Response to Peer Review Comments on Proposed Site-Specific Water-Effect Ratios for Copper and Recalculated Lead Objectives in the Los Angeles River Watershed and Revised TMDL for Metals in the Los Angeles River and Tributaries

Peer Review Commenters

1. Philip Bachand, Ph.D Tetra Tech R&D, Principal Environmental Engineer
2. Tom Young, Professor, Department of Civil and Environmental Engineering

Comment #	Comment	Response
1.1	In my review of the above WER report, I found the report comprehensive, clear and thorough in detailing and discussing the approach (e.g. calculation methods and assumptions, Biotoc Ligand Model – BLM, chemical analyses), the supporting sampling analyses and QAQC, the toxicity testing methods, the results, the WER analysis and the BLM analysis. The EPA guidelines are well referenced and assumptions used during this effort are well reasoned and scientifically defensible. Based upon this review, I believe the analysis is consistent with EPA guidelines.	Comment noted.
1.2	The analyses conducts a final analyses of the calculated WERs using historical comparisons, a discussion of WERs greater than five, and checking the protectiveness of the calculated fWERs against estimated No Observed Effects Concentrations (NOECs). The analyses are well considered, reasonable and defensible. The fWER based upon the Streamlined Procedure seems conservative and defensible.	Comment noted.
1.3	The spike recovery specifications detailed in Section 5.1, Table 11, do not seem very stringent. Is this specification typical of laboratories your office has used for these purposes?	The spike recoveries were within the range of labs available at the time the study Work Plan was developed in 2009.
1.4	Section 6.5 provides a comparison to historical data. The tables shown provide the necessary data to make the historical comparison. However, the presentation provided makes that comparison difficult without considerable effort. By default, the reader accepts the word of the author regarding the comparability of	The Interim Guidance suggests parameters collected during WER Study sampling events should be compared to historical concentrations of these same parameters, but does not indicate how such a comparison

	the data used to historical data. A better presentation using well designed figures and appropriate statistics would provide a stronger, more clear and more defensible assessment of the comparability between data used and historical data. Consistency between data used in the WER analyses and historical data is an assumption behind the calculations of the WER.	should be conducted. In this case, the approach chosen to perform this comparison was to present summary statistics. Additional information (confidence limits) was added in response to comments raised by the independent Technical Advisory Committee (TAC) on the initial draft WER report. The comparison of the data in a table format (Tables 25 through 34) allows for a detailed comparison as it provides the reader with detailed and comprehensive data.
1.5	Section 7.2 shows error bars in the figures with no definition of what those bars represent (e.g. 95% CI, SD, SE, other).	The error bars in the figures in Section 7.1 indicate the 95% confidence limits for the Least Square Means comparisons of the ANOVA model effects (*Event Type* and *Waterbody*). They are a visual complement to the p-values presented for the Effect Tests in the tables.
1.6	The BLM method for determining WERs seems less conservative with a higher predicted range and standard deviation than the WERs calculated from the Streamlined Procedure. Thus, the BLM WER should be used as an indicator but not in replacement of the WER calculated from the Streamlined Procedure.	Comment noted.
1.7	I have not conducted these calculations [lead recalculations] in the past using the referenced USEPA guidelines. I reviewed the above report and found the authors are very thorough in their documentation of their methods, their citing the USEPA guidelines, and their discussion of the datasets including modifications to the 2008 dataset. Based upon the detail and rigor in the document, I conclude they followed the recalculation methods for the lead criteria.	Comment noted.
1.8	The Acute-chronic-ratio (ACR) for Daphnia is not what I would expect as far as the dependence on hardness. The ratio is about 50 at hardness of 50 and 150, but only about 8 at a hardness of 100. What mechanisms would cause this trend? More consideration of the underlying science and assumptions are needed.	The data used for Daphnia in the recalculated criteria were used consistent with USEPA's criteria development guidance and these data were similarly used in USEPA's 1985 lead criteria document [the basis of the California

		Toxics Rule (CTR)] as well as the draft USEPA 2008 lead criteria document. The decreasing, then increasing trend of the ACR relative to hardness is a result of the magnitude of the acute value (LC50; which increases with increasing hardness) relative to the chronic value (EC20; which varies with increasing hardness) as the ACR is calculated as the quotient of the two.
1.9	The pooled slope method for calculating the dependence of toxicity on hardness for different species seems very rough (even if it is the approved EPA method). The hypothesis behind this method is that the slope is the same for all species and the method determines the slopes are not statistically different. However, the N values for each specie assessed is very low (4 to 8) and thus the resulting standard errors and confidence intervals relatively large. In developing the slope, data on that relationship is discarded for given species if the range of hardness is deemed insufficient. It seems more data is now available since this approach was developed. A more rigorous approach that is more inclusive of the larger data set could potentially be pursued by the EPA with regard to this guideline. Moreover, this report does not graph the raw data used for this analyses. Oftentimes, graphing raw data showing results in a better understanding of relationships and provide justification and validation for statistical approaches.	The approach for calculating the hardness slope was consistent with USEPA's criteria development guidance. As acknowledged by the reviewer, potential changes to the scientific approach would have to be initiated by the USEPA. The data used to recalculate the slope included all available data at the time of study initiation (N=24) and are an threefold increase over USEPA's 1985 lead criteria document (N=8), which is the basis for the CTR.
1.10	The approach used to determine downstream protectiveness is reasonable and rigorous. The approach uses Monte Carlo simulations to quantitatively show when TMDLs in upstream reaches and tributaries are met how downstream reaches are affected and their compliance with the TMDL. The analyses assesses both concentration and flow considerations, and looks at both wet and dry periods. The purpose of this approach was to assess the effects of upstream reaches with high WERs on the likelihood that downstream reaches would comply with TMDL requirements. The analyses concludes that if upstream TMDLs are met, even for	Comment noted.

	reaches or tributaries with high WERs, that downstream TMDLs will be met. This conclusion is defensible. I had a few comments on this analysis. First, it was not clear to me what the basis of the 90% confidence specification was. Is that a regulatory requirement or a statistical criteria? Some explanation and justification on that 90% number would be informative.	The 90% confidence level was selected in coordination between Regional Board staff, the Technical Advisory Committee, and the consultant conducting the analysis as an appropriate statistical threshold by which to evaluate confidence with attaining the threshold of zero exceedances within a three year period, which was defined as attaining the TMDL. The 90% level was selected to provide reasonable confidence and a consistent definition of attainment for upstream inputs and downstream outcomes of the Monte Carlo simulations.
1.11	Second, the report states that based upon the criteria of no exceedances in the upper reaches, the lower reaches that were studied would meet their TMDL with 99.2% confidence. The reasoning here is flawed. The upstream TMDLs are based upon a 90% confidence for the WERs that are calculated for those reaches. The mean and SD of the copper distributions in those reaches are adjusted such that the TMDL is met for a 90% confidence. Thus, the simulation shows that if the TMDL is imposed for a WER target, there are still going to be exceedances. In the downstream analyses, the confidence level is assessed with and without those exceedances. Without those upstream exceedances, the report states the downstream reaches will meet the TMDL with a 99.2% confidence (Section 4.3). This last jump in reasoning is problematic. You can only not have those exceedances if you impose a more strict standard and further reduce the mean and SD of the copper distribution. With the targeted WERs for those upstream stretches, there will be copper exceedances in those upstream stretches (90% confidence level is the TMDL design criteria) and they will cause exceedances in the downstream stretches as would be expected under the mass balance approach used. The good news is that the	The intent of parsing the downstream results this way was to provide insight into the upstream conditions that resulted in the exceedances and non-exceedances, not to re-define upstream attainment conditions. The re-presentation of the information was done to help contextualize the results based on discussions between Regional Board staff, the Technical Advisory Committee, and the consultant conducting the analysis. Using Tujunga Wash as the example, the relevant findings of this evaluation are that: • if both upstream reaches attain TMDL targets with at least 90% confidence of no exceedances, then the downstream reach will also comply with at least 90% confidence, i.e., it will meet the defined TMDL compliance target, as stated by the commenter.

	TMDL will still meet the 90% confidence level criteria in the downstream stretches that are being studied. But if the goal is to have or show a higher confidence level, more stringent upstream WERs are required.	• when both upstream reaches met TMDL targets, the downstream reach complied with ~99.8% confidence. • *if downstream exceedances were observed*, they were caused by upstream exceedances ~98% of the time (87 of the 89 downstream exceedances in 1,095,000 simulations). In addition, exceedances of the TDML target based on the WER are unlikely. In fact, the simulation shows that if the upstream reaches *never* exceed their TMDL targets, then downstream reaches will also attain their TMDL targets *with an extremely high level of confidence.* Indeed, the reviewer acknowledges that the TMDL will still meet the 90% confidence level criteria in the downstream stretches that are being studied.
1.12	I have reviewed Staff Report Section 5 and relevant sections in the report *Implementation of Results of the Los Angeles River Copper Water-Effect Ratio and Lead Recalculation Studies.* In the Staff Report Section 5, the staff recommends the following to track changes in water quality conditions that could affect copper toxicity and to ensure the protectiveness of the copper WER: • Monitoring water quality constituents in the river and its tributaries that affect the WER to ensure the water quality is consistent with water quality measured during the earlier studies that were used to determine the copper WER in the different reaches. • Assess that copper waste load allocations (WLAs) are achieving downstream water quality standards through various regulatory tools (e.g. POTWs NPDES permits, County MS4 permits).	Comment noted.

1.13	The Staff Report states the BLM model will use the water quality monitoring to estimate the WER. The approach planned appears to be the method identified in the report *Implementation of Results of the Los Angeles River Copper Water-Effect Ratio and Lead Recalculation Studies* provides more background on the suggested monitoring. The BLM model can be used as an indicator of when the WER may have changed. The BLM calculates the WER using easily measured water quality parameters. If the calculated BLM WER moves away from current calculated BLM WER values, than the fWER may be changing as well. In this way, the BLM model can be used to trigger test regarding if the WER in a reach or reaches should be changed using the methods used for this study. A 75% change in BLM WER from current BLM WER is suggested for that determination and that trigger value seems reasonable and defensible. Additionally, the Staff Report states than increasing trends in WLAs will be require the MS4 Permittee to identify and correct the causes. For NPDES permittees, levels of water quality treatment will meet the facility's capabilities. The monitoring and analyses suggested seem reasonable and defensible in ensuring the TMDL continues to protect against copper toxicity. Importantly, the monitoring checks that the water quality being measured is similar to that during 2008 and 2014 when the WER was developed validating the appropriateness of the calculated WERs, provide the BLM WER as an indicator of changing WERs, and guards against backsliding and degradation. If the WER is reassessed, expecting that one WER might only be needed (as suggested in the WER calculation report) seems overly optimistic.	The TMDL has been revised to clarify the proposed approach to use the BLM to track changes in water quality and trigger WER monitoring if changes are observed. However, the criteria for triggering WER monitoring proposed in the Implementation Report have not been approved by the Regional Board. Thus, the TMDL has been revised to require responsible parties to propose such criteria as part of their monitoring plans.
1.14	The Staff Report does not state the frequency of monitoring. For the MS4 and NPDES permittees, the frequency required for those permits would seem sufficient and acceptable. For the water quality sampling to assess changes in water quality that affect the WER, water quality sampling frequency similar to that used in developing the WER and conducted during the dry seasons would seem appropriate.	The frequency of monitoring for the MS4 and POTW permittees will be specified in POTWs' NPDES permit monitoring and reporting programs and the Los Angeles County and Long Beach MS4 Permits' monitoring and reporting programs or the Integrated Monitoring Programs and/or Coordinated

		Integrated Monitoring Programs, where approved by the Executive Officer of the Regional Board in lieu of the MS4 permits' monitoring and reporting programs, or other Regional Board required monitoring programs.
2.1	Overall, the document "Final Report: Copper Water-Effect Ratio Study to Support Implementation of the Los Angeles River and Tributaries Metals TMDL" describes procedures that are consistent with the letter and spirit of the EPA guidance embodied in both the Agency's "Interim Guidance" and its "Streamlined Procedures" documents. In cases in which professional judgment is required or is exercised in developing the WER values, the choices seem to be fully protective of designated beneficial uses within this watershed. The most significant deviations from published procedures are discussed below.	Comment noted.
2.2	WER values are experimentally determined only for dissolved copper rather than for both dissolved and total copper as indicated in the Interim Guidance document. This choice is based on a broad scientific understanding that dissolved metal concentrations are most closely related to observed aquatic toxicity and are thought to be the bioactive species. This point could be better supported in the Final Report by the inclusion of references to studies that support this conclusion. The choice to focus attention on dissolved metals is consistent with the approach taken in the WER analysis completed in 2008 and has been approved by technical advisors and the regional board. There is also no reason to believe that WERs developed for total copper would be systematically lower than those for dissolved metals.	References that support the conclusion that dissolved metal concentrations are most closely related to observed aquatic toxicity will be included in the administrative record for the proposed amendments, including the California Toxics Rule and EPA's 2007 aquatic life freshwater quality criteria for copper.
2.3	More samples were used to develop the fWERs than is required under the EPA guidance (6 instead of 3 in most cases). These samples were deployed in sensible ways to (re-)examine the critical conditions (dry vs. wet, winter vs. summer) in the watershed. The increased sample number enhances the representativeness of the WERs derived from the toxicity testing and supports more complete statistical analysis of the results.	Comment noted.

2.4	Decisions about rescheduling sampling events seem to have been made to enhance the representativeness of the resulting WERs. In particular, the decision to reschedule sampling in Arroyo Seco because of anomalous background water quality caused by the Station Fire clearly enhanced the general applicability of the resulting WERs. Although wildfires occur with some regularity within the watershed, they result in changes to a wide range of background water quality parameters. These changes collectively might have a large impact on resident aquatic species beyond changes in the relative toxicity of dissolved metals and should, if they are of concern, be the subject of a broader targeted investigation.	Comment noted. Ongoing monitoring of water quality parameters that could be affected by wildfires is required by the proposed TMDL revision in part to address this issue.
2.5	Laboratory waters used for dilution in toxicity tests had hardness values no greater than 220 mg/L as $CaCO_3$ following the EPA guidelines. Although site waters frequently had hardness values somewhat higher than 220 mg/L, during calculation of sWERs all EC50 values are hardness normalized so this should not have a significant impact, especially over the limited range of observed hardness deviations from 200 mg/L (the normalization value).	Comment noted.
2.6	No secondary species was employed during the toxicity testing. The single species approach is recommended under the Streamlined Procedure and seems appropriate.	Comment noted.
2.7	Following the addition of copper spikes to the site and laboratory water samples, a 3-hour equilibration period was employed. Although this equilibration period is both appropriate and reasonable, and was applied consistently to both site and laboratory waters, the purpose of the equilibration is described as being "intended to avoid exposure of the test organisms to the ionic form of the metal of interest" (p. 27). Since the free metal ion is, as discussed above, thought to be the primary bioavailable form, and toxicity was observed in the tests, exposure to the ionic form of the metal of interest presumably occurred. The phrasing here suggests a desire to "avoid observing toxicity" which might be misconstrued as an effort to elevate sWERs. Better phrasing would indicate that the equilibration period was observed to more closely approximate the metal distribution among freely dissolved, ligand-bound and adsorbed forms that would be observed in environmental samples	Staff agrees with this comment. The administrative record for the proposed amendments will reflect that the intent of the 3-hour equilibration period was to more closely approximate the metal distribution among freely dissolved, ligand-bound and adsorbed forms that would be observed in environmental samples with elevated copper concentrations.

	with elevated copper concentrations.	
2.8	The fWERs derived in this study are remarkably similar to those derived in the 2008 study (Table 45) and this should provide a substantial degree of confidence that these factors are authentically related to background water quality and expected differences in species sensitivities under these conditions.	Comment noted.
2.9	Problems observed with the DOC/TOC blanks and the resulting failure of a relatively large fraction of these samples to pass QA/QC checks are obviously of some concern. There is nothing in the methods employed (other than possibly the initial use of HDPE bottles for OC sample collection that was subsequently revised) that would explain these difficulties. These problems do not directly impact the resulting fWERs, however, and therefore do not affect the fundamental study conclusions.	Comment noted. Staff agrees that these data were not used in the calculation of the final WERs included in the proposed Basin Plan amendments. While the contamination did not jeopardize the calculation of the WERs and is not an uncommon occurrence, based on discussions with Regional Board staff, the study proponents, and the TAC, it was decided that the laboratory should be changed and the investigation documented in the WER Study Report.
2.10	The methodology for recalculation of the lead criteria included in the lead recalculation report, Section 3, is consistent with USEPA guidelines. The lead recalculation report is fundamentally different from the copper report since it does not involve independent or site specific experiments or data analysis and because it employs all toxicity data in an approved USEPA dataset. In the absence of experimental work or data selection, the only question is the appropriateness of the statistical analysis performed. All aspects of these procedures appear to be appropriately selected and performed but statistical analysis of toxicity datasets is not my specialty area.	Comment noted.
2.11	The approach to determine downstream protectiveness of upstream copper WERs included the Site Specific Objective Report, Section 4. And Appendix A, ensures that water quality objectives will be attained in all reaches.	Comment noted.

	To obtain the high degree of confidence sought by regulations that a very small (one or fewer) exceedance of water quality criteria will be observed over an extended time period requires a very large number of observations. Such a large number would be prohibitively expensive to obtain experimentally and would require at least the number of years specified in a compliance period (e.g., 3+) to complete. Therefore, a statistical approach, as employed in the Implementation Report (aka Site Specific Objectives Report), is essential. The approach taken here, combining predicted tributary metal loads that are minimally compliant with TMDL requirements with expected flows and concentrations in downstream reaches using mass balance principles, within a stochastic (Monte Carlo) framework, is an appropriate way to obtain the required information. These simulations support the idea that, although the WERs developed for upstream sites are higher than those derived for downstream sites, the resulting standards are protective throughout the portion of the watershed under examination. This analysis appears to have been carefully and thoroughly done and results in scientifically defensible conclusions.	
2.12	Background water quality conditions within the watershed (e.g., hardness, pH, DOC composition and concentration, copper concentrations and physical/chemical forms, suspended solids concentrations, etc.) can change significantly over time and can affect the speciation and toxicity of metals, resulting in changes to the appropriate WER value. The staff report proposes ongoing monitoring to address this issue, but the report does not provide specifics about how frequently the monitoring would be conducted or what would be done with the monitoring results. The Implementation Report, section 10, provides specifics in this regard. One of the proposed provisions, the requirement to conduct additional WER testing if three consecutive monitoring events (conducted at a rate of 3 events/year) result in BLM-derived WERs (calculated) that are less than 75% of those derived for that reach from the data collected during the copper WER study. If an average BLM-derived WER is below this threshold, then a single additional	Comment noted. The criteria for triggering WER monitoring proposed in the Implementation Report have not been approved by the Regional Board. As a result, the proposed monitoring to assess ongoing protectiveness of the WER is not specific. The TMDL has been revised to clarify that responsible parties shall propose the trigger value, the monitoring frequency, and the associated follow-up actions as part of their monitoring plans.

	WER measurement must be conducted. The rationale provided for the 75% trigger and the associated follow-up actions are not strong. The first reason provided is that a deviation in WER of this magnitude is "sufficiently large to be of concern relative to changes that might impact beneficial uses." However, there is no explanation of the particular "impacts" that might occur if this particular trigger were to be violated and no provision of supporting references or documentation. Why is 75% the right threshold for action rather than 50% or 90%? The second supporting statement for the 75% value is that it is appropriate because it is within the 95% confidence interval of BLM-derived WER values (at least that is what I understood this section to say, since it was not completely clear). The recommended follow-up to a calculated WER below this threshold is to measure a single WER value, presumably by comparing the experimental result to the fWERs developed as part of the current special study process. It seems difficult to determine whether the underlying water quality parameters and copper partitioning have fundamentally shifted from a single value. My concern is that this threshold may be triggered rather frequently, resulting in significant disruption and expense that does not advance water quality or species protections in the watershed. Perhaps more thought has gone into the monitoring provisions, but as presented they seem to have received substantially less scrutiny than other portions of this package. I recommend that the trigger value (as a percent of the BLM-derived WER or in some other form), the monitoring frequency, and the associated follow-up actions all be reconsidered or at least better justified.	
2.13	In general, the Staff Report adopts the recommendations of the Copper WER report, but there are two important exceptions. The first relates to changes in the wet-weather TMDL that I found difficult to follow (Implementation Report section 5.3, page 12). It appears that this involves the application of a single WER for wet-weather. The reasoning behind this recommendation is not clear. A second similar issue is that the Staff Report recommends the application of a single WER to the Burbank Western Channel rather	Due to the complicated application of the WER for wet weather described in the Implementation Report, the proposed changes are not recommended for the proposed Basin Plan amendments. The proposed application of two separate WERs for Burbank Western Channel is not

	than the two values contemplated in the Copper WER and Implementation Reports. This is explained in regulatory language that was not clear and that did not sufficiently explain the underlying reasons. These decisions may well be reasonable and well-supported, but the existing explanation provided does not make that clear to outside parties.	recommended because it would necessitate complicated changes to the TMDL that are outside the scope of the proposed action. Furthermore, applying a higher WER in the channel segment above the Burbank Water Reclamation Plant (BWRP) as proposed is not needed. The existing copper data (2003-2013) collected above the BWRP show that copper concentrations are lower than the adjusted numeric target calculated using the WER of 4.75. The median copper concentration of samples collected above the BWRP is 14 ug/L and the maximum is 95 ug/L, while the adjusted numeric target using the WER of 4.75 is 123 ug/L.