# Technical Memorandum

*To:* *Stormwater Quality Standards Task Force - SAWPA*

*From:* *CDM Smith*

*Date:* *October 13, 2013*

*Subject:* *Calculation of Anti-degradation Targets for REC2 Only Freshwaters*

The purpose of this memorandum is to present antidegradation analysis calculation methods for bacteria during dry weather flows in waterbodies proposed to be re-designated as REC2 in the Santa Ana Basin Plan Amendment. These waterbodies are Mill-Cucamonga Creek, Temescal Creek, Santa Ana Delhi Channel, and Greenville-Banning Channel. This memo supplements calculations provided in a previous technical memorandum (Calculation of Antidegradation Targets for REC2 Only Freshwaters - April 24, 2012) that contains additional background information and the raw data. Analysis via three methods was performed, each of which can be used to estimate an anti-degradation target for ambient water quality in streams. The calculations for Method 1 and Method 2 are based on guidelines provided in EPA's Technical Support Document for Water Quality Based Toxics Control (1991)[i]. Method 1 may be found in Appendix E of the Technical Support Document, and Method 2 within Box 3-2 in Section 3. Method 3 may be found in EPA's Data Quality Assessment: Statistical Methods for Practitioners (2006)[ii], in Section 3.2.1.5. Each method is described below.

***Method 1 - Percentiles of Log-Normal Distribution***

For waterbodies with larger data sets (n>20), it may be appropriate to fit the existing dataset to a log-normal distribution. Most pollutant data, especially bacteria concentration data, are commonly found to most closely fit a log-normal distribution. Fitting the data to a distribution characterizes the full range of potential bacteria concentrations. The standard deviation of the log-transformed data is used to estimate deviations from the mean for a target frequency of occurrence. Estimating a bacteria concentration for a targeted percentile ($C_p$) from a fitted distribution involves use of a z-score for a standard normal distribution (z=1.645 for 95$^{th}$ percentile; z=2.326 for 99$^{th}$ percentile). The value is then equal to the exponentiation of the log-mean ($\bar{y}$) plus the deviation from the log-mean ($z_p\sigma$), as follows:

$$C_p = e^{(\bar{y} + z_p\sigma)}$$

i EPA, 1991. Technical Support Document for Water Quality Based Toxics Control, EPA/505/2-90-001, Office of Water, March 1991.
ii EPA, 2006. Data Quality Assessment: Statistical Methods for Practitioners, EPA/240/B-06/003, Office of Environmental Information, February 2006.

***Method 2 - Reasonable Potential Multiplying Factor***

This method calculates a maximum expected single sample bacteria concentration. This is accomplished by multiplying the maximum value in a data set by a factor. The factor is based on the assumption that concentrations would be log-normally distributed, and therefore uses the log-mean and log standard deviation to estimate the reasonable potential multiplying factor. EPA developed matrices of factors for varying coefficients of variation and sample sizes, so that the estimated maximum concentration would equal the upper bound of the expected lognormal distribution at a target confidence level (Technical Support Document for Water Quality Based Toxics Control (1991) Table 3-1 and 3-2).

***Method 3 – Land's Method for Confidence Limits for a Mean***

This method involves setting an anti-degradation target using the geomean of subsets of samples from all data. Future monitoring data may have a different average than historical data simply as a result of the natural variability of bacteria concentrations and the small sample size in the initial years of a monitoring program. Using confidence limits around the mean estimates the variability of the mean of a dataset when subjected to changing sample size. One approach to estimating confidence limits for a mean would be to evaluate the distribution of means from numerous (~10,000) random subsamples of the historical dataset. According to the central limit theorem, the mean of future bacteria data would tend to approach the mean of historical data with increasing number of observations, if conditions remain the same.

Land's Method provides a parametric statistical method to estimate the confidence interval of the mean (*y*) of a dataset that is assumed to be log-normally distributed. The upper confidence limit (UCL) for the mean of a data set can be estimated as a function of the standard deviation ($\sigma$) and a factor ($H_{1-\alpha}$), as follows:

$$\bar{y}_{UCL} = e^{\left[\bar{y} + \left(\sigma^2/2\right) + \left(H_{1-\alpha}\sigma/\sqrt{n-1}\right)\right]}$$

Values for H for determining the one-sided confidence limit ($H_{1-\alpha}$) provided for $\alpha=0.05$ are provided in Table A-17 of EPA Data Quality Assessment: Statistical Methods for Practitioners. This table involves a matrix of both standard deviation and number of samples in the dataset, with the highest H values for the smallest sample size and largest standard deviation.

Table 1 presents the results of the analysis methods.

**Table 1 Analysis of Existing Water Quality Data***

| Parameter | Fecal Coliform | | | *E. coli* | | |
|---|---|---|---|---|---|---|
| | Mill-Cucamonga | Temescal [1] | Delhi [2] | Mill-Cucamonga | Temescal | Delhi |
| **Dry Weather[3]** | | | | | | |
| n | 229 | 12 | 503 | 168 | 108 | 56 |
| Geomean (cfu/100 mL) | 434 | 3,259 | 854 | 218 | 192 | 411 |
| 75rd Percentile of Data (cfu/100 mL) | 2,000 | 8,750 | 2,300 | 805 | 300 | 1,160 |
| 90th Percentile of Data (cfu/100 mL) | 5,410 | 10,200 | 11,902 | 2,960 | 832 | 2,430 |
| 95th Percentile of Data (cfu/100 mL) | 10,650 | 13,100 | 17,000 | 5,140 | 1,352 | 4,523 |
| Maximum value (cfu/100 mL) | 50,000 | 16,000 | 241,920 | 23,000 | 9,200 | 12,590 |
| **Anti-Degradation Objective** | | | | | | |
| Method 1 (cfu/100 mL) - 75th Percentile | 1,817 | 7,915 | 2,634 | 871 | 359 | 1,104 |
| Method 1 (cfu/100 mL) - 95th Percentile | 14,230 | 28,333 | 13,282 | 6,362 | 886 | 4,557 |
| Method 2 (cfu/100 mL) | 90,000 | 22,400 | 314,496 | 41,400 | 11,960 | 16,367 |
| Method 3 (cfu/100 mL) - 95% UCL for Geomean | 6,645 | 29,763 | 4,293 | 2,873 | 357 | 2,121 |

* Raw data contained within CDM Smith technical memo - Calculation of Anti-degradation Targets for REC2 Only Freshwaters (April 24, 2012)
1) Outlier samples collected from Temescal Creek on 9/8/2007 were removed (Max Fecal Coliform = 1,800,000; Max *E. coli* = 410,000)
2) All samples values include historical records when no flow data was available to determine hydrologic condition, these samples are not included in dry weather calculations
3) Dry weather is determined by daily flow <60 cfs at Mill-Cucamonga, <25 cfs at Temescal, and <5 cfs at Delhi

The following provides a brief explanation of each calculated value in the table:

- n – The total number of samples in the dataset.
- Geomean – The central tendency of the dataset, determined by multiplying the series of sample values together then taking the "nth" root of the product, where n is the number of samples in the dataset.
- 75th Percentile of Data (cfu/100 mL) - The sample value that is greater than or equal to 75% of all the sample values.
- 90th Percentile of Data (cfu/100 mL) - The sample value that is greater than or equal to 90% of all the sample values.
- 95th Percentile of Data (cfu/100 mL) - The sample value that is greater than or equal to 95% of all the sample values.
- Maximum value (cfu/100 mL) – The maximum, or single highest value in the dataset.

- Method 1 (cfu/100 mL) - 75th Percentile. This is the 75th percentile from a log-normal distribution fitted to historical data (Method 1 above).

- Method 1 (cfu/100 mL) - 95th Percentile. This is the 95th percentile from a log-normal distribution fitted to historical data (Method 1 above).

- Method 2 (cfu/100 mL) – This method takes the maximum historical concentration and uses variability in historical data to estimate how much higher a future maximum concentration could be without suggesting there is some cause beyond natural variability (Method 2 above).

- Method 3 (cfu/100 mL) - 95% UCL for Geomean –The value is determined based on Land's Method, which involves a statistical analysis of variability for the geomean of subsets of samples from all data.