Comments on the Proposed California Ocean Plan Amendment

Dr. Kenneth H. Reckhow and Dr. Song Qian
K. H. Reckhow
Duke University

March 24, 2005

The Proposed Ocean Plan Amendment suggested three statistical procedures for determining whether a discharge will cause, or have a reasonable potential to cause, an excursion above an Ocean Plan Water Quality Objective.

The first statistical methodology is applied when facility-specific monitoring data are available. This method originated from the US EPA (TSD) to "project an estimated maximum concentration" for an effluent, based on a small number of concentration measurements. The proposed method is different from the TSD method in two ways. First, the proposed method uses the sufficient statistics of the data (the sample mean and standard deviation), and the TSD method does not. Second, the proposed method will result in a consistent conservative estimate of the 95%-percentile while the TSD method often fails to be conservative when the population standard deviation is unknown. However, the proposed method has the tendency to overestimate the potential risk.

We performed a Monte Carlo simulation to illustrate this weakness of the proposed method. We randomly drew concentration samples of size n = (5, 10, 20, 50, 120) from a lognormal distribution with log mean 0 and log standard deviation 1. For each set of samples, we estimated the upper confidence bound equation for a lognormal distribution (Ocean Plan, page 9). This process was repeated 150,000 times, and we calculated the frequency of the estimated 95th-percentile exceeding 2 (and 3) times the true 95th-percentile, along with the mean and median of the estimated 95th-percentile (Table 1). For the usual sample sizes (≤ 20), the chance of substantially overestimating (estimated 95th-percentile exceeding 3 times the true value) is not small.

This overestimate of an upper percentile is a result of the requirement that the estimated maximum concentration has a 95% chance of being larger than a bound that itself has a 95% chance of exceeding each concentration measurement. Because the proposed method is based on an assumed lognormal distribution for individual concentration measurements, it will give misleading results if this assumption fails. The lognormal distribution assumption could fail, for example, if there is a small chance of a data entry mistake. The proposed method will be applied to data sets with small sample sizes, so the method's robustness can be an issue. The log-scale sample standard deviation used is relatively robust, but the log-scale sample mean is not. One may consider using the sample median, instead of the sample mean.

On average, the proposed method yields much higher than a 95th-percentile for concentration measurements, closer to a 97th or 99th-percentile -- at the high cost of substantially overstating the concentrations. Less extreme bounds would still meet the conservative concern of ensuring that the "projected maximum concentration" would be at an acceptable level, without unnecessary overestimates. For example, the estimated 98th-percentile $\hat{C}_{98} = \exp(\hat{\mu} + 2.054\hat{\sigma})$ gives more than a 95th-percentile on average for sample sizes over 10. A Monte Carlo simulation (150,000 iterations) provided average percentiles when estimating the 96, 97, 98, 99, and 99.5 percentiles assuming the log concentration follows a normal distribution and using the sample mean and sample standard deviation as known (Table 2). According to these simulation results, we may be able to find these less extreme bounds for selected sample sizes. For example, under our simulation, we can use the rough estimator to estimate the 96th-percentile and ensure coverage of at least 95.5% when the sample size is above 50.

Our simulation indicated that the proposed method is adequate in that it ensures the estimated 95th-percentile is rarely below (less than 5% of the time) the true 95th-percentile of the underlying concentration distribution. Our simulation also indicated that alternative methods, with the same level of protection and without substantially overestimating the potential risk, are available.

The second method is the regression on order statistics (ROS) approach for estimating the mean and variance of the underlying distribution when there are censored values. The ROS method for censored data is well documented. It has been shown that the ROS method is robust against the departure from the log-normality assumption. For a small sample size, the ROS method is less flexible when there are ties in the uncensored data, yet the Ocean Plan proposed to reduce the sample size when there are ties in the data. This is a weakness in the approach. In addition, it is our understanding that the ROS method is less efficient than the traditional MLE when the underlying distribution is known to be lognormal, especially when there are multiple detection limits. In practice, uncertainty associated with the ROS estimates is rarely discussed in literature (and not in the Ocean plan). If the estimated log normal mean and variance are to be used for estimating the 95% upper bound of the 95th-percentile, it is difficult to claim that the estimated upper bound will be 95% confident to be larger than the true 95th-percentile.

The third statistical method is hypothesis testing. The test is poorly presented and may be mischaracterized. For example, the Ocean Plan suggests that the null hypothesis of the test is the exceedance rate being larger than 18% and the alternative is the exceedance rate being less than 3%. This is a misrepresentation of the test. The alternative hypothesis should be that the exceedance rate is less than 18%. The value 3% represents an effect size of 15%, that is, the test will have a small type II error rate (< 20%) when the true exceedance rate is below 3%. The objective of this arrangement is the balance of the type I and type II error rates. A traditional hypothesis test will only ensure a small type I error rate, and the type II error rate is usually associated with the sample size and the unknown true exceedance rate. It is important that the concept of effect size be interpreted correctly. The current presentation would mislead people to believe that the proposed test procedure will keep both the type I and type II errors in check. This can be only conditionally true, and the condition is not spelled out explicitly. When we fail to reject the null hypothesis, we conclude only that the exceedance rate is most likely to be larger than 18%. The type II error (failure to reject the null when the null is wrong) rate will not be reduced to the claimed 20% unless the true exceedance rate is below 3%. The type II error rate will be larger when the true exceeding rate is between 3 and 18%. The small type II error rate is almost meaningless because the odds of having a higher than 20% type II error rate is 5 to 1. The high exceedance rate of the null (18%) is not explained at all. We suspect that the number is used such that a reasonable type I and type II error rate can be used for a small sample size of 16. Curiously, the hypothesis testing approach is designed for data with large (> 80%) censorship. It is unclear why a binary transformation is particularly useful for this situation. For effluent from a given facility, it is likely that the same lab is used for analyzing all samples. Therefore, it is likely that there will be only one detection limit. If the detection limit is above the water quality criterion, it is impossible to carry out the binary transformation. If the detection limit is below the water quality criterion, given the data set has at least 80% censored values, it is almost impossible not to reject the null hypothesis. It will be helpful if an example can be presented to illustrate a typical situation where such hypothesis testing is useful.

When evaluated separately, the first two methods are reasonable and well presented. The third method (nonparametric binomial approach) is, however, questionable. We feel that the document did not present the connection between the third method and the other two methods. A flow chart would be helpful. Because the Plan limits water quality assessment to effluent from a particular facility, the effectiveness of any assessment will often be limited by small sample size and high censorship. The state should consider developing a metadata set to pool information from similar facilities. Pooling information can provide a better estimate of the between and within facility variances for better quantifying the reasonable potential. In addition, when evaluating facilities with few data points or high censorship, decisions can be made based on information from similar facilities with adequate information. We recommend a Bayesian hierarchical modeling approach similar to the model developed for EPA's six-year drinking water regulatory review (Qian et al, 2004, which also addresses the censored data problem). The Bayesian approach was shown to be more efficient than the system-by-system approach. Perhaps the most notable advantage of the Bayesian approach is its capability of better handling censored data.

Reference:

Qian, SS; Schulman, A; Koplos, J.; Kotros, A.; and Kellar, P. (2004) A hierarchical modeling approach for estimating national distribution of chemicals in public drinking water systems. *Environmental Science and Technology.* 38:1176-1182.

Table 1: Performance of the MLE method on synthetic data with σ^2 unknown.

n	$\Pr[\hat{C}_{95} > 2 \cdot C_{95}]$ (Overshoot)	$\Pr[\hat{C}_{95} > 3 \cdot C_{95}]$ (Gross Overshoot)	$E[\hat{C}_{95}]$ (Mean)	$Md[\hat{C}_{95}]$ (Median)	$\Pr[\hat{C}_{95} < C_{95}]$ (Should be 0.05)
5	0.856	0.777	201.852	47.027	0.0516
10	0.738	0.534	22.797	16.570	0.0505
20	0.519	0.197	11.775	10.572	0.0496
50	0.128	0.004	8.044	7.770	0.0507
120	0.002	0.000	6.728	6.642	0.0503
Exact	0.000	0.000	5.18	5.18	0.050

Note: $\hat{C}_{95}$ is the estimated upper bound of the 95-percentile, C_{95} is the true 95-percentile, Pr[] represents probability, E is the mean, and Md is the median.

Table 2: Comparison of the performance of proposed method and the simple alternative.

N	$\Pr[X \le \hat{C}_{96}]$	$\Pr[X \le \hat{C}_{97}]$	$\Pr[X \le \hat{C}_{98}]$	$\Pr[X \le \hat{C}_{99}]$	$\Pr[X \le \hat{C}_{99.5}]$	$\Pr[X \le \hat{C}_{95}]$
5	0.9077646	0.9197998	0.9322928	**0.9489302**	0.9602513	0.9904869
10	0.9355184	0.9469408	**0.9589702**	0.9730578	0.9817357	0.9891722
20	0.9480749	**0.9589307**	0.9702955	0.9825327	0.9894728	0.9848082
50	**0.9552819**	0.9656643	0.9762082	0.9871804	0.9929998	0.9767864
120	0.9581336	0.9682813	0.9784544	0.9888671	0.9942109	0.9694472

Note: $\Pr[X \le \hat{C}_q]$ represents the probability of a concentration less than the estimated q-percentile. The numbers in the table, thus, represent the true percentiles of the estimated $\hat{C}_q$.

File translated from TEX by TTH, version 3.61
On 20 Mar 2005, 23:44.