|\bar{z},S_z^2)\le q\}}\right],$$

respectively, where $\bar{z}\sim n(0,1/n), S_z^2\sim\chi^2_{n-1}$ *and* I_A *an indicator function. Here*

$$P_\alpha(H_1|\bar{z},S_z^2) = 1 - E^Y\left\{\Phi\left[\sqrt{n}\left(\sqrt{\frac{Y}{S_z^2}}(p_\xi-\bar{z})-p_\xi\right)\right]\right\},$$

and

$$P_\beta(H_1|\bar{z},S_z^2) = 1 - E^Y\left\{\Phi\left[\sqrt{n}\left(\sqrt{\frac{Y}{S_z^2}}(p_{\xi_1}-\bar{z})-p_\xi\right)\right]\right\},$$

where $Y\sim\chi^2_{n-1}$ *and* $\Phi(x)$ *is the c.d.f. of the standard normal distribution.*

Note that the computational formulae for the above expectations may be obtained using either numerical integration or simulation. In this section, we will use simulation along with the Monte–Carlo method to derive the above probabilities when the underlying distribution is normal. In Figs 2 and 3 we compare these probabilities to those derived using the raw score and Binomial methods.

It is known that EPA's raw score method tends to over-regulate in the sense that a site is easier to be declared impaired than using the Binomial method (Smith *et al.*, 2001). Comparing the Type I and Type II error between these two methods in the graphs (Figures 2 and 3), the statement is clearly shown to be true. On the other hand, it is interesting to notice that the Type I error probability using the raw data approach is quite close to that of using the Binomial method (note that the Type I error probability using the Binomial can not always be controlled at a fixed level because of its discreteness). This means that although the Bayes rule is used here, namely rejecting H_0 when $P(H_r|\underline{x}) > 0.9$, the results are approximately equivalent to the usual classical results when the level of significance is controlled at 0.1. This is no surprise since a non-informative prior analysis usually yields results quite similar to frequentist analysis.

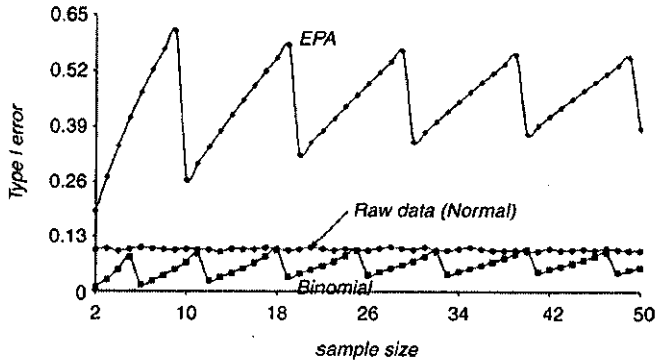

Figure 2. Comparisons of the Type I error probability using the three procedures.

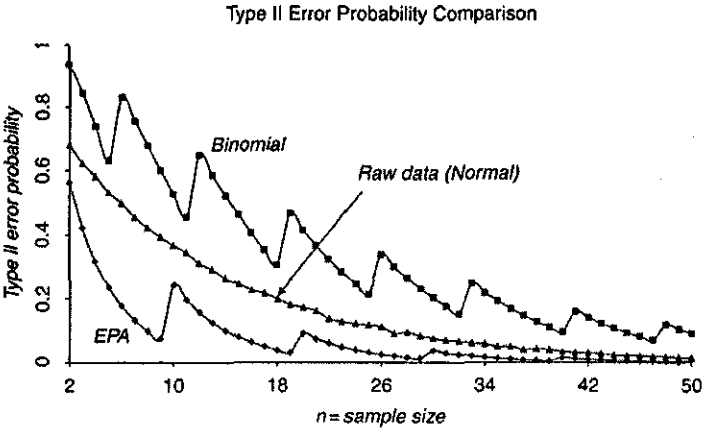

Figure 3. Comparisons of the Type II error probability using three procedures using $\alpha = 0.10$.

The Type II error probability using the raw data Fig. 3 shows more satisfactory results since it is always better than that of using the Binomial method. Although EPA's raw score method always results in a lower Type II error probability, it is at the expense of the Type I error probability. However, it can be seen that when the sample size tends to be large, the Type II error probabilities using the raw score method and the raw data method tend to agree. Based on these observations, we conclude that using the raw data approach has advantages in terms of the error probabilities discussed above when the sample size is moderate to large. The results presented here are perhaps biased in favor of the raw data method as the data were generated with underlying normal distributions. Clearly the robustness of the method needs to be evaluated.

4.2 *Rejection rules*

An explicit form of the rejection rule as in Equation (4) can be easily obtained when the underlying population distribution is normal. We have the following result whose proof is presented in the Appendix.

Result 6 *Given the setup as in Section 2 under a normal population with an upper tail standard, the rejection criterion (4) is equivalent to the criterion of rejecting the null hypothesis when*

$$T = \sqrt{n}\frac{\overline{X} - 0_0}{S_X} > t_0, \tag{9}$$

where $\overline{X}$ is the sample mean, S_x the sample standard deviation and t_0 a cutoff value. Furthermore, the statistic T follows a non-central student t-distribution (cf. Johnson et al., 1995, page 508) with non-centrality parameter $\sqrt{n}(\mu - 0_0)/\sigma$ and degrees of freedom $n - 1$ when the original normal population has mean μ and standard deviation σ.

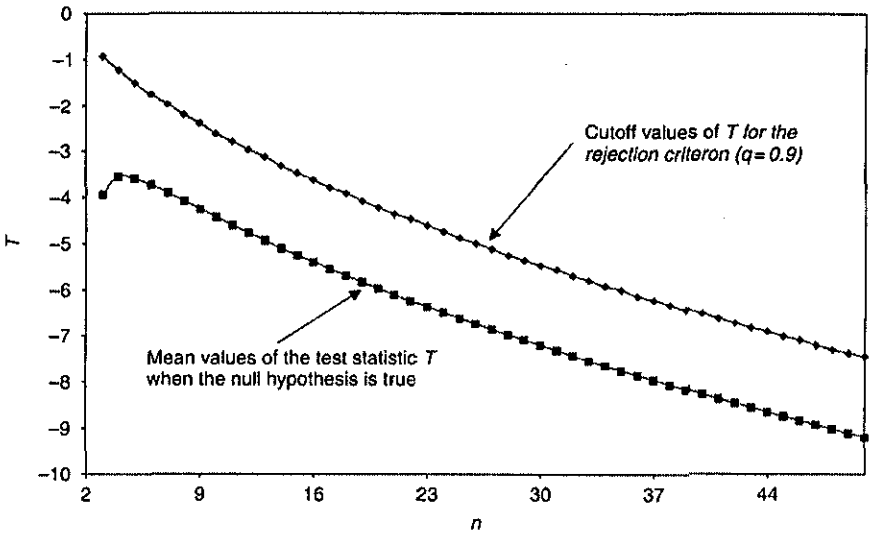

Figure 4. Simulated cutoff values and means for the statistic for the rejection rule $P(H_1|\underline{x}) > 0.9$ under a normal population.

Using simulation, we can find the cutoff value of the T statistic for rejecting the null hypothesis. Fig. 4 below provides the simulated cutoff values of T when q is given as 0.9 for sample sizes ranging from 2 to 50. Also, the associated means of T are given in the figure.

Table 1 provides the critical values of the statistic T for various sample sizes when q is 0.9.

5. Discussion

In this paper, we discussed the problem of using raw measurements to monitor water conditions for impairment checking instead of using "threshold data". Other than the simplicity and perhaps some other reasons to use threshold data, we feel that the use of raw measurements should be more beneficial in making accurate estimation and prediction.

Although the Bayesian methodology is employed in this article, traditional non-Bayesian analysis can be done similarly. The reason we use Bayesian analysis here is because we feel it's more flexible in terms of incorporating historical information into the current data by assigning certain informative prior distribution to the parameters. For instance, we only studied reference priors over the parameters. It is quite reasonable to assign a prior distribution of θ to be more concentrated around the standard level θ_0. Furthermore, if a piece-wise prior preference of the hypotheses is given, a piece-wise informative prior distribution can be elicited for θ. Though the computation will be not as easy as given in this paper, MCMC Gibbs sampling provides an approach for computation.

Finally, although in this article we investigate the situation when the alternative hypothesis is $\theta > \theta_0$, to consider an alternative hypothesis $\theta < \theta_0$ we only need to use

the same method discussed in the article for the negative values of X since only the location-scale model is studied here. Also there would be problems with this approach for two-sided tests due to the improper prior (Berger and Pericchi, 1996). As pointed out in Berger, Pericchi and Varshavsky (1998) this is not a problem for our location-scale model.

Appendix

A.1

Proof of Result 1: First, it can be shown that the Fisher information matrix for the parameter setting (μ, σ) is given as follows,

$$I(\mu,\sigma) = \frac{n}{\sigma^2}\begin{pmatrix} a_0 & a_1 \\ a_1 & 3+a_2 \end{pmatrix},$$

where $a_i = \int z^i f'^2(z)/f(z)dz$, provided their existence for $I = 0, 1, 2$. To make a transformation of (μ, σ) to (θ, σ), the Jacobian matrix of the transformation is

$$J = \begin{pmatrix} 1 & -p_\xi \\ 0 & 1 \end{pmatrix}.$$

Therefore, the Fisher information for (θ, σ) can be expressed as

$$I(\theta,\sigma) = \frac{1}{\sigma^2}\begin{pmatrix} b_0 & b_1 \\ b_1 & b_2 \end{pmatrix},$$

where b_i's are constants, for $I = 0, 1, 2$. The result then follows using the algorithm for calculating the reference priors. □

A.2

Proof of result 2: The Type I error probability using the rejection rule $P(H_1 \mid \underline{x}) > q$ is

$$\alpha = P_{\theta_0,\sigma_0}(P(H_1|\underline{x}) > q) = \frac{1}{\sigma_0^n}\int_{P(H_1|\underline{x})>q} \prod_{i=1}^{n} f\left(\frac{x_i - \theta_0}{\sigma_0} + p_\xi\right) d\underline{x}$$

$$= \int_{P_\alpha(H_1|\underline{z})>q} \prod_{i=1}^{n} f(z_i)d\underline{z},$$

where z_i's are defined in the result for $I = 1, \ldots, n$. The probability

$$P_\alpha(H_1|\underline{z}) = \frac{\int_{\theta>\theta_0}\int_0^\infty \frac{1}{\sigma^{n+1}}\prod_{i=1}^n f\left(p_\xi + \frac{\theta_0+\sigma_0(z_i-p_\xi)-\theta}{\sigma}\right)d\sigma\, d\theta}{\int_{-\infty}^{\infty}\int_0^\infty \nu^{n-2}\prod_{i=1}^n f\left(p_\xi + \frac{\theta_0+\sigma_0(z_i-p_\xi)-\theta}{\sigma}\right)d\sigma\, d\theta}.$$

After making a transformation of the variables $u = (\theta - \theta_0)/\sigma$ and $\nu = \sigma_0/\sigma$, the result is straightforward. For Type II error probability, the proof is similar. □

A.3

Proof of Result 4: For the normal population, $\prod_{i=1}^{n} f(p_\xi + v(z_i - p_\xi) - u)$ in (8) can be expressed as

$$\prod_{i=1}^{n} f(p_\xi + v(z_i - p_\xi) - u) = \frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{n}{2}\left[u - (p_\xi + v(\bar{z} - p_\xi))\right]^2 - \frac{v^2}{2} S_z^2 \right\},$$

where

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n} z_1, \quad \text{and} \quad S_z^2 = \sum_{i=1}^{n} (z_i - \bar{z})^2.$$

Making the transformation $t = \sqrt{n}\left[u - (p_\xi + v(\bar{z} - p_\xi))\right]$ and $y = v^2 S_z^2$, the result for the Type I error probability follows after a little algebra. Similarly we can get the result for Type II error probability. □

A.4

Proof of Result 5: Transform (θ, σ) to (u, v) by $u = \sqrt{n}(\theta - \bar{X})/\sqrt{n-1}S_x$ and $v = \sqrt{n-1}S_x/\sigma\sqrt{n}$, the posterior probability of the alternative hypothesis can be written as

$$P(H_1|\underline{x}) = \frac{\int_0^\infty \int_{u > \sqrt{n}(\theta_0 - \bar{x})/\sqrt{n-1}S_x} v^{n-1} \exp\left\{ -nv^2\left[1 + \left(u - p_\xi/v\right)^2\right]/2 \right\} du\, dv}{\int_0^\infty \int_{-\infty}^\infty v^{n-1} \exp\left\{ -nv^2\left[1 + \left(u - p_\xi/v\right)^2\right]/2 \right\} du\, dv}.$$

Let t be defined in (9). After integrating u out and letting $y = nv^2$, the posterior probability can be written as

$$P(H_1|\underline{x}) = \int_0^\infty g(y)\Phi(\sqrt{n}p_\xi + t\sqrt{y/(n-1)})dy,$$

where $g(y)$ is the density of a χ^2-distribution with degrees of freedom $n - 1$. Clearly, the above formula shows an increasing function of t. Hence, $P(H_1|\underline{x}) > q$ is equivalent to $t > t_0$. On the other hand,

$$T = \frac{\bar{X} - \theta_0}{S_x/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \theta_0)/\sigma + \sqrt{n}(\mu - \theta_0)}{S_x/\sigma}$$

follows a non-central student t-distribution as defined in Johnson *et al.*, 1995. □

Acknowledgments

The work was partially funded by the Virginia Water Resources Research Center and United States Environmental Protection Agency's Science to Achieve Results (STAR) (Grant No. R82795301). Although the research described in the article has been funded wholly or in part by the United States Environmental Protection Agency's STAR programs, it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred. The

authors would like to thank the referees for their insightful comments and helpful suggestions.

References

Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis*, second edition, Springer-Verlag, New York.

Berger, J. and Bernardo, J. (1992) On the development of reference priors (with discussion). In *Bayesian Statsitics* 4, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds), Oxford University Press, pp. 35–60.

Berger, J. and Pericchi, L. (1996) Intrinsic Bayes factor for model selection and prediction. *Journal of the American Statitsical Association*, **91**, 109–22.

Berger, J., Pericchi, L., and Varshavsky, J. (1998) Bayes factors and marginal distributions in invariant situations. *Sankhya A*, **60**, 307–21.

Bernardo, J. (1979) Reference posterior distributions for Bayesian inference (with discussion). *Journal of the Royal Statistical Society Series B*, **41**, 113–47.

Commonwealth of Virginia (1997) *Review of the Department of Environmental Quality*. House Document No. 67.

Darken, P.F., Holtzman, G.I., Smith, E.P., and Zipper, C.E. (2000) Detecting changes in trends in water quality using modified Kendall's tau. *Environmentrics*, **11**, 423–34.

Datta, G.S. and Ghosh, M. (1995) Some remarks on noninformative priors. *Journal of the American Statitsical Association*, **90**, 1357–63.

Duncan, A.J. (1974) *Quality Control and Industrial Statistics*, Irwin, New York.

Johnson, N.O., Kotz, S., and Balakrishnan, N. (1995) *Continuous Univariate Distributions, Vol. 2*, John Wiley & Sons, New York.

Kass, R.E. and Wasserman, L. (1996) The selection of prior distribution by formal rules. *Journal of the American Statistical Association*, **91**, 1343–70.

Smith, E.P., Ye, K., Hughes, C., and Shabman, L. (2001) Statistical assessment of violations of water quality standards under Section 303(d) of the Clean Water Act. *Environmental Science and Technology*, **35**, 606–12.

Wild, P., Hordan, R., Leplay, A., and Vincent, R. (1996) Confidence intervals for probabilities of exceeding threshold limits with censored log-normal data. *Environmetrics*, **7**, 247–59.

Wilks, D.S. (1997) Resampling hypothesis tests for autocorrelated fields. *Journal of Climate*, **10**, 65–82.

Biographical sketches

Dr Keying Ye is associate professor of statistics, Virginia Polytechnic Institute and State University. His research focuses using Bayesian analysis to help solve problems in environmetrics and bioinformatics.

Dr Eric P. Smith is professor of statistics and director of the Statistical Consulting Center, Virginia Polytechnic Institute and State University. His research focuses on the development and application of statistical methods for analysis of environmental and ecological problems.