ENVIRONMETRICS
Environmetrics 2006; **17**: 95–106
Published online 13 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/env.752

Evaluating water quality using power priors to incorporate historical information

Yuyan Duan*,†, Keying Ye and Eric P. Smith

Department of Statistics, Virginia Tech, Blacksburg, VA 24061-0439, U.S.A.

SUMMARY

To assess water quality standards, measurements of water quality under the Clean Water Act are collected on a regular basis over a period of time. The data are analyzed to evaluate the percentage of samples exceeding the standard. One problem is that current data are limited by the time range and consequently the sample size is inadequate to provide necessary precision in parameter estimation. To address this issue, we present a Bayesian approach using a power prior to incorporate historical data and/or the data collected at adjacent stations. We develop a modified power prior approach and discuss its properties under the normal mean model. Several sets of water quality data are studied to illustrate the implementation of the power prior approach and its differences from alternative methods. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: Bayesian analysis; historical data; power prior; water quality standards

1. INTRODUCTION

One important problem in environmental statistics is the evaluation of air or water quality standards. Issues include the definition of standards (Barnett and O'Hagan, 1997), trend assessment (Hirsch *et al.*, 1982) and the evaluation of data from locations to determine compliance. A standard for a chemical or pollutant is a qualitative or quantitative description of expectation for the chemical or pollutant. To implement such a standard a numerical criterion is often required. The numerical criteria might be different for a lake used for drinking water than for a lake used for fishing. Also associated with the standard are expectations related to frequency, magnitude and duration. Air quality evaluation often involves the expected frequency of violation (for example, the ozone standard, Thompson *et al.*, 2002). However, evaluation of water quality standards often involves a percentile view. For example, for dissolved oxygen, a site is expected to have 10% or fewer samples in violation.

*Correspondence to: Yuyan Duan, Department of Statistics, Virginia Tech, Blacksburg, VA 24060-0439, U.S.A.
†E-mail: yuduan@vt.edu

Contract/grant sponsor: U.S. Environmental Protection Agency—Science to Achieve Results (STAR); contract/grant number: RD 83136801-0.

Copyright © 2005 John Wiley & Sons, Ltd.

Received 10 December 2004
Accepted 23 May 2005

To assess water quality standards, measurements of water quality under the Clean Water Act (e.g. pH, dissolved oxygen, biological oxygen demand) are collected on a regular basis (e.g. quarterly) over a two year period and analyzed to evaluate the percentage of samples in violation of the standard. A common approach accepted by the U.S. Environmental Protection Agency (USEPA) is the raw score approach that simply calculates the proportion of violations and declares the water segment impaired if this proportion exceeds 10%. Smith *et al.* (2001) noted that this is essentially a statistical hypothesis testing problem assuming a *binomial* population without controlling for error rates. They suggested to use a binomial test and discussed both Type I and Type II error probabilities. They showed that the tests using a statistical method have greater power than the raw score approach by USEPA. Furthermore, Ye and Smith (2002) proposed to use a Bayesian test on the desired percentile of the distribution (e.g. 10th percentile) that is of interest, to check for impairment of the monitoring site (see also McBride and Ellis, 2001). By using Bayesian methodology, the quantity of interest (e.g. the percentile of the measurement distribution) can be naturally treated as a parameter and thus its posterior distribution can be used to make necessary decisions.

However, because all the approaches mentioned above do not fully use all the information provided by the data in the sense that only the data with 'standard violation' or not are used in the analysis, Smith *et al.* (2003) suggested an approach using a tolerance limit, pointing out that this would reflect the magnitude of violation (see also Smith, 2002).

Suppose a water quality measurement follows a certain distribution and a small value indicates a violation. Then the raw data can be used to test a hypothesis

$$\mathrm{H}_0 : L \geq L_0 (\text{no violation}) \text{ versus } \mathrm{H}_1 : L < L_0 (\text{violation}) \tag{1}$$

where L denotes the true lower percentile of the population distribution, and L_0 is the standard. To test the hypotheses in (1), one may consider rejecting the null hypothesis and hence declaring impairment when the posterior probability of the null hypothesis given the data is small (e.g. <0.05).

Unfortunately, because decisions are based on data from a limited time period (for water monitoring data, many current samples have only two years of data), the sample size is often inadequate to provide necessary precision in parameter estimates. In such situations, 'historical' data, a data set from similar studies or a data set from previous time periods can be very helpful in interpreting the current status of water quality. Owing to the nature of updating information sequentially, it is straightforward to use a Bayesian approach with an informative prior on the model parameters to incorporate the historical data into the current study. A traditional approach to incorporating historical data is to construct an informative prior using the historical data and then to combine the prior with the likelihood to yield the posterior distribution for statistical inference. This implies a simple pooling of current data and historical data, since the two data sets are equally weighted, and can be well justified by assuming current and historical data are from the same population. However, the population parameters may change over time, or over different sites, although current and historical data are usually assumed to follow distributions in the same family. If the sample size of the historical data is much larger than that of current data and heterogeneity exists between the current and previous studies, historical data would dominate the analysis and the data pooling may result in misleading conclusions.

To address this issue, Ibrahim and Chen (2000) proposed the concept of the *power prior*, based on the notion of the availability of historical data. The basic idea is to let a power parameter $\delta\ (0 \leq \delta \leq 1)$ tell us how much historical data are to be used in the current study. Ibrahim and Chen (2000) and Chen *et al.* (2000) demonstrated how to construct power priors and discussed the

Copyright © 2005 John Wiley & Sons, Ltd.

general conditions for propriety. They also examined the power prior approach for generalized linear models, generalized linear mixed models, semiparametric proportional hazards models, and cure rate models with real data examples. Ibrahim *et al.* (2003) gave a formal justification of the power prior as an optimal class of informative priors, and showed that the power prior is a 100% efficient information-processing rule in the sense that the ratio of the output to input information is equal to 1. However, in their approach, the power parameter δ always has a tendency to be close to zero, which suggests that much of a historical data set is not used (see Duan and Ye, A modified Bayesian power prior approach, manuscript in preparation). Here we propose a modified power prior approach with applications in water quality assessment. In Section 2, the general development of a modified power prior approach is given and certain properties of the approach under a normal population are discussed. More development of this modified approach can be found in Duan and Ye (A modified Bayesian power prior approach, manuscript in preparation). In Section 3, we will demonstrate the application of power priors in water quality assessment with two examples. Finally, in Section 4, discussion of this Bayesian power prior approach and its usage in environmental and ecological studies is given.

2. POWER PRIOR BAYESIAN ANALYSIS

2.1. Power prior approach

Suppose that θ is the parameter of interest in a water quality measurement, for instance, concentration of a chemical. Suppose that such a measurement follows a distribution and $\mathrm{L}(\theta|D_0)$ is the likelihood function of θ based on the historical data, denoted by D_0. In this article we assume that, given θ, historical data (D_0) and current data (D) are independent random samples from an exponential family. $\pi(\theta)$ is taken as the initial prior before any historical information is gathered, and usually it is a non-informative prior. Given δ, Ibrahim and Chen (2000) define the power prior of θ for the current study as

$$\pi(\theta|D_0, \delta) \propto \mathrm{L}(\theta|D_0)^{\delta}\pi(\theta) \tag{2}$$

The parameter δ measures the portion of historical information needed in the current study and is described using the prior in (2). The case $\delta = 0$ means that no historical data should be used, while $\delta = 1$ gives equal weight to $\mathrm{L}(\theta|D_0)$ and the likelihood of the current study $\mathrm{L}(\theta|D)$, resulting in full incorporation of the historical data. Therefore, (2) can be viewed as a generalization of the usual Bayesian update of $\pi(\theta)$ (see discussion in Ibrahim and Chen, 2000). The power parameter δ can be interpreted as a precision parameter. For example, consider the case of a normal sample with known variance. Suppose that D_0 consists of n_0 observations, $X_1, X_2, \ldots, X_{n_0}$, from the normal population with unknown mean parameter θ and known variance σ^2. If the prior $\pi(\theta)$ is assumed to be uniform (non-informative), (2) implies a prior distribution of θ for the current data set D, $\theta|D_0, \delta \sim \mathrm{N}(\bar{x}_0, \sigma^2/\delta n_0)$, where $\bar{x}_0$ is the sample mean of the historical data. Hence, δ can be viewed as part of the precision parameter, because smaller δ implies larger power prior variance, while larger δ means smaller power prior variance.

The power prior $\pi(\theta|D_0, \delta)$ in (2) was initially elicited for fixed δ. However, since δ is not necessarily predetermined, we may extend it further to the case that δ is random. Thus the power prior $\pi(\theta|D_0, \delta)$ in (2) is treated as a distribution of θ conditional on δ and historical information. The power

Copyright © 2005 John Wiley & Sons, Ltd.

prior specification on (θ, δ) is then completed by specifying a prior distribution for δ. A random δ gives the investigator more flexibility in weighting the historical data. A natural prior for δ would be a Beta(α, β) distribution, or simply a uniform distribution, since $0 \leq \delta \leq 1$.

We prefer a modification of the original approach of Ibrahim and Chen (2000) for two reasons. First, in their original approach, Ibrahim and Chen (2000) constructed the joint power prior of (θ, δ) as

$$\pi(\theta, \delta | D_0) \propto \mathrm{L}(\theta | D_0)^{\delta} \pi(\theta) \pi(\delta) \tag{3}$$

Notice that multiplying the likelihood function $\mathrm{L}(\theta | D_0)$ by a positive constant would change the joint prior of (θ, δ) and consequently the posterior, which violates the likelihood principle. Another problem with the original approach comes up when we investigated the application of power priors on normal mean models. The influence of historical data is small no matter how compatible current and historical data are, and the inference on θ is not much different from that without considering historical data (see more discussions in Duan and Ye, A modified Bayesian power prior approach, manuscript in preparation). Therefore, we propose the following power prior:

$$\pi(\theta, \delta | D_0) = \frac{\mathrm{L}(\theta | D_0)^{\delta} \pi(\theta) \pi(\delta)}{\int_{\Theta} \mathrm{L}(\theta | D_0)^{\delta} \pi(\theta) \mathrm{d}\theta} \tag{4}$$

in the region of δ such that the denominator in (4) is not infinity. In this article, Θ is used to denote the range of parameter θ. This new approach is free of the problems mentioned above and it needs little to ensure the propriety of the joint power prior for (θ, δ). The more discussion of this prior is given in next section.

2.2. *General development of the power prior*

For ease of exposition, first we develop a power prior and the consequent posterior with only one historical data set, then follow with the extension to multiple historical data sets in Section 2.4.

The modified power prior distribution of (θ, δ) is given in (4). The assumptions and conditions of this prior are $0 \leq \delta \leq 1$, $\mathrm{L}(\theta | D_0) \geq 0$, $\pi(\theta) \geq 0$ and $P_{\theta}(\mathrm{L}(\theta | D_0) > 0) > 0$. Consequently, we have $\int_{\Theta} \mathrm{L}(\theta | D_0)^{\delta} \pi(\theta) \mathrm{d}\theta > 0$. Define A as the following:

$$A = \left\{ \delta : 0 < \int_{\Theta} \mathrm{L}(\theta | D_0)^{\delta} \pi(\theta) \mathrm{d}\theta < \infty \right\} \tag{5}$$

Furthermore, define A^{-} as the complement of A in [0,1]. The assumption

$$\int_{\Theta} \mathrm{L}(\theta | D_0) \pi(\theta) \mathrm{d}\theta < \infty$$

guarantees that $\pi(\theta)$ can be used as a prior in a coherent Bayesian updating scheme due to the consequent proper posterior of θ. This assumption constrains our discussion to a sensible range, and implies $1 \in A$.

Copyright © 2005 John Wiley & Sons, Ltd.

We propose a joint power prior distribution for (θ, δ) of the form

$$\pi(\theta, \delta|D_0) = \begin{cases} M \frac{\mathrm{L}(\theta|D_0)^{\delta}\pi(\theta)\pi(\delta)}{\int \mathrm{L}(\theta|D_0)^{\delta}\pi(\theta)\mathrm{d}\theta}, & \text{if } \delta \in A \\ 0, & \text{if } \delta \in A^{-} \end{cases} \tag{6}$$

where M is a normalization constant. This modified joint power prior is built on the conditional power prior in (2) and the following proposition.

$$\pi(\delta|D_0) = \begin{cases} M\pi(\delta), & \text{if } \delta \in A \\ 0, & \text{if } \delta \in A^{-} \end{cases}$$

It is easy to check that the joint prior $\pi(\theta, \delta|D_0)$ defined above is always proper, which also ensures the propriety of the joint posterior for (θ, δ). Such a joint posterior distribution for (θ, δ) can be written as

$$\pi(\theta, \delta|D_0, D) \propto \mathrm{L}(\theta|D)\pi(\theta, \delta|D_0) \propto \frac{\mathrm{L}(\theta|D)\mathrm{L}(\theta|D_0)^{\delta}\pi(\theta)\pi(\delta)}{\int_{\Theta} \mathrm{L}(\theta|D_0)^{\delta}\pi(\theta)\mathrm{d}\theta} I_A(\delta)$$

where $I_A(\delta) = 1$ if $\delta \in A$ and 0 otherwise.

The marginal posterior distributions of δ and θ can be derived as follows:

$$\pi(\delta|D_0, D) \propto \pi(\delta) \frac{\int_{\Theta} \mathrm{L}(\theta|D)\mathrm{L}(\theta|D_0)^{\delta}\pi(\theta)\mathrm{d}\theta}{\int_{\Theta} \mathrm{L}(\theta|D_0)^{\delta}\pi(\theta)\mathrm{d}\theta} I_A(\delta) \tag{7}$$

and

$$\pi(\theta|D_0, D) \propto \pi(\theta)\mathrm{L}(\theta|D) \int_A \frac{\mathrm{L}(\theta|D_0)^{\delta}\pi(\delta)I_A(\delta)}{\int_{\Theta} \mathrm{L}(\theta|D_0)^{\delta}\pi(\theta)\mathrm{d}\theta} \mathrm{d}\delta \tag{8}$$

We can make inferences on θ, e.g. in hypothesis testing, Bayesian interval calculation, or model selection, based on the marginal posterior distribution of θ in (8). On the other hand, using the marginal posterior distribution of δ shown in (7), the properties of the power parameter δ can be studied.

We next discuss the case of a normal population. Then in Section 3 environmental applications are demonstrated.

2.3. Normal population

Suppose we are interested in a normal population and inference for the normal mean with unknown variance, by incorporating both current and historical data. Although there may be many observations available for study, not all are used. For instance, in one location there may be more than 50 years of data available; however, it is not reasonable to treat observations 50 years ago as providing the same information as just one prior year of information about the water quality at the same location. If the data set can be assumed approximately normal, we can use the approach discussed below, the power prior analysis.

Suppose that current data $D = (x_1, \ldots, x_n)$ come from a normal $n(\mu, \sigma^2)$ population with unknown μ and σ^2, and $D_0 = (x_{01}, \ldots, x_{0n_0})$ is a historical data set. Let $\bar{x}_0 = \frac{1}{n_0}\sum_{i=1}^{n_0} x_{0i}$, $\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$,

Copyright © 2005 John Wiley & Sons, Ltd.

$\hat{\sigma}_0^2 = \frac{1}{n_0}\sum_{i=1}^{n_0}(x_{0i} - \bar{x}_0)^2$ and $\hat{\sigma}^2 = \frac{1}{n}\sum_{j=1}^{n}(x_i - \bar{x})^2$. Furthermore, suppose that we use the prior $\pi(\mu, \sigma^2) \propto (1/\sigma^2)^a$ as the initial prior of (μ, σ^2), where $a > 0$ is a pre-determined constant. Note that $a = 1$ corresponds to the reference prior (Berger and Bernardo, 1992), while $a = \frac{3}{2}$ results in the Jeffreys prior (Jeffreys, 1946). Also, the prior distribution of δ is a Beta(α, β), where hyper-parameters α and β are all known. Following (6), the joint power prior distribution of (μ, σ^2, δ) can be expressed as

$$\pi(\mu, \sigma^2, \delta | D_0) \propto \frac{\delta^{\frac{\delta n_0}{2} + a + \alpha - 2}(1-\delta)^{\beta - 1}}{\left(\frac{2\sigma^2}{n_0\hat{\sigma}_0^2}\right)^{\frac{\delta n_0}{2} + a}\Gamma\left(\frac{\delta n_0 - 3}{2} + a\right)} \exp\left\{-\frac{\delta n_0}{2\sigma^2}[\hat{\sigma}_0^2 + (\mu - \bar{x}_0)^2]\right\}$$

where the range of δ is $(b, 1]$ for $b \geq 0$ or $[0, 1]$ for $b < 0$ and $b = \frac{2}{n_0}(\frac{3}{2} - a)$. Hence, if the reference prior of θ is used, the set A defined in (5) is $(\frac{1}{n_0}, 1]$ for the normal mean model, while if the Jeffreys prior of θ is used, $A = (0, 1]$. The lower bound b suggests that the information in historical data is automatically taken into account to a certain extent, depending on the availability of historical data. However, such a case may be changed once the original prior is changed.

Combining the joint power prior with the likelihood based on the current data $\mathrm{L}(\mu, \sigma^2 | D)$, we obtain the joint posterior distribution of (μ, σ^2, δ). Integrating μ and σ^2 out of $\pi(\mu, \sigma^2, \delta | D_0, D)$, the marginal posterior distribution of δ leads to

$$\pi(\delta | D_0, D) \propto \frac{\delta^{\frac{\delta n_0}{2} + a + \alpha - 2}(1-\delta)^{\beta-1}\Gamma\left(\frac{\delta n_0 + n - 3}{2} + a\right)}{\left[\frac{\delta n}{\delta n_0 + n}\frac{(\bar{x}_0 - \bar{x})^2}{\hat{\sigma}_0^2} + \delta + \frac{n}{n_0}\frac{\hat{\sigma}^2}{\hat{\sigma}_0^2}\right]^{\frac{\delta n_0 + n - 3}{2} + a}\Gamma\left(\frac{\delta n_0 - 3}{2} + a\right)}$$

with the range described above. The behavior of the power parameter δ can be studied from this marginal posterior distribution.

Similarly, the marginal posterior distribution of (μ, σ^2) can be derived by integrating δ out of $\pi(\mu, \sigma^2, \delta | D_0, D)$, but it does not have a closed form. Combined with $\pi(\delta | D_0, D)$, the conditional posterior distribution of (μ, σ^2) given δ indirectly reflects the behavior of $\pi(\mu, \sigma^2 | D_0, D)$. Both the posterior distribution of μ conditional on δ and the posterior of σ^2 conditional on δ turn out to be commonly used distributions. From standard calculations of Bayesian analysis using a normal population (see Gelman *et al.*, 1995), we find that the conditional posterior distribution of μ, given δ and data, follows a Student-*t* distribution with location parameter $(\delta n_0 \bar{x}_0 + n\bar{x})/(\delta n_0 + n)$, scale parameter

$$\sqrt{\frac{2}{C(\delta)}\frac{1}{(\delta n_0 + n + 2a - 3)(\delta n_0 + n)}}$$

and degrees of freedom $(\delta n_0 + n + 2a - 3)$, where

$$C(\delta) = \frac{2}{\frac{\delta n_0 n (\bar{x}_0 - \bar{x})^2}{\delta n_0 + n} + \delta n_0 \hat{\sigma}_0^2 + n\hat{\sigma}^2}$$

Copyright © 2005 John Wiley & Sons, Ltd.

Furthermore, the conditional posterior distribution of σ^2, given δ and the data, follows an inverse-gamma distribution with parameter $(\delta n_0 + n + 2a - 3)/2$ and $C(\delta)^{-1}$ (again, see Gelman *et al.*, 1995).

The role of the power parameter δ is to control the influence of the historical data on the current study. Using simulation studies, we found that this power control parameter is adjusted automatically based on the compatibility between the historical and current data, and also based on the sample sizes of the two studies. In the case of the normal mean model, the compatibility between the historical and current data can be measured by differences in sample means and sample variances. If historical and current data are fully compatible, i.e. two data sets have the same sample mean and sample variance, the marginal posterior mode of δ is always 1, no matter how high the ratio n_0/n is. This is very rational since when the historical data can contribute necessary information into the current study, we would like to use it as much as possible to achieve higher precision. How quickly the mode reaches 1 also depends on the extent of compatibility between the historical and current samples. When the discrepancy between historical and current samples is large, the posterior mode of δ goes to zero quickly. The posterior mode of δ decreases with n_0/n, and attains 1 with very small n_0/n. These trends imply that the random δ responds to data in a sensitive and desirable way.

2.4. *Extension to multiple historical data sets*

The priors defined in (6) can easily be generalized to multiple historical data sets. Suppose there are m historical studies. We define D_{0j} to be the historical data based on the jth study, $j = 1, \ldots, m$, and $D_0 = (D_{01}, \ldots, D_{0m})$. Chen *et al.* (2000) suggested defining a different weight parameter δ_j for the jth historical study and taking the δ_js to be i.i.d. Beta random variables with hyperparameters (α, β). Let $\underline{\delta} = (\delta_1, \ldots, \delta_m)$. Under the new approach, the power prior in (6) can be generalized as

$$\pi(\theta, \underline{\delta}|D_0) \propto \frac{\left(\prod_{j=1}^{m} \mathrm{L}(\theta|D_{0j})^{\delta_j}\pi(\delta_j|\alpha,\beta)\right)\pi(\theta)}{\int_\Theta \left(\prod_{j=1}^{m} \mathrm{L}(\theta|D_{0j})^{\delta_j}\right)\pi(\theta)\mathrm{d}\theta} I_B(\underline{\delta})$$

where $B = \{(\delta_1, \ldots, \delta_m) : 0 < \int_\Theta \left(\prod_{j=1}^{m} \mathrm{L}(\theta|D_{0j})^{\delta_j}\right)\pi(\theta)\mathrm{d}\theta < \infty\}$

This framework can accommodate potential compatibility among different sites or different time periods. For example, we could take data collected over time at different but adjacent sites as different historical data sets. Moreover, data collected over a long period may be divided into several historical data sets instead of being treated as one big historical data set. In such a way, the role of historical data can be more accurately evaluated. In Section 3.2, we will discuss an example implementing the modified power prior approach using multiple site information.

3. POWER PRIOR APPLICATIONS IN EVALUATING SITE IMPAIRMENT

When applying Bayesian analysis with power priors to water quality data, two kinds of additional information could be incorporated: past information or information from adjacent sites. We compare the modified power prior approach to other methods, including the USEPA's raw score method, the binomial test and a traditional Bayesian approach using the reference prior. Suppose that

Copyright © 2005 John Wiley & Sons, Ltd.

the measurements of water quality follow a normal distribution, and, for ease of comparison, the normal model with a simple mean is considered. Based on the equations derived in Section 2.3, we use *Winbugs* (Release 1.4) to simulate the marginal posterior distribution for (μ, σ^2), and then the percentile of interest.

3.1. Using past information to build the prior

In this example, we use measurements of pH to evaluate impairment of four sites in Virginia individually. Of interest in these data sets is the determination of whether the pH values at a site indicate that the site violates a (lower) standard of 6.0 more than 10% of the time. For each site, larger sample size is associated with the historical and smaller with the current data. In this example, pH data collected over a two-year or three-year period are treated as the current data, while pH data collected over the previous nine years represents one single historical data set. The current data and historical data are plotted side by side for each site in Figure 1. In the power prior approach, a violation is evaluated using a Bayesian test of

$$\mathrm{H}_0 : L \geq 6.0\,(\text{no impairment, don't list})$$

$$\mathrm{H}_1 : L < 6.0\,(\text{impairment, list})$$

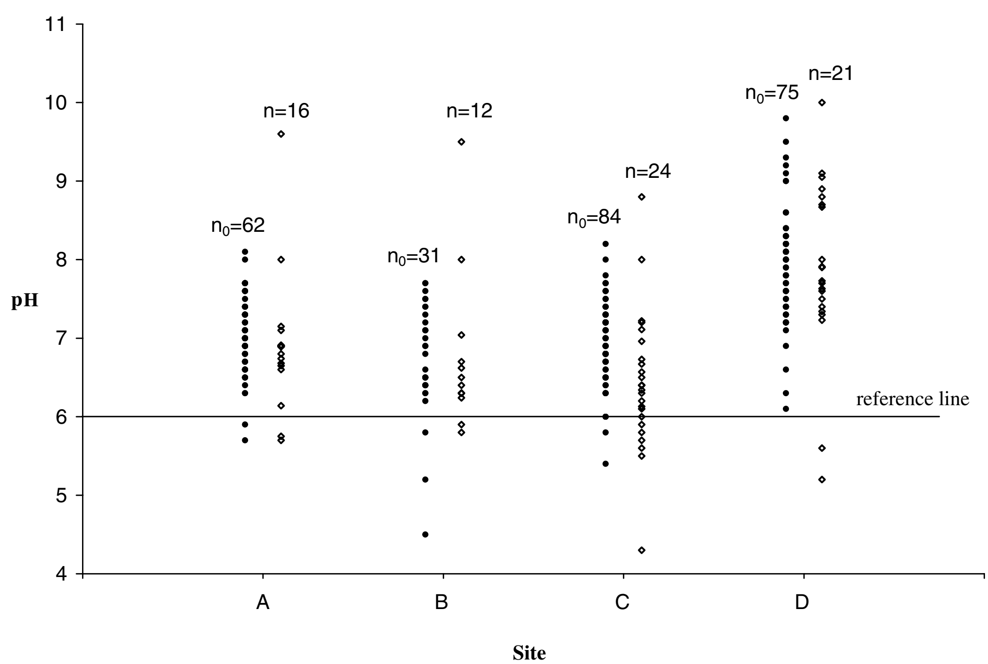

Figure 1. pH data collected at four stations. For each site, historical data are on the left (circle) and current data on the right (diamond)

Copyright © 2005 John Wiley & Sons, Ltd.

where L is the lower 10th percentile of the distribution for pH. Comparison of results from different methods is presented in Table 1.

In Table 1, the summarization of the current and historical data are given, along with the percentage of samples below the standard (raw score values). Furthermore, the test results using a binomial population as well as the test results using both the reference prior analysis (without incorporating historical data) and power prior analysis (with reference prior, i.e. $a = 1$ in Section 2.3) are presented. For sites C and D, the raw data are used in the Bayesian analysis, while for sites A and B, a log transformation of data is applied, because the lognormal distribution fits the data better than normal.

Sites A, B and C have greater than 10% of observations below the standard of 6.0. Therefore, we would declare that all sites, except D are impaired using EPA's raw score approach. Note that this raw score approach results in higher Type I error probability (see Smith *et al.*, 2001), which means that it would declare more impaired sites than it should have. On the other hand, if the 0.05 significance level is used, the binomial test would only indicate site C as impaired. The Bayesian test using the reference prior results in a similar conclusion, although the p-values are smaller, compared to the binomial tests, in all cases but one. Here we use the posterior probability of H_0 as equivalent to the p-value (Berger, 1985) for testing a one-sided hypothesis. Using historical data does lead to different conclusions for site B. The test using a power prior results in significance for sites B and C. In the case of site B, there are around 10% of historical observations below 6.0. Hence our prior opinion of the site is suggestive of impairment. Less information is therefore required to declare impairment relative to a reference prior, and the result is a smaller p-value.

Another notable advantage of the power prior method is that it improves the estimation of L by using past information. This can be shown by the consistently smaller posterior standard deviation of L with the power prior than with the reference prior for all four sites.

3.2. *Borrowing information from adjacent sites*

Alternate sources of information from other locations can also be used to aid inference on environmental quantity at the site of interest. In this section, measurements of dissolved oxygen

Table 1. Comparison of the power prior method with alternative methods in evaluating site impairment when one historical data set is available. In the table, n and n_0 are sample sizes, mean (s.d.) refers to sample mean (sample standard deviation), and s.d. of L is the posterior standard deviation of L. Percent below is the percentage of samples below the EPA standard (6 for pH)

Site	Current data		Historical data		% Below	Binomial P-value	Posterior probability of H_0 (s.d. of L)	
	n	Mean (s.d.)	n_0	Mean (s.d.)			Reference prior	Power prior
A	16	6.91 (0.90)	62	7.05 (0.47)	0.13	0.4853	0.2074 (0.27)	0.6027 (0.21)
B	12	6.78 (1.03)	31	6.73 (0.71)	0.17	0.3410	0.0627 (0.34)	0.0294 (0.19)
C	24	6.43 (0.88)	84	6.95 (0.49)	0.25	0.0277	0.0003 (0.26)	0.0017 (0.24)
D	21	7.87 (1.11)	75	7.88 (0.67)	0.10	0.6353	0.8673 (0.36)	0.9831 (0.25)

Copyright © 2005 John Wiley & Sons, Ltd.

(DO) are used to evaluate impairment of four sites in the Philpott reservoir in Virginia. Of interest here is to determine whether DO values violate a (lower) standard of 5.0 more than 10% of the time at each site. DO values are plotted in Figure 2.

Since the four sites are in the same reservoir, the geographical connection suggests similar water quality at four sites due to the spatial correlation among their data. So data collected at one site should partially reflect the water quality at the other three. For each site, DO data collected over a three-year period are treated as the current data, while DO data collected over the same period but at the other three sites are referred to three 'historical' data sets. This section illustrates the utility of power priors with multiple historical data sets as discussed in Section 2.4. In the power prior approach, the problem is set up as a Bayesian test of

$$\mathrm{H}_0 : L \geq 5.0\,(\text{no impairment, don't list})$$

$$\mathrm{H}_1 : L < 5.0\,(\text{impairment, list})$$

where L is the lower 10th percentile of the distribution for DO. A comparison of results from different methods is presented in Table 2.

Both binomial and reference prior Bayesian tests, which are only based on each individual site's data, showed the same results on sites A, C and D, and different results on site B. However, the power prior Bayesian analysis results in significance of all tests and hence a decision to declare all sites as impaired. Coincidentally, this conclusion consistently matches the EPA's raw data approach.

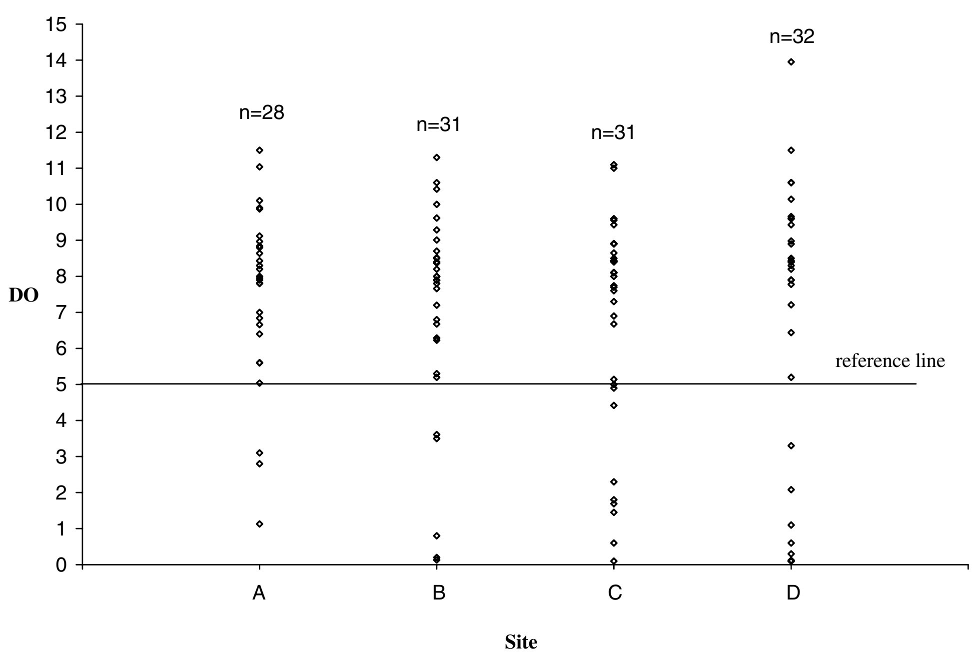

Figure 2. DO data collected at four stations on Philpott reservoir (years 2001, 2002, and 2003)

Copyright © 2005 John Wiley & Sons, Ltd.

Table 2. Comparison of the power prior method with alternative methods for evaluating site impairment when multiple historical data sets are available. In the table, n is sample size, mean (s.d.) refers to sample mean (sample standard deviation), and s.d. of L is the posterior standard deviation of L. Percent below is the percentage of samples below the EPA standard (5 for DO)

Site	n	Mean (s.d.)	% Below	Binomial P-value	Posterior probability of H_0 (s.d. of L)	
					Reference prior	Power prior
A	28	7.55 (2.41)	0.11	0.5406	0.1640 (0.65)	0 (0.55)
B	31	7.10 (2.88)	0.16	0.1932	0.0038 (0.73)	0 (0.54)
C	31	6.66 (3.09)	0.26	0.0096	0 (0.79)	0 (0.54)
D	32	6.67 (4.02)	0.28	0.0033	0 (1.01)	0 (0.62)

Clearly, the power prior approach improves the estimation of L by borrowing information from adjacent sites, which can be seen by the consistently smaller posterior standard deviation of L using the power prior relative to the use of the reference prior for all four sites. Other sites' information does affect the inference substantially. For example, based on each site's data individually, site A seems to have higher water quality than the other three. However, when we use the power prior, the low water quality at sites B, C and D drags down the estimate of L for site A. So we would not declare site A impaired based on the posterior probability of H_0 without considering the other three sites' information, but we would do so with a power prior.

4. DISCUSSION

The power prior is elicited to take into account the heterogeneity between historical and current data when we are not be able to describe or adequately model the heterogeneity explicitly. To increase the estimating efficiency, we should model the heterogeneity explicitly whenever it is possible. The power prior approach may be viewed as a non-parametric method to incorporate the historical data into the current study, as the likelihood of the historical data is always raised to a fractional power regardless of the specific form of heterogeneity. The fact that we often don't have enough knowledge to model such heterogeneity or to specify a fixed power makes this power prior with random δ especially attractive in practice. In addition, we could combine subjective information about δ by using different $\pi(\delta)$. An extreme case would be when researchers believe that the current and historical data are from the same population, in which case $\delta = 1$ should be used even when two samples seem incompatible.

As mentioned before, the power prior is most useful when the size of the current data set is small but ample historical data are available. If the current data set is large enough to achieve a satisfactory precision for parameter estimates, the benefit from historical data is diminished. On the other hand, the accuracy of estimates lost would be more than the precision gained if the historical data are used in such a situation. A useful approach then might be to incorporate common information through a linear

Copyright © 2005 John Wiley & Sons, Ltd.

model (Smith *et al.*, 2003). This may be of considerable value in water quality monitoring in which many sites are sampled but sample sizes for individual sites may be small.

ACKNOWLEDGEMENT

This research was funded in part by U.S. EPA—Science To Achieve Results (STAR) Grant RD 83136801-0. Although the research described in the article has been funded wholly or in part by the U.S. Environmental Protection Agency 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.

REFERENCES

Barnett V, O'Hagan A. 1997. *Setting Environmental Standards*. Chapman & Hall: London.

Berger JO. 1985. *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. Springer Verlag: New York.

Berger JO, Bernardo JM. 1992. On the development of reference priors. In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Clarendon Press: Oxford; 35–60.

Chen M-H, Ibrahim JG, Shao Q-M. 2000. Power prior distributions for generalized linear models. *Journal of Statistical Planning and Inference* **84**: 121–137.

Gelman A, Carlin JB, Stern HS, Rubin DB. 1995. *Bayesian Data Analysis*, 2nd edn. Chapman & Hall and CRC Press: Boca Raton, Florida.

Hirsch RM, Slack JR, Smith RA. 1982. Techniques for trend analysis for monthly water quality data. *Water Resources Research* **18**: 107–121.

Ibrahim JG, Chen M-H. 2000. Power prior distributions for regression models. *Statistical Science* **15**: 46–60.

Ibrahim JG, Chen M-H, Sinha D. 2003. On optimality properties of the power prior. *Journal of the American Statistical Association* **98**: 204–213.

Jeffreys H. 1946. An invariant form for the prior probability in estimation problems, *Proceedings of the Royal Statistical Society of London, Series A* **186**: 453–461.

McBride GB, Ellis JC. 2001. Confidence of compliance: a Bayesian approach for percentile standards. *Water Research* **35**: 1117–1124.

Smith EP, Ye K, Hughes C, 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**: 606–612.

Smith EP, Zahran A, Mahmoud M, Ye K. 2003. Evaluation of water quality using acceptance sampling by variables. *Environmetrics* **14**: 373–386.

Smith RW. 2002. The use of random-model tolerance intervals in environmental monitoring and regulation. *Journal of Agricultural, Biological and Environmental Statistics* **7**: 74–94.

Thompson ML, Cox LH, Sampson PD, Caccia DC. 2002. Statistical hypothesis testing for U.S. environmental regulatory standards for ozone. *Environmental and Ecological Statistics* **9**: 321–339.

Ye K, Smith EP. 2002. A Bayesian approach to evaluating site impairment. *Environmental and Ecological Statistics* **9**: 379–392.

Copyright © 2005 John Wiley & Sons, Ltd.