



PERGAMON

www.elsevier.com/locate/watres

## CONFIDENCE OF COMPLIANCE: A BAYESIAN APPROACH FOR PERCENTILE STANDARDS

GRAHAM B. McBRIDE<sup>1\*</sup> and JULIAN C. ELLIS<sup>2</sup>

<sup>1</sup>National Institute of Water & Atmospheric Research (NIWA), P.O. Box 11-15, Hamilton, New Zealand and <sup>2</sup>Water Research Centre, Medmenham, UK

(Received 20 October 2000)

**Abstract**—Rules for assessing compliance with percentile standards commonly limit the number of exceedances permitted in a batch of samples taken over a defined assessment period. Such rules are commonly developed using classical statistical methods. Results from alternative Bayesian methods are presented (using beta-distributed prior information and a binomial likelihood), resulting in “confidence of compliance” graphs. These allow simple reading of the consumer’s risk and the supplier’s risks for any proposed rule. The influence of the prior assumptions required by the Bayesian technique on the confidence results is demonstrated, using two reference priors (uniform and Jeffreys’) and also using optimistic and pessimistic user-defined priors. All four give less pessimistic results than does the classical technique, because interpreting classical results as “confidence of compliance” actually invokes a Bayesian approach with an extreme prior distribution. Jeffreys’ prior is shown to be the most generally appropriate choice of prior distribution. Cost savings can be expected using rules based on this approach.  
© 2001 Elsevier Science Ltd. All rights reserved

**Key words**—percentile standards, supplier’s risk, consumer’s risk, bayesian approach, exceedance probability, prior distribution

### NOMENCLATURE

Be( <i>a</i> , <i>b</i> )	two-parameter beta distribution
CC	confidence of compliance
CF	confidence of failure
<i>e</i>	number of exceedances of a percentile limit
<i>e</i> <sub>bd</sub>	number of permissible exceedances under a benefit-of-doubt stance
<i>e</i> <sub>fs</sub>	number of permissible exceedances under a fail-safe stance
<i>F</i>	posterior distribution function
<i>g</i> ( <i>x</i> )	prior probability density function
<i>h</i> ( <i>x</i> )	posterior probability density function
<i>l<sub>X</sub></i> ( <i>a</i> , <i>b</i> )	incomplete beta function ratio
<i>L</i>	likelihood function
<i>n</i>	number of samples
<i>p</i>	<i>p</i> -value of a one-sided hypothesis test
<i>s</i> <sup>2</sup>	variance of sample exceedance rate
<i>x</i>	probability of exceedance of a single random sample
$\bar{x}$	sample mean exceedance rate
<i>X</i>	critical exceedance probability ( <i>X</i> = 0.05 for a 95 percentile standard)

### Greek symbols

$\alpha$	significance level
----------	--------------------

### INTRODUCTION

Percentile standards are increasingly used in environmental management, e.g., for drinking-water supplies (MoH, 1995), waste discharge permits (Ellis, 1989) and toxicity standards (ANZECC & ARMCANZ, 1999). Rules for compliance with these standards are generally derived from classical one-sided hypothesis tests in which the *p*-value is used as a weight-of-evidence measure against the tested hypothesis. In developing and in applying such rules we need to consider both risks of reaching false conclusions: the supplier’s risk (falsely inferring a breach of standard) and the consumer’s risk (falsely inferring compliance). The former risk is faced by the water supplier or by the waste discharger; the latter is faced by the water drinker or by the receiving environment.

The tested hypothesis usually posits compliance, in which case the supplier’s risk is kept small—because this is the Type I error risk and the testing procedure always restricts this to be below a small value, typically significance level  $\alpha = 5\%$ . Accordingly, the consumer’s risk is rather large in cases, where the standard actually is in breach (it would be 95% for marginal non-compliance). However, though much less common, the tested hypothesis may be chosen to posit breach (non-compliance), in which case the Type I error risk is now the consumer’s risk and so this is the risk that is kept small. The Type II error

\*Author to whom all correspondence should be addressed.  
Tel.: +64-7-856-1726; fax: +64-7-856-0151; e-mail: g.mcbride@niwa.cri.nz

10636

risk is then the supplier's risk and it can be large for marginal compliance.

While this procedure is objective (inasmuch as choosing a value of  $\alpha$  is so) it has a number of inherent problems. First, the assessor is often not aware of the possibility of reversing the role of the two error risks, by choosing to test either the hypothesis of compliance or the hypothesis of breach. Second, there is debate in statistical literature (e.g., Goodman, 1993; Harlow *et al.*, 1997; Royall, 1997) as to the role of  $p$ -values as weights-of-evidence (the debate is strongest for two-sided tests of point hypotheses). Third, and most importantly, the procedure does not appear to provide direct statements about the probability of compliance, as it considers instead the probability of gaining ranges of data (e.g., more exceedances than were actually obtained) *if* the tested hypothesis (e.g., compliance) were true. A compliance assessor or a Court may find this somewhat tangential to the central issue, i.e., balancing the probability of breach vs. the probability of compliance.

Compliance assessment based on a Bayesian approach has the potential to address such issues directly, without the need to consider significance levels, Type I error risk and Type II error risk—indeed these are irrelevant in a Bayesian framework (Lee, 1989). We argue herein that they can give simpler and more direct information to a compliance assessor or to a Court, even though they do call for a subjective element of information—the assessor's belief as to the probability of compliance *before* data come to hand. This belief is represented as a "prior" distribution. If a "reference prior" is specified, reflecting the assessor's lack of prior knowledge, one obtains a simple statement of "confidence of compliance". Using the most appropriate (Jeffreys') prior gives less stringent compliance rules than those based on classical methods. If an informative prior is used—reflecting the assessor's prior belief as to likely exceedance rates—there can be considerable discrepancies between Bayesian and classical results, particularly for small numbers of samples.

#### COMPLIANCE WITH PERCENTILE STANDARDS—CLASSICAL APPROACH

Percentile standards require that a certain limit (e.g., of a chemical's concentration) should not be exceeded for more than a percentage of an assessment period. This requirement is properly stated as a percentage of *time* in which the limit may be exceeded (which is always unknown), not as a percentage of *samples* in which that limit may be exceeded (which is only known once data are reported—Ellis and Lacey, 1980). A compliance rule derived from such a standard generally requires that no more than a stated number of exceedances of the percentile limit should occur in a given number of samples over a compliance assessment period. In formulating this

rule we do not know whether sampling error will cause future assessments to be in error in their inferences as to compliance or breach, nor is it known which error—supplier's or consumer's—might occur. Accordingly the rule should be based on a consideration of both error risks (Ellis, 1989). In doing so one must note that once the number of samples has been decided only one of these risks can be controlled (i.e., kept small). If compliance is in fact marginal, the other risk will be large. Furthermore, the risk to be controlled is determined by the assessor's choice of the hypothesis to be tested. If that hypothesis posits compliance (as is usual) it is the supplier's risk that will be kept small. But in a precautionary approach one could invert the tested hypothesis so that it posits that the standard has been breached. In that case it is the consumer's risk that will be controlled. These have been characterised by Ellis (1989) as the *benefit of doubt* and *fail-safe* approaches.

The error risks can be quantified using classical one-sided hypothesis tests. The key calculation therein is the probability of obtaining data at least as extreme as was obtained *if* the tested hypothesis is (only just) true, i.e., compliance is "borderline". This is the " $p$ -value". If this is less than the *a priori* significance level (usually taken as  $\alpha = 5\%$ ), one has a "statistically significant" result and infers that there has been breach or compliance (depending on which hypothesis has been tested). That is, the  $p$ -value is used as a weight-of-evidence against the tested hypothesis. If that hypothesis posits compliance, then  $p$  is the supplier's risk; if the hypothesis posits breach, then  $p$  is the consumer's risk. Because the tested hypothesis posits borderline compliance, these risks are maximum values.

For example, consider a permit requiring (*inter alia*) that the total inorganic nitrogen concentration ("TIN") in the effluent from a coastal sewage treatment plant should be below  $10 \text{ g m}^{-3}$  for at least 98% of a three-month summer period, based on two samplings every outgoing tide. Compliance with this 98 percentile standard is therefore to be assessed using about  $n = 350$  samples. Denote by  $e$  the number of exceedances of the  $10 \text{ g m}^{-3}$  limit in these samples. If the supplier's risk is chosen to be the one that is controlled then, as above, the tested hypothesis must posit compliance, i.e.,  $x \leq 0.02$ , where  $x$  is the probability of exceedance for any single random sample. We can then use the following formula to identify the maximum permissible number of exceedances under this benefit-of-doubt approach ( $e_{bd}$ ) while still reaching a verdict of "compliance":

$$p = \Pr(e > e_{bd} | n, x = X) \leq \alpha \quad (1)$$

where  $X$  is the value of  $x$  where the tested hypothesis is only just true (i.e.,  $X = 0.02$  in this example) and  $\alpha$  is the maximum acceptable supplier's risk. The left-hand-side of equation (1) gives the probability of obtaining more than  $e_{bd}$  exceedances when in fact the

10638

effluent was just complying (i.e.,  $x = X = 0.02$ )— $e_{bd}$  is to be selected as its lowest value satisfying  $p \leq \alpha$  ( $p = \alpha$  will seldom be attained because exceedance data are discrete, not continuous). Because the tested hypothesis is compliance and we are demanding a high standard of proof before it will be rejected (e.g., only a 5% significance level), the proportion of allowable exceedances (i.e.,  $e_{bd}/n$ ) will *always* exceed  $X$ .

The required probability can be calculated by noting that for random sampling  $e$  is distributed as cumulative binomial (regardless of the distribution of the underlying process), i.e.,

$$p = \Pr(e > e_{bd} | n, x = X) \\ = \sum_{e=e_{bd}+1}^n C_e X^e (1-X)^{n-e} \quad (2)$$

where  $C_e$  is the binomial coefficient. Taking  $n = 350$ ,  $X = 0.02$  and  $\alpha = 5\%$  we can use available numerical algorithms (e.g., Press *et al.*, 1986) to obtain  $p = 5.15\%$  for  $e_{bd} = 11$  and  $p = 2.57\%$  for  $e_{bd} = 12$ . Accordingly, from equation (1), we must allow up to  $e_{bd} = 12$  exceedances of the  $10 \text{ gm}^{-3}$  limit in those 350 samples (i.e., about 3.4% of the samples).

However, if the consumer's risk is to be controlled (in a precautionary fail-safe approach), the tested hypothesis must now posit breach, i.e.,  $x \geq 0.02$ . Because we are demanding a high standard of proof before this hypothesis will be rejected, the proportion of allowable exceedances (i.e.,  $e_{fs}/n$ ) will *never* exceed  $X$ . The appropriate formula is

$$p = \Pr(e \leq e_{fs} | n, x = X) \\ = \sum_{e=0}^{e=e_{fs}} C_e X^e (1-X)^{n-e} \leq \alpha \quad (3)$$

where  $e_{fs}$  is its highest value satisfying  $p \leq \alpha$ . Note that testing this fail-safe approach means that  $\alpha$  is now the maximum acceptable consumer's risk. Taking  $n = 350$ ,  $X = 0.02$  and  $\alpha = 5\%$ , we obtain  $p = 2.85\%$  and  $p = 7.97\%$  for  $e_{fs} = 2$  and 3 respectively. Accordingly we can allow up to only  $e_{fs} = 2$  exceedances of the  $10 \text{ gm}^{-3}$  limit in those samples (i.e., about 0.6% of the samples).

#### COMPLIANCE WITH PERCENTILE STANDARDS—BAYESIAN APPROACH

In a Bayesian analysis we no longer assume one particular exceedance probability value (i.e.,  $x = X$ ) in computing probabilities; it is regarded as a continuous variable about which we want to make confidence statements. In doing so we use the actual data obtained (and not any data more extreme than that obtained) to update a *prior* belief to obtain a *posterior* belief, which we call "confidence of compliance". This uses Bayes' theorem. The prior belief or knowledge is stated as a probability density function (pdf) over the range of its possible values,

and is to be decided before data come to hand (similarly, the significance level for a classical test should be stated before data are collected). Accordingly the posterior information gained is also a pdf. We can then get the required probabilities by integrating that function up to  $X$ .

The equation for the posterior pdf, from Bayes' theorem (Press, 1989), takes the form

$$h(x|e, n) = \frac{L(e|n, x)}{\int_0^1 L(e|n, x)g(x) dx} g(x) \quad (4)$$

where  $e$  is the number of exceedances in  $n$  samples;  $h(x|e, n)$  is the posterior pdf of  $x$  for a given value of  $e$  and  $n$ ;  $L(e|n, x)$  is the "likelihood function" for any  $n$  and  $x$ ; and  $g(x)$  is the prior pdf of  $x$  (independent of  $n$  and  $e$ ). This equation shows that the *a priori* belief about the true exceedance rate  $x$ , measured by  $g(x)$ , is updated by the "standardised likelihood" (the term in brackets) to derive the required posterior probability density  $h$ .

For a set of  $n$  samples the likelihood function is the probability mass function for a particular value of  $e$ , with  $x$  (rather than  $e$ ) being regarded as the function's parameter. It is the term being summed in equations (2) and (3), for any value of  $x$  (not just for  $X$  as in those equations), i.e.,

$$L(e|x, n) = {}^n C_e x^e (1-x)^{n-e} \quad (5)$$

(the sum of  $L$  over all possible values of  $e$  is unity, but the integral of  $L$  over all  $x$  is not—hence it is called a likelihood rather than a density). Therefore, noting that the binomial coefficients cancel in the standardised likelihood function, the posterior probability density is given by

$$h(x|e, n) = \left[ \frac{x^e (1-x)^{n-e}}{\int_0^1 x^e (1-x)^{n-e} g(x) dx} \right] g(x) \quad (6)$$

and the probability of compliance (i.e., "confidence of compliance", denoted as CC) is calculated from the distribution function

$$CC = F(x \leq X | e, n) = \int_0^X h(x|e, n) dx \quad (7)$$

We also define the "confidence of failure" as  $CF = 1 - CC$ .

#### Choosing a prior distribution

The choice of the prior density  $g(x)$  may be approached in two general ways. In either case it is highly desirable to use a "conjugate distribution". This guarantees that the prior and posterior probability densities belong to the same family of distributions, making the calculations required by equations (6) and (7) much simpler. Therefore, noting that the likelihood function follows a binomial distribution, we choose  $g(x)$  to follow the versatile beta distribution,  $Be(a, b)$  (see the appendix),



because these distributions are conjugate to a binomial likelihood, Lee (1989).

First, we can use a "reference prior", reflecting the views of someone with no particular *a priori* beliefs, as will be shown in Fig. 1. Noting that an optimal prior does not exist (Lee, 1989), perhaps the most obvious selection is the uniform prior, also known as "Bayes' postulate", in which  $g(x) = 1$  for all possible values of the exceedance rate (i.e.,  $0 \leq x \leq 1$ ). It implies that all values of  $x$  are equally likely and is a special case of the beta distribution, i.e.,  $Be(1, 1)$ . However the U-shaped prior derived from "Jeffreys' rule" (Lee, 1989)— $Be(\frac{1}{2}, \frac{1}{2})$ —can be argued to be a better choice (it is invariant under any change of scale of measurement). It implies that extreme values of the exceedance rate (i.e.,  $x \rightarrow 0$  or  $x \rightarrow 1$ ) are more likely than intermediate values.

Second, we may feel that, for example, it is very unlikely that  $x$  would be too much above zero, and so specify our own (informative) prior. For example, based on past experience, we could take the expected value of the true exceedance rate as  $\bar{x} = 0.1$  and its variance as  $s^2 = 0.01$  (corresponding to meeting a 90 percentile standard). Or we could hold that the true exceedance rate is  $\bar{x} = 0.01$  with variance  $s^2 = 0.001$  (corresponding to meeting a 99 percentile standard). We can then calculate  $a$  and  $b$  from the known expressions for the mean and variance of the beta distribution (Lee, 1989), i.e.,  $\bar{x} = a/(a+b)$  and  $s^2 = ab/[(a+b)^2(a+b+1)]$ , from which we obtain

$$a = \bar{x} \left[ \frac{\bar{x}(1-\bar{x})}{s^2} - 1 \right] \text{ and } b = a \left[ \frac{1}{\bar{x}} - 1 \right] \quad (8)$$

We therefore obtain  $Be(0.8, 7.2)$  and  $Be(0.089, 8.811)$  as our prior distributions for these two cases. As we shall be considering 95, 98 and 99 percentiles, these two functions will be denoted as the pessimistic prior and the optimistic prior, respectively.

#### Calculating posterior probabilities

Because  $h$  is a beta density, it follows that  $F$  follows an incomplete beta function ratio, written as  $I_X(a, b)$  (see the appendix). Inserting the beta pdf (see the

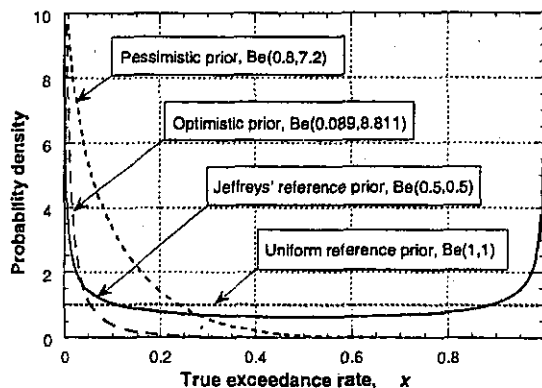


Fig. 1. Four prior densities compared.

appendix) into equations (6) and (7) we see that  $a_{\text{posterior}} = a_{\text{prior}} + e$  and  $b_{\text{posterior}} = b_{\text{prior}} + n - e$  (Lee, 1989). Accordingly the posterior distribution function is given by

$$CC = F(x \leq X | e, n) = I_X(a_{\text{prior}} + e, b_{\text{prior}} + n - e) \quad (9)$$

This can be calculated using numerical methods (e.g., Press *et al.*, 1986).

#### Confidence of Compliance graphs

Figure 1 shows the shape of the four prior distributions discussed above. Each curve encloses a unit area. Figures 2 and 3 show the sensitivity of confidence of compliance (and confidence of failure) to the choice of prior distribution, for the 99 and 95 percentile, using both reference priors and both informative priors. These graphs, and those that follow, include marker lines at  $CC = 5\%$  and  $95\%$ , for reasons to be explained. The confidences are plotted for a range of numbers of samples and exceedances. We base the main confidence of compliance graphs (Figs 4-6) on Jeffreys' prior,  $Be(\frac{1}{2}, \frac{1}{2})$ , as will also be explained. These figures show graphs for three percentiles in common use: 99, 98 and 95 percentile.

## DISCUSSION

#### Choice and influence of priors

As shown in Fig. 1, the uniform reference prior affords equal probability to all (equal) increments of  $x$ . This is not entirely reasonable, as lower exceedance rates can most often be expected to have a higher probability than do higher rates. Jeffreys' reference prior explicitly allows this to occur. However, because this (and other) reference priors are symmetrical on the unit  $x$  interval (Lee, 1989), it also allows high exceedance rates to be more likely than moderate values. This apparent problem turns out to be minimal, as will be seen. That is, if data indicate only low values of  $x$ , the influence of the prior near  $x = 1$  is quickly extinguished. For this reason we choose Jeffreys' as the reference prior. A further possibility, Haldane's  $Be(0, 0)$  prior, acts rather similarly but has the unfortunate property that  $CC = 100\%$  when no exceedances occur (because  $I_X(0, n) \equiv 1$ ). For this reason we do not favour this prior, even though it is recommended by Lee (1989).

The two informative (pessimistic and optimistic) priors are more reasonable in that they accord highest probabilities to the lowest exceedance rate, and decrease monotonically from there (this is a property of all beta distributions with  $a < 1$  and  $b > 1$ ). These priors are of course subjective, in that individual assessors may nominate different priors and so reach somewhat different conclusions.

Figures 2 and 3 illustrate the effect of different priors on the computed confidences. These show





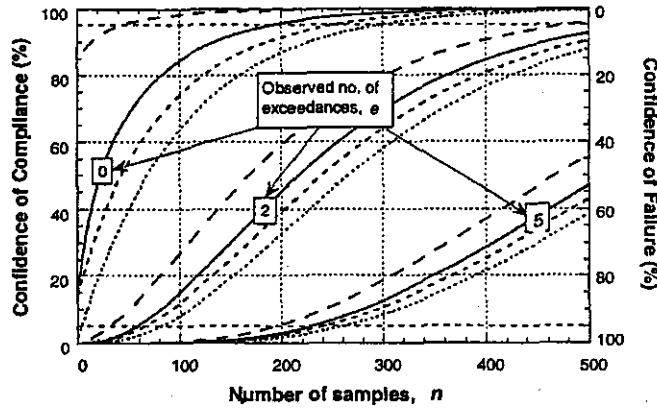


Fig. 2. Confidence of Compliance for a 99 percentile standard, using four priors: (—) Jeffreys'  $Be(\frac{1}{2}, \frac{1}{2})$ , (---) optimistic  $Be(0.089, 8.811)$ , (- -) pessimistic  $Be(0.8, 7.2)$ , (· · ·) uniform  $Be(1, 1)$ .

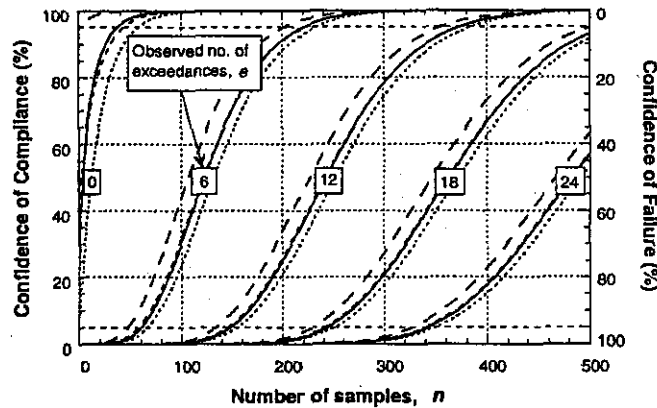


Fig. 3. Confidence of compliance for a 95 percentile standard, using four priors: (—) Jeffreys'  $Be(\frac{1}{2}, \frac{1}{2})$ , (---) optimistic  $Be(0.089, 8.811)$ , (- -) pessimistic  $Be(0.8, 7.2)$ , (· · ·) uniform  $Be(1, 1)$ .

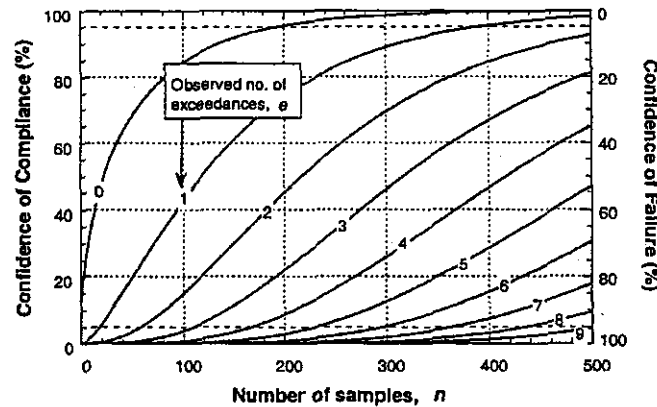


Fig. 4. Confidence of compliance for a 99 percentile standard, using Jeffreys' prior.

that, as expected, the optimistic prior produces the highest CC values. The uniform prior gives the lowest CC values, lower even than the pessimistic prior. So while the uniform prior has some appearance of impartiality, in fact it can be argued to produce overly pessimistic results. The influence of the priors is most pronounced at smaller numbers of samples

and exceedances. The divergence between the results for these priors is most pronounced at the highest percentile (99 percentile), which is where the priors are most different. For a 95 percentile (Fig. 3) the divergence is quite small, except for zero exceedances and  $n < 20$  where the optimistic prior gives much higher CC values.



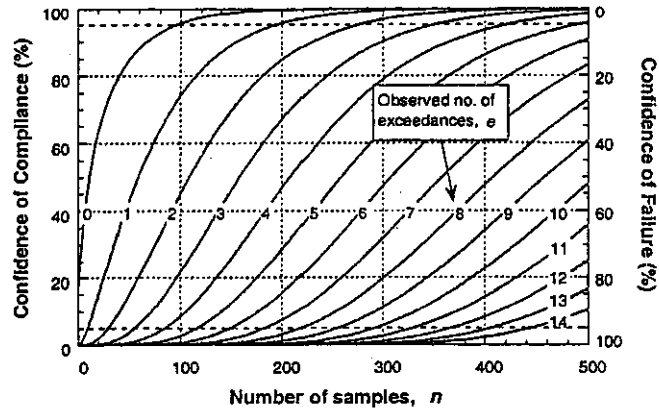


Fig. 5. Confidence of compliance for a 98 percentile standard, using Jeffreys' prior.

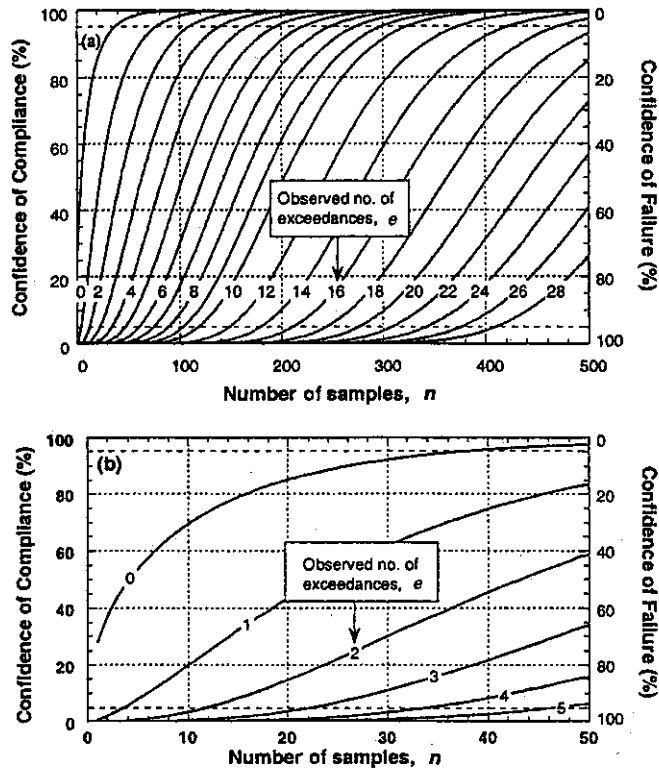


Fig. 6. Confidence of compliance for a 95 percentile standard, using Jeffreys' prior.

Compliance of confidence graphs can be used when considering sampling requirements for compliance rules seeking to control either the consumer's risk or the supplier's risk. Tables 1 and 2 demonstrate the results for a 95 percentile, with maximum risks constrained to be less than 5%. Results for the consumer's risk using the Bayesian approach were obtained by reading  $n$  values for which CC first exceeds 95% (i.e., for which CF first falls below 5%). For Jeffrey's prior and a 95 percentile this may be done using the top dashed line on Fig. 6. (Complete CC figures for other priors are not shown, except that values for  $e = 0$  & 6 can be read from Fig. 3). Results for the supplier's risk using the Bayesian approach

were obtained by reading  $n$  values for which CF first exceeds 95% (i.e., for which CC first falls below 5%), using the bottom dashed line on the figure. The classical results were obtained from equations (2) and (3) as the  $n$  values where the appropriate  $p$ -values first fell below 5%. (Software has also been developed to do these calculations for any percentile, risk and prior parameters.) In these tables we see that the uniform prior produces the most onerous sampling requirements of the four Bayesian options, with the classical approach being slightly more so. Note that the classical supplier's risk (equation (2)) may also be written as  $p = I_X(e_{bd} + 1, n - e_{bd})$  (see appendix), whereas the CC value for the uniform prior (equation



Table 1. Numbers of samples and maximum permissible exceedances needed to keep the maximum consumer's risk below 5% when assessing compliance with a 95% percentile standard

Maximum permissible exceedances	Classical approach	Bayesian approach using the following priors			
		Uniform	Jeffreys'	Optimistic	Pessimistic
0	59-92 <sup>a</sup>	58-91 <sup>a</sup>	38-76 <sup>a</sup>	2-53 <sup>a</sup>	44-79 <sup>a</sup>
1	93-123	92-122	77-108	54-87	80-110
2	124-152	123-151	109-138	88-118	111-139
3	153-180	152-179	139-166	119-147	140-167
4	181-207	180-206	167-193	148-174	168-195
5	208-233	207-232	194-220	175-201	196-221
6	234-259	233-258	221-246	202-228	222-247
7	260-285	259-284	247-272	229-254	248-273
8	286-310	285-309	273-298	255-279	274-298
9	311-335	310-334	299-323	280-304	299-323
10	336-360	335-359	324-348	305-329	324-348

<sup>a</sup>It is not possible to keep the consumer's risk below 5% if less than 59, 58, 38, 2 or 44 samples are to hand (classical approach and using the four Bayesian priors, respectively).

Table 2. Numbers of samples and minimum permissible exceedances needed to keep the maximum supplier's risk below 5% when assessing compliance with a 95% percentile standard

Minimum permissible exceedances	Classical approach	Bayesian approach using the following priors			
		Uniform	Jeffreys'	Optimistic	Pessimistic
0	1 <sup>a</sup>	b	b	b	b
1	2-7	1-6	1-3	1	1
2	8-16	7-15	4-11	2	2-7
3	17-28	16-27	12-22	3-10	8-18
4	29-40	28-39	23-34	11-21	19-30
5	41-53	40-52	35-46	22-33	31-43
6	54-67	53-66	47-60	34-46	44-57
7	68-81	67-80	61-74	47-60	58-71
8	82-95	81-94	75-88	61-74	72-85
9	96-110	95-109	89-102	75-89	86-100
10	111-125	110-124	103-117	90-103	101-115

<sup>a</sup>The risk is exactly 5% in this case.

<sup>b</sup>It is not possible to keep the supplier's risk below 5% in these cases.

(9), with  $e = e_{bd}$ ) is  $CC = I_X(e_{bd} + 1, 1 + n - e_{bd})$ . Therefore, if we had chosen  $Be(1, 0)$  as the reference prior the classical and Bayesian results would have been identical. Such correspondences are known to occur for particular choices of priors in one-sided tests (Edwards *et al.*, 1963; Pratt, 1965; DeGroot, 1973; Casella and Berger, 1987; Lee, 1989). Because  $Be(1, 0)$  concentrates all prior probability at the most extreme exceedance (its pdf is everywhere zero except at  $x = 1$ ), we see that the classical approach is the most pessimistic of all cases.

Table 1 shows that using the Bayesian approach with Jeffreys' prior, the minimum number of complying samples that need to be collected to give 95% confidence of meeting a 95 percentile standard is 38, compared with the classical approach's requirement of 59 samples. However if one transgression of the percentile limits occurs at least 77 samples are needed (compared with 93). Therefore, using this reference prior, substantial reduction in sampling effort can be expected while still meeting the same confidence criteria.

Returning to the example of the coastal sewage TIN discharge 98 percentile standard we note from Fig. 5 (using Jeffreys' prior) that for  $n = 350$  samples the confidence of failure first exceeds 95% when  $n = 12$

exceedances are allowed (in agreement with the classical approach). However, the confidence of compliance first exceeds 95% when 3 exceedances are observed in (vs. 2 in the classical approach). So again, a more reasonable prior assumption (than given by the  $Be(1, 0)$  distribution implicit in the classical approach) results in less onerous compliance requirements.

## CONCLUSIONS

Bayesian approaches give direct answers to questions of confidence of compliance with percentile standards, and so enable the supplier's risk and the consumer's risk to be identified. This facility comes at the "cost" of having to state one's prior belief as to likely exceedance rates. However, the comparison between classical and Bayesian results demonstrates the strong similarity in their results if one uses a uniform reference prior distribution, and that the classical approach in fact gives the most pessimistic results. Indeed, in making a confidence statement using that approach one has (perhaps unwittingly) adopted the most pessimistic prior. Using the proposed Bayesian technique (with Jeffreys' reference prior) makes compliance rules less onerous, particularly for



smaller numbers of samples, while still affording the desired degree of protection. Informative prior belief as to likely performance (good or bad) can easily be incorporated into the technique, as shown. At larger numbers of samples results for all techniques become more similar—the information in the data increasingly overwhelms that in the chosen prior distributions, a well-known property of Bayesian analyses.

*Acknowledgements*—This work was funded in part from the New Zealand Foundation for Research, Science & Technology (contract C01819). Dr M.E.U. Taylor (New Zealand Ministry of Health) made useful suggestions regarding Tables 1 and 2.

#### REFERENCES

- Abramowitz M. and Stegun I. A. (1970) *Handbook of Mathematical Functions*. Dover, New York.
- ANZECC & ARMCANZ (1999) *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*, Vol. 2. Public Comment Draft. Australian & New Zealand Environment and Conservation Council, Agricultural & Resource Management Council of Australia & New Zealand.
- Casella G. and Berger R. L. (1987) Reconciling Bayesian and frequentist evidence in the one-sided testing problem. *J. Am. Stat. Assoc.* **82**, 106–111 (with discussion following subsequent paper, by Berger and Sellke).
- DeGroot M. H. (1973) Doing what comes naturally: interpreting a tail area as a posterior probability or as a likelihood ratio. *J. Am. Stat. Assoc.* **68**, 966–969.
- Edwards W., Lindman H. and Savage L. J. (1963) Bayesian statistical inference for psychological research. *Psychol. Rev.* **70**, 193–242.
- Ellis, J. C. (1989). *Handbook on the Design & Interpretation of Monitoring Programmes*. Report NS 29, Water Research Centre plc. Medmenham, UK.
- Ellis J. C. and Lacey R. F. (1980) Sampling: defining the task and planning the scheme. *Water Pollut. Control* **79**, 452–467 (discussion 482–484).
- Goodman S. N. (1993) *p* values, hypothesis tests and likelihood: implications for epidemiology of a neglected historical debate. *Am. J. Epidemiol.* **5**(1), 485–496.
- Harlow, L. L., Muliak, S. A., Steiger, J. H. (Eds) (1997) *What If There Were No Significance Tests?* Lawrence Erlbaum Ass., Mahwah, New Jersey.
- Lee P. M. (1989) *Bayesian Statistics: An Introduction*. Edward Arnold, London.
- MoH. (1995) *Drinking-Water Standards for New Zealand 1995*. Ministry of Health, Wellington, New Zealand.
- Pratt, J. W. (1965). Bayesian interpretation of standard inference statements. *J. Roy. Stat. Soc. B* (27), 169–203.
- Press S. J. (1989) *Bayesian Statistics: Principles, Models, and Applications*. Wiley, New York.
- Press W. H., Teukolsky S. A., Vetterling W. T. and Flannery B. R. (1986) *Numerical Recipes in FORTRAN*. Cambridge University Press, Cambridge, UK.
- Royall R. M. (1997) *Statistical Evidence: A Likelihood Paradigm*. Chapman & Hall/CRC, Boca Raton, FL.

#### APPENDIX

##### *Beta and binomial functions*

The pdf of  $x \sim \text{Be}(a, b)$  is  $f(x; a, b) = x^{a-1}(1-x)^{b-1} / B(a, b)$ , where  $0 < x < 1$ ,  $a > 0$ ,  $b > 0$ ,  $\text{Be}(a, b)$  denotes the two-parameter beta distribution and  $B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$  is the beta function. The incomplete beta function ratio is  $I_X(a, b) = \int_0^X f(t; a, b) dt$ , so that  $I_{X=1}(a, b) = 1$ . It is related to the cumulative binomial distribution by  $\sum_{i=e+1}^n {}^n C_i X^i (1-X)^{n-i} = I_X(e+1, n-e)$  (Abramowitz and Stegun, 1970). Limiting cases are  $I_X \times (a, 0) \equiv 0$  and  $I_X(0, b) \equiv 1$ .

