Peer Review of the Proposed TMDL for Toxic Pollutants in Dominguez Channel and Greater Los Angeles and Long Beach Harbor Waters

This scientific review of the Proposed TMDL for Toxic Pollutants in Dominguez Channel and Greater Los Angeles and Long Beach Harbor Waters is based on the information provided in the Draft Dominguez Channel and Greater Los Angeles and Long Beach Harbor Waters Toxic Pollutants Total Maximum Daily Loads document prepared and provided by the California Regional Water Quality Control Board, Los Angeles Region. It includes the main document and three appendices, as well as a Tentative Basin Plan Amendment to Incorporate the TMDL.

The nature of this TMDL, which considers several toxic pollutants of very different nature, strongly dominated by various non-point sources in multiple watersheds and a coastal waterbody with an open boundary, is quite complex. It requires a profound understanding of the environmental behavior and toxicity of every one of these toxic pollutants in many different media; a detailed quantification of the sources and current levels of contamination in the watersheds, marine water column and sediments; an accurate estimate of the daily loads transported through the watersheds to the receiving water body; and an accurate yet protective estimate of the reduction needed in those loads to achieve the objective of the TMDL, namely to restore the beneficial uses.

Given the nature of this TMDL and its complexity, a large amount of information is needed to make a scientifically sound determination of the total maximum daily load for each of these pollutants, and the subsequent allocation to the various point and non-point sources. However, it is clear from the documents provided that the information available is rather limited, and in some cases insufficient to make a scientifically valid estimate. The large data gaps, to be discussed in more detail below, result in significant uncertainty in the determination of the TMDLs. Although this is sometimes acknowledged in the documents, the assessment of the actual uncertainty is inadequate. The proposed margin of safety is unlikely to be sufficiently protective, and may result in continued non-attainment of the beneficial uses.

Another important issue is the assumption that these various toxic pollutants do not have any synergistic or antagonistic effects. The numerical targets have been determined based on the individual toxicity of each pollutant. However, it is quite likely that the organisms that will be exposed to these pollutants as a complex mixture will not be adequately protected by the individual numeric targets. The toxicological information simply does not exist to make an accurate determination of numeric targets that would take into account the temporally-varying nature of the complex mixture of pollutants. Therefore, instead of assuming that there is no potential synergistic effect, an additional margin of safety for the numeric targets would result in a more protective TMDL and should thus be considered. While precedent in other TMDLs may have led others to assume that there are no synergistic effects, a risk assessment of this nature should be conservative and thus assume that there are likely some synergistic toxic effects, when an aquatic organism is exposed throughout its entire life to several metals and a cocktail of toxic organic pollutants.

In addition to these major points, there are a number of major and minor concerns with regards to the development of the TMDL, regarding the use of sound scientific knowledge, methods and practices. The review of the document is complicated due to the fact that a number of important documents were not made available for the review, for example the documents that describe the implementation, calibration and use of the LSPC model for the Los Angeles River (LAR), San Gabriel River (LAR) and Dominguez Channel (DC). While those documents may have been reviewed for a different TMDL, the validity of their assumptions and the quality of the implementation could not be assessed. In addition, the Draft TMDL document is not always clearly written, and important information is omitted or not clearly presented. These issues are discussed below in detail under the headings for each issue outlined in the request for a peer review.

1. Appropriateness of selected sediment, fish tissue and water numeric targets for pesticides, PCBs, PAHs and metals.

Although the overarching question is whether the numeric targets were adequately selected, it is first important to determine whether the pollutants that are being targeted are the correct ones. Therefore, this peer review also provides some pertinent remarks with regards to Section 2.

In assessing the impairment a number of water quality, sediment and fish tissue observations were considered (Table 2-8). However, in most cases little or no information is given in the Draft TMDL for each dataset reviewed with regards to the number of observations considered, the number of exceedances, and a sense of the magnitude, frequency and duration of the exceedances. The best example is Table 2-16, but most other tables are lacking in this important information. The information on the magnitude, frequency and duration of the exceedances could be provided within the text, to put into context the magnitude of the impairment. It is important to know whether the objectives are always exceeded, frequently exceeded or only during very short periods; whether the short periods are frequent or only once in a decade; or whether the exceedance is 10% above the objective or 200%. For many of the datasets reviewed, little or no information is actually provided (e.g. 2.4.3.1, 2.4.3.3, 2.4.3.4, 2.4.3.5); they are essentially just mentioned but with no analysis. Given that the POLA & POLB 2006 sediment survey (2.4.3.5) is apparently of high quality, it would have been extremely useful to provide a detailed analysis. Same for the SCCWRP 2006 (2.4.3.6) study. This is a clear example of important information omitted from the Draft TMDL. It should at least be provided in an appendix, but a serious scientific report would have included a detailed analysis of this information within the main text. The level of credibility decreases when the information and its analysis are not provided. The summary provided in section 2.5 is inadequate, due to its lack of specificity.

Monitoring data for some individual PAHs is available (e.g. Table 2-12). However, criteria are based on either Total PAHs or benzo[a]pyrene (e.g. Table 2-3), which does not adequately reflect the toxicity or bioaccumulation of individual PAHs. The State of California explicitly considered in the 2006 303(d) listing for these waterbodies the individual PAHs as opposed to the general category of PAHs, yet this is not reflected in the assessment of NTs.

It is difficult to understand how a regulatory agency in California would allow an NPDES discharger to report the concentration of toxic pollutants using analytical methods that do not have adequate detection limits to assess whether in fact the discharge meets the objectives (Section 2.4.3.2). What is the point of allowing NPDES dischargers to report the required information based on a method of analysis that is useless for the intended use of the information?

With respect to the Numeric Targets, as stated in the document (page 43), they should be “guided by the Basin Plan and the California Toxics Rule (CTR)”, but in order to use the most current scientific knowledge one should look into more recent studies. The authors have taken the rather conservative approach of using the CTR for most of the numerical targets, without considering more recent information. This is particularly concerning in those cases where the CTR provides no information. For example, in Table 3-1 there are several numerical targets which are indicated as n/a or ‘-’ (which is confusing to the reader, since n/a is not defined and it is not clear what the difference is between n/a and ‘-’). There is no mention of an effort to review other studies or sources of information that may be used to establish Numeric Targets for these pollutants.

While it useful to list the CTR values for acute, chronic and organism only (human health), the document should be explicit as to which CTR value has been designated as the Numeric Target. One cannot use three different values for one pollutant in a given water matrix (freshwater or seawater). Therefore, Table 3-1 should be simplified, presenting only the specific Numeric Target for each pollutant in each type of water.

The list of Numeric Targets for water (Table 3-1) is incomplete, given the scope of the TMDL. As indicated above and in Table 2-18, there are several other waterbody-pollutant combinations that require a TMDL to be developed, which are not included here, such as several individual PAHs (e.g. pyrene, chrysene, etc.), dieldrin, toxaphene, Cd, and Cr. Thus, the Numeric Targets are incomplete. There is no explanation for the omission, and in fact the text at the beginning of page 43 indicates that the intent was to consider all of these compounds. Also, in Table 3-1 staff explicitly designates a Numeric Target for just 4,4’-DDT. However, it is likely that the transformation products of DDT, namely DDD and DDE, are also present in the sediments and water column, and may be of concern. Either one considers each explicitly, or as the sum of DDT compounds, which is generally considered to be DDT + DDD + DDE.

Staff also used a “translator” to adjust the Numeric Targets for three metals, to account for water hardness. While the text provides an indication of the rationale for selecting the conversion factors, the calculations are not provided within the document or the appendices. It is important to present this calculation somewhere within the document or appendices, so that the method can be reviewed. There is also no explanation for the selection of only the acute values for this calculation. A likely explanation is the short residence time of the metals in the water column in freshwater bodies within this region, but this assumption should be made explicit.

For Water Toxicity, staff defines the use of the Toxicity Unit Chronic. While this is adequate, there is no mention of the methods that will be used to determine toxicity. Specific testing protocols/assays should be defined, so that it is not ambiguous and subject to interpretation. In the case of Sediment Toxicity, staff clearly defines the organisms to be used and the specific criteria for interpreting the test results (Table 3-4, p. 48). A similar approach should be used for water toxicity.

The sediment concentration Numeric Targets are based on the sediment quality guidelines of Long and MacDonald (1995 and 2000). The use of the Effects Range Low and Threshold Effects Concentrations is scientifically valid, since as noted by staff, these are more applicable to the prevention of impairment, which is the objective of the TMDL. However, the application of these sediment numeric targets is inconsistent. For the toxic organics, Numeric Targets are set only for Marine Sediment. As will be established later in the Draft TMDL document, many or all of these toxic organics are still present in freshwater sediments which are transported through the various freshwater bodies to the harbor waters. Therefore, a Numeric Target should also be set for the freshwater sediment. In the absence of toxicity data (if indeed none is available), the default should be the marine sediment Numeric Target, such that the sediments delivered to the harbor do not enter at concentrations that will continue to impair these waters. For clarity, in Table 3-7 the labels “TECs” and “ERLs” should be removed. These are now Numeric Targets, and there should be no confusion with other terminology.

For the EFDC modeling effort (Appendix I), partitioning or distribution coefficients were determined for seawater in contact with the marine sediments, based on a comparison of observed concentrations in seawater and sediments. These provide a solid scientific basis for establishing the concentration of toxic pollutants in the seawater that will be in equilibrium with the concentration in marine sediments. This information should be used to determine whether in fact the marine sediment Numeric Targets are in concordance with the seawater Numeric Targets. If this is not the case, then achievement of one of the targets may not be feasible, since there will be continuous exchange between these two environmental compartments. The partitioning coefficients could also be used to develop seawater Numeric Targets for those PAHs and pesticides which were not listed in Table 3-1.

The use of Fish Contamination Goals (FCGs) for fish tissue Numeric Targets (Table 3-8) is scientifically valid, since the FCGs have been based on scientific knowledge. However, Table 3-8 lists the associated sediment “targets”, which in two cases are higher (less protective) than the Numeric Targets presented in Table 3-7 (e.g. chlordane and Total DDT), and in one instance is below (Total PCBs). This will lead to unnecessary confusion, since there shouldn’t be two (or more) targets for a pollutant in a given environmental compartment, in this case sediments. The Draft TMDL document should be clear as to which one of the values IS the Numeric Target (either the one in Table 3-7 or Table 3-8). In addition, since there are Numeric Targets for dieldrin, and the PAHs in Table 3-7, the notation “n/a” is terribly confusing. How can it be that there are Numeric Targets in one table for these pollutants, and not in the next table?

The Numeric Targets for tissue residues are based on scientific knowledge.

Overall, the document as currently written is confusing as to the specific Numeric Targets for water, sediments and tissues. Staff should separate the presentation of the underlying toxicity values (acute, chronic, human based on organism; sediment ERLs and associated sediment targets for fish tissue) from the final presentation of the Numeric Targets, which should be one value for a pollutant-matrix combination (i.e. pollutant-freshwater, pollutant-seawater, pollutant-freshwater sediments, pollutant-seawater sediments, pollutant-fish tissue, pollutant-tissue residues). In addition, partitioning coefficients (e.g. sediment-water, fish tissue-water, fish tissue-sediments) should be considered to ensure that there is consistency in the various combinations of Numeric Targets.

2. Appropriateness of the selection and calibration of the numerical models to estimate load capacity and load reductions

Two modeling frameworks were selected for modeling the fate and transport of the toxic pollutants. The LSPC model was used for the watershed and river/channel transport from the sources to the receiving waterbody. The EFDC model was used for the fate and transport within the harbor waters. The documents provided describe in detail the implementation of the LSPC model for the near shore watersheds and the implementation of the EFDC model. The implementation and calibration of the LSPC models for the LAR, SGR and DC watersheds was not provided. Therefore, this review can only provide an assessment of the scientific appropriateness of the implementation of the EFDC and LSPC model for near shore watersheds.

Source assessment

Source assessment is a very important component of the linkage analysis. Section 4.1 essentially lists or mentions the point sources with NPDES permits. However, after more than 10 pages of generic descriptions, no specific information is provided on the results of monitoring by these important sources. Information is provided about some of the difficulties in monitoring. For example, one learns that the Los Angeles County stormwater monitoring has been of no use to date since they are using analytical methods with insufficient sensitivity to detect the pollutants of concern. Thus, even though taxpayer or ratepayer resources are being used to monitor these waters, the information cannot be used at all. The omission of NPDES monitoring results reduces the credibility of this document.

As indicated in the report, these watersheds include some highly industrialized sections. In particular, the area around the ports includes several heavy industry facilities. There does not appear to be any consideration of the difference in types of industry in the source assessment. An acre of light manufacturing (e.g. clothing) is considered the same as an acre of heavy manufacturing (e.g. refinery). This assumption is not supported by any evidence that suggests that there is no difference in stormwater quality surrounding these different types of facilities.

The summary of results for the point sources (Table 4-2) is severely lacking in any detail that helps to determine the magnitude of the fluxes of the pollutants of concern. The last column reports on the "Potential for significant contribution", but there is no information in the entire document that supports this assessment. The lack of transparency is not scientifically adequate.

The assessment of direct atmospheric deposition is an interesting analysis, in that there is an attempt to link the emitters to the atmosphere to the watersheds where they operate, at least for three of the metals. However, airsheds and watersheds don't have the same boundaries. An emitter just outside a watershed may contribute significantly to the actual deposition in the watershed. While this may be captured in the more general deposition analysis, it may be better from a scientific perspective to determine what the potential radius of influence is for major emitters to the atmosphere that are in the vicinity of these watersheds. The estimated atmospheric deposition presented in Table 4-5 appears to be based on sound scientific knowledge and methods. The only issue is that it does not include PCBs and the pesticides that are also part of the TMDL, and is limited to three metals. Thus, the analysis is incomplete.

The assessment of the loads in the freshwater bodies is based on model output from LSPC. Given the concerns with the model calibration discussed below, there is low confidence in these estimates. The estimates have a large uncertainty associated with them, which is not evaluated anywhere in the report. A table in Section 4.3.1 should provide the estimates, and a more thorough analysis of the loads (temporal variations, contribution from different regions, estimate of the uncertainty, evaluation of assumptions) should be included within the main TMDL document. In fact, this level of analysis is not available anywhere within the documents provided.

The assessment of the amount of pollutant present in the marine sediments is based on EFDC model output. Again, based on the major concerns with model calibration discussed below, there is low confidence in these estimates. Use of EFDC model output introduces considerable uncertainty in the calculation, and this uncertainty has not been evaluated or taken into consideration. The estimates presented in Table 4-6 are given with an apparent high level of precision, in some cases 7 significant digits. In reality, these estimates can only be given with 1 or 2 significant digits; the data should be presented in scientific notation and only to the level of precision justified by the uncertainty in the estimate. Otherwise one is misrepresenting the precision. It is also unclear as to whether the estimated loadings presented in Table 4-6 represent the mean value from 2002 to 2005, or the final value at the end of the simulated period (2005). In any case, this information does not reflect the current concentrations in 2010 or 2011. Given the significant bias in model output, observed data would provide a better estimate of the pollutants present in the marine sediments. Since the model provides output as of 2005 (simulation is from 2002-5), it is more dated than the 2006 and 2007 studies that collected observed data. As indicated before, these estimates do not cover all the metals and toxic organics which have been identified, so the analysis is incomplete.

Section 4.4 (Sources Summary) indicates that the major sources of metals are stormwater and urban runoff. Since no information was provided previously about the contribution from NPDES dischargers, this statement is not supported by the evidence. The statements also are restricted to the Dominguez Channel freshwater, but in fact there are contributions from other major watersheds (LAR and SGR), which are not discussed in any meaningful detail. The summary also indicates that there are a number of activities that contribute pollutants to the harbor, and in particular discusses the "re-suspension of contaminated sediments from propeller wash". While this is a valid source, it was not discussed in the previous analysis, and there is no additional information provided here, so it is incorrect to bring up additional sources at this late stage with no justification. If the section intends to highlight those activities that were not considered, that should be made more explicit. One should then add that dredging is likely an important activity that was not considered in the assessment. It is also odd and confusing to be referred in this section to two tables in a later section of the report (Tables 6-9 and 6-11); that is poor scientific writing. Those tables present Waste Load Allocations, and do not thus pertain to the source assessment. Overall, the source assessment does not present sufficient information for a correct assessment of the sources, and relies too heavily on very uncertain modeling results, as discussed below.

LSPC model

Since the implementations of the LSPC model for the LAR, SGR and DC were not provided, they are not discussed in this peer review to any detail. The implementation of the LSPC model for the near shore watersheds followed generally standard procedures for setting up the model. The level of discretization into 67 subwatersheds is appropriate. The land use dataset used is somewhat dated (2000), but more importantly there is no distinction between different types of industrial activities. As indicated above, the emissions from heavy industry will be quite different than those for medium and light industry. If the same approach was used for the LAR, SGR and DC, that would introduce considerable uncertainty in all these models. There should also be a consideration of known hotspots that may be contributing more than the average load of a given pollutant. It is important to mention that the information provided in the TMDL document is very incomplete with regards to the implementation and calibration of the LSPC model for the near shore watersheds, so the comments below refer to the information provided in Appendix II. The lack of transparency in the TMDL document with regards to the relatively poor calibration of the model is not acceptable scientific practice.

The original LSPC (and underlying HSPF) model is capable of handling in a continuous simulation both dry and wet weather conditions. Since there can be significant accumulation of pollutants on the landscape of these watersheds, and the antecedent soil moisture conditions play an important role in the hydrologic response, the current approach where the wet weather is simulated separately from the dry weather deviates from the original model assumptions. No evidence was provided that this approach (separating dry and wet weather) is scientifically better in terms of the representation of the system.

A scientifically defensible approach for calibration is to consider objectives goodness of fit measures that can indicate whether the parameter values being adjusted are actually resulting in an improvement of the simulated results in comparison to observed data. Simple visual comparison is insufficient for determining whether the simulated response adequately reflects the true system. However, in all cases (hydrology, sediment transport, and pollutant transport) the approach used for the LSPC modeling was based on this inadequate visual comparison. In all cases, it appears that only one storm event within the 10-year period of simulation was actually used for the calibration of hydrology, sediment transport, and pollutant transport. This is a very limited basis for calibration. In addition, the "calibration" was done only at one location, and then it was "validated" at two other locations. While the text in Appendix II seeks to lead the reader to believe that the calibration results are "well within acceptable modeling ranges", the reality is that most of the simulated results are poor representations of the observed values, for hydrology, sediments and pollutants. More significantly, the worst match is for the location with the highest flow and loads, that is the one that is most significant. Thus, the credibility of the results presented in the TMDL report in Tables 5-1 and 5-2 is low for the near shore watersheds. If the same poor fits were obtained with the LAR, SGR and DC watershed models using LSPC, then the linkage analysis for this section is not scientifically acceptable. However, the TMDL report does not provide sufficient information to make this determination.

The linkage analysis for the freshwater loads also does not consider the entire list of pollutants. Thus, the analysis is incomplete. Since this information is the basis for the EFDC model, it introduces a significant amount of uncertainty in the harbor model, since the loads into the harbor are not adequately simulated. If a scientifically defensible approach had been used to estimate the uncertainty in the watershed loads, then at least one could make use of that information for the EFDC model.

EFDC model

The EFDC model as implemented considers a three-dimensional representation of the harbor waters and sediments. This is appropriate for this relatively complex system; a 1-D or 2-D representation would be inadequate. The use of several marine sediment layers is also an adequate and useful representation of the system. The grid and model configuration are appropriate.

The temporal simulation period considered is January 2002 to December 2005. The statement is made that "this period encompasses the greatest density of observational data for model calibration." However, the TMDL document indicates that the most extensive study of pollutant concentrations was the POLA/POLB study performed in 2006. Thus, the temporal simulation period is inappropriate. This important dataset should be considered for calibration and modeled explicitly (i.e. at least to the end of 2006). As discussed below, this dataset should NOT be used for setting initial conditions. This is an important error (i.e. not simulating to the end of 2006 to use this important dataset correctly) which reduces the credibility of the modeling effort.

The model considers the correct boundary conditions for freshwater and associated sediments, as well as exchange with the San Pedro Bay waters. However, it is unclear whether sediments can be transported in and out of the Harbor through the open boundary condition; omitting this exchange can introduce error and increases the uncertainty of the calculations.

While the contaminants of interest include six metals and at least a dozen toxic organics (see Tables 2-18 and 3-7), the actual modeling considers only three metals (Cu, Pb and Zn) and three organics (DDT, Total PAH and Total PCBs). Thus, the modeling is incomplete in this regard. Given the significant differences in fate, transport and toxicity among these pollutants, it is not scientifically appropriate to use the subset of pollutants modeled as representative of the larger set of pollutants that need to be addressed in the TMDL.

The partitioning of pollutants among seawater and marine sediment compartments is adequate for simulating the equilibrium distribution; however, it is not clear that under dynamic conditions the pollutants are truly at equilibrium. While this is a common and convenient assumption, it leads to some uncertainty in the calculations, which is not assessed or discussed in the document. The use of partitioning coefficients based on actual observed concentrations in seawater and sediments is a very good choice and reduces some of this uncertainty. However, the method for selecting values (visual best) is not scientifically appropriate. A statistical method should be used for this.

The EFDC model does not appear to take into consideration efflux of PAHs, PCBs and the other toxic organic compounds to the atmosphere. While these compounds have a low volatility, they can transfer from the marine environment to the atmosphere since they are very hydrophobic. A number of studies have quantified this efflux for different waterbodies around the world. Without an estimate by the modelers, this introduces another source of uncertainty into the EFDC modeling. A simpler model could have been used to perform a calculation to determine the relative magnitude of this flux, and decide whether it is significant enough to use a model that takes it into consideration. Along the same lines, there appears to be no consideration of the slow but continuous transformation via reaction of these toxic organics, which occurs mostly in the water column. Ignoring this transformation is a conservative assumption from a risk assessment perspective, but this is not explicitly stated in the report. Again, a more scientific approach would be to do an assessment of the magnitude of this process to establish how significant it is, and thus determine whether to include it or not in the model. The model as implemented does not appear to take into consideration these processes, which increases the uncertainty in model output.

The initial conditions for the pollutant concentrations in marine sediments were based on a substantial dataset. However, there is no explanation of the methodology used to consider data from different years. This lack of transparency in this important step reduces the credibility of the modeling effort. In addition, the modelers apparently used the dataset from 2006 to set the initial conditions in 2002. There is no discussion about how this was done. Scientifically this approach is not acceptable. The 2006 dataset should be used for calibration and validation, not to set the initial conditions for a simulation that starts in 2002.

The input functions for the load of sediments and associated pollutants (dissolved and adsorbed) are based mostly on LSPC simulation output. Given the significant issues associated with the calibration of the near shore watersheds LSPC model, there is a significant level of uncertainty in this important model input. The values from LSPC are considered deterministic. No apparent effort was made to consider the uncertainty associated with these inputs and how this may affect EFDC model output. This lack of rigor in the evaluation of this important aspect seriously reduces the credibility of the EFDC model as implemented.

Although there is a complete section in Appendix I that discusses “Model Performance Measures” in considerable detail, the document fails to present any quantitative assessment of the EFDC model performance with respect to scientifically acceptable measures of “goodness of fit”. Although clearly the modelers produced a lot model output, all the comparisons between simulation (“predicted”) and observations is visual. For the hydrologic calibration, there appears to be a noticeable difference in the tidal amplitude (e.g. Figures 5 and 6 in Appendix I), but without an objective measure it is difficult to determine whether this is an acceptable fit. The match between simulated and observed phase and amplitude of the tidal current velocities seems to be even lower (Figures 8-11 in Appendix I). Since the hydrologic calibration is key for model performance, this mismatch is likely to result in significant error in the simulation of sediment and pollutant transport. The authors of this appendix consider the match “reasonably good” (p. 12 in Appendix I), but that is strictly subjective and not based on a scientifically defensible performance measure.

A significant amount of effort appears to have been placed in calibrating the salinity. It is not clear that this is very relevant to the issues considered in the TMDL. Again, no model performance measures are reported. In any case, the simulation of salinity appears from a visual perspective to be quite good at the bottom, which should not be very surprising since these waters are at a fairly constant salinity and are not diluted significantly by the incoming freshwater. However, there is considerable scatter in the data for the surface seawater salinity (Figure 5-3 in TMDL report). This is further corroborated upon visual inspection of Figures A-1 to A-20 in Appendix I. As the authors indicate in p.23 of Appendix I, “point wise agreement is not always good”. Surprisingly, the TMDL report indicates that “the hydrodynamic model provides a good foundation for the simulation of sediment and contaminant transport”. Given the previous findings, the use of the word “good” seems unwarranted.

The next step in the calibration is adjusting the sediment transport parameters. While there is a discussion in the TMDL document of the approach that should be taken to perform this calibration, there is no presentation of results. There is also no analysis of the results. Pages 78 and 79 of the main TMDL report fail to provide any serious discussion of the results for the sediment or the pollutant concentrations. This lack of transparency is not acceptable. If the results are not good, this should be made clear. The reader is referred to Appendix I for the bad news. In page 60 of Appendix I, the modelers note that “model predicted concentrations are reasonable, however a quantitative measure of agreement would be extremely low”. While this is an honest assessment, it indicates that the EFDC is not adequately predicting sediment transport. Only three graphs (Figures 40-41) are presented within this Appendix, and the simulated results show significant variability. A major issue is that the simulated results are for a temporal period (2002-5) that does not correspond to most of the observed data (2006 and 2007). The omission of the presentation of more results, and of quantitative "Model Performance Measures" is not scientifically acceptable. The credibility of the output of this implementation of the EFDC model with regards to sediment transport is thus very low.

The final step in the calibration is the adjustment of parameters related to the various pollutants. Again, no results are presented in the main TMDL report, and there is no discussion of the results of the calibration. Although several figures are presented in Appendix I, no quantitative "Model Performance Measures" are presented. The authors remark that "the comparison show extensive scatter, but model predicted levels are within the range of observations". Clearly, the EFDC model as implemented does not adequately simulate the concentration of these pollutants. The comparison of the copper concentrations (Fig. 42) indicates that the model tends to over predict the concentrations in general. The over prediction is by a factor ranging from around 1.5x to 2x, based on visual inspection (since all we have is a graphic). The authors could have provided such analysis in their report, to be more quantitative in the comparison. The over prediction is more pronounced for lead (Fig. 43) and zinc (Fig. 44) concentrations, where the factors are 2x to at least 5x, if not more. This is a substantial difference, and is not truly "within the range of observations". The best correspondence appears to be for DDT concentrations (Fig. 45) although there are very few observations. For total PAHs, the over prediction is again around 3x to possibly 10x. The observations and simulation for PCBs indicate that these toxic organics are below detection levels (although the detection level considered to make this assessment is not reported). No temporal trends are presented for any of the toxic compounds modeled (metals or organics), so it is not possible to assess whether there is also a temporal bias (accumulating or depleting the reservoirs). The presentation of results is seriously lacking, with diminished scientific integrity. Overall, the calibration of the EFDC model is not adequate, since it has a clear bias towards over predicting concentrations of toxic pollutants in the harbor. While this may result in a more protective TMDL, a model should not have a bias.

Overall the implementation of the EFDC model for the harbor waters had several important deficiencies, and the calibration of the various components needed to predict the concentrations produce inadequate results. The outcome is that the simulated concentrations of toxic pollutants in the harbor are biased and may not reflect the actual concentrations. Thus, the linkage analysis is seriously deficient.

Section 5.3 (Summary of Linkage Analysis) makes no mention of the problems with calibrating the LSPC and EFDC models. Scientific integrity requires one to report and discuss the problems with the calibration, but this is not done. The summary introduces the presentation of pollutant load reduction scenarios; this should not be done in a summary, but rather in an earlier section. In any case, while Appendix III, Section 8 does present a "no upland loading scenario", there is no mention in the appendix of the "reduction of contaminated sediments in receiving waters to attain desired sediment target concentrations" scenario. Thus the summary is misleading, or the results of the scenario were omitted. Since this information is used for the determination of the Waste Load Allocations (WLAs), the omission is significant.

3. Appropriateness of the estimate of load capacity and load reductions

Toxicity TMDL in freshwater

There is no presentation of a load capacity for toxicity. The discussion is not very clear, but one can gather that the intent is to assume that the load capacity is 1 TUc, that is that each discharger must reduce the concentrations in their discharges to less than or equal to the chronic concentration of each pollutant. The interim allocation is ≤ 2 TUc, which apparently is currently being achieved, although the data presented in the TMDL report is insufficient to make this assessment. The final allocation is ≤ 1 TUc, which would be protective of freshwater organisms within the Dominguez Channel. Presumably similar determinations were made for the SGR and LAR. It is unclear why this section does not make it more explicit that this TMDL, WLA and LA will be applicable to all watersheds draining into the harbor waters, including LAR, SGR, DC and the near shore watershed, even if those actions have been or are being taken as part of separate TMDLs. Since no modeling was needed to arrive at this TMDL and the corresponding allocations, this TMDL is not affected by the issues discussed in the previous sections.

According to staff, an implicit margin of safety (MOS) is included in these TMDLs. There is no significant discussion of how this implicit MOS is determined. Although the NOEC were used, it would be useful to evaluate the methods used by the CTR to estimate the chronic criteria, to see whether an MOS is truly implicit in the determination of these criteria. In addition, as mentioned earlier, the assumption that the freshwater organisms can be exposed to a mixture of pollutants all at the chronic toxicity NOEC may not be warranted, and thus to be protective an explicit margin of safety should be included.

Wet weather metals TMDL in DC

The approach taken by staff is to consider the daily storm volume and the numeric target to calculate the maximum daily load acceptable in DC. The numeric target considered for the calculation is the acute criterion for each metal. However, as stated by staff earlier, “the Basin Plan narrative toxicity does not allow acute or chronic toxicity in any receiving waters”. Therefore, to meet the narrative toxicity and the Toxicity TMDL, the numeric target must be the chronic criterion, not the acute one. Otherwise, a discharge at the acute level would immediately violate the chronic criterion. Table 6-2 should consider the chronic numeric targets, not the acute criteria.

In addition, the daily storm volumes were estimated using LSPC. Given the issues with the calibration of this model, there is likely a significant amount of error in the estimate of daily volumes. Thus, the estimated allowable load has significant uncertainty. A 10% explicit MOS is insufficient for capturing the uncertainty in the LSPC estimates. Table 6-3 should be revised considering a higher MOS. Given that there is also considerable uncertainty in the estimate of the existing load, the percent reduction should be considered a rough estimate, rather than a very precise value. Certainly it is not known to three significant digits, as currently indicated in Table 6-3.

An additional concern is that since no exceedances have been observed during dry weather, then the decision by staff is that no TMDL is needed under these conditions. The rationale makes sense for freshwater organisms within DC, although it is possible that these waters can exceed the toxicity thresholds as the water volume decreases during dry weather. More importantly, the most severe problem is in the estuary and harbor waters. The cumulative load during dry and wet weather has an impact on the amount of metals present in the harbor. Thus, since the DC freshwater organisms are already protected by the Toxicity TMDL, the focus of the reductions should be the protection of the marine organisms, and the load capacity should reflect the maximum capacity of the receiving TMDL zones in the estuary and harbor. If the maximum capacity of the receiving waters is greater than allowed by the Toxicity TMDL, then the default should be the Toxicity TMDL for the freshwater loads.

The approach used for the WLA and LA calculations is scientifically sound, except that a 10% MOS is extremely small given the uncertainty in the load capacity estimates. To be clear, the explanation in Sections 6.2.2.2 and 6.2.2.3 should indicate that the allocation is done by area, as presented in Appendix III and Table 6-4. Good scientific writing practice is to refer to the section in an appendix or other supporting document where more details are presented, so that the reader can easily follow the calculations.

There is a significant difference between the "Allowable Loads" in Table 6-3 and the TMDL in Table 6-4. For example, for Cu the allowable load in Table 6-3 is only 234 kg/yr or 640 g/d. The TMDL in Table 6-4 is for 1,416.6 g/d of Cu (clearly the TMDL cannot be calculated to 5 digits of precision!). Even if one considers only the wet days, there is no explanation of how the calculation goes from the Allowable Annual Loads in Table 6-3 to the TMDLs in Table 6-4, and Appendix III does not provide any information. Since this is a crucial calculation for the TMDL, it should be more transparent.

The interim metal allocations are presented in Table 6-5. In the preceding text, staff indicates that these are calculated "based on the 95th percentile of total metals concentration from January 2006 to January 2010." Where was this information presented in the entire report (TMDL document and appendices)? In addition, these values are substantially above the interim toxicity allocations. A reconciliation of these interim allocations (toxicity vs. individual metals) is needed, to ensure they can be met.

Wet weather metals TMDL in Torrance Lateral

The approach taken by staff is different than for DC. In this case, the staff has not taken into consideration the LSPC model results. This may be a good decision. In this case, water and sediment "allocations" are based on concentrations. The approach is scientifically sound, with the exception that these are based on acute concentrations, so it again does not follow the Basin Plan: "the Basin Plan narrative toxicity does not allow acute or chronic toxicity in any receiving waters". Thus, the chronic toxicity values must be used to be protective. Rather than assume an implicit MOS, it would be scientifically more defensible to assume that an explicit MOS is needed if more than one of the metals is present at concentrations near the chronic criterion.

The Waste Load Allocations for the ExxonMobil refinery are based on a stormwater flow rate of 3.7 MGD for only 7 days/yr. While this flow rate may be reasonable, no data was presented to support the calculation. The Numeric Targets used are not indicated; if the acute targets were considered, this would not meet the Basin Plan.

Marine Sediment interim allocations

Interim sediment allocations for metals are based on observed concentrations. Staff considered the 95th percentile values of the observed values for this interim allocation. There is no specific justification for the use of the 95th percentile, as opposed to a lower level; it is likely set at a level that will not be easily exceeded. It would be better to have a justification for this choice, other than it being consistent with NPDES permitting, since this is not an NPDES permit. More importantly, the underlying data for this choice is not presented anywhere in the document, and there is no explanation of how data from different years was combined to produce a single value. It is possible that the 95th percentile values reflect samples from 1998, while the current condition may be much better, or it could be the inverse. In either case, the scientific basis is not transparent so that one can clearly understand the selection of the values in Table 6-8. For the PAHs, instead of using a value for Total PAHs, the interim and final allocation should be based on individual PAHs, as presented in Table 3-7. There is no mention of interim allocations for pesticides other than DDT, which indicates that this is not a complete set of allocations.

It should be noted that in some cases, using the 95th percentile value means that the Numeric Target is exceeded by almost two orders of magnitude, particularly in the LA Harbor Consolidated Slip which apparently is heavily polluted. Thus, higher priority must be given to these areas in terms of reducing their concentrations to the Numeric Targets.

Marine Sediment TMDL and final allocations

The TMDL, WLA and LA are presented in Table 6-10. The description of the methods in Section 6.4.3.1 (page 90) is quite vague, and thus hard to evaluate whether these critical calculations are scientifically sound. The short description of the approach in Appendix III (Section 1) is also rather limited. This lack of transparency is not appropriate for building credibility.

It should be mentioned somewhere in this section that the "Current Load" in Table 6-10 is calculated based on the sediment concentrations in the table in Appendix III that lists "Sediment Concentration Information per model zone (top 5 cm)", which was generated using EFDC. The current loads are presented in Table 4-6, but again the connection is not made clear in the document. Again, it is not clear if these predicted concentrations are at the end of the simulation (2005) or the average from 2002-2005. In any case, the current situation by 2010 may be quite different, so the observed values would have provided better estimates of the current load. Given the uncertainty associated with EFDC output, discussed above in the Linkage Analysis question, these sediment concentrations may not reflect the actual values. Note the significant difference between the values in Table 6-8 and the values in Appendix III. The depth of sediment considered for the Current Load is not clear – just the top 5 cm? There is also no mention of whether the load in the water column was considered or not.

For the TMDL calculation, the Numeric Target (ERL) was presumably multiplied by the mass of sediments up to the same depth. That is a scientifically sound approach, assuming that the mass of pollutant (dissolved and associated with suspended sediments) in the water column is very small relative to the mass in the sediments.

The air deposition estimates are explained in Appendix III Section 6. Those follow scientifically sound methods. It is important that the TMDL document make reference to the section in the appendices where such calculations are provided, so that the reader can easily follow them. One important issue with the air deposition estimates is that there is no estimate of the uncertainty or variability in these values. Since these calculations are based on a few data points in a relatively short timeframe, some allowance for uncertainty should be taken into account in an explicit MOS.

There is no explanation of how the Load Allocation for “Bed Sediments” was done. Are these based on the total sediment deposition rates presented in Appendix III, multiplied by the pollutant concentration calculated by EFDC? Or the pollutant concentration calculated by the corresponding LSPC models? Given this lack of information, the scientific validity of these estimates cannot be determined. In any case, the total sediment deposition rates in Appendix III have considerable uncertainty and may be in error, based on the relatively poor calibration results; they are certainly not known to 5, 6 or 7 significant digits as presented in the table in the appendix. There is also considerable uncertainty in either of the models with respect to pollutant concentrations, so again the estimated LA for these bed sediments has considerable uncertainty.

Waste Load Allocations are apparently determined based on the freshwater input estimated for each permittee and waterbody based on their area (a well known value) and the LSPC flow rates (a value with potentially significant uncertainty and bias as indicated by the calibration results of the near shore watersheds model). There is no mention of the pollutant concentrations used to estimate the WLAs. Given this lack of information, the scientific validity of these estimates cannot be determined.

Although the text mentions that “refineries which have provided discharge flow data along with monitoring results receive mass-based allocations”, Table 6-10 does NOT list any refinery explicitly. In fact, only the TIWRP is identified explicitly as a point source, other than the MS4 permittees (LA County, City of Long Beach and CalTrans). Throughout the TMDL document, information about these point sources (i.e. refineries and other major sources) is at best obscure. It is possible that these are indeed minor sources, but the lack of transparency is a major issue.

The use of concentration-based limits, applied as daily average limits, for minor or temporary sources (e.g. construction), is a scientifically sound approach. The problem is that the values in Table 6-9 don't correspond to the Numeric Targets in Table 3-1 for Cu, Pb, and Zn, and that the value for benzo[a]pyrene is being used for Total PAHs, when the impairment is by individual PAHs, not the total. This lumping of PAHs is not as protective, since PAHs have distinctly different toxicities and bioavailabilities.

Staff mentions that "an implicit margin of safety exists in the final allocations." Since the method for calculating the TMDL and allocations is not transparent, this statement cannot be evaluated. However, given the uncertainties, it is unlikely that an unquantified "implicit" MOS is protective. The assumption that the LA in bed sediments and air deposition is calculated with significant certainty does not seem warranted, given the issues with modeling. Even if the information is not based on modeling (i.e. observed sediment concentrations in a given volume), there is some uncertainty in the determination of the pollutant concentrations in these sediments, which should be reflected in an explicit MOS.

The other three metals that had not been considered in any of the modeling or previous calculations are finally considered in Table 6-11. If there is no effort to reduce their loading from the watershed, then a much longer time may be needed to achieve the Numeric Targets. It is unclear why these values do correspond to the Numeric Targets in Table 3-7, but those in Table 6-9 do not.

The proposal by staff to achieve the Direct Effects TMDL either by meeting the final sediment allocations or by demonstrating the desired qualitative condition via multiple lines of evidence is a scientifically sound approach, IF the final sediment allocations are truly protective of the aquatic organisms. As mentioned before, the lack of transparency in the calculations reduces their credibility, and the implicit MOS may not be protective enough.

TMDL for Bioaccumulatives

The term bioaccumulatives is used incorrectly in this TMDL document, since PAHs and some metals are also bioaccumulated and thus should be considered here. It would be best to either use the term toxic organics (and move the PAHs to this section) or just organochlorines.

As mentioned in the response to the first question, the use of numeric targets for different pollutant-media combinations requires a consideration of the partitioning coefficients, otherwise a numeric target could contradict another one. Thus, staff considered the ERLs in some cases and the BSAFs in other cases. The most protective value was used, which is scientifically sound. It would be best if this problem was resolved at the moment the numeric targets are set, so that it is clear what the target is.

Although there is a better description of the method used to determine the TMDL in this section (Equation 3), and the method is scientifically sound, the approach for allocating the loads to LA and WLA is not clear. The lack of transparency does not permit the evaluation of the method used to determine mass-based WLAs. The approach used for minor and temporal sources is scientifically sound. The implicit approach for determining the MOS is not scientifically sound; an explicit calculation of the uncertainty should be done to determine the MOS. A 10% MOS is unlikely to be protective. The selection of multiple numeric targets is not by itself a determination of an implicit MOS. The most conservative target must be used, but there are uncertainties in the calculation of the loads, so an additional MOS is needed. The concentration based WLAs for chlordane, dieldrin and toxaphene require a better assessment of the sources to be useful for the TMDL.

One issue with concentration-based load allocations is that it could lead to a total load greater than the TMDL under some circumstances. Therefore, monitoring of the actual loads will be needed to ensure that the TMDL is actually being met.

Critical condition

The critical condition would be a large wet weather event that produces extensive contaminated sediment transport through the channels as well as contaminated sediment redistribution in the estuary and harbor waters. Thus, although the report indicates that the "critical condition is not identified based upon flow or seasonality", there is clearly a seasonal nature to the critical condition, i.e. high precipitation events during the rainy season. The concern is that areas that achieve attainment of the beneficial uses may again become impaired due to such events. As such, the current analysis does not contemplate what to do in this case. A solution would be to implement a monitoring program after such events, to reassess the situation and determine whether the TMDLs and allocations are adequate.

4. Sufficiency of proposed monitoring program to assess effectiveness of the TDML and attainment of water quality standards

The proposed monitoring program is generally scientifically sound. The samples should be analyzed for all the pollutants listed in Table 2-18. The current text is unclear as to the metals to be considered. The text also does not indicate that any future samples MUST be analyzed using analytical techniques with detection limits low enough to indicate whether the Numeric Targets are being met. While this seems an obvious requirement for any QAPP, it is still distressing to have read that so many samples have been and are still analyzed with unsatisfactory analytical instruments. The proposed frequency is appropriate, except that as noted in the Critical Conditions section above, after an extreme wet season a round of sediment sampling should be conducted to assess the situation and make adjustments to the TMDL and allocations as needed. Since eliminating toxicity is the primary goal of this TMDL, toxicity testing should be required of all stations in Table 7-1, and should include both water and sediment toxicity. Hopefully, the reduction of the pollutants targeted by this TMDL will eventually eliminate toxicity, but such a monitoring program would ensure that toxicity does not continue due to new pollutants not targeted here.

5. Evaluation of the implementation plan and allocations

The narrative for the implementation plan is generally scientifically sound. The proposed phase approach, where some more immediate actions are taken along with a more detailed monitoring program, makes sense. Given the large uncertainties in the source terms and modeling results, in addition to these steps, a full revision of the TMDL and allocation calculations should be done before beginning Phase II.

It is surprising (in a bad way) that the Superfund sites present in this area, which are likely major contributors, are only mentioned at this late stage in the document. These potentially major sources should have been considered in the Linkage Analysis and the TMDL. How can such hot spots not be taken into account?

The timeline for the implementation (Table 7-2) is reasonable, although the deadlines for Tasks 12 and 13 have considerable uncertainty. Key will be to (1) have a much better monitoring program; and (2) have much better models that can help to make a better assessment.

Minor comments for Draft TMDL document:

Page 21: what is meant by “Some areas changes also occurred.”?

Page 32: The statement is made that “From 1994 to 2004, sampling frequency has decreased and now only occurs only in years when there is a discharge, such as 2005.” The first part of this statement refers to a particular period, yet the second part refers to a year outside this period. Did the sampling frequency return to normal after 2004? Apparently not. We are in 2010, so it would be useful to know what is happening today, not 6 or more years ago.

Page 50: The document states that “the chlordane, dieldrin, toxaphene, DDT and PCBs sediment targets presented in section 3.1.2 may need to be revised”. Section 3.1.2 refers to water numeric targets, not sediment.

Page 55: Lots of information is provided, for example the requirements of Storm Water Management Plans, but this is not relevant to the TMDL. The document should not be padded with such information.

Page 56: A map of all these permittees would be quite useful. A table indicating the monitoring data collected by each of the permittees is necessary, as well as an appendix with the actual data. Table 4-1 is too general. When did they start monitoring? What parameters? Are the results indicating that these are important sources? For example, at the end of the page it is mentioned that the City of Long Beach received a permit since 1999, but no monitoring results are reported.

Page 64: The correct units are $\mu g/m^2$-day and ng/m^2-day, not $\mu g/m^2/day$ or $ng/m^2/day$. Also, the acronyms of the water bodies should be provided in a footnote.

Page 65: The heading of Section 4.3.2 is incorrect. This is not an analysis of the existing “sediment”, but rather of the pollutants within the sediment.

Page 82: The equation for TUc was already introduced in page 45. It is not good practice to be repetitive within a report. However, in this case there is an example provided, in which the authors state that “if the NOEC is estimated to 25% using hypothesis testing”. What does 25% refer to? Percent of what? Presumably 25% of the NOEC, but this is unclear. The definition actually should be revised. It should be the sample concentration divided by the NOEC. Thus, a sample concentration which is twice the NOEC would have a TUc of 2.

Page 94: The paragraph in this page does NOT correspond to the Margin of Safety discussion. A separate heading is needed.

Page 114: The information that separate TMDLs are being implemented for LAR and SGR should be mentioned earlier in the document, and in an earlier section numerical information should be provided to be able to determine how the joint actions of the various TMDLs will eventually result in achievement of the beneficial uses.

Comments on Appendix I (EFDC model)

Discussed above.

Comments on Appendix II (LSPC watershed model)

This appendix intends to present the methodology utilized to setup the LSPC watershed model, for calibrating and validating the model, and its subsequent use for developing the loads associated with the various sources in the Los Angeles and Long Beach Harbors. While the appendix does provide many important elements of the model setup, there are some important gaps in the information provided. More significantly, the calibration of the LSPC model for the near-shore watersheds is not scientifically supportable. The analysis relies on previous implementations of the LSPC model of the LAR and SGR for the load calculations; insufficient information is provided in this report to determine whether those calibrations were adequate, but if the same approach was undertaken, the scientific validity would be questionable. Although the authors attempt to "validate" the model, the results of the validation are not adequate, particularly for TSS. Since the transport of the metals and toxic organic compounds studied here depends considerably on the flow and TSS calibration, those results are questionable as well. Quite frankly, the answer to the question of whether this study is scientifically adequate is no.

Additional comments:
Page 1: reference to a modeling approach for metals is given, but the citation is SCCWRP "unpublished results". Since this reference is not available, it is not useful at all. Need to provide date and number of any report cited in the document.
Page 2: the authors indicate that they use two different approaches for wet and dry weather loads. They justify this indicating that other TMDLs in the LA region have been done that way (without indicating which ones, so the statement is not backed by a proper citation). It is unusual that they have chosen to use an approach that appears not to be able to handle a continuous simulation of dry and wet weather. One could incorporate source functions for the dry weather flows into LSPC to account for them, so it is unclear why the authors have chosen this more complicated and less scientifically defensible approach. Antecedent conditions can be important for the simulation of hydrology, TSS and pollutant transport, and the current approach seems to *a priori* discount their influence.
Page 3: The authors correctly point out that although non-point sources are distributed throughout these watersheds, there are likely some hot spots. For example, although PCBs can come from several land uses, there are likely some electrical transformer locations which are hot spots. However, this observation is then completely disregarded. There is no effort to identify hot spots since "their presence and impact to receiving waters are difficult to identify/characterize." Since these may be the most significant sources for specific pollutants, a proper study would have made the effort to identify them and consider them in the model. Management actions will have to be specific for

these hot spots, so ignoring them (or averaging them into the land use coefficients) is not useful.

Page 4: It is mentioned that the LSPC model "has been successfully applied and calibrated" for the LAR and SGR. There is a need to objectively define "successfully". What was the goodness of fit measure used to determine success? Was there any other statistical approach used to evaluate this? Since this is also a major problem with the current analysis, one is left to wonder how "successful" the development of the LAR and SGR models was.

Page 4: The criterion used to discriminate between wet and dry weather was the 50th percentile observed flow. Was there observed flow for every day at all monitoring locations, to be able to make this determination? Or was this based on flow at a particular location? Was it consistent across all the watersheds, or were wet days specific to a watershed (LAR, SGR, etc.)? How are the antecedent conditions handled for this discontinuous approach?

Page 5: Define CALWTR and provide a reference.

Page 5-6: The drainage of Machado Lake was not considered in the analysis, even though the authors indicate that it may be connected to the Harbors during extremely large and rare meteorological events. While these events may be rare, they are large, and could represent a significant fraction of the cumulative load, since they tend to wash the landscape more intensely. The exclusion does not seem justified without additional analysis to show how rare they are (some measure of frequency) and whether they can be a relevant fraction of the cumulative load to the Harbors.

Page 6: The gaps in the rainfall data were patched. While this was done only for less than 5 percent of the records, there is no information on how significant these patches were. Since rainfall is sparse in Southern California, 5% may be a significant number of rain days. There needs to be a table indicating all of the meteorological stations, the number of records per station, the number (or frequency) of missing records, and an indication of the station used to fill in the gaps. This is particularly important since the authors considered hourly precipitation, and the approach indicate in the last paragraph of page 6 tends to reduce the validity of using "hourly data", if it is going to be spread out throughout the day.

Page 7: While it was adequate to discretize the land uses as indicated, the parameter values associated with each land use were assumed deterministic, with a single value for a given land use. For the sensitivity analysis, it would be important to allow the most important parameters vary to have a better idea of the real sensitivity, and then to be able to determine the uncertainty in the load estimates. The current approach for the sensitivity analysis is overly simplistic.

Page 11: Most point sources do not have a constant outflow, and their concentrations are also quite variable. In particular, there is no reason to believe they control the metals or toxic organics in their discharge, so these values are likely to vary considerably from day to day. Table 2 does not indicate which dischargers have limited data. The authors indicate that the average flows are in the model database, but that information should also be provided in this appendix. If the majority of the NPDES dischargers are being treated as constant flow and loads, then this is likely to be an incorrect representation of the point source loads.

Page 14: The authors indicate that "after comparing the results, key hydrologic parameters were adjusted". Using what goodness of fit measure? Nash-Sutcliffe? It appears from the later sections that it was all done visually, which is an unscientific approach. Even if the goodness of fit is not good, it is important to know how bad it is, not just whether it "looks" good or bad.
Page 15: The statement is made that "During low flow conditions, the model is unable to predict dry urban runoff". If the authors had considered adequate source functions for the various landuses, this could be modeled using LSPC. Their ad-hoc approach is not as defensible.
Page 15, Figure 4: The model clearly over-predicts flow at all times during this event, perhaps by over 25-30%. The authors should also look at the cumulative flow. They would then see that the simulated pulse is much bigger than the measured pulse. This would have significant implications for TSS and toxics transport, and also affects the simulated concentrations, if more water is available for diluting the load. The authors indicate that "this small discrepancy in flow is well within acceptable modeling ranges." Based on what? This statement is very misleading. In reality, this error is significant and most modelers would continue calibrating to reduce the bias (over-prediction). Since the authors are only using a visual comparison, they feel they have done an acceptable job, but in reality this is a poor fit.
Page 15: It appears that only one storm event was used for the calibration, out of the 3 years of simulation. There is no basis to think that this one storm event is representative of typical events. If no additional storm event data was available, this should be stated clearly. One obvious solution would have been to collect a few more events. In addition, it is surprising to see that the Forest Subwatershed which has the lowest flow is used for "calibration". It is the least representative of the three locations. Thus, the baseline for the calibration was poorly chosen.
Page 16: Most modelers would use data from the same location at a different time to do a proper validation. The authors have chosen to use two different sites for their validation. However, the underlying parameter values are different, given the different land uses, so this approach has much lower scientific validity.
Page 16, Figure 5: The match is poor even by visual standards. The authors indicate that "the initial peak was low; however the second peak was fairly close." Again, only a visual comparison. The authors fail to state that they miss the size of the first peak by around 75%, and that their overall pulse is much broader so that they are simulating a much larger total flow than was observed, by a significant factor. Thus, stating that it was "fairly close" is rather inaccurate. An analysis of the cumulative flow would have shown that flow at this location is also seriously over predicted. This site has about 8 to 10 times more flow than the Forest Subwatershed, so the over predicting is quite significant.
Page 16, Figure 6: At least the authors acknowledge that "the validation results did not match the measured flow". In this case, the model seriously under predicts flow, both the peaks and the cumulative flow. This is the most important subwatershed in terms of flow, and it is the worst in terms of model output. Clearly the choice of subwatershed for calibration was a poor one. The authors also mention that they did not adjust the LAR watershed parameters "outside of recommended ranges." Who recommended the range of parameter values? Is there a basis for these ranges? Are the studies or literature values to refer to?

Page 17: The authors mention a “robust calibration and validation process” for Ballona, LAR and SGR. What is the basis for saying it is “robust”? Is there a more objective quantification of the quality of the fit? If the current implementation of the LSPC is an example, then one has to wonder what the authors consider as “robust”. Clearly, the parameter values were not just transferable, but the authors go ahead and assume this is OK, even after a poor outcome in the calibration and validation process.
Page 17: The paragraph that starts with “Similar to …” should be the first paragraph in section 3.2.2.1.
Page 18, Table 4: This is a very good table. A similar table should have been presented in the hydrologic calibration section, with all the hydrologic parameters, showing the adjusted values and the ones from the previous (LAR, SGR) models.
Page 19, Figure 7: The model clearly over predicts the TSS pulse even in this small subwatershed. The authors indicate that “these discrepancies are well within acceptable modeling ranges.” This indicates that either the authors (1) have no significant previous modeling experience; or (2) have no significant scientific integrity. Either way, it is not good. The match is poor, and if this is the best they can obtain, then the resulting load calculations, which rely to a great extent on TSS concentrations are going to be incorrect. If they consider the difference in cumulative TSS load in this pulse between the simulation and the observed values, they will realize that they are simulating a pulse that is probably an order of magnitude greater. That is clearly not “within acceptable modeling ranges.” This in addition to the notion that a model can be calibrated based on a single event at one location.
Page 20, Figure 8: Things get worse. The authors claim that it is similar to the Forest subwatershed, but given that the highest observed concentration is 200 mg/L vs. 800 mg/L in the simulation, the error is much larger. In addition, the over prediction of total sediment flux is much greater.
Page 21, Figure 9: And even worse. In this case, the simulation does not even resemble the observed data at all. The model under predicts the sediment load significantly. This of course has to do with the poor hydrologic match.
Page 21, Figures A-2 to A-15: Without a scientifically valid measure of goodness of fit, it can’t be stated whether the model predicts the TSS well or not, but in general it appears that the model over predicts them substantially.
Page 21: Amazingly, the authors have the audacity to state: “Overall, the model appears to reproduce the magnitude of the observed data well.” This model has clearly been poorly implemented. Another option is that this model is not applicable to these conditions. But to fool oneself into thinking that the output of the model is valid is incorrect.
Page 21: What is the significance of Jan 1995 to July 2005? Why not extend the simulation period to cover the time frame where very good observed data is available for the harbors, in 2006? What is the temporal resolution of the LAR and SGR models for this longer period? Still hourly?
Page 22: What was the source of observed data for the concentrations of toxic organics used? There should be a table summarizing the datasets (source, period of record, number of records per toxic, detection limits, etc.) How representative is this data of the entire watershed?

Page 22: The authors indicate that the previously calibrated models (assume it is LSPC models, but should be explicit) of the LAR and SGR were expanded in some way. How were they "expanded"? What does this do to the calibration?

Page 22: Does the POTFW parameter depend on pH for the metals? Or fraction of organic content for the toxic organics? If not, then this parameter does not truly represent the relationship between sediments and these toxics, and should be improved before using it this way. Are any reactions taken into consideration? If not, state this.

Page 23: It is unclear whether the model output is total metal, dissolved metals, or metals in particulate. If only dissolved, how do you account for the load on the TSS? When comparing to observed data, are you comparing the correct fraction? This would make a huge difference.

Page 23: The authors mention that the comparison was graphical. They really mean visual, which as indicated above, is not scientifically acceptable. The authors mention that for these three metals the predicted concentrations are "slightly lower" than the observed concentrations. For Cu the simulated peak concentration is significantly less than half of the observed value. For Pb, there is a larger discrepancy. The least difference is for Zn, but there is still a significant error. The cumulative loads (integrating Figure 11) are seriously over predicted, which is not surprising given the error in flow and TSS. Thus, it is unclear what the authors consider to be "slightly lower" or "fairly close". Again, there is a statement that "these model results are within acceptable modeling ranges" which is rather unnerving. Just like in the previous "calibrations" only one storm event at one location was used to "calibrate" the model. Scientifically this is unacceptable.

Figure 10: The small negative values for the simulation are an artifact of the graphing software, but should not be presented. They are not real.

Page 26, Table 6: The percent differences for the largest subwatershed are around 84 to 87%. Clearly this model is not predicting the correct toxic metal concentrations or loads during wet weather. Given that it under predicts the concentrations, it would result in a higher risk to the environment and humans, since one would be misinformed in the actual levels.

Page 26: Are these EMCs flow-weighted? Unclear, and very important.

Page 28: These results clearly indicate that this model is not valid. The results are not "well within the ranges of observed data."

Figures A-16 to A-27: Without a scientifically valid measure of goodness of fit, it can't be stated whether the model predicts the metal concentrations well or not, but in general it appears that the model under predicts them substantially at most locations, most of the time.

Page 33: This sensitivity analysis is terribly simplistic. These two sediment parameters are important, but there are many others that may play a role in determining the metal concentrations. A thorough review of the hydrologic, sediment and metals parameters in LSPC should be done, and then those that result in the highest sensitivity should be considered. The error bars for the EMC are of interest, but the most important calculation is the load for each metal, not the EMC. The current sensitivity analysis is not scientifically acceptable. The authors are referred to Chapra's book on "Surface Water Quality Modeling", to learn how a sensitivity analysis is performed.

Page 36: The authors indicate that “Final EMC values for SGR and Coyote Creek were obtained by averaging the three storms EMCs and their respective standard deviations for each reach.” Frankly, this sentence makes no sense.
Page 36: The authors indicate that McPherson et al. (2006) “state that in most cases, the total load estimated using EMCs for long-term simulation can have similar accuracy as more complex models.“ While this is a statement, this has not been proven. The use of EMCs has its place where insufficient data is available, in which case using a more complex model is not going to improve the result. If that is the case, then what was the point of setting up LSPC/HSPF for these watersheds when a simpler calculation could be performed?
Page 36: These sensitivity analyses were again based on just a perceived “most sensitive parameter” without any formal evaluation of other parameters at all. While one can generate different values using different EMCs, it is not valid to assume that this represents the widest range of probably values.
Figure 17: Most of the observed values are outside the plus/minus one standard deviation range according to the model. This indicates that the model does not adequately predict the actual range of concentrations that will be observed. Again, only one storm event is evaluated.
Figure 18: In this case, most of the observed values are below the lower range based on one standard deviation, so the model over predicts at this location.
Figure 19: No observed data, so no way to know if the model over or under predicts.
Page 40: These results for “total PAHs” are only valid for the aggregate, and not for specific PAHs. Since each PAH has its own toxicity and fate and transport, the results are not useful for predicting the actual toxicity of the discharges. The reader should be made aware of this.
Page 46: The method is acceptable, except that the TSS values used for the calculations are incorrect, so the results are not valid.
Figure 24: Define “Port DL”.
Page 50-51: As far as one can gather, for the LAR the authors used observed flow data, but for the SGR they used LSPC modeled flows. Given that the LSPC cannot model dry-weather flows, it is unclear how one can use it for some but not all. There is no clear explanation for the inconsistent approach.
Page 51: The dry weather flows are apparently based on 1 or 2 days of flow monitoring. How do we know those were typical days? The load analysis is being extrapolated to thousands of dry days based on this sample size?
Table 13: Are the data log-normal? This should be made explicit. The standard deviation seems to be much larger than the mean, so if the data are normal, then the mean minus one standard deviation would be a negative value. Are these total metal or dissolved metal concentrations? Was there enough flow at these locations to mobilize sediments during dry weather flows?
Table 14: Given the scarcity of data, this approach is adequate for dry weather flows. However, it should be clearly stated that these estimated loads have a high degree of uncertainty, given that they are based on very few observations. The high range may not reflect the variability in flow, and the “mean” value is not known to the degree indicated in these values (3 to 4 significant digits). At best it is an order of magnitude estimate with one significant digit.

Page 54: Is there any study that can support the assumption of the sediment composition? Surely the soils in this area have been studied by others in the past.
Page 55: the “sensitivity analyses” performed are not true sensitivity analyses.
Page 55: to the dry-weather flow and load predictions, add that these are based on data from one day only.
Page 55: There was no presentation of the estimated the point source and non-point source loads separately. Since this will be needed in the TMDL, this is an important flaw in the presentation. It is not clear that these were actually calculated separately.
Page 55: Similarly, the final results do not separate the dry and wet weather loads for each pollutant. Instead, only “average daily” loads are presented in the figures. Since management actions may be different during these days, lack of this information is a major flaw in the presentation of results. A table presenting the average dry and wet weather loads is needed.
Page 55: There is no formal estimate of the uncertainty in the loads. Figures 30-35 should present the error bars that reflect the uncertainty in load estimates. Clearly, given the poor calibration basis (one storm event) and the poor calibration results (as discussed above), there is a considerable amount of uncertainty in the estimated loads. This information is very important for the TMDL. These sensitivity analyses do not adequately reflect the uncertainty in the calculations.
Page 55: The EMCs and other land use based load estimates have been considered for the “industrial” land use as if this was a typical mix of industries. However, near the harbors there are many facilities which are clearly “heavy industry”, including refineries and other chemical processors, which are likely to generate much higher loads than light industry, or even a mix of industrial sources. One could look at the Toxic Release Inventory information for the facilities in this area to have a much better idea of the types of sources. These sources are very close to the waterways and harbors, so the transport pathway is short. Since this has not even been mentioned in the report, or particularly in these modeling assumptions, it is likely that this was not taken into consideration by the authors. Thus, the load estimates are likely to be incorrect.
Figure 29: The label is hard to read.

Minor typos:
Page II-i: heavily “rely” not “reply”
Page 51: change “verses” to “versus”

Comments on Appendix III

Appendix III.1

Page 3: The time period for the EFDC model was 2002 to 2005, while the best observed data is from 2006. This does not make sense, and the explanation for truncating the simulation in 2005 is that the LSPC models were simulated from 1995 to 2005. Why not extend the LSPC simulations to 2006?

Page 4 and others: as in the rest of the report, tables and figures should be numbered so that they can be referenced in the text, and some interpretation of the information in each table and figure should be provided in the manuscript.

Page 4, Waterbody Information table: the deposition rates are not known to such precision, and should thus be reported only to the degree that the calculations justify. I doubt there are more than 2 or 3 significant figures, but not 7 (e.g. 1,564,089 kg/yr). Good scientific practice requires one to report the correct precision. In the caption it mentions “TMDL waterbody”, but in reality these are “TMDL zones”.

Page 4, Sediment Concentration Information table: It is unclear if these are total metal concentrations, dissolved or adsorbed. For the toxic organics it is clear that these are total concentrations, so it is even more confusing.

Page 5: The text reads “The areas and percentages below are…” Percentages of what? Should be clear that these are percentages of freshwater inputs.

Appendix III.2

Page 7: The threshold for wet weather days is inconsistent. For the Dominguez Channel the 90th percentile flow is used, while for the near shore watersheds, the 50th percentile flow is used (Appendix II, page 4). This can make a significant difference in the load considered for different watersheds, given the different approaches used for dry and wet weather. The load duration curves were apparently developed using only the wet days. Given the relative few wet days in this region, this may bias the analysis. No effort seems to have been made to determine the impact of this decision. Although the flow and loads in dry days is smaller, the cumulative contribution to the harbor waters can be quite significant over time. There does not appear to be any indication of the relative contribution of dry and wet days to the total load.

Page 8: The “Allowable Loads” presented in the table do not match those presented in page 11. The difference is quite significant. There is reference to a section in the entire report where these Allowable Loads are calculated. Since this is critical for the TMDL calculation, it should be a transparent presentation. What is the uncertainty in the calculation of these Allowable Loads? Clearly there are many data gaps, so there must be some sense of the major uncertainties. Again, reporting these values to a high precision gives the false impression of certainty.

Pages 9 and 10: The y-axis labels are unreadable, and the numbers in the x-axis are also unreadable.

Page 11: A 10% explicit MOS is considered. No justification is given. Given the data gaps, it is very hard to justify such a small MOS.

Appendix III.3

Page 13: The method for determining the initial concentrations is not discussed at all. Given that this is a complex calculation based on data from several years and locations, and that it is crucial for an adequate estimate of the concentrations of toxics over time, it is a major deficiency in the report. Was equal weight given to all data? If not, what were the weights? How can one use data from 2006, past the simulation period, to determine the initial concentrations in 2002? There is no scientific basis for doing this, since the only method for back calculating the concentrations from 2006 to 2002 is the model that is being calibrated. The authors have a serious problem with circular logic. In addition, there is no scientific basis for reporting the concentration to such a high precision. Laboratory results do not have such precision. Perhaps the authors could take a look at a few lab reports to understand the actual precision of such data.

Appendix III.6

Page 50, Table 2: Correct notation is $\mu g/m^2$-day, not $\mu g/m^2/day$. Also, it is better to present a range of values, or some other measure of their variability. Clearly, their sources are different and meteorological conditions play a major role. There is no discussion of these considerations; one must assume that this was not taken into account.

Page 51: There is no calculation of the uncertainty in these estimates. Since these loads are an important part of the TMDL calculation, it is important to determine the uncertainty.

Appendix III.8

Page 2, Fig. 1: Hard to read label.

Page 2: A four-day average was considered. Is there a regulatory or scientific basis for this selection? What is the objective of such averaging? This tends to smooth out peaks in concentrations, which may under protect organisms that are exposed to such peaks.

Page 4, Table 2 and Figure 3: While the comment is made that almost all of the TMDL zones exceed the criterion (34 mg/kg) even when all the upland sources are eliminated, it is clear from Figure 3 that this is a matter of time. The simulation ended after 4 years, but if additional time was taken into consideration, in fact most zones would eventually meet the criterion. Some may take too much time, and thus additional actions may be needed. However, the current analysis does not point out this important finding. Eliminating or reducing upland sources does have a very positive effect, as expected. The current text implies that there is little or no value in doing so, since the criterion is still exceeded. The authors could have done further analysis to determine why there are some locations that respond very rapidly and some that almost do not respond, to guide the development of the TMDL.

Page 5, Fig. 2: Why use negative values in the x-axis? What is the significance of starting day 0 in the middle of the simulation? Why not use actual dates?

Page 11, Figs. 8 and 9: These two figures are very similar, if not identical. Is this a mistake? It is hard to understand how the PAH concentrations would decrease so rapidly in the Base Scenario. If this was the case, one would not need to do anything but wait.

Page 11, Figs. 10 and 11: Same issue.

Page 15: The authors mention that “copper hot spots within all zones were reduced”. How were they reduced? There was no mention of this previously in the report. This would be quite important to know. How was this information considered in the TMDL calculation?