1-11

Vuke

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
REGION 6
1445 ROSS AVENUE, SUITE 1200
DALLAS, TEXAS 75202-2733

RECEIVED
OCT 10 1991
Dillon
Oklahoma Water Resources Board

October 8, 1991

L.E. GIGLIO

REPLY TO: 6W-P

L.E. GIGLIO

Mr. Dave Dillon
Chief, Water Quality Division
Oklahoma Water Resources Board
P.O. Box 150
Oklahoma City, Oklahoma 73101-0150

Re: Determining the Need for Water Quality -Based Permit Effluent Limitations

Dear Mr. ~~Dillon~~, Dave

The Region 6 Permits Branch has developed a procedure for effluent data analysis that we will use in FY92 to determine when a water quality based permit limitation is necessary. Our regulations call for the imposition of a permit limit if there is a "reasonable potential" to exceed a water quality standard. The limited effluent data obtained with the permit application may not represent a complete picture of the actual range of pollutant concentrations.

Assessing the potential to cause a water quality violation is one of many points which need to be covered in water quality standard implementation documents. To date, the only state permitting implementation to address "reasonable potential" is that developed by the Texas Water Commission. The Region 6 staff has worked up a sound and straightforward method that we will use in writing permits for the other states in the region, providing us with a workable alternative to the method described in the Technical Support Document for Toxics.

Our letter of January 3, 1991 described a statistical approach that would allow us to use a single piece of data or a small number of effluent measurements to estimate the upper range of concentrations that could be discharged and cause an exceedance of a standard. This procedure can be used to estimate the 95th percentile of an effluent data set, or the value that would be expected to exceed 95% of effluent concentrations in a discharge. The estimate of the 95th percentile is obtained by the following relationship:

pollutant concentration * 2.13 = 95th percentile pollutant concentration

11-1

pollutant concentration * 2.13 = 95th percentile pollutant concentration

The procedure is based upon the relationship of the geometric mean to the 95th percentile in a lognormal distribution, assumes a constant coefficient of variance and is independent of the number of data points considered.

In the derivation, they used the daily average instead.

A single measurement of pollutant concentration or the geometric mean of multiple measurements may be used to estimate the upper range value. The upper range estimate of the pollutant is then used to calculate the concentration of that toxic parameter after dilution in the receiving stream. For example, if a permittee reported an effluent measurement of 4.0 ug/l of cadmium, the upper range of cadmium expected for that discharge would be estimated as 8.6 ug/l. The permit writer would determine if a discharge of 8.6 ug/l of cadmium would cause an exceedance of the applicable water quality criteria.

Our permit writers will begin using the above procedure in writing FY92 permits to examine the potential of a discharge to cause an excursion above a water quality standard. For Texas permits, reasonable potential to violate a standard will be assessed in the manner described in the TWC implementation policy. A permit limit will be imposed on Texas dischargers if the effluent pollutant concentration is within 85% of the allowable value. The permittee will measure and report that parameter if within 70% of the limit.

All of our states should address the "reasonable potential" of a discharge to cause excursions above water quality standards in an implementation document or their Continuing Planning Process. They may reference the method Region 6 has developed, or adopt something of equivalent stringency.

Accommodating the uncertainty in effluent data will be protective and will likely result in a higher number of permits containing water quality based limits. We believe our approach will provide the permit writers with a consistent, clean and equitable technique of implementing water quality standards. Please let met know if you have any questions on this. If your staff has questions on the underlying statistics, they may speak with Jane Watson of my staff at (214) 655-7175.

Sincerely yours,

Jack Ferguson
Chief
Permits Branch (6W-P)

Attachment

Attachment

Region 6 Approach
Determining Reasonable Potential

Region 6 has developed a procedure to extrapolate limited datasets to better evaluate the potential for the higher effluent concentrations to exceed a State water quality standard. Our method yields an estimate of a selected upper percentile value. We believe that the most statistically valid estimate of an upper percentile value is a maximum likelihood estimator which is proportional to the population geometric mean. If one assumes the population of effluent concentrations to fit a lognormal distribution, this relationship is given by:

$$C_p = C_{mean} * \exp(Z_p * \sigma - 0.5 * \sigma^2)$$

where: Z_p = normal distribution factor at p^{th} percentile

$\sigma^2 = \ln(CV^2 + 1)$

To calculate the maximum likelihood estimator of the 95th percentile, the specific relationship becomes:

$$C_{95} = C_{mean} * \exp(1.645 * \sigma - 0.5 * \sigma^2)$$

if CV is assumed = 0.6,
$\sigma^2 = .307$

The ratio of the estimated 95th percentile value to the mean (C_{95}/C_{mean}) is calculated :

$$C_{95}/C_{mean} = 2.13$$

A single effluent value or the geometric mean of a group of values is multiplied by the ratio to yield the estimate of the 95th percentile value.

The following table shows the ratio of the upper percentile to the mean for the 90th, 95th, and 99th percentiles

Ratio of Upper Percentiles to Geometric Mean

Percentile	Z	C_{90}/C_{mean}
90	1.283	1.74
95	1.645	2.13
99	2.386	3.11

[illegible] information [illegible] which are indicative of [illegible] hazardous [illegible] shall [illegible] ground or surface water [illegible].

[illegible] depth. The waste loading [illegible] of the soil and vegetation [illegible] capacity or need [illegible] the waste [illegible] the assimilative capacity of the soil. [illegible] nitrogen [illegible] the amount of nitrogen [illegible] soil [illegible] shall be determined [illegible] uptake rate of the vegetation [illegible] to minimize the formation [illegible] concentrations which may leach [illegible]

[illegible] Application methods. The following [illegible] land application [illegible]

(A) Uniform application [illegible]

[illegible] land application [illegible]

(B) Application rate. [illegible]

[illegible]

[illegible] and [illegible] be maintained [illegible] technical [illegible]

[illegible] waste [illegible]

[illegible] shall be [illegible]

[illegible] shall be [illegible] and aesthetics [illegible] by the [illegible]

[illegible] the waste [illegible] exchangeable [illegible] should be maintained [illegible] the tolerable [illegible] capacity of the soil [illegible]

(E) The soil bulk density shall be maintained [illegible]

[illegible] the Board as [illegible]

[illegible]

`Date:   8/29/91`

Attachment

Region 6 Approach
Determining Reasonable Potential

Region 6 has developed a procedure to extrapolate limited datasets to better evaluate the potential for the higher effluent concentrations to exceed a State water quality standard. Our method yields an estimate of a selected upper percentile value. We believe that the most statistically valid estimate of an upper percentile value is a maximum likelihood estimator which is proportional to the population geometric mean. If one assumes the population of effluent concentrations to fit a lognormal distribution, this relationship is given by:

$$C_p = C_{mean} * \exp(Z_p * \sigma - 0.5 * \sigma^2)$$

where: Z_p = normal distribution factor at p^{th} percentile

$$\sigma^2 = \ln(CV^2 + 1)$$

To calculate the maximum likelihood estimator of the 95th percentile, the specific relationship becomes:

$$C_{95} = C_{mean} * \exp(1.645 * \sigma - 0.5 * \sigma^2)$$

if CV is assumed = 0.6, $\sigma^2 = .307$

The ratio of the estimated 95th percentile value to the mean (C_{95}/C_{mean}) is calculated :

$$C_{95}/C_{mean} = 2.13$$

A single effluent value or the geometric mean of a group of values is multiplied by the ratio to yield the estimate of the 95th percentile value.

The following table shows the ratio of the upper percentile to the mean for the 90th, 95th, and 99th percentiles

Ratio of Upper Percentiles to Geometric Mean

Percentile	Z	C_p/C_{mean}
90	1.283	1.74
95	1.645	2.13
99	2.386	3.11

Attachment 1

EXAMPLE
DETERMINING REASONABLE POTENTIAL
REGION 6 PROTOCOL

The outcome of this approach is illustrated in the following example:

Assume a discharger has reported 3 effluent concentrations of cadmium [9 ug/l, 12 ug/l, 15 ug/l]. The discharge flow is 3 mgd, the receiving stream critical flow is 6.4 mgd. The ambient chronic standard for cadmium is 6 ug/l as total metal. Assume 100% mix at the point of discharge and that the upstream concentration of cadmium is nondetectable. Evaluate the potential of the discharge to exceed water quality standards by assessing the impact of the 95th percentile effluent cadmium concentration.

1. Estimation of 95th percentile (Regional Approach)

The geometric mean effluent concentration of 12 ug/l is used as a parameter to estimate the 95th percentile value, assuming a lognormal distribution and a coefficient of variation of 0.6.

$C_{95} = C_{mean} * \exp(1.283 * \sigma - 0.5 * \sigma^2)$

$\sigma^2 = \ln(CV^2 + 1)$

$C_{95}/C_{mean} = 2.13$

12 ug/l *2.13 = 25.6 ug/l

The 95th percentile effluent value is used to calculate the Instream Waste Concentration:

2. Determination of Instream Waste Concentration

Cd = [(Qr * Ca) + (Qe*Ce)]/(Qr + Qe)

where
Cd= ambient concentration of cadmium after mix (Instream Waste Concentration)
Qr=river flow
Qe=effluent flow
Ca=upstream concentration of cadmium
Ce= maximum effluent concentration of cadmium

Cd = [(6.4 x 0) + (3 mgd*26 ug/l)]/(6.4 mgd + 3 mgd) = 8.2 ug/l

The IWC of 8.2 ug/l exceeds the ambient standard of 6.0 ug/l, a limit would be placed in the permit.

Use of other Upper Percentiles

The 90th percentile effluent value would be estimated as follows:

12 ug/l * 1.74 = 21 ug/l cadmium

The IWC would be calculated:
[(6.4 x 0) + (3 mgd * 21 ug/l)]/(6.4 mgd + 3 mgd) =6.6 ug/l cadmium

The 99th percentile effluent value would be estimated as follows:

12 ug/l * 3.11= 37 ug/l cadmium

The IWC would be calculated
[(6.4 x 0) + (3 mgd *37 ug/l)]/ (6.4 mgd + 3 mgd) =12 ug/l cadmium

As one selects more extreme tail values at which to evaluate potential water quality exceedances, the reported effluent concentrations must decrease to conclude that the potential to exceed the standard is not present.

Dealing with Highly Variable Datasets

The example above assumes that the coefficient of variation, defined as the ratio of the standard deviation to the mean, is 0.6.
If multiple effluent concentrations are reported which exhibit a large range between the highest and lowest values, the statistical variance of this population of numbers may well be greater than 0.6.

One can calculate the geometric mean of a group of numbers as follows:

1. Take the logarithm of each pollutant value.

2. Sum the logarithmically transformed values.

3. Divide the sum of transformed data by the number of measurements.

4. Express the geometric mean pollutant value by determining the antilog of the average of the logarithmically transformed values.

Dealing with Large Datasets

When a larger dataset of pollutant measurements is available, one may not need to statistically estimate the upper range or 95th percentile as described above. It is suggested that the 95th percentile be determined from the data and compared to the statistical estimation, the larger of these values should be assumed as the reasonably potential concentration of the discharge.

Attachment 2

Technical Support Document for Water Quality-Based Toxics Control
Determining Reasonable Potential

The procedure assumes that the concentrations of chemical parameters in wastewater fit a lognormal frequency distribution. Assuming some coefficient of variation (CV), the ratio between any two percentiles of the distribution may be calculated. An upper percentile of the distribution is selected, and a level of uncertainty in the confidence of estimation of the upper percentile is also selected. The procedure for estimating the extreme tail value consists of five steps.

1. The upper percentile for a sample is calculated given some level of confidence that the data set has captured the maximum discharge concentration.

$p_n = (1\text{-confidence level})^{1/n}$

where: p_n is the upper percentile for n samples
n is the number of samples

If one selects a confidence level of 99% and is evaluating a dataset consisting of three effluent analyses of a given substance,
the TSD states that one can predict that the maximum of the 3 values reported is greater than the 21.8th percentile of all potential samples from the same population with 99% confidence.

2. The normal distribution factor (Z) at the p^{th} percentile is determined from tabulated values of the areas of the normal curve.
For example, if the p_{th} percentile is determined to be the 21.8th percentile, Z is equal to -0.823.

3. The ratio of the concentration at the p_{th} percentile to the average concentration is calculated based on the CV and the relationship between these statistics in a lognormal population.
$C_p = C_{mean} * \exp(Z_p * \sigma - 0.5 * \sigma^2)$

where: Z_p = normal distribution factor at p^{th} percentile

$\sigma^2 = \ln(CV^2 + 1)$

4. The ratio of the 99th percentile concentration to the average concentration is calculated assuming some CV and a lognormal distribution.

$C_{99} = C_{mean} * \exp(2.326 * \sigma - 0.5 * \sigma^2)$

5. The ratio of the 99th percentile concentration to the p_{th} percentile concentration is calculated by dividing C_{99} by C_p. This ratio is the multiplication factor presented in the TSD by which the maximum data value reported is multiplied to calculate the estimate of the potential extreme effluent value.

The following table contains multipliers expressing the ratio of the 99th percentile to the p_{th} percentile for a range of sample sizes, selecting a confidence level of 99% about the estimate of the p_{th} percentile and assuming the population to be characterized by a coefficient of variation (CV) of 0.6. The TSD recommends the use of a CV of 0.6 where site specific data are not available to determine variance and estimation of the 99th percentile of effluent values to evaluate the potential to exceed water quality standards.

Multipliers: 99% Confidence Level and 99th Percentile

Number of Samples	Upper Percentile	Probability Factor	Ratio to Mean	Ratio 99th to Upper Percentile
1	.01	-2.326	.236	13.2
2	.100	-2.326	.421	7.4
3	.2154	-0.823	.543	5.6
4	.3162	-0.48	.657	4.7
5	.3981	-0.261	.742	4.2
6	.4641	-0.092	.815	3.8
7	.5179	0.045	.879	3.6
8	.5623	0.157	.935	3.3
9	.5994	0.253	.987	3.2
10	.6309	0.334	1.032	3.0
11	.6579	0.406	1.074	2.9
12	.6812	0.473	1.114	2.8

Column 2 is calculated from the equation $p_n = (1-0.99)^{1/n}$.
To illustrate, one may state that the maximum of 2 effluent concentrations reported is greater than 10% of all potential samples from the same population with 99% confidence. In using the TSD protocol, one assumes that in datasets of less than 7 samples, the maximum value reported will be less than the median effluent concentration.

Column 3 are tabulated values of Z corresponding to the p_{th} percentile.

Column 4 is calculated by the equation $C_p / Cmean = \exp(Z_p*\sigma-0.5*\sigma^2)$

Column 5 is calculated by the equation

$$C_{99}/C_p = \frac{\exp\,(2.326*\sigma-0.5*\sigma^2)}{\exp\,(Z_p*\sigma-0.5\sigma^2)}$$

EXAMPLE
DETERMINING REASONABLE POTENTIAL
TSD PROTOCOL

The scenario is the same as that described in Attachment 1. Assume a discharger has reported 3 effluent concentrations of cadmium [9 ug/l, 12 ug/l, 15 ug/l]. The discharge flow is 3 mgd, the receiving stream critical flow is 6.4 mgd. The ambient chronic standard for cadmium is 6 ug/l as total metal. Assume 100% mix at the point of discharge and that the upstream concentration of cadmium is nondetectable.

1. Federal regulations require that a permit limit be derived if a discharge causes or has the reasonable potential to cause an exceedance of water quality standards. One first would determine the ambient concentration of cadmium after mixing to assess if there is an exceedance of the water quality standard.

   Cd = [(Qr * Ca) + (Qe*Ce)]/(Qr + Qe)

   where
   Cd= ambient concentration of cadmium after mix (Instream Waste Concentration)
   Qr=river flow
   Qe=effluent flow
   Ca=upstream concentration of cadmium
   Ce= maximum effluent concentration of cadmium

   Cd = [(6.4 x 0) + (3 mgd*15 ug/l)]/(6.4 mgd + 3 mgd)

   Cd = 4.8 ug/l

The Instream Waste Concentration of 4.8 ug/l is less than the ambient standard of 6.0 ug/l, the discharge as described does not cause a violation of water quality standards. Does the reasonable potential exist to cause such an exceedance? Consistent with the TSD recommendations, assume that the coefficient of variation of cadmium in this effluent is 0.6, and evaluate the potential of the projected 99th percentile effluent value to exceed the standard. Attach a level of uncertainty about your estimate of the 99th percentile by constructing a 99% confidence level above the estimate.

2. Determine potential to exceed chronic standard.

   Select highest effluent concentration reported:
   15 ug/l cadmium
   multiply by 5.6 - ratio of C_{99}/C_p n=3
   15 * 5.6 = 84 ug/l cadmium potentially discharged

   Determine potential cadmium concentration after mix:
   (3 mgd * 84 ug/l) /(3 mgd + 6.4 mgd) =26 ug/l
   26 ug/l exceeds the chronic standard of 6.0 ug/l, a limit must therefore be placed in the permit.

3. A permit limit based on the chronic standard and the 90% percentile would be computed:

WLA= 18.7 ug/l
LTA = 18.7 * .77=14.4
Daily max = 14.4 *3.11= 45 ug/l
Daily ave.= 14.4*1.77= 21 ug/l

Comparison of TSD Approach and Regional Approach

In the Region 6 method described above, the mean value is used to estimate the 90th percentile effluent concentration. Considering the example presented, any single effluent value or the mean of multiple values greater than 11 ug/l would trigger an effluent limit, if the 99th percentile concentration is considered any single value or mean greater than 6.5ug/l would trigger an effluent limit. The decision to impose a permit limit becomes more likely as higher percentile values of the effluent concentration are evaluated.

Using the TSD protocol, the evaluation of reasonable potential is to some degree dependent on sample size. Using the same scenario, any single value greater than 1.3 ug/l would trigger a limit. If two effluent values were reported, any single value greater than 2.5 ug/l cadmium would trigger a permit limit. A daily maximum permit limit computed in accordance with the TSD would be 45 ug/l cadmium.

Effluent Concentration Which
Triggers Limit

	n=1	n=2	n=3
TSD Protocol (maximum of values reported)	1.4 ug/l	2.5 ug/l	3.3 ug/l
Region 6 Approach Assess IWC from 90th percentile value (single value or mean of n values)	11 ug/l	11 ug/l	11 ug/l
Assess IWC from 95th percentile value	9 ug/l	9 ug/l	9 ug/l
Assess IWC from 99th percentile value	6.5 ug/l	6.5 ug/l	6.5 ug/l

The Region 6 staff objects to the TSD reasonable potential protocol because it uses a highly biased statistical estimator of an extreme tail effluent value. In using the TSD method, one is at least 99% sure that the estimated value of the upper percentile is greater than the true population value of this statistic. There also seems to be some confusion in the relationship between the "confidence level" of the estimator and the percentile being estimated. The two numbers are unrelated. Indeed, if one is to proceed in this manner, it should be remembered that the confidence level indicates the confidence that one has in <u>over-estimating</u> the true percentile.

The Region has developed this alternative to the TSD approach because we have concluded that this method yields a less biased statistical estimate of the population upper percentile value. The upper percentile estimate is proportional to the mean, the mean may be estimated by calculating the arithmetic average of a group of values or assuming a single value to be equivalent to the mean. This introduces considerably less bias to the estimate of the upper percentile value than the TSD method of estimating the p_{th} percentile from the maximum of values reported. Since the more statistically valid estimate of an upper percentile value is proportional to the population mean, calculating a confidence interval for the population mean will yield a less biased confidence interval for the upper percentile.

Developed by: Jane Watson (6W-PM).

Copies of this document were sent to the following people by Jack Ferguson on 8/29/91:

Linda Korn-Levy, LA
Dick Quinn, AR
Jim Piatt, NM
Quang Pham, OK
Dave Dillon, OK
Ann McGinley, TX TWC
Windle Taylor, TX RRC

Reading File (6W-P)
Huffman (6W-PI)
Stenger (6W-PM)
Vickery (6W-PT)
Roger Hartung (6W-E)
James Steibing (6E-S)
Richard Hoppers (6W-Q)
Ferguson (6W-P)
Craig Weeks (6W-PI)
Huffman (6W-PI)
Mike Morton (6W-PT)
Jane Watson (6W-PM)