G. Fred Lee & Associates

27298 E. El Macero Dr.
El Macero, California 95618-1005
Tel. (530) 753-9630 • Fax (530) 753-9956
e-mail: gfredlee@aol.com
web site: http://members.aol.com/gfredlee/gfl.htm

June 4, 1999

Supplemental Comments on Consolidated Hot Spot Cleanup Plan

Mary Jane Forster
State Water Resources Control Board
PO Box 100
Sacramento, CA 95801

Dear Mary Jane and Members of the Board:

I wish to follow up on the June 3rd BPTCP hearing devoted to review of the Draft FED Toxic Hot Spot Consolidated Plan to provide some additional comments based on comments made at the hearing and subsequent workshops. These are supplemental to the comments that I delivered to the State Board on June 3rd. The comments made by others at the hearing strongly reinforce the comments that I made in my June 2nd dated comments about the lack of technical validity in the toxic hot spot designation and ranking approach that was adopted by the State Board upon recommendations of the BPTCP staff in implementing the toxic hot spot legislation. I want to strongly recommend that the State Board members critically review the transcript of the workshop sessions that took place in the afternoon of June 3rd. These sessions brought out, in a resounding way, some of the significant technical problems that I have been discussing over the past nine years with this program.

The discussions regarding San Diego Regional Board using one level of statistical significance in toxicity testing vs. the San Francisco Board, causing a significant difference in the number of high priority toxic hot spots, points to the arbitrariness of the designation and ranking process as it is now being implemented. Up until now the problems with this program have been largely abstract and poorly understood by most of those involved. These problems are now becoming well documented, where California cities and industries are going to be spending millions to tens of millions of dollars in toxic hot spot remediation based on inappropriately developed designation and ranking criteria and assessments.

With many millions of dollars of public and private funds resting on a decision as to the appropriate statistical significance level used in a toxicity test, it is important to properly assess aquatic life toxicity and its meaning to beneficial uses of a waterbody. The current BPTCP approach is obviously not a valid basis for designating and ranking toxic hot spots. There should be a more reliable approach developed for assessing the need for remediation than is now being used. It is well documented that toxicity in sediments

depends to a considerable extent on the test organisms used and the characteristics of the reference sediments used. Dr. Anne Jones-Lee and I addressed this issue in,
Lee, G.F. and Jones-Lee, A., "Evaluation of the Water Quality Significance of the Chemical Constituents in Aquatic Sediments: Coupling Sediment Quality Evaluation Results to Significant Water Quality Impacts," In: WEFTEC '96, Surface Water Quality and Ecology I & II, Vol 4, pp 317-328, Proc. Water Environ. Fed. Annual Conference (1996).

As we discussed, the issue should not be whether a sediment is toxic to a single test organism, as is now being done, but whether this toxicity is significant to the beneficial uses of a waterbody of concern to the public who must ultimately pay for the cleanup. This type of review is totally missing from the current BPTCP toxic hot spot designation and ranking.

Further, for sediments that contain elevated concentrations of potentially bioaccumulatable chemicals, the issue, as is now being followed under current BPTCP guidance Policy, should not be whether there is elevated concentrations of some constituents in sediments that bioaccumulate under certain conditions from some type of sediments in certain types of organisms, but whether a specific sediment should be listed as a toxic hot spot because of elevated concentrations of chlordane, PCBs, DDT, etc., is the source of the constituents that are bioaccumulating to excessive levels in edible organisms. This type of review is totally missing under the current BPTCP guidance Policy and in its implementation by the Regional Boards. Instead, this Policy follows an obviously technically invalid approach of assuming that an elevated concentration of a constituent in sediments bioaccumulates to excessive levels in aquatic life, irrespective of how the sediment associated constituent of concern is bound in the sediments.

Readily available testing procedures have been available since the late 1970s from the Corps of Engineers and the US EPA for making these types of evaluations. These procedures evolved out of an over-$30 million research effort conducted by the Corps in the 1970s, which resoundingly demonstrated that the chemical approach for assessing the water quality significance of chemical constituents in aquatic sediments is unreliable and non-predictive. These issues have been discussed in detail in our paper,
Lee, G.F. and Jones, R.A., "Water Quality Aspects of Dredging and Dredged Sediment Disposal," In: Handbook of Dredging Engineering, McGraw Hill pp. 9-23 to 9-59, (1992). An updated version of this chapter which will appear in the second edition of this handbook is available from http:// members.aol.com/gfredlee/gfl.htm. (1998).

Further information on the importance of focusing on biological effects rather than on chemical concentrations in assessing biological effects of chemical constituents is provided in our paper,
Lee, G.F. and Jones-Lee, A., "Sediment Quality Criteria: Numeric Chemical- vs. Biological Effects-Based Approaches," Proc. Water Environment Federation National Conference, Surface Water Quality & Ecology, pp. 389-400, (1993).

These papers contain numerous references to the literature on these issues.

I strongly recommend that the Board pay particular attention to the discussions by Cynda Maxon of Arthur D. Little regarding the validity of the so-called ERM values in reliably predicting sediment toxicity or bioaccumulation as impacted by CSOs and stormwater runoff in waterbodies of concern (Mission Creek and Islais Creek) to the city of San Francisco. Her discussions of these issues were appropriate and in accord with what is well known today by those who understand the chemistry/toxicology of constituents in aquatic sediments. Rather than the ERM values themselves or quotients being predictive of aquatic life toxicity or bioaccumulation, these values have been shown with unbiased datasets to be less reliable than flipping a coin.

I have appended to these comments two recent discussions of the unreliability of the Long & Morgan (MacDonald) co-occurrence based sediment quality guidelines that were developed by Tom O'Connor of NOAA. Dr. O'Connor is associated with the NOAA National Status and Trends Program. He has been reporting on the unreliability of the Long & Morgan/MacDonald co-occurrence based sediment quality guidelines since the mid 1990s. These issues were discussed in,
Daskalakis, K.D. and O'Connor, T.P., "Inventory of Chemical Concentrations in Coastal and Estuarine Sediment," NOAA Technical Memorandum NOS ORCA 76, Silver Spring, MD, January (1994).

I have previously brought to the State Board's attention, through my comments, this work in my previously published papers. However, as I have pointed out, when a critical review of the references provided by the WRCB BPTCP staff used in support of their proposed Policy is conducted, it is found that none of the substantial literature that shows that the approaches advocated, and unfortunately adopted by the State Board, are referenced. Instead, only references are provided to those papers that support the staff's position. This is the kind of biased, unreliable, technically invalid approach that has lead to the problems that are now beginning to surface with the implementation of this Policy to designating and ranking toxic hot spots.

Dr. O'Connor's discussions of the unreliability of the co-occurrence based sediment quality guidelines are particularly pertinent to the appropriateness of using these guidelines for predicting sediment toxicity. As he points out, in unbiased datasets the predictive capability is poor and unreliable. This means that the fundamental basis of the current BPTCP Policy, which uses exceedances of Long & Morgan/MacDonald co-occurrence based guidelines as values for associating toxicity or bioaccumulation with concentrations of chemical constituents present in a sediment, is fundamentally flawed. As you know, I have been informing the Board of this issue for about nine years. You are now beginning to see the kinds of significant problems that are going to develop as this BPTCP Policy is implemented.

The Board should now be aware, based on the June 3rd workshop, that the State Board BPTCP staff and their associated "Department of Fish and Game" contractors are playing games with chemical concentration data to contrive systems for using co-occurrence based values as a basis for associating biological effects such as toxicity with concentrations of chemicals. Even with all of their contrived values of ERM quotients, these quotients still now have only slightly better predictive capability of toxicity than flipping a coin.

I repeatedly heard, during the June 3rd workshop discussions, the BPTCP staff/"Fish and Game" contractors talk about the importance of PCBs and chlordane in sediments, etc., and how they must be included in an assessment that is related to aquatic life toxicity. This, of course, is foolishness, since these constituents are of concern because they bioaccumulate, not because they are toxic at the concentrations present in sediments of concern. The discussions by Cynda Maxon about the importance of TOC affecting toxicity is the point that I have repeatedly made about how the bulk constituents of the sediments impact how the constituents of concern impact aquatic life or other beneficial uses. It has been well known since the early 1990s that TOC has to be used to normalize the availability of organic and, for that matter, some inorganic species, yet the BPTCP association program ignores this issue in designating and ranking toxic hot spots. This is not a new issue. It is well documented in the sediment quality literature, yet, based on this kind of what I term sloppy science, cities like San Francisco could be spending many millions of dollars likely unnecessarily to cleanup a transitory toxic hot spot that was toxic one time a measurement was made but is not toxic in subsequent measurements.

Another statement that was made by several of those at the June 3rd workshop who are attempting to defend the technically invalid approaches that have evolved out of the State Board staff's BPTCP Policy was that the recommended procedures for designating and ranking toxic hot spots were designed to be a screening procedure. Ed Long has repeatedly stated that the co-occurrence based values should not be used as regulatory tools. However, he asserts that they are reliable screening procedures. Tom O'Connor and others have substantial data to show that they are not reliable, even for screening. As discussed in the attached May 1999 discussion by Dr. Tom O'Connor, exceedance of ERM values is resulting in a situation such as is occurring in the BPTCP toxic hot spot designation and ranking, *"where authors gratuitously bestow biological significance upon their chemical data by invoking ERMs."* This is exactly what is happening in the implementation of the BPTCP.

This problem was reported by Dr. Jones-Lee and myself, based on the situation that had occurred in 1990 in San Diego Bay and in 1994 in the Santa Monica Bay Restoration Project, where large amounts of public funds were to be spent based on co-occurrence type assessments of the chemical constituents in aquatic sediments, in,

Lee, G.F. and Jones-Lee, A., "Equilibrium Partitioning-Based Values: Are They Reliable for Screening Contaminated Sediment?" Letter to the editor of Environ. Sci. & Technol., 27:994 (1993).

We pointed out that unreliable screening procedures cannot be used since once a sediment is classified as "polluted" or a "toxic hot spot" it is extremely difficult, if not impossible, to remove that designation. A review of this situation is provided in the transcript of the June 3rd workshop discussions, where it is clear that even though extremely limited data was used to designate two waterbodies of concern to the city of San Francisco as toxic hot spots, there was a reluctance to accept a more comprehensive, reliable dataset as a more appropriate assessment of the situation. Based on my 40 years of work on sediment quality issues, in many waterbodies throughout the US and other countries I know that some sediments at some times will be toxic to some organisms. The next time the measurements are made, however, they may be non-toxic or have a significantly different toxicity. As discussed in my previous comments, it is extremely

important that the State Board establish a more reliable approach for designating toxic hot spots and a readily implementable approach for removing sites that have been inappropriately designated.

A repeated theme at the June 3rd workshop in the discussions of the State Board BPTCP staff and their contractors was that "this is the best we could do with the information we have." That situation simply points to the mismanagement of this program by the BPTCP staff, which was allowed by previous Boards. These issues were pointed out in the early 1990s about how this program needed to focus on developing a proper assessment of the cause of aquatic life toxicity. Unfortunately, millions of dollars were spent in inadequately developed and implemented studies which did not obtain the data necessary to properly designate and rank toxic hot spots, as well as to identify the constituent(s) responsible for toxicity.

It is now totally inappropriate to force San Francisco, San Diego, or any other city, county, industry, or agricultural interest to become responsible parties for toxic hot spot cleanup and NPDES or WDR permit modification because the State Board's BPTCP staff adopted the co-occurrence based approach as a substitute for proper chemical studies in incorporating chemistry into the BPTCP Policy. Until this significant error is corrected, this and future State Boards will be spending significant resources in justified litigation where cities like San Francisco will have no alternative but to take this matter to the courts in order to obtain judicial relief from a technically invalid Policy that was adopted by the State Board for designating and ranking toxic hot spots.

You should ask yourself, after reading the transcript of the June 3rd afternoon workshops, whether, if you were a judge who heard this testimony, you would force the people in San Francisco or other areas to spend the large amounts of money that will be needed to remediate improperly designated toxic hot spots. I am sure that if you critically examine the situation you will come to the conclusion that the Board made a serious error last fall by accepting the staff recommendations on toxic hot spot designation, and especially the association with chemicals responsible for the designation. You can see the situation developing where the Regional Board staff will be testifying that they followed the State Board guidance for designating and ranking toxic hot spots. The testimony will be overwhelming that the designation and ranking approach set forth in the guidance has significant technical problems that were well understood at the time that the Board adopted the Policy.

If you have questions on these comments, please contact me. Again, as in the past, these comments are unsponsored and are being made in the name of trying to incorporate better quality science and engineering into water quality management policy within the state.

Sincerely yours,

G. Fred Lee

G. Fred Lee, PhD, DEE

Learned Discourses: Timely Scientific Opinions

Sediment Quality Guidelines Do Not Guide

*Thomas P. O*Connor*
NOAA National Status and Trends Program
<tom.oconnor@noaa.gov>

It would be very useful if Sediment Quality Guidelines (SQGs) could reliably convert data on chemical concentrations in sediment to information on biological hazard. However, they cannot. Continued use of SQGs in scientific journals to provide "perspective" for data on chemical contamination is misleading and conveys more credibility than SQGs deserve. That credence is leading to the unfortunate and inappropriate use of SQGs in sediment management. Exceedance of any particular SQG means very little with regard to sediment toxicity.

These conclusions are based on comparing chemical data with measured amphipod toxicity in 1,824 sediment samples collected in two monitoring programs: 1,373 from the EPA EMAP/Estuaries program in the Virginian, Carolinian, and Louisianian Provinces, and 451 samples from NOAA Bioeffects Surveys in Boston Harbor, Long Island Sound, Hudson Raritan Estuary, Tampa Bay, Los Angeles, and estuaries in South Carolina, Georgia, and the Florida Panhandle. Most of these results have been summarized by O*Connor et al. (1998). Recent additions are from EMAP/E in the Carolinian Province and the NOAA data for SC, GA, and FL. Samples were deemed toxic if less than 80% of amphipods survived 10-day exposures to whole sediment.

The SQGs tested were those based on equilibrium partitioning of neutral organics, those based on equilibrium partitioning of sulfide-insoluble metals, and empirically derived ERMs and ERLs. The neutral organic SQGs were for the five chemicals (endrin, dieldrin, phenanthrene, acenaphthylene, and fluoranthene) with SQGs published by EPA (1993). The same equilibrium theory has been applied to PAH compounds in general by Swartz et al. (1995) who predict sediments to be toxic when the sum of toxic units (calculatedpore-water concentration/10 day LC50) is greater than one. This calculation was made for the sum of toxic units of the 13 PAH compounds with LC50s provided by Swartz et al. (1995). DiToro et al. (1992) predict that sediments will not be toxic due to metals if the molar concentration of acid-volatile sulfide (AVS) extracted with 1 M HCI is greater than total molar concentration of simultaneously extracted sulfide-insoluble metals (Zn, Cu, Ni, Pb, Cd, Hg, and Ag). In this comparison we have 1,435 samples with AVS measurements and tested whether samples were not toxic when AVS was greater than Total Metals (i.e. the same metals extracted in concentrated HF). The concentration of total metals is greater than SEM, so whenever AVS exceeds Total Metals it also exceeds SEM. The empirically based concentrations are the 25 bulk-chemical concentrations listed by Long et al. (1995). Sediments are said to be likely to be toxic if one of the Effects-Range Mean (ERM) concentrations is exceeded and, conversely, if no Effects-Range Low (EEL) concentration is exceeded the sediments are unlikely to be toxic.

The SQGs that were supposed to pick nontoxic samples fared well but were basically untested because 1,602 samples were not toxic. Even a random selection would find nontoxic samples 88% (1,602/1,824) of the time. Of the 590 samples without an EEL exceedance, 563 were not toxic. Of the 962 samples where AVS exceeded Total Metals, 857 were not toxic. It is worth noting that the median Total Metal concentrations among samples with and without excess AVS were 1.73 and 1.72 :M/g, respectively, or essentially identical. The corresponding median AVS concentrations were very different, 6.49 and 0.43 :M/g, respectively. This guideline is very sensitive to AVS and leads to the conclusion that managing trace metal toxicity requires adjusting AVS but has little to do with metal contamination.

Within the 1,824 sample data set, 222 of the samples were toxic and the SQGs that were supposed to select toxic samples did poorly. Predictions of toxicity based on equilibrium partitioning of nonionic organic compounds were very few and usually wrong. In particular the five-chemical SQGs were exceeded only 13 times and 7 of those were not toxic. The similar SQG based on PAH toxic units was exceed only 8 times and 5 of those were not toxic. ERM exceedances were more common (287) and about as frequent as actual toxicity (222) but only 32% of the ERM exceedances coincided with toxicity. Most (68%) of the sediments withchemical concentrations high enough to exceed an ERM were not toxic.

There is a rough connection between sediment contamination and toxicity. A higherproportion of toxic samples were collected in the NOAA Bioeffect Surveys that emphasize urban areas than in the EMAP/E sampling that israndom over large areas. The ERM concentrations are in the high end of the spectrum of chemical concentrations in sediments and the SQGs based on non-ionic organics are very rarely exceeded because they occur only in extremely contaminated sediments. All these criteria are better than random selections in identifying toxic sediment but they are not reliable. They are all more often wrong than right and should not be used, by themselves, to imply anything about biological significance of chemical data.

References

DiToro DM, MaboneyJD, Hansen DJ, Scott KJ, Carlson AR, Ankley GT. 1992. Acid-volatile sulfide predicts the acute toxicity of cadmium and nickel in sediments. *Environ Sci Technol* 26:96-101.

Long ER, MacDonald DD, Smith SL, Calder FD. 1995. Incidence of adverse biological effects within ranges of chemical concentrations in marine and estuarine sediments. *Environ Manage* 19:81-97.

O*Connor TP, Daskalakis KD, Hyland JL, Paul JF, Summers JK. 1998. Comparisons of measured sediment toxicity with predictions based on chemical guidelines. *Environ Toxicol Chem* 17:468-471.

Swartz RC, Schults DW, Ozretich RJ, Lambertson JO, Cole FA, DeWitt TH, Redmond MS, Ferraro SP. 1995. PAH: A model to predict the toxicity of polynuclear aromatic hydrocarbon mixtures in field-collected samples. *Environ Sci Technol* 14:1977-1988.

U.S. Environmental Protection Agency. 1993. Technical basis for establishing sediment quality criteria for nonionic contaminants for the protection of benthic organisms by using equilibrium partitioning. EPA 822-R-93-011. U.S. EPA, Office of Science and Technology, Health and Ecological Criteria Division, Washington, DC.

Learned Discourses: Timely Scientific Opinions

Sediment Quality Guidelines Reply-to-Reply

Thomas R O*Connor

National Status and Trends Program, NOAA
<tom.oconnor@noaa.gov>

Long and MacDonald*s reply to my "Sediment Quality Guidelines Do Not Guide" is that they "...Do Guide Those Willing to See..." This should not be a matter of faith. If there is no empirical verification for a guideline, it should be abandoned. Since the Long and MacDonald reply dealt only with the ERM/ERL guideline (or the TEL/PEL equivalent), I will do so here. My original Discourse was based on an 1,824-sample dataset of which 222 samples were deemed toxic because ten-day survival of amphipods was less than 80% of control. Such a dataset cannot test the validity of a guideline that predicts non-toxicity (e.g., no ERL exceedance) because even a random choice would find a non-toxic sample with an 88% (1,602/1,824) frequency. The dataset could serve to verify whether an ERM exceedance predicts toxicity, but it turned out that only 92 (32%) of the 287 samples with an exceedance were actually toxic. I concluded nothing about ERLs, and that ERMs should not be used, by themselves, to imply anything about sediment toxicity. Long and MacDonald have presented no evidence to counter that conclusion. The dataset has grown. My version now has 2,475 samples with 17% toxic (they write of 2,600 samples with 18% toxic). Among the 453 samples with an ERM exceedance, 186 (41%) were toxic. Still more often wrong than right.

Long and MacDonald do not address this simple point. They do indicate that ERMs and PELs (they wrote ERLs and TELs but must have meant otherwise) "correctly classified samples as either significantly or highly toxic in about 90% of samples." In Table 1 of the cited paper (Long and MacDonald 1998) they list a 1,068-sample dataset in which 291 samples had an ERM exceedance. Only 39% were toxic in the usual sense (less than 80% amphipod survival relative to controls). Allowing samples to be deemed toxic if they only showed statistically less survival than control raised the toxic frequency among the ERM-exceedances to 52%. A small improvement for a price that 1) increased the total number of "toxic" samples in the 1,068-sample dataset from 230 to 426, and 2) put samples in the toxic category when their only sin was to be paired with particularly precise controls. The less-than-80% survival to establish toxicity stems from the Mearns et al. (1986) intercomparisons showing that samples with more survival did not yield reproducible results among laboratories. Nonetheless, the Long and MacDonald results based just on statistically-less-than-control survival do not validate ERMs.

Table 1 in the same paper also lists a 437-sample dataset where toxicity was determined by the conventional amphipod tests and by measures of sea urchin egg fertilization in pore water, mollusk embryo development in elutriates, or bacterial bioluminescence in organic solvent extracts. With a toxic response (less than 80% of control) in any one test sufficient to declare the sediment toxic, fully 78% of the 225 samples with an ERM exceedance were toxic. This became 86% if any statistically less than control response would do. What happens here, though, is that by those definitions either 70% or 80% of all the samples were toxic. Thus, creating the situation where almost all sediments are toxic and using ERMs to find them is hardly better than a random selection.

In Table 2 of the same paper, Long and MacDonald prioritize the predictability of ERMs by returning to the less than 80% amphipod survival as the definition of toxicity and assigning highest predictability to samples with more than 10 ERM exceedances, medium-high predictability to samples with 6 to 10 ERM exceedances, and medium-low predictability to samples with 1 to 5 exceedances. In the 2,475-sample dataset mentioned above there are 25, 46, and 453 samples in each category with corresponding frequencies of toxicity of 68%, 63%, and 40%, respectively. It is certainly true that limiting the prediction to cases with 6 or more ERM exceedances does increase the frequency of co-occurrence with actual toxicity. However, samples with 6 or more exceedances are so rare as to be self-evidently special cases. Furthermore, while no one claims that any Sediment Quality Guideline (SQG) will identify all toxic samples, a guideline requiring at least 6 ERM exceedances would have found only 46 of the 388 (12%) toxic samples in the 2,475-sample dataset.

My objection to the use of ERMs remains. While Long and MacDonald apparently no longer consider an ERM exceedance to mean very much, the scientific literature contains more and more papers where authors gratuitously bestow biological significance upon their chemical data by invoking ERMs. The papers would invariably stand without ERMs but the fact they are mentioned contributes to a growing and ill advised use of ERMs among environmental managers. By themselves, they are misleading not guiding.

References

Long, ER, MacDonald, DD. 1989. Recommended uses of empirically-derived, sediment quality guidelines for marine and estuarine ecosystems. *Human Ecol RiskAssess* 4:1019-1039.

Mearns AJ, Swartz RC, Cummins, JM, Dinnel, PA, Plesha, P. Chapman PM., 1986. Inter-laboratory comparison of asediment toxicity

test using the marine amphipod, *Rhepoxynius abronius. Mar Environ* Res 19:13-37.