Appendix A

Peer Review Comments and Responses

The technical portions of the proposed Basin Plan amendment to incorporate TMDLs for toxic pollutants in sediment were peer reviewed by Mr. Robert Brown Ambrose, Jr., P.E., environmental engineering consultant and former Environmental Engineer with the Ecosystems Research Division of the U.S. Environmental Protection Agency and by Professor Ashish Mehta, Professor Emeritus of Coastal and Oceanographic Engineering at the University of Florida. External scientific peer review of the technical portion of a proposed rule (in this case, the proposed Basin Plan amendment) is mandated by Health and Safety Code section 57004. This statute states that the reviewer's responsibility is to determine whether the scientific portion of the proposed rule is based upon sound scientific knowledge, methods, and practices.

The San Diego Water Board provided the peer reviewers with the draft Technical Report, the draft Basin Plan amendment, and a list of key issues with discussion for the peer reviewers to address. The list of key issues with discussion provided to the peer reviewers is given below in the first section of this appendix. The peer reviewers' comments and the San Diego Water Board's responses follow in subsequent sections.

I. Issues for Peer Review

1. Use of the Logistic Regression Model Threshold 20 percent (LRM T20) values as the numeric targets.

Numeric targets are established to evaluate attainment of the narrative sediment quality objectives for chlordane, total PAHs, and total PCBs. The 20 percent threshold (T20) values of the California Sediment Quality Objectives, Southern California Logistic Regression Model Approach (CA LRM) were used to set the numeric targets for chlordane, PAHs, and PCBs. The Southern CA LRM T20 values are derived using a logistic regression equation of paired toxicity data and sediment chemistry data from estuaries in Southern California to predict the probability of observing sediment toxicity that corresponds to the 20 percent threshold.

The rationale for choosing the T20 value is that 20 percent mortality is commonly used as the threshold at which a sample is considered toxic for the standard 10-day amphipod toxicity test. This 20 percent mortality (80 percent survival) value was statistically determined to be the appropriate significant value from the control to use for this type of analysis (Thursby et al. 1997). Further explanation for using the T20 as the point at which a significant difference can be seen in toxicity test results is that 1) the data used in the SQO analysis was taken from estuaries in California only, 2) the analyses are used to determine impairment in each of the creek mouths, and 3) the LRM approach in the SQO analysis was developed to determine the condition of California estuarine sediment.

2. **Reasonableness of assumptions for wet-weather watershed modeling.**
Several assumptions are relevant to the modeling system used to simulate the fate and transport of wet-weather sources of organic pollutants that are associated with sediment. Another major assumption not in the general list is that the regression equation for TSS and total PAHs, and for TSS and Chlordane was used to estimate concentrations of total PAHs and Chlordane.

   This model was used to estimate both existing pollutant loads and total maximum daily loads. Please comment on the validity of these assumptions.

3. **Use of the LSPC Model to calculate pollutant loading during wet-weather conditions.**
The TMDLs were developed using a computer modeling system that includes a watershed and a receiving water model. The models provide an estimation of loadings of the pollutants from the watersheds based on rainfall events, and a simulation of the response of the receiving waters to these loadings based on transport and fate of suspended sediment loading and dynamic effects of tidal flushing.

   U.S. EPA's Loading Simulation Program in C++ (LSPC) was used to simulate watershed hydrology and transport of sediment in the streams and storm drains conveying pollutants to the impaired areas of the San Diego Bay shoreline. Please comment on the use of this modeling system for the purpose of calculating TMDLs.

4. **Use of the EFDC Receiving Water Model to simulate fate and transport of toxic pollutants and to calculate TMDLs.**
The receiving water models, based on Environmental Fluid Dynamics Code (EFDC), were developed to simulate the assimilative capacity of the waterbodies, the transport and fate of suspended sediment loading, and dynamic effects of tidal flushing.

   The structure of the EFDC model includes four major modules: (1) a hydrodynamic sub-model, (2) a water quality sub-model, (3) a sediment transport sub-model, and (4) a toxics sub-model. The modeling effort for San Diego Bay included the hydrodynamic, sediment transport, and toxic sub-models.

   Instead of developing a bay-wide sediment transport and toxics modeling system based on the hydrodynamic model, individual sediment transport and toxic models were developed for the impaired areas. This was done to focus on the depositional zones at the mouths of each of the creeks and to reduce computational time. Three separate models were constructed: the Paleta Creek mouth, Chollas Creek mouth, and Switzer Creek mouth toxics models.

The focus of this question is whether the overall approach of using these sub-models was appropriate and appropriately used.

5. **Calculations of waste load allocations, load allocations and TMDLs during wet-weather.**
   Data and model limitations require that assumptions be made to calculate the wet-weather waste load allocations. Assumptions included boundary conditions, state variables, initial environmental conditions, and the generation of the model grid cells.

   Calibration refers to the adjustment or fine-tuning of modeling parameters to reproduce observations. After the model was configured, model calibration and validation were performed. This was a two-phase process, with hydrodynamic model calibration and validation completed before evaluating the performance of sediment transport and toxic modeling. Calibration and validation were performed for all three models: the hydrodynamic model, sediment transport model, and toxic model.

   This question focuses more on the details of the modeling and whether the appropriate assumptions were used when developing or running the receiving water model; as well as whether the calibration and validation results are sufficient.

6. **Identification of critical locations and use of average sediment pollutant concentrations across all grid cells in each creek mouth area for TMDL calculation.**
   For TMDL-related calculations, the water quality at a critical location in an impaired waterbody was compared to numeric targets for assessment of required reductions of pollutant loads to meet WQOs.

   For San Diego Bay shorelines, the critical locations for meeting numeric targets include the entire length of impaired shoreline, extending to the end of the piers. For model development, receiving waters at impaired shorelines were represented in the model with multiple grid cells (see Modeling Report Figures 4.4 and 4.5). Compliance with the TMDL target for each pollutant was assessed based on the results from the receiving water models for the creek mouth areas. Predicted sediment concentrations from the end of the modeling period were averaged across all grid cells within each creek mouth area to determine if watershed reductions were needed.

   The EFDC model was incapable of providing an accurate prediction in measured receiving water sediment pollutant concentrations at the small, individual grid cell-level scale. The model generated predicted concentrations that varied within each grid cell. The model is capable of prediction at a slightly larger scale, but has most likely reached its capabilities at the grid cell-level scale (see Figures 4.4 and 4.5).

7. **Use of conservative assumptions to comprise an implicit Margin of Safety.** The conservative assumptions built into both the wet weather and receiving water models are considered sufficient to account for any uncertainties. The implicit MOS was thus generated by incorporating a series of conservative assumptions regarding current source loading of pollutants from the watersheds, as well as assumptions regarding the critical location in the creek mouths, and Bay pollutant concentrations.

8. **Implementation of the TMDL.** The Implementation Plan will include a range of regulatory alternatives available to the San Diego Water Board to require the responsible parties to meet the waste load reductions. Specifically, the San Diego Water Board plans to incorporate requirements of the TMDL upon issuance or reissuance of permits. The primary TMDL implementation action to be required of the responsible parties may include development of a Comprehensive Load Reduction Plan (CLRP), if necessary. If needed, the San Diego Water Board may issue Water Code section 13267 investigative orders to direct responsible parties to complete required plans in the interim, while further implementation may also include increased emphasis, including compliance assessments.

Overarching Questions

Reviewers are not limited to addressing only the specific issues presented above, and are asked to contemplate the following “big picture” questions.

(a) In reading the staff technical report and proposed implementation language, are there any additional scientific issues that are part of the scientific basis of the proposed rule not described above? If so, please comment with respect to the statute language given above.

(b) Taken as a whole, is the scientific portion of the proposed rule based upon sound scientific knowledge, methods, and practices?

The San Diego Water Board has a legal obligation to consider and respond to all feedback on the scientific portions of the proposed rule. Because of this obligation, reviewers are encouraged to focus feedback on the scientific issues that are relevant to the central regulatory elements being proposed. Some proposed actions may rely significantly on professional judgment where available scientific data are not as extensive as desired to support the statutory requirement for scientific rigor. In these situations, the proposed course of action is to be favored over no action.

References Cited:

RWQCB. 2007a. Waste Discharge Requirements for Dischargers of Urban Runoff from the Municipal Separate Storm Sewer system (MS4s) Draining the Watersheds of the County of San Diego, the Incorporated Cities of San Diego County, the San Diego Unified Port District, and the San Diego County Regional Airport Authority. Order No. R9-2007-0001. NPDES No. CAS0108758. January 24, 2007. Available at: http://www.waterboards.ca.gov/sandiego/water_issues/programs/stormwater/docs/sd_permit/r9_2007_0001/2007_0001final.pdf

SWRCB. 1999. National Pollutant Discharge Elimination System (NPDES) Permit for Storm Water Discharges from the State of California, Department of Transportation Properties, Facilities, and Activities. Order No. 99-06-DWQ.

SWRCB. 2008. Draft Staff Report: Water Quality Control Plan for Enclosed Bays and Estuaries. Part 1. Sediment Quality. CA Environmental Protection Agency, State Water Resources Control Board, Sacramento, CA. July 2008. On State Water Board's website at the following link: http://www.swrcb.ca.gov/water_issues/programs/bptcp/docs/sediment/071808_draftstaffreport.pdf

Thursby, G.B., J. Heltshe, and K.J. Scott. 1997. Revised Approach to Toxicity Test Acceptablility Criteria Using a Statistical Performance Assessment. Environmental Toxicology and Chemistry. Vol. 16, No. 6, pp.1322-1329.

II. **Comments from Robert Brown Ambrose, Jr.**

Comment: Overall, my judgment is that this TMDL is based upon sound scientific knowledge, methods, and practices. Overall, the management actions proposed are based on the available evidence and seem reasonable, i.e., controllable sources meet the WLA and remediate the marine sediments. Keep in mind, however, that the uncontrollable sources constitute the majority of the TMDL, and pretty much all of it for PAHs and PCBs. These sources could possibly lead to episodic violations due to infrequent high watershed loadings that can occur during years when rainfall exceeds the design year of 2004-2005.

Response: Comments noted and the San Diego Water Board thanks Mr. Ambrose for his remarks.

Comments to the posed questions are provided below.

1. ***Use of the Logistic Regression Model Threshold 20 percent (LRM T20) values as the numeric targets.***

   Comment: I have no particular expertise on the use of toxicity data to establish target concentration levels. In general, the method strikes me as reasonable.

   Response: Comment noted.

2. **Reasonableness of assumptions for wet-weather watershed modeling.**

   Comment: I agree with the authors that good precipitation coverage is a prerequisite for good hydrology simulation, and that good hydrology simulation is a prerequisite for good sediment transport simulation. The hydrology and solids simulations here seem to me to be reasonable in light of the available meteorological, flow, and water quality data available for calibration. One must recognize that the uncertainties in pollutant loadings are high, driven as they are by uncertainties in hydrology, sediment transport, and solids partitioning. I address estimated uncertainties in response to Question 7 below.

   It is reasonable to assume that the largest loadings will be delivered to San Diego Bay during wet weather conditions, and that basing loadings on the wettest year monitored (out of 6) will produce reasonably conservative TMDLs. But it is also reasonable to expect that the TMDL will be violated periodically, when rainfall totals exceed the design year used here. The report described the design year as “one of the highest rainfall years on record” but did not give any statistics on the expected return interval (e.g., 10, 20, 50, or 100 years) or the length of the rainfall record. To reinforce the large effect of inter-annual variability in rainfall, note that Appendix D Table 6-1 shows annual pollutant loadings that vary widely. For the five winters simulated in Paleta Creek, the runoff volumes from the highest year were 4 times higher than in the second highest year, while solids and pollutant loadings were 50 times higher.

Assumptions for the sediment modeling are listed in Appendix E, section 2.2. The uniformity of detachment across land use types was of concern to me initially, since there are great differences between, say, row crop agriculture and forest. Looking at the land uses in these urban basins, however, I believe this assumption is reasonable for this TMDL. The assumptions about non-road and road sediment composition (sand-silt-clay) seem reasonable, but would be better if backed by some observed data.

Sediment buildup in storm drains between events is not represented in LSPC, but is thought to be important in San Diego. Consequently an extra processing step was developed outside of LSPC to represent this process. This step as described is a reasonably justified addition.

It is reasonable to use regression to link PAHs and chlordane to TSS. Note that these are log-log regressions, and that the uncertainty is relatively high.

Response: The year selected for the simulation was chosen based on nearly 100 years of rainfall data (WY 1911-2010). The selected year (WY2005, 22.6 in), was in the 99th percentile with only 1941 having more rainfall. Choosing an extreme rainfall year provides an implicit margin of safety and a reasonable expectation that foreseeable conditions will not exceed the modeled year. The model represents a continuous time period which includes storms with different intensities and return periods; therefore, it was not designed to explicitly simulate a particular design storm.

3. Use of the LSPC Model to calculate pollutant loading during wet-weather conditions.

Comment: As noted, the LSPC model was developed from HSPF to facilitate watershed loading calculations for TMDLs. The hydrology module is well established, and can be reasonably calibrated to individual watersheds. The sediment transport module is based on accepted detachment algorithms. The pollutant loading calculations for PAH and chlordane are based on correlations with solids. All of the algorithms are reasonable, but of course their actual performance depends to a great degree on the supporting data.

Response: Comment noted.

4. Use of the EFDC Receiving Water Model to simulate fate and transport of toxic pollutants and to calculate TMDLs.

Comment: EFDC is an appropriate model for toxicant TMDLs. It has been used widely for this kind of analysis. The three submodels used here – hydrodynamics, sediment transport, and toxicant fate – are appropriate for this analysis. I support the selection and use of EFDC for this TMDL. That said, I want to correct a misstatement about WASP. In Appendix D (page 6), the authors claim that WASP only simulates 4 layers of bed sediment. This is false; the number of bed layers has never been limited since the first version of TOXIWASP back in the early 1980s. This statement should be corrected if possible. Sometimes errors like this are propagated in reports without anyone checking the original sources. Another limitation of WASP stated in this section was true at the time Appendix D was written, but is no longer true. WASP did offer only simple descriptive solids settling and resuspension routines. Since Spring 2009, however, WASP also includes mechanistic solids routines similar to those in EFDC.

I was skeptical about dividing the overall San Diego Bay computational network into individual subnetworks for the individual TMDLs. In particular, my concern focused on whether the open water boundary conditions (BCs) for solids and toxicant would unduly affect the area of concern. In reviewing the application, it seems to me that the BCs do not, in fact significantly affect the area of concern. Furthermore, the boundary assumptions appear to be mildly conservative (as intended).

State-of-the-art PCB modeling break out total PCBs into constituent fractions, usually homolog groups, for more accurate representation of their sediment partitioning and air-water exchange (volatilization). One example is the Delaware River TMDL for PCBs. While this more detailed approach is more scientifically defensible, it requires site-specific data on the homolog groups. If that is not available, then PCBs must be represented as a single toxicant using averaged or lumped chemical properties (e.g., partition coefficient or Henry's Law constant) calibrated to site-specific data. Recognizing that the partition coefficient in fact represents a distribution of values, it would be wise to conduct sensitivity analyses to investigate the effects of higher and lower values on the TMDL. It is likely that the average PCB composition in sediments will shift over a long period of time, probably to the higher chlorinated fractions with generally higher partition coefficients.

Response: The model results indicate the Bay boundary contribution is negligible under existing conditions because the watershed and sediment concentrations have a much greater influence on water quality in the creek mouth areas. As expected, the Bay boundary has a greater effect under the TMDL scenario after reducing the watershed and sediment contributions.

The partitioning coefficients for total PCBs were 0.0033 for clay and silt fractions and 0.00001 for sand.

The modeling framework used to develop these TMDLs was based on several factors including data availability. A more complex homolog-based approach was not used primarily due to the lack of available data and increased uncertainty in the results. A lumped representation typically provides a more reliable representation of the general trend and magnitude than a more detailed representation that is based on sparse data.

5. **Calculations of waste load allocations, load allocations and TMDLs during wet-weather.**

Comment: Calibration of the hydrodynamics was done with an appropriate bay-wide grid. Results were then used as boundary conditions for the individual networks. This seems reasonable to me. The calibration focused on matching observed tidal elevations, which is standard. Apparently predicted tidal current velocities were not checked against observations. Since velocities are very important in the calculation of shear stress and consequent sediment erosion and deposition, it would be good to provide some predicted current velocity plots (preferably at the creek mouths) and check them against data. If no data are available, the plots could be assessed qualitatively.

Calibration of solids was done for each of the computational networks. Appropriate open-water solids boundary concentrations were derived iteratively, and were very low (0.001 mg/L). I'm surprised that there were no ambient solids concentrations from monitoring programs in the Bay to use for boundary concentrations. It would be more defensible to use measured TSS values, particularly since the solids calibration focused on reproducing TSS data in the areas of concern. It seems (at least anecdotally) that ambient Bay TSS is quite low, so this uncertainty in boundary concentration is probably minimal. While calibrating to TSS data helps constrain model coefficients, it would have been better to also include sediment solids in the calibration. Perhaps this was done; some of the text was ambiguous on this point. In particular, calculated net burial (or scour) rates and calculated fractions of sand, silt, and clay should be compared with available observations. This would better constrain model coefficients and provide a higher level of confidence in projections.

The toxicant modules were calibrated with site-specific partition coefficients (K_p). The PCB K_p values seem a little low to me – 1.2×10^3 and 1.9×10^3 for clay and silt, 3.2×10^3 for water column clay and silt (using units of L/kg). These are not completely outside of reported ranges. By way of contrast, the KOC value for penta-PCB used in the Delaware River PCB TMDL was $10^{6.26}$. In addition, that study modeled partitioning to DOC, with a KDOC value of $10^{5.26}$. That study modeled particulate detrital carbon (PDC) as a state variable, and used the K_{OC}

to partition penta-PCB to PDC. For the sediment layer, PDC represented about 5% - 20% of the sediment solids, and so the apparent bulk K_p for the sediment layer would have been on the order of 105.

Specifying higher K_p values for PCB and PAH to silt and clay in the water column versus the sediment layers will cause "artificial scavenging" of pollutant from the water column to the sediment layer. While this constitutes another conservative factor for this TMDL, I prefer that calibration be done without this discrepancy. For PCBs, the water column K_p was about 3 times the sediment K_p, but for PAHs, the difference was greater than a factor of 50, which accentuates the artificial scavenging. Observed bulk partition coefficients in the sediment layer are often lower than the water column, but this may be due to differences in organic matter, with more sand (and generally lower foc and K_p) in the sediment bed than in the water column. At this point in the TMDL process, I do not recommend recalibration of the toxicant module. Perhaps this calibration could be revisited in future work when more data might be available.

From the simulated sediment concentrations generally show the short term effects of watershed loading as well as long-term effects inherent in sediment pollutants slowly equalizing with the water column. It is clear that the modeled response times of interest are greater than the 3 or 6 year length of the historical calibration period. The procedure for using EFDC to calculate TMDLs was to run the model for 3 years and see if sediment concentrations are building up or declining. This will work OK as long as this 3-year period includes the high flow design conditions. That was not stated explicitly in Appendix D Section 6.7, but elsewhere it was emphasized that the high flow year is being used for the TMDL.

Response: Tidal velocity data were not available for model calibration.

TSS data were not available in the outer Bay to help derive the open boundary condition; however, TSS data were available for the creek mouth areas for model calibration. Also, no data were available to quantify net burial or scour rates.

The partitioning coefficients were derived based on data presented in the following report (cited in the EFDC modeling report):

Chadwick, B., J. Leather, K. Richter, S. Apitz, D. Lapota, d. Duckworth, C. Katz, V. Kirtay, B. Davidson, A. Patterson, P. Wang, S. Curtis, G. Key, s. Steinert, G. Rosen, M. Caballero, J. Groves, G. Koon, A. Valkirs, K. Meyers-Schulte, M. Stallard, S. Clawson, R. Streib Montee, D. Sutton, L. Skinner, J. Germano, R. Cheng. 1999. Sediment quality characterization Naval Station San Diego. Technical Report 1777, pp. 152.

The Bay boundary condition was set based on data presented in the following report (cited in the EFDC modeling report):

Katz, C.N. 1998. Seawater polynuclear aromatic hydrocarbons and copper in San Diego Bay. Technical Report 1768. SPAWAR Systems Center San Diego

The critical high flow year was repeated for the three year simulation period and is explicitly stated in the Technical Report, Section 7.6 Calculation of TMDLs and Allocation of Loads.

6. Identification of critical locations and use of average sediment pollutant concentrations across all grid cells in each creek mouth area for TMDL calculation.

Comment: I agree with this conclusion about the predictive resolution of EFDC. Accurate calculations of sediment concentrations for specific grid cells are not to be expected from this (or other) models.

Response: Comment noted.

7. Use of conservative assumptions to comprise an implicit Margin of Safety.

a. Comment: It is difficult to assess whether the conservative assumptions are sufficient to account for all uncertainties. It must be recognized that the uncertainties in pollutant loadings are high, driven as they are by variability and uncertainties in hydrology, sediment transport, and solids partitioning. Before looking at the conservative assumptions, I qualitatively review the uncertainties in the major model components.

(1). Watershed runoff: The quality of the hydrology calibration depends on the spatial coverage of meteorological stations to represent precipitation patterns accurately. While the calibration was reasonable, there are some discrepancies between observed and predicted flows, particularly in Switzer and Cholas Creeks. The simulated cumulative water volume in Switzer was about 50% of the data, while the simulated cumulative volume in Cholas South is about 2.5 times the data. Simulations of Paleta and Cholas North are closer. Some simulated events compare quite well with observations. Qualitatively, I would judge the uncertainty in seasonal simulated hydrology to be greater than 20% but less than 100% depending on the creek and the year.

Response: The differences in cumulative flow volume are a function of the rainfall heterogeneity and other weather conditions in the area. Available meteorological data were not sufficient for a rigorous calibration and validation of each watershed model. For consistency with other watershed modeling efforts in the San Diego area, regionally-derived model coefficients were used. The graphs generally show a reasonable representation of the rainfall-runoff response. For example, the first major storm in the Switzer watershed underpredicts storm volume, which shifts the cumulative curve down, but for the remainder of the simulation period, the model accurately mimics the watershed response.

(2). Watershed solids loadings: There is a lot of expected variability in the TSS data, and wide 95% confidence intervals particularly for the largest concentrations. The model calibration for solids had mixed results. It was noted that at most of the stations, the 95% confidence limits for simulated solids concentrations overlapped with the 95% confidence limits for observed solids concentrations, and so the model could be considered acceptable. The null hypothesis here is that the model is correct, and the overlap in confidence intervals only means that one cannot statistically reject the model as being wrong at the 95% level. This says little about whether and with what confidence one can assert that the model is correct. It seems to me that the uncertainty in simulated solids loadings is high, qualitatively perhaps a factor of 2 or more.

Response: The model performed reasonable well across the majority of land uses for the modeled constituents comparing both the event mean concentrations, as well as throughout the storm pollutographs. Having independent datasets for calibration on a small land use scale and validation on a larger scale provides additional confidence of the model's overall performance.

(3). Watershed pollutant loadings: While it is reasonable to use regression to link PAHs and chlordane to TSS, note that these are log-log regressions, and that the uncertainty is relatively high. Looking at the regression plots, the actual PAH and chlordane loadings could easily be 50% higher (or lower).

Response: Available monitoring data for TSS and the other constituents is relatively sparse. A more rigorous analysis would require additional monitoring data, as a result, a log-log scale was used to better identify the relationship to TSS concentrations.

(4). Water body solids balance: The water body solids balance is driven by simulated hydrodynamics and watershed solids loadings, and utilizes process based solids settling and resuspension algorithms. This module was calibrated to water column TSS data, by adjusting such parameters as critical shear stress for deposition and scour. I have already noted that current velocities were not verified, leading to some uncertainty in the scour calculations. Because the calibration seems to be unconstrained by data on net burial and benthic solids composition (fraction sand-silt-clay), the uncertainty in the calibrated mix of parameters is a little higher than it otherwise might be. Still, the parameter values are reasonable, and extrapolation of the solids module in the TMDL should not add too much to the overall uncertainty, at least in terms of the predicted pollutant concentrations in the sediment bed. It is quite possible that the calibrated balance of solids loading and sediment exchange components (deposition, scour, burial) will be too fast or slow, leading to over or underprediction of pollutant removal rates. This, however, should not unduly affect this TMDL. I encourage the collection of sediment composition data at some point in the future when the TMDL might be revisited.

Response: Comment noted. Also refer to the response to Comment No. II.5., above.

(5). <u>Water body pollutant fate</u>: The water body pollutant fate module is driven by simulated hydrodynamics, watershed pollutant loading, and solids transport. This module was calibrated primarily by adjustment of the partition coefficients for sand, silt, and clay in the water column and in the sediment bed. Partitioning is quite variable among the PCB and PAH fractions, and so the lumped PCB and PAH partition coefficients will be quite uncertain, even with site-specific data. Different combinations of values could result in reasonable calibration to observed water column and sediment data. My experience, for example, is that partitioning of PCBs to silt and clay is higher, and to sands is lower. Like the solids parameters, the calibrated combination of partitioning values may affect the overall kinetics of pollutant buildup and removal, but should not unduly affect the target sediment concentrations for this TMDL. My guess is that the uncertainty here in sediment concentration levels is within a factor of 50%.

Response: Comment noted. Also refer to the response to Comment No. II.5., above.

b. ***Comment***: While some components of this modeling analysis have relatively high uncertainty, the calibration process should keep the uncertainties from building on each other too much. For example, if calibrated watershed solids concentrations and loadings are too high, the calibrated water body deposition and scour rates could compensate and still produce reasonable solids and pollutant concentrations in the sediment. A formal uncertainty analysis would be too computationally intensive for these models. A series of sensitivity analyses could shed more light on the ultimate uncertainty in sediment pollutant levels. My judgment is that the uncertainty would be within a factor of 2.

Response: The uncertainty in simulated watershed loadings is expected to be much less than is inherent in the choice of the critical model year to simulate. The receiving water response to the terrestrial inputs will also be masked because of the large amount of freshwater input to the Bay in WY2005. Model TSS validation was, on average, within 58 percent of the measured value. When the exception of one storm that showed poor agreement (Chollas South, Storm 1), the average accuracy improved to within 17 percent of the measured. This degree of accuracy is well within the range of acceptable model results as discussed in Donigian, 2000.

	Percent Difference Between Simulated and Recorded Values		
	Very Good	**Good**	**Fair**
Hydrology/Flow	< 10	10 - 15	15 - 25
Sediment	< 20	20 – 30	30 – 45
Water Temperature	< 7	8 – 12	13 - 18
Water Quality/Nutrients	< 15	15 – 25	25 - 35
Pesticides/Toxics	< 20	20 - 30	30 - 40

Donigian, Jr., A.S., 2000. HSPF Training Workshop Handbook and CD. Lecture #19. Calibration and Verification Issues, Slide #L19-22. EPA Headquarters, Washington Information Center, 10-14 January, 2000. Presented and prepared for U.S. EPA, Office of Water, Office of Science and Technology, Washington, D.C.

c. ***Comment***: Next I review the six conservative assumptions listed for the implicit margin of safety.

(1). <u>High rainfall design conditions</u>: Choosing a high rainfall year as the design condition for calculating watershed loads for the TMDL makes sense. Its effectiveness as an implicit margin of safety, however, will vary year to year. All things being equal, it is reasonable to expect that actual watershed pollutant loadings will be less than the watershed loads allocated in the TMDL for the majority of years where rainfall totals are less than the design year. During low to average rainfall years, this design condition should provide a MOS of a factor of 50 or more. But for higher rainfall years, this MOS declines. All other things being equal, it is reasonable to expect that the MOS will be zero during those years when rainfall totals exceed the design year. So the choice of design rainfall conditions can be considered as part of the implicit MOS only for the years when rainfall falls below these design conditions. The report described the design year as “one of the highest rainfall years on record” but did not give any statistics on the expected return interval (e.g., 10, 20, 50, or 100 years) or the length of the rainfall record.

Response: See response to Comment No. II.2.

(2). <u>Use of highest sediment grid for TMDL</u>: The text is inconsistent (or I am confused) on how the sediment grids will be used to calculate TMDL reductions. Section 7.5 states that "predicted sediment concentrations… were averaged across all grid cells within each creek mouth area to determine if watershed reductions were needed…" This section also states that "individual grid cell results were also considered in determining the final percent reduction required." But since the maximum grid cell concentration will always be higher than the average, I don't understand how the averaged concentrations will actually be used. If the sediment standard applies to the average concentration in the creek mouth, then using the maximum grid concentration will provide an implicit MOS. I did not see any analysis of the maximum versus the average sediment concentration. I would guess that this would provide a MOS of greater than 50%, and perhaps up to a factor of 10. It would be useful to add a little text to this section to clarify how the highest grid concentration will be used, and what percent MOS this would represent.

Response: The average of the maximum predicted concentration for all grid cells within each impaired shoreline area was used to determine the percent reduction needed to meet the TMDL targets. The use of individual grid cells to determine final percent reduction required was done in previous runs, but was determined to not be appropriate because of limitations in the accuracy of the results at individual grid cell level. This language has been removed from the Technical Report. However, individual grid cell were generally reviewed for the purpose of determining the appropriate explicit MOS.

(3). <u>Use of historic data for San Diego Bay boundaries</u>: Using historic data for the Bay boundary concentrations should be at least somewhat conservative, but if the boundaries are properly drawn, the open water BCs should affect the maximum creek mouth sediment concentrations only a minimally. It would be easy to conduct a set of sensitivity runs to assess this implicit MOS. Absent those, my guess is that this boundary assumption will provide less than 5% MOS.

Response: The model results indicate the Bay boundary contribution is negligible under existing conditions because the watershed and sediment concentrations have a much greater influence on water quality in the creek mouth areas. However, as expected, the Bay boundary has a greater effect under the TMDL scenario after reducing the watershed and sediment contributions.

In addition to the implicit margins of safety, explicit margins of safety have been assigned for each TMDL.

(4). No loss of pollutants through the Bay to the ocean: I do not understand this assumption. First, the grids used for the TMDL only cover the individual creek mouths and a small stretch of the Bay. These grids do not extend to the ocean. Furthermore, the model will be exchanging water, sediment, and pollutant across the open water boundaries. Because the model is using historical data for BCs, and because the TMDL loads will be less than the historical loads, there will likely be no net transfer of pollutant out of the individual model grids. This would not be an assumption, but rather a consequence of the specified BCs (see previous paragraph). Perhaps I don't understand the point here, but my judgment is this provides no extra MOS.

Response: The model was setup to not allow for losses outside the system; therefore, this aspect of the model development was considered to be a conservative assumption.

(5). Use of half detection limit for watershed PCB non-detects: While this is a reasonable approach, and probably somewhat conservative, it is not clear how much present low watershed loads contribute to historical PCB sediment concentrations in the creek mouths. Sediment PCB concentrations can persist for a long time, and it is possible that present concentrations are due mainly to large historical discharges and spills. It would be easy to run a set of sensitivity simulations to assess this implicit MOS. Absent that, my guess is that would provide less than 50% MOS for PCBs, and the real MOS could be much less than that.

Response: The model development is only considering existing concentrations and impacts to the creek mouth area. Sufficient data on historical conditions are not available to assess the relative impact from existing and historical loads. The emphasis here is using half the detection limit to represent the watershed concentration of PCBs, which is a typical conservative approach.

(6). Conservative pollutant degradation kinetics: While this is a reasonable approach, and somewhat conservative, the actual degradation rates for PCBs and PAHs are low anyway.

Biodegradation of PCBs, for example, often convert one congener to another, not reducing the total. Lighter PCBs and PAHs can be lost to the atmosphere through volatilization, and excluding this process can be considered conservative, especially under existing conditions when the net volatile flux appears to be from water to atmosphere. Following remediation, however, it is possible that the net flux will be neutral or even reverse. So for the TMDL, I don’t think that the exclusion of volatilization kinetics contributes much to the MOS.

Response: Yes, we agree that in some cases, the degradation of environmentally persistent pollutants may not be conservative. Sabin et al. (2010) found that San Diego Bay has a notably high net loss to the atmosphere of both low and high molecular weight PAHs. Overall, assuming that pollutants do not degrade over time is a conservative assumption.

d. **Comment**: My judgment is that the first two assumptions provide most of the implicit margin of uncertainty for this TMDL. For most years, the design high rainfall conditions should provide a margin of safety exceeding a factor of 50, and thus, in my judgment, sufficient to account for modeling uncertainties. During the infrequent high rainfall years, however, this assumption provides little or no implicit MOS. For those years, the second assumption, using the highest sediment grid concentration for the TMDL should provide an implicit MOS of a factor of 1.5 to perhaps 10. On the mid to high end of this crude estimate, the sediment grid assumption should be sufficient to account for modeling uncertainties. At the low end, however, this assumption provides a margin of safety on the same order as my estimated model uncertainty. It would take a set of sensitivity analyses to shed more light on these estimates.

Because any potential infractions of the TMDL should be infrequent, I recommend proceeding with the implementation of the TMDL as now envisioned.

Response: The average of the maximum predicted concentration for all grid cells within each impaired shoreline area was used to determine the percent reduction needed to meet the TMDL targets. Use of the maximum concentration simulated for each grid cell incorporates a significant implicit MOS factor. Other implicit MOS factors included using the high flow year for the TMDL calculations. In addition, an explicit MOS was included for these TMDLs to offset uncertainty in the analysis.

8. Implementation of the TMDL.

Comment*:* Overall, the management actions proposed are based on the available evidence and seem reasonable, i.e., remediate the marine sediments and make sure that controllable sources meet the WLA. Keep in mind, however, that uncontrollable sources constitute the majority of the TMDL, and pretty much all of it for PAHs and PCBs.

Response: Comment noted.

9. Overarching Questions.

Comment: Overall, my judgment is that this TMDL is based upon sound scientific knowledge, methods, and practices. Because the uncontrollable sources constitute the majority of the TMDL, it is possible that infrequent high watershed loadings during high rainfall years could lead to episodic violations.

Response: It is not expected that watershed loads will exceed those used to develop the TMDL during most years, as the critical period was defined by an extremely high rainfall period. It is possible that there may be future wet periods that exceed the rainfall total that was reported during the TMDL critical period; however, this occurrence will be very low. Sediment toxicity is primarily related to the build-up of toxic pollutants over a long period of time which can affect sensitive species. The TMDL calculations should be sufficiently conservative to protect against long-term accumulation of these toxic constituents given that they were derived using the critical wet period and other implicit MOS assumptions. In addition, an explicit MOS was used to provide additional assurance that TMDL conditions, if met in the future, will result in the improvements needed.

III. Comments from Professor Ashish Mehta

A. Review Material

The following is a list of the draft technical publications examined by the reviewer and the associated comments:

1. *Technical Report: Total Maximum Daily Loads for toxic pollutants in sediment at San Diego Bay shorelines - Mouths of Paleta Creek, Chollas Creek and Switzer Creek:*

Comment: This report reads well. Only two comments:

a. On P. 60, 6th line from the bottom, change “above the tidal prism” to “upstream of the tidal prism”.

b. Do not use fps units; e.g. p. 69 change depths to meters.

Response: Both recommended changes were incorporated into the report.

2. *Appendix C-1: Watershed monitoring & modeling report (Schiff & Carter 2007)*

Comment: Embedded comments were made to the document.

Response: These comments are summarized and addressed in section III.C.1.a., below.

3. *Appendix C-2: Watershed modeling report (Tetra Tech 2008)*

Comment: Embedded comments were made to the document dated June 30, 2008.

Response: These comments are summarized and addressed in section III.C.1.b., below.

4. *Appendix D: EFDC receiving water modeling report*

Comment: Embedded comments were made to the document dated June 30, 2008..

Response: These comments are summarized and addressed in section III.C.1.c., below.

5. *Appendix E: Watershed modeling for Chollas, Switzer and Paleta Creek watersheds for simulation of loadings to San Diego Bay report (Tetra Tech 2011)*

Comment: Embedded comments were made to the document.

Response: These comments are summarized and addressed in section III.C.1.d., below.

6. *Appendix F: Compilation of sediment, storm water and water quality data summaries for mouths of Paleta, Chollas and Switzer Creeks*

Comment: Embedded comments were made to the document.

Response: These comments are summarized and addressed in section III.C.1.e., below.

7. *Appendix G: Facilities in the watersheds of Paleta, Chollas and Switzer Creeks enrolled in Industrial Storm Water General Permit (Order No. 97-03-DWQ)*

Comment: No comments.

8. ***Chollas Creek storm drain characterization study: Final report***

Comment: No comments.

9. ***Chollas Creek storm drain characterization study: Final report***

Comment: No comments.

10. ***California Water Quality Control Board tentative resolution***

Comment: No comments.

B. General Observations

The following comments are general observations provided by the reviewer.

1. ***Comment***: In the broad sense the four relevant components of this study are: (1) water quality data collection and analysis, (2) watershed sediment load modeling, (3) estuarine hydrodynamic and sediment transport modeling and (4) other supportive modeling and analyses. The focus of my review is mainly related to items 2 and 3, with emphasis on item 3. After going through the entire documentation sent to me I have concluded that in general this is an impressive effort, especially with respect to modeling items 2 and 3. Modeling based analysis strategy is presented well, and with due (significant) revision should make this a worthwhile study. Having said so I must note the following:

   1). Appendix C-1 requires some revision. Appendix D is the weakest link in the overall analysis, mainly because input data are sparse; much of the data collection effort being related to sediments (and contaminants) in the watershed, with the exception of SSC (suspended sediment concentration) time-series collected for the creek mouths. In general inadequate attention has been paid to catalogue sediment transport at the sites. Appendix E appears to have been written hurriedly. However, the required changes are mainly editorial. Sections 5 and 6 in particular must be revised for grammar.

   Response: Comments noted and the San Diego Water Board thanks Professor Mehta for his remarks. Responses to comments on the individual appendices are provided below.

2. ***Comment***: 2). I assume Appendix C-1 will be revised to reflect the modeling correction for Chollas Creek done in Appendix C-2?

Response: No, Appendix C-1 will not be revised. Appendix C-2 will continue to be provided because it documents a correction to the delineation of the reach and watershed boundary of the Switzer Creek model. Appendix C-1 contains watershed monitoring data for three storm events and the base model configuration, which has since been slightly revised, as presented in Appendix E, for the final TMDL calculations.

3. ***Comment***: 3). In Appendix D at numerous places simple statistical measures must be introduced to support subjective inferences such as “good”, “very good” etc. when comparing model output with data.

   Response: Given the lack of extensive data for model calibration, it is unlikely that simple statistical measures would be meaningful. In such cases, visual comparison is used to evaluate model performance. It is the normal practice in water quality modeling to use visual comparison and qualitative judgments to describe the relative accuracy of the model.

4. ***Comment***: 4). The use of fps units, even though an unfortunate fact of U.S. data-collection practice, must be eliminated throughout, or at least appended with corresponding values in SI units. (Total elimination of fps is the best solution). Mixed-unit representations, such as plots in which one axis has fps units and the other axis is in SI (or metric), are not commensurate with reporting scientific analyses.

   Response: The units shown are consistent with other similar technical reports; however, this comment was noted to help guide the development of future reports.

C. Specific Comments

The following specific comments were provided by the reviewer.

1. ***Comment***: 1). I have introduced comments in the text margins and in some cases within the text areas in the accompanying documents (Appendices C-1, C-2, D, E, and F).

   a. **Appendix C-1** – Many comments are related to adding clarifications and correcting grammar.

      Response: While the model configuration and results presented in this report have been updated as documented in the additional modeling reports, the report is still relevant because it documents the monitoring data collected in the 2005-06 wet season and initial watershed model configuration. These suggested corrections will not be made directly to this document. The responses are provided below.

The following specific comments were provided as embedded comments by the reviewer in the report titled, "Monitoring and Modeling of Chollas, Paleta, and Switzer Creeks," dated May 2007:

(1). The sentence that reads, "Rainfall was measured using a standard tipping bucket gauge that measures rainfall in 0.01 increments." (Page 3, 5th paragraph, 2nd sentence)

Comment: Is "0.01 increments" in inches?

Response: The increments are in inches.

(2). The sentence that reads, "Water quality was collected either as pollutographs or flow-weighted composites." (Page 3, 6th paragraph, 1st sentence)

Comment: Include a brief description of what a pollutograph is.

Response: A pollutograph is comprised of multiple grab samples that are analyzed separately for pollutant concentrations and collected across the hydrograph. The first line on page 4 provides this information.

(3). The sentence that reads, "Flow weighted composites were individual samples collected at set storm volume intervals and placed into the same container." (Page 3, 6th paragraph, 2nd sentence)

Comment: Explain "storm volume intervals."

Response: Storm volume interval is a term used for sample pacing using automated equipment to collect flow-weighted composites. The equipment is set to sample for every preset volume of storm discharge. Ideally, if one wants to collect a flow-weighted composite sample using 10 aliquots, you would set the storm volume interval to one-tenth the expected discharge volume of the entire storm.

(4). The sentence that reads, "Therefore, between 10 and 12 individual grab samples were collected per storm event at each site and analyzed separately." (Page 4, top of page, 1st full sentence)

Comment: This is somewhat confusing as grab samples are usually collected from the bottom, whereas the context here appears to be TSS.

Response: Grab samples were collected using the autosampler's pump system whose intake was on or near the channel bottom.

(5). The sentence that reads, "It integrates comprehensive data storage and management capabilities, a dynamic watershed, and a data analysis/post-processing system into a PC-based windows interface." (Page 4, 2nd full paragraph, last sentence)

Comment: In reference to "dynamic," briefly explain how the equations of flow momentum and continuity are modeled in this case.

Response: Dynamic refers to a time-variable model as opposed to static seasonal or annual model algorithms. For greater details on algorithms of this model, please consult EPA's web site (http://water.epa.gov/scitech/datait/models/basins/index.cfm) or see Bicknell et al (2001) from the literature cited section.

(6). The sentence that reads, "The Manning's roughness coefficients varied for each representative reach and ranged between 0.045 and 0.060 based on substrate." (Page 4, 4th full paragraph, last sentence)

Comment: These values are somewhat high. Briefly indicate what the bottom was like.

Response: Much of the Paleta, Chollas, and Switzer creeks watersheds have unlined bottoms composed of sands and gravel.

(7). The sentence that reads, "Insufficient information was currently available or necessary to calibrate flow or water quality parameters for each of the 19 land uses at the present time." (Page 5, 3rd paragraph, 1st sentence)

Comment: Delete "currently".

Response: Comment noted.

(8). The sentence that reads, "Hydrologic parameters examples include interception, infiltration, evaporation, transpiration, groundwater flow, etc." (Page 5, 5th paragraph, 2nd sentence)

Comment: Change "parameters" to "parameter."

Response: Comment noted.

(9). The sentence that reads, "This process was considered uniform regardless of the land use type or season." (Page 6, 2nd bullet, 2nd sentence)

Comment: Do you mean spatially uniform?

Response: Both spatial and temporal uniformity was intended.

(10). The sentence that reads, "Trace metals were bound to a particle during wet-weather washoff until they disassociated upon reaching the receiving waterbody." (Page 6, 5th bullet)

Comment: Based on equilibrium partitioning? Explain.

Response: The emphasis of the assumption is on particle-binding while within the modeling domain. Disassociation occurred in stream during transport to the mouth of the watershed.

(11). The sentence that reads, "For example, correlations coefficients for rainfall volume ranged from 0.64 to 0.80 among the three rain gauges." (Page 7, 1st paragraph, 5th sentence)

Comment: Change "correlations" to "correlation."

Response: Comment noted.

(12). The sentence that reads, "Chollas North had the single greatest peak flow of the season at 378 cfs (February 27, 2006)." (Page 7, 2nd paragraph, 3rd sentence)

Comment: Report value instead in cubic meters per second.

Response: Chollas North had the single greatest peak flow of the season at 10.7 cubic meters per second.

(13). Regarding the word "greatest" in the sentences that read, "No single storm generated the greatest concentrations (Tables 7 through 10). At Chollas North and Paleta Creeks, the greatest concentrations were generally seen in the first storm (February 19) event. In contrast, the greatest concentrations were generally seen in the last storm (March 10) at Chollas South and Switzer Creeks." (Page 7, 4th paragraph, 1st through 3rd sentences)

Comment: "Highest"?

Response: Yes, highest.

(14). The sentence that reads, "Concentrations in this study were similar to concentrations measured by the municipal stormwater NPDES copermittees at the Chollas Creek sites (Figures 5 and 6)." (Page 7, 5th paragraph, 1st sentence)

Comment: Should be "co-permittees."

Response: Comment noted.

(15). The sentence that reads, “Although the concentration ranges overlapped for lead at Chollas South, the current data were skewed towards the lower end of the range compared to historical data.” (Page 8, partial paragraph at top of page, 1st full sentence)

Comment: "Current data" can be confusing in the present study context. Change to "recent."

Response: Comment noted.

(16). The sentence that reads, “In a complete reversal, the historical lead data were skewed towards the lower end of the range compared to the current study.” (Page 8, partial paragraph at top of page, 2nd full sentence)

Comment: Change to "present".

Response: Comment noted.

(17). Regarding the words “storm events” in the sentences that read, “Individual pollutographs indicated a large variability in COPC concentrations during each storm event (Figure 7; Appendix B). In nearly all storm events at all sites, changes in COPC concentrations commonly varied from one to two orders of magnitude.” (Page 8, 1st full paragraph, 1st and 2nd sentences)

Comment: Delete "event" in the first sentence. Change "storm events" to "storms" in the second sentence.

Response: Comment noted.

(18). The sentence that reads, “As a result, cumulative mass distribution curves indicated that first flush during these storm events were moderate (Figure 8).” (Page 8, 1st full paragraph, 5th sentence)

Comment: What might be the reason for a weak first flush in the present case? Provide a possible reason.

Response: The moderate first flush could be due to several reasons including, but not limited to: precipitation intensity, precipitation volume, precipitation location in the watershed, antecedent dry period, and/or upstream mixing.

(19). The sentence that reads, “Simulated flows were compared to observed flows to assess calibration accuracy and precision.” (Page 8, last paragraph, 1st sentence)

Comment: What is the distinction between the terms “accuracy” and “precision” in the present context?

Response: Accuracy refers to how much variability the model could account for observed conditions (see 2nd sentence, same paragraph). Precision refers to any bias associated with the model prediction (see 3rd sentence, same paragraph).

(20). The sentence that reads, “Across all storms in all four watersheds, the model predicted 84% of the variability observed in average storm flows, 76% of the variability observed in storm volume, and 75% of the variability observed in peak flow.” (Page 8, last paragraph, 3rd sentence)

Comment: How "accurate and precise" is this?

Response: This accuracy and precision is comparable to other models developed in southern California, as noted on page 12, 3rd paragraph, last sentence.

(21). The sentence that reads, “This model currently represents the best optimization of hydrologic parameters (i.e., timing, volume, average flow, peak flow) across all four watersheds.” (Page 9, 1st partial paragraph, last sentence)

Comment: Provide some additional information on how it was decided this optimization was "the best". Was it a statistical approach?

Response: The optimization was based on maximum accuracy, minimal bias, while keeping model variables the same across watersheds.

(22). Regarding the word “similar” in the sentences that read, “Simulated storm event mean concentrations (EMCs) for TSS, copper, lead, and zinc were similar to modeled values at Chollas North indicating reasonable accuracy (Figure 12). Simulated EMCs were considered similar to observed EMCs if the 95% confidence intervals from the comparison overlapped one another. In the case of Chollas Creek North, all three of the simulated storms had similar EMCs to observed values for TSS, copper, and lead; two of the three storms had similar modeled and observed zinc EMCs. What’s more, in no case was the simulated consistently greater than, or consistently lesser than, the observed EMC. Similar results were observed for the bar charts for each station, storm and constituent combination (Appendix C).” (Page 9, 2nd full paragraph)

Comment: Use of the word "similar" is a purely subjective assessment. Indicate statistical basis of judging whether model and observed values are "similar".

Response: The basis for similarity was based on overlapping 95% confidence intervals as described in the paper by Robert W. Smith titled, “Visual Hypothesis Testing with Confidence Intervals” (see Attachment 1 to this Appendix).

(23). The sentence that reads, “What’s more, in no case was the simulated consistently greater than, or consistently lesser than, the observed EMC.” (Page 9, 2nd full paragraph, 4th sentence)

Comment: Change to "simulated value."

Response: Comment noted.

(24). The sentence that reads, “The precision for Chollas North averaged 31% relative percent difference, which compared favorably to the data quality objectives for laboratory precision of trace metal analysis (25%).” (Page 10, partial paragraph at top of page, last sentence)

Comment: Provide relevant citation.

Response: The relevant citation is EPA Method 200.8, Project Quality Assurance Project Plan.

(25). The sentence that reads, “The sensitivity of the watershed model to the sediment potency factor was a linear function (Figure 13).” (Page 10, 2nd full paragraph, 1st sentence)

Comment: Provide definition of potency factor.

Response: Potency factor is the conversion factor of TSS concentration to trace metal or PAH concentration based on regressions of empirical data from throughout southern California and verified from the study watershed.

(26). The section heading titled, “Model Results” (Page 10, preceding 3rd full paragraph)

Comment: Explain what the model results are for.

Response: The model results section presents the model results from the watershed modeling analysis.

(27). The sentence that reads, "Nine year simulations demonstrated that this variability was due to large differences in year-to-year loading (Figure 14; Appendix D)." (Page 10, last paragraph, 2nd sentence)

Comment: Would there be any other reason? Explain.

Response: Inter-annual co-variates could be responsible for the observed variation (i.e., precipitation).

(28). The sentence that reads, "Decadal simulations predicted that there was large within year variation in loading (Figure 15; Appendix E)." (Page 11, 1st paragraph, 1st sentence)

Comment: Do you mean nine-year simulations?

Response: The model output was from January 1, 1996 to December 31, 2005.

(29). The sentence that reads, "In general, high density residential areas were typified as highly impervious with large pollutant build-up maxima." (Page 11, 2nd paragraph, 5th sentence)

Comment: Please define "build-up maxima."

Response: High density residential areas were assigned high pollutant build-up maximum values.

(30). The sentence that reads, "Chollas Creek, which was the largest of the three watersheds, generally had the greatest emissions." (Page 12, 1st paragraph, 2nd sentence)

Comment: Change to "loads". I think emission usually means the state is gaseous.

Response: Comment noted.

(31). The sentence that reads, "Assumptions regarding sorbtion to transported sediment appear warranted since TSS significantly correlated to copper, lead, zinc and total PAH all in all four creek systems." (Page 12, 2nd paragraph, 3rd sentence)

Comment: Do you mean "sorption"?

Response: Comment noted.

(32). The sentence that reads, "Interestingly, the variability between modeled and measured volumes from this study approximated the variability observed by Ackerman et al. (2005)." (Page 12, 3rd paragraph, last sentence).

Comment: This seems entirely fortuitous, unless you know of a physical basis for this. I suggest deletion.

Response: The intent of the sentence was to demonstrate the reasonableness of the model calibration and validation based on work elsewhere.

(33). The sentence that reads, "The accuracy and precision of the water quality model developed for Chollas, Switzer and Paleta Creek watersheds was similar to the accuracy and precision of the model developed in the Los Angeles Region (Ackerman and Weisberg 2006)." (Page 12, last paragraph, 1st sentence)

Comment: What is the distinction between the terms "accuracy" and "precision" in the present context?

Response: Please see comment C.1.a.(20), above.

(34). The sentence that reads, "The variability in modeled EMC estimates were not significantly different than measured estimates over 70% of the time." (Page 12, last paragraph, 2nd sentence)

Comment: Change to "was".

Response: Comment noted.

(35). The sentence that reads, "Ultimately, however, the only way to truly assess if the extrapolation of Los Angeles calibration terms was appropriate would be to sample additional land use sites in San Diego." (Page 13, 1st full paragraph, last sentence)

Comment: It would seem that ultimately only (not just additional) data based on measurements in San Diego would be used.

Response: At the time that this report was written, it was not likely that site-specific data would be collected until after implementation of adopted TMDLs. Fortunately, site-specific data was collected and used to re-run the watershed modeling, presented in Appendix E of the Technical Report, which replaces the analysis presented in this report.

(36). The sentence that reads, "Dynamically modeled total PAH concentrations were attempted, but the inaccuracy and bias between modeled and measured concentrations was too large." (Page 13, 2nd full paragraph, 3rd sentence)

Comment: Change to "were".

Response: Comment noted.

(37). The sentences that read, "We assume this was due to a strong first flush in total PAH observed at the start of most storm events that was not linked to either TSS or land uses. No attempt was made to model chlorinated hydrocarbons because contributions of these compounds were not based on land use. Instead, compounds such as PCBs, chlordane, and others are a result of specific locations in the watershed where these legacy constituents were used. Empirical estimates of organic constituents could be improved with additional sample events. Using the estimates of variance from the three storm events captured during this study, power analysis could be used to determine the approximate number of storm events needed to estimate average concentration with a known level of confidence." (Page 13, 2nd full paragraph, 4th through 8th sentences)

Comment: Change use of the term "storm events" to "storms."

Response: Comment noted.

(38). The sentence that reads, "Alternative modeling approaches could also be attempted by examining other potential covariates besides TSS such as total organic carbon." (Page 13, 2nd full paragraph, last sentence)

Comment: What sorts of differences would you expect in the outcome?

Response: Total organic carbon may provide a better covariate than TSS, thereby improving the accuracy and reducing the bias in modeling organic pollutants.

(39). The sentences that read, "Regardless of what scenarios could be selected, it appears that BMPs focusing on capturing particle would be helpful at reducing total loads. Since trace metals and total PAHs were significantly correlated to TSS, BMPs that focus on removing TSS would necessarily reduce these COPCs. Design of TSS-reducing BMP's should explore unknown variables in TSS delivery of COPCs including partitioning to various particle size fractions." (Page 14, last 3 sentences)

Comment: However this inference does not require the model.

Response: Comment noted.

(40). Regarding the use of the phrase “Current Study” in the legend of Figure 5 and referencing all other occurrences. (Page 19)

Comment: Here and elsewhere change "current study" to "present study."

Response: Comment noted.

(41). Regarding graphs in Figure 7. (Page 20)

Comment: Flow should be in cubic meters per second.

Response: Comment noted.

(42). Regarding graphs in Figure 9. (Page 22)

Comment: Concentration units are missing.

Response: Comment noted.

(43). Regarding graphs in Figure 11. (Page 24)

Comment: Flow should be in cubic meters per second.

Response: Comment noted.

(44). Regarding the table endnotes on Tables A-2 and A-4 in Appendix A. (Pages A-4 and A-8)

Comment: Eliminate the use of fps units and use SI system only. Change inch to meter.

Response: Comment noted.

(45). Regarding the table endnotes on Table A-3 in Appendix A. (Pages A-6)

Comment: Use the Celsius unit only.

Response: Comment noted.

(46). Regarding the table header row on Table A-6 in Appendix A. (Page A-10)

Comment: Use kg per metric ton instead.

Response: Comment noted.

(47). Regarding the flow units on all graphs in Appendix B. (Pages B-2 through B-17)

Comment: Plot discharge in SI units (m^3/s) only.

Response: Comment noted.

b. Appendix C-2 –

The following specific comments were provided as embedded comments by the reviewer in the report titled, "Watershed Modeling for Simulation of Loadings to San Diego Bay," dated June 30, 2008:

(1). The sentence that reads, "The transport of metals and organic pollutants during wet-weather events is generally believed to be associated with the detachment and transport of sediment (Buffleben et al., 2002; CALTRANS, 2003; Hoffman et al., 1982; Lau and Stenstrom, 2005; Logonathan, et al. 1997; Stein et al., 2005; Yunker et al. 2002)." (Page 4, 1st paragraph, 1st sentence)

Comment: Change to "Loganathan."

Response: This citation correction was made in the referenced sentence and in the "References" section for the document.

(2). The sentence that reads, "These data are used to calculate hourly potential evapotranspiration, which can be incorporated into the modeling process." (Page 6, 1st paragraph, 7th sentence)

Comment: If evapotranspiration is calculated each hour and used in the calculation of runoff, why is it "potential?"

Response: The actual amount of evapotranspiration is calculated internally in the model. The time series of potential evapotranspiration details the amount of water that could be lost. For example, in the dry, hot summer, a great deal of water could be lost through evapotranspiration but little is because the soils don't hold the full amount of water that could be lost.

(3). The sentence that reads, "The Manning's roughness coefficients varied for each representative reach and ranged between 0.045 – 0.060." (Page 9, 2nd paragraph, last sentence)

Comment: It would be useful to provide justification for the use of these values. What was the bottom like?

Response: Mannings roughness coefficients were varied based on channel type and, where available, monitored hydrograph data.

(4). The sentence that reads, "This process was considered uniform regardless of the land use type or season." (Page 10, 2nd bullet, 2nd senentence)

Comment: Consider inserting "over the land area of interest" to describe what is "uniform."

Response: This recommendation was incorporated into the document.

(5). The sentence that reads, "Sediment in the watershed consisted of 5% sand, 40% clay, and 55% silt." (Page 10, 3rd bullet)

Comment: Based on what information? Include citation.

Response: Sediment fractionation was based on best professional judgment and similar studies in regional watersheds.

(6). The sentence that reads, "Use of flow-weighted mean concentrations assumes no variability in storm concentrations, first flush, and indication of sediment association." (Page 11, last sentence of partial bullet at top of page)

Comment: Do you mean "or"?

Response: Comment is unclear; however, the flow-weighted event mean concentrations incorporated the varying pollutograph concentration and flow data throughout the sampled events.

(7). The sentence that reads, "The sediment-associated land uses relied upon the sediment parameters calibrated for the Los Angeles Region." (Page 25, 2nd paragraph, 2nd sentence)

Comment: This is a drawback.

Response: The sediment parameters from a nearby geographic region were the best available information for use at the time that the model was being configured. The San Diego Water Board agrees with the reviewer's comment. Fortunately, the City of San Diego was able to collect site-specific data for the three watersheds, making it possible to re-configure the models using site-specific data (see Appendix E).

c. Appendix D –

The following specific comments were provided as embedded comments by the reviewer in the report titled, "Receiving Water Model Configuration and Evaluation for the San Diego Bay Toxic Pollutants TMDLs," dated June 30, 2008:

(1). The sentences that read, "The Loading System Program in C++ (LSPC) was selected to simulate the watershed loadings (Shen et al. 2004; USEPA 2003a). The mouths of the five watersheds (estuaries) and the bay were represented by the Environmental Fluid Dynamics Code (EFDC) (Hamrick 1992). The EFDC hydrodynamic model incorporates flow and loading from the watershed models (see watershed modeling reports) and subsequently determines their impact on the five impaired shorelines as the pollutants are transported through the bay." (Page 2, 1st partial paragraph, 2nd and 3rd sentences)

Comment: This presupposes (?) that waves are not important at any time in the study area. If so please provide justification based on wave data.

Response: Waves may or may not be significant in the study area; however, wave/current data were not available at the time of the study. The model was developed considering various data limitations (e.g. boundary conditions, bed sediment characteristics, toxics sources, etc). Wave impact can be one source of uncertainty in the model. TMDL development, however, was based on a long-term simulation of the Bay (3 year period), and therefore, wave impacts, which are generally event-based, might not be a significant factor on such a long time scale.

(2). The sentence that reads, "Representation of sediment and adsorptive contaminant transport. Ideally multiple classes of sediment size and species of toxics need to be represented in a single analytical framework to efficiently address the differential settling velocity of different classes of sediment as well as their different capability of adsorbing toxics." (Page 4, 2nd bullet)

Comment: "Differential settling velocity" has a specific meaning in cohesive sediment transport; it means the velocity of one falling particle (floc) relative to another. Change to "different settling velocities".

Response: This recommendation has been incorporated into the report.

(3). The sentence that reads, "Transport within the bed should include pore water diffusion of the dissolved phase and mixing and burial of the particulate phase." (Page 5, 1st partial paragraph, 1st full sentence)

Comment: Indicate how this has been addressed in your study.

Response: These methods are referenced in the EFDC sediment transport/toxic manual (available at http://www.epa.gov/ceampubl/swater/efdc/index.html).

(4). The sentence that reads, "Particulate organic material is assumed to be associated with the fine sediment class." (Page 15, 2nd paragraph, 3rd sentence)

Comment: Biogenic organic matter reduces the overall particle density. What was the organic fraction and how did you account for it?

Response: For this study, organic matter was not considered in the sediment-toxic interaction due to the lack of available data.

(5). The sentence that reads, "The computational grids of the four local models were developed based on the bathymetry of the Navy's CH3D grid (Figure 4-3)." (Page 15, last paragraph, 1st sentence)

Comment: Only three are shown in Fig. 4-3.

Response: There is a separate model for each of the impaired shoreline areas (Chollas, Paleta, and Switzer). The fourth model includes the impaired shorelines for B Street/Broadway Piers and Downtown Anchorage, which are part of a separate TMDL study.

(6). The sentence that reads, "Are the Paleta Creek grids bigger than rest of the bay grids because it is more time consuming to run the water/sediment quality component of the model as opposed to the hydrodynamic component of the model?" (Page 17, sentence at top of page)

Comment: Where did this comment come from? Reviewer question?

Response: The sentence appears to be a comment that was made by the modeling staff on the internal draft version of the report. This sentence has been deleted from the document.

(7). ***Comment***: For the reader it would be good to include plots showing the shorelines of the creek mouth areas enlarged. Terms such as creek mouth mentioned later in reference to TMDL need to be identified." (Page 20, at Figure 4-7)

Response: The San Diego Water Board agrees that it would be helpful; however, this report and the other modeling reports are supporting documents to the project. Terms related to and identifying the creek mouth are identified and discussed in the Technical Report.

(8). The sentence that reads, "Each of the sediment transport models was configured to simulate two cohesive sediment classes: clay (with a diameter < 3.9 micrometers) and silt (with a diameter > 3.9 micrometers and < 63 micrometers); and one non-cohesive sediment class: sand (with a diameter > 63 micrometers)." (Page 20, 1st paragraph, 1st sentence)

Comment: 1) Cite or include plot showing the size distribution used to derive these classes.

2) Is the fine sediment component flocculated? If so what is the floc size distribution?

Response: In response to the first comment, Chadwick et al. (2008) was used as the basis to derive the distribution of sediment classes. In response to the second comment, flocculation is not considered in the modeling since there are no data available to support including this component.

(9). The sentence that reads, "This allowed the model to represent up to 1.2 meters of active bed, which was deemed sufficient for representing the bed dynamics in San Diego Bay." (Page 20, 1st paragraph, last sentence)

Comment: Define "active bed." Would you expect 1.2 m to be eroded, ever?

At this point it is important to briefly describe what the bed is like: 1) what are the clay and non-clay minerals? 2) What percent is organic? 3) what is the density structure? 4) Is fluid mud present? 5) What density value did you use to characterize the bed?

Response: The term "active bed" was used simply to mean that the depth of bed was configured as active grid layers in the sediment transport model. The value was assumed to be large enough and, in the 3 year scenario runs, the bed was never completely eroded. We agree that to quantify the detailed characteristics of the bed, percent organic matter and vertical density structure would be useful; however, no data were available to further define these detailed structures/physical properties.

(10). The sentence that reads, "No water column data were available for the other contaminants, therefore, they were derived using TPAH concentrations and the ratio of concentrations between TPAH and each contaminant in the bed." (Page 21, 2nd full paragraph, 5th sentence)

Comment: If one has to assume that the ratios are fixed, how would that constrain the dynamic simulations of the constituents relative to each other? Indicate if this is reasonable because the background concentrations are low compared to watershed inputs.

Response: Data were not available to specify the concentration of these contaminants; therefore, the method discussed in the report was used to define these concentrations.

(11). The sentence that reads, "The flows and toxic contaminant concentrations were directly extracted; however, additional effort was required to divide the TSS concentration among the three modeled sediment classes in EFDC (i.e., clay, silt, and sand)." (Page 22, 1st paragraph, 3rd sentence)

Comment: By "effort' do you mean treatment based on additional assumptions?

Response: The additional effort refers to the use of the sediment ratios that were derived from Chadwick et al (2008) to split the sediment into clay, silt, and sand fractions.

(12). The sentence that reads, "These ratios were used as the basis for TSS division." (Page 22, 1st paragraph, 5th sentence)

Comment: Were bed sediment ratios available for these creeks? How did they compare with those based on TSS?

Response: At the time of model development bed sediment ratios were not available for these creeks.

(13). The sentence that reads, "The initial sediment concentration in the water column was set to 1.0 mg/L for each of the three classes." (Page 22, 3rd paragraph, 1st sentence)

Comment: Is there no background level of suspended matter in the bay? If so what is it? 1 mg/L seems low(?)

Response: This is an assumed value. Since initial condition usually is washed away very quickly, the impact to the overall model result is negligible. The modeling period begins with a dry period, therefore, a very low sediment concentration was assumed for initial condition.

(14). The sentence that reads, “Initial bed sediment compositions for the Paleta, Chollas, and Downtown/B-Street models were specified based on data reported in SCCWRP and SPAWAR (2005).” (Page 22, last sentence on page)

Comment: Indicate exactly what is meant by initial bed sediment composition.

Response: Initial bed sediment composition means the fractionation of silt, clay, and sand in the sediment bed at the beginning of the model run.

(15). The paragraph that reads as follows:

“These data were collected in 2001, which corresponds to the modeling period selected. Since no data were available to set the initial bed composition at the Switzer Creek mouth for 2001, the data for 2003 collected by the CRG Marine Lab were used. Initial bed toxic concentrations for the Paleta and Chollas models were specified based on data reported in SCCWRP and SPAWAR (2005). Initial conditions for the Switzer and Downtown/B-Street models were specified based on data reported in Anderson et al. (2004, 2005). Since data were available at multiple locations at the mouths of the creeks, the initial bed toxic conditions were specified on a spatially-variable basis. Where data were available, the values were directly applied. Where data were not available, conditions were set using the minimum values of the data available at the hot spots. The initial bed condition for these cells does not have a significant impact on the simulation results, because these locations are generally outside the area of the incoming tributary mouths and are generally deep. Therefore, resuspension is not expected to occur and contribute significantly to re-distribution of toxics among cells.” (Page 23)

Comment: Still, please summarize what specific values were assigned for each location.

Response: This information is better viewed in the EFDC model input file. The model input file is available for review as needed.

(16). The sentence that reads, “The hydrodynamic model of the bay was calibrated using observed surface elevation data from the bay.” (Page 24, 2nd paragraph, 1st sentence)

Comment: This suggests that current data were not used. Since calibration tends to be more sensitive to current than water level, and since sediment is transported by current, the absence of current-based calibration may be questioned.

Response: The hydrodynamic model was calibrated using both surface elevation and salinity data. This is a widely accepted calibration approach for hydrodynamic models. A model that is capable of reproducing observed surface elevation, as well as, salinity at the freshwater-saltwater interface area is generally considered well calibrated. Current data were not available during model development.

(17). The sentence that reads, “As shown, the simulated elevation matches the observed elevation very well, indicating a reasonable representation of tidally-influenced water movement in the bay.” (Page 24, 4th paragraph, last sentence)

Comment: The term "very well" is vague; please introduce error statistics.

Response: See comment III.B.3., above.

(18). The sentence that reads, “No additional parameter adjustment was made for this simulation, and the resulting surface elevation was again plotted against the observed data at the same location for three additional, randomly chosen months.” (Page 24, 5th paragraph, 2nd sentence)

Comment: Change "made" to "required".

Response: The recommended change was incorporated into the report.

(19). The sentence that reads, “The simulated elevation matches the observed data very well (see Figures A-3 and A-4 in Appendix A), which suggests that calibration of the hydrodynamic model is reliable for periods beyond the calibration period.” (Page 25, paragraph/sentence at top of page.

Comment: Include error statistics.

Response: See comment III.B.3., above.

(20). The paragraph that reads as follows for section 5.1.2 Freshwater-Saltwater Interaction at the Watershed Mouths:

“In addition to the calibration and validation for water surface elevation, model performance was also evaluated by checking the ability of the model to predict local freshwater-saltwater interaction at the mouths of the watersheds. In February 2001, the Navy conducted a survey before and during a storm event, and multiple trackline data were collected at the mouths of Chollas and Paleta Creeks and described in Chadwick et al. (2008). Figure 5-1 presents the locations of the trackline data collection sites. Since the trackline data were collected at different times and locations along the tracklines, the data are not suitable for conducting a time series type of model evaluation. Therefore, an approach that compares general statistics for model-simulated results against those for observed data along the tracklines was adopted. For this effort, five statistical measures were used to compare the model results against the data: minimum, 25th percentile, median, 75th percentile, and maximum." (Page 25, 2nd paragraph)

Comment: If current data are available for any location within the modeled domain during the study period, or from another period, a similar analysis should be carried out.

Response: See comment III.C.16, above. Current data was not available at the time of the analysis in this report.

(21). The sentence that reads, "As shown, the model-simulated salinity at the mouth of Chollas Creek matches the observed temporal-spatial distribution well." (Page 25, 3rd paragraph, last sentence)

Comment: There needs to be quantification of comparison based on statistics.

Response: See comment III.B.3., above.

(22). The sentence that reads, "The model results match the data at these temporal-spatial locations as well." (Page 25, 4th paragraph, 2nd sentence)

Comment: Please provide statistics.

Response: See comment III.B.3., above.

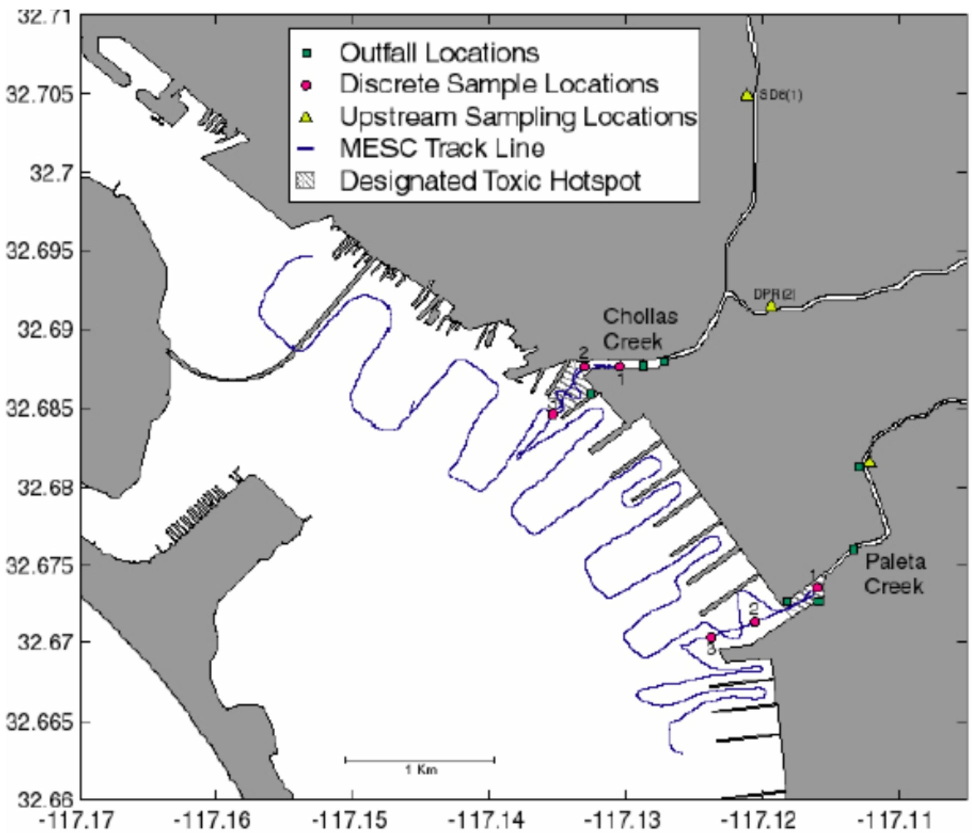
(23). The sentence that reads, "Although there are some deviations between the observed data and model results, these discrepancies can largely be explained by the fact that the hydrodynamic model is driven by inputs from the LSPC watershed model, which is not exactly the same as the real values at the inflow locations." (Page 25, 4th paragraph, 4th sentence)

Comment: What is meant by "some" in "there are some deviations?"

What are the differences between the inputs from the LSPC watershed model and the real values at the inflow locations?

Response: The term "some deviations" was used because a judgment on the number of deviations between observed and model results is dependent on individual interpretation. Differences between the LSPC model results and observed values are shown in the watershed modeling report. No model can match reality exactly; therefore, some differences between model results and observed values are expected.

(24). In reference to Figure 5-1, Location of the trackline data collection sites:



Comment: How is this information related to the tracklines and stations mentioned in the output plots given in the appendices?

Response: The tracklines in Figure 5-1 cover both the Paleta Creek and Chollas Creek watersheds. The Paleta Creek model results were compared to the data that were collected at the stations that correspond with the Paleta Creek section of the trackline. The same was done for the Chollas Creek section.

(25). The sentence that reads, "The same data-model comparison approach used for the hydrodynamic calibration was adopted to evaluate the sediment transport model." (Page 26, 1st paragraph, last sentence)

Comment: A statistics-based approach is required.

Response: See comment III.B.3., above.

(26). The sentence that reads, "In addition, the open boundary sediment concentration was adjusted during the calibration process to obtain reasonable estimates of the background conditions." (Page 26, 2nd paragraph, last sentence)

Comment: What do you mean? Background TSS?

Response: It was assumed that the outer bay TSS concentration is not significantly impacted by the contribution from the creeks; therefore, a background TSS concentration was used to represent the Bay contribution.

(27). The sentence that reads, "For this study, the settling velocities for clay and silt were set to be the same as in Chadwick et al. (2008): 0.048 m/day for clay and 9.5 m/day for silt." (Page 27, 1st paragraph, 1st sentence)

Comment: Please indicate on what physical basis did Chadwick et al. arrive at these values of the settling velocities. This needs to tie in with the state of flocculation of the suspended sediment.

Settling velocity should be reported in m/s.

Response: Chadwick et al. used settling velocities from Stoke's Law, which were also used in this study. Flocculation was not considered in the evaluation. No data was available to address the issue of flocculation. Note that settling velocity rates are reported in m/day and m/s in the literature.

(28). The sentence that reads, "The settling velocity for sand is internally calculated based on the assumed median size of 75 micrometers." (Page 27, 1st paragraph, 2nd sentence)

Comment: What is this assumption based on? No size distributions were available?

Response: This assumption was used for sand modeling and is within the range reported in literature. No data were available to represent particle size distributions in the bay and the creek mouth areas, which can change significantly over time.

(29). The sentence that reads, "This is consistent with the value used in Chadwick et al. (2008)." (Page 27, 1st paragraph, 3rd sentence)

Comment: Consistent but not the same? What equation is used in EFDC?

Response: The settling velocity formula in EFDC is based on Rijn's 1984 formula.

(30). The sentence that reads, "Hwang and Mehta (1989) reported that the critical shear stress for cohesive sediment resuspension varied from 0.125 N/m2 to 0.525 N/m2." (Page 27, 2nd paragraph, 2nd sentence)

Comment: In the SI system, stress (tau) has the units of Pa. Change from N/m2 to Pa. The method of selection of the critical stress tau_c is perfunctory (given the lack of data). In any event, their relevance depends on $(tau - tau_c)/tau$. What were the (maximum or mean) tau values?

Response: While it would be nice to provide the maximum and/or mean tau values, there are limited resources to do so and the current analysis provides the necessary information for the study. This comment is valuable input for future studies. The units will not be converted to Pa.

(31). The sentence that reads, "The critical bottom shear stress for deposition was set to the value of 0.1 N/m2 based on the ratio applied in the previous study (Ji et al. 2002)." Page 27, 2nd paragraph, last sentence)

Comment: Please clearly state what you mean by "based on the ratio...."

Response: This is a typo. It should read, "… based on the rate applied in the previous study," which means use the value in the literature. This correction has been made to the document.

(32). The sentence that reads, "The final values were 0.001 mg/L, which indicates that the background water in the San Diego Bay is very clear and almost free of suspended solids." (Page 27, 3rd paragraph, 3rd sentence)

Comment: Does this mean that underwater visibility is high?

Response: Not necessarily. Besides TSS, underwater visibility can be impacted by other factors such as organic matter, color, algae, etc.

(33). The paragraph that reads as follows:

"In general, the model results reproduce the observations. Timing associated with the sediment plume's entrance and dissipation during and after the storm, is accurately represented. The model predicts the range of observed TSS concentrations well, i.e., the minimum and maximum, and reasonably simulates the 25th percentile and median values. Prediction of the 75th percentile level is not as accurate as for the other percentiles." (Page 27, 5th paragraph)

Comment: Please provide supportive statistics.

Response: See comment III.B.3., above.

(34). The sentence that reads, "In reality, sediment exhibits much less homogeneity and more variable behavior." (Page 27, last paragraph, last sentence)

Comment: What evidence is there for this assertion? Any published data?

Response: It is generally known that sediment size classifications vary widely in reality; however, the model is only capable of representing three distinct size classes.

(35). Table 5-1, Estimated bed partitioning coefficients used in the TMDL models (Page 28)

	Clay (L/mg)	Silt (L/mg)	Sand (L/mg)
TPAH	0.0011	0.0011	0.0001
TPCB	0.0019	0.0012	0.0001
Zinc	0.024	0.024	0.01

Comment: Would it make a difference if sand values were assumed zero?

Response: Model sensitivity analyses would need to be run to determine if there would be a significant difference.

(36). The paragraph as follows:

Any idea why we are seeing PAHs on sand? Maybe the car repair shops (or home owners) have just dumped oil behind their shops (houses) for years (not all of it), and it's soaked into the sand really well, and coated the sand. Remember the Navy finding the solid sand-sized lead particles in Chollas sediments? Another case of some small businesses doing something similar – maybe they are just sweeping filings out their back door for the last 25 years.

Comment: Leftover comment?

Response: This paragraph appears to be comments that were made by the report authors on the internal draft version of the report. This text has been deleted from the document.

(37). Table 5-2, Estimated water column partitioning coefficients used in the TMDL models. (Page 29)

	Clay (L/mg)	Silt (L/mg)	Sand (L/mg)
TPAH	0.0543	0.0593	0.0223
TCHLOR	0.0033	0.0033	0.0001
Zinc	0.024	0.024	0.01
PCB	0.0033	0.0033	0.0001
Lindane	0.0033	0.0033	0.0001

Comment: Include the definition of partitioning coefficient as a foot-note.

Response: The partition coefficient is a number that quantifies the tendency of the contaminant to associate with solid matter. A large partition coefficient indicates the contaminant has a greater association with solid matter.

(38). The sentence that reads, “Sediment bed concentration data were found to vary significantly even for cores collected from locations in close proximity to one another.” (Page 30, 2nd paragraph, 3rd sentence)

Comment: Delete "close."

Response: The recommended change was incorporated into the report.

(39). The sentence that reads, “This is because the data in bed sediment are localized and can vary significantly between closely adjacent locations.” (Page 30, 2nd paragraph, 9th sentence)

Comment: Delete "closely".

Response: The recommended change was incorporated into the report.

(40). Table 6-1, Annual loading comparison at the mouths of the five watersheds for 2001-2006 (Page 31 and 32)

19001	***Chollas Creek***				
Hydrologic Year	**October 2001-September 2002**	**October 2002-September 2003**	**October 2003-September 2004**	**October 2004-September 2005**	**October 2005-September 2006**

Flow (m3)	681,238	4,859,085	2,101,106	15,896,657	3,032,667
TSS (kg)	62,300	1,555,357	318,297	20,846,674	519,669
Cu (g)	9,565	253,311	51,677	2,951,606	84,123
Pb (g)	6,895	208,934	37,803	2,045,280	60,863
Zn (g)	81,565	2,076,595	437,861	19,122,090	709,279
PAH (mg)	443,537	11,073,242	2,266,094	148,416,293	3,699,743
PCB (mg)	269	6,719	1,375	90,054	2,245
Chlordane (mg)	11,915	297,469	60,876	3,987,021	99,389
Lindane (mg)	269	6,719	1,375	90,054	2,245

19040	***Switzer Creek***				
Hydrologic Year	**October 2001-September 2002**	**October 2002-September 2003**	**October 2003-September 2004**	**October 2004-September 2005**	**October 2005-September 2006**
Flow (m3)	265,715	1,370,857	624,526	4,647,085	863,561
TSS (kg)	27,489	483,422	83,035	8,713,469	117,226
Cu (g)	6,819	128,042	21,617	2,047,204	30,944
Pb (g)	5,253	110,465	17,255	1,395,397	24,532
Zn (g)	66,348	1,165,674	205,186	13,812,521	292,428
PAH (mg)	40,311	708,923	121,769	12,778,005	171,908
PCB (mg)	38	662	114	11,926	160
Chlordane (mg)	3,557	62,555	10,745	1,127,527	15,169
Lindane (mg)	38	662	114	11,926	160

19042	***Paleta Creek***				
Hydrologic Year	**October 2001-September 2002**	**October 2002-September 2003**	**October 2003-September 2004**	**October 2004-September 2005**	**October 2005-September 2006**
Flow (m3)	180,227	650,505	315,446	2,622,822	463,797
TSS (kg)	16,736	151,033	39,163	7,697,295	56,125
Cu (g)	3,415	33,922	8,369	1,752,745	12,154
Pb (g)	2,202	22,520	5,376	814,764	7,809
Zn (g)	25,467	247,889	62,392	6,674,765	89,981
PAH (mg)	85,826	774,533	200,835	39,473,546	287,823
PCB (mg)	50	455	118	23,171	169
Chlordane (mg)	4,080	36,817	9,547	1,876,361	13,682
Lindane (mg)	50	455	118	23,171	169

19044	***B Street/Broadway Pier***				
Hydrologic Year	**October 2001-September 2002**	**October 2002-September 2003**	**October 2003-September 2004**	**October 2004-September 2005**	**October 2005-September 2006**
Flow (m3)	158,825	619,009	306,167	2,148,606	336,770
TSS (kg)	18,435	159,339	43,063	4,866,390	44,910
Cu (g)	4,224	40,203	10,058	904,042	10,490
Pb (g)	3,422	33,692	8,127	649,293	8,504

Zn (g)	45,821	418,672	109,150	7,371,398	113,845
PAH (mg)	27,034	233,666	63,151	7,136,395	65,859
PCB (mg)	25	218	59	6,661	61
Chlordane (mg)	2,385	20,619	5,572	629,713	5,811
Lindane (mg)	25	218	59	6,661	61

19046	***Downtown Anchorage***				
Hydrologic Year	**October 2001-September 2002**	**October 2002-September 2003**	**October 2003-September 2004**	**October 2004-September 2005**	**October 2005-September 2006**
Flow (m3)	87,693	301,341	153,426	1,111,244	179,217
TSS (kg)	7,501	66,759	17,432	2,944,081	18,146
Cu (g)	2,293	22,807	5,517	810,949	5,641
Pb (g)	1,870	19,241	4,508	505,964	4,596
Zn (g)	20,063	193,576	48,123	4,597,110	49,413
PAH (mg)	11,000	97,900	25,564	4,317,395	26,610
PCB (mg)	10	91	24	4,030	25
Chlordane (mg)	971	8,639	2,256	380,966	2,348
Lindane (mg)	10	91	24	4,030	25

Comment: Should these numbers be rounded off? For example, 1.8×10^5 rather than 180,227. To what degree of accuracy is there confidence in the values?

Response: While a high degree of accuracy is not assumed, these are the values mathematically calculated by the model.

(41). The sentence that reads, "Figure 6-2 shows the time series of surface bed layer TPAH results at three randomly selected sites in the Paleta model." (Page 34, 2nd paragraph, 1st sentence)

Comment: For the three selected sites it would seem that their locations would dictate the nature of the time-series in Fig. 6-2. Please describe the relationship between site location and the respective output in Fig. 6-2.

Response: While it would be nice to investigate the relations between the site locations and the model output, additional model runs would be needed in order to provide this comparison. The three sites were selected randomly and provided enough information to assess spatial variability in the model results.

(42). The sentence that reads, "The watershed loading results used to formulate lateral boundary conditions for the EFDC models were based on a combination of LSPC model predicted flow and TSS concentrations and assumptions based on mean organics and TSS concentrations observed within Chollas, Paleta, and Switzer Creeks, as reported in the Watershed Modeling for Simulation of Loadings to San Diego Bay (Watershed Modeling Report), prepared by Tetra Tech, Inc. (2008)." (Page 35, 1st paragraph, 1st sentence)

Comment: Details on this should be summarized earlier in this report (where sediment is described). I do not believe you mean petroleum products(?)

Response: This sentence was moved to section 4.2.3.2 to supplement the description of the lateral boundary conditions for the sediment transport and toxics models. The use of the term, "mean organics," refers to mean concentration values of the organic constituents being evaluated in the study – PAHs, PCBs, chlordane, and lindane.

(43). The sentence that reads, "The relative sensitivity depends on the absolute magnitude of the concentration." (Page 35, 2nd paragraph, 2nd sentence)

Comment: It would help to define and provide illustrative values. Presently the comparison is entirely qualitative.

Response: While it would be nice to provide illustrative values, the model would need to be re-run in order to produce the values. The comparison is intended to describe the general trend observed in the evaluation of sensitivity.

(44). The sentence that reads, "Note the surface bed is defined as the top two layers, which consist of the fixed thickness layer of 20 centimeters below the top layer, and the top layer with variable thickness." (Page 35, last paragraph on page, 1st sentence)

Comment: What are the densities of these two layers?

Response: The bulk density of all the layers changes over time. The model would need to be re-run to obtain this information.

(45). The sentence that reads, "Therefore, even though the sediment toxicity might reach the target at this layer, it is not guaranteed that sufficient depth of compliance is achieved." (Page 36, 1st partial paragraph, last sentence)

Comment: I take it that two layers would "guarantee" compliance? On what basis has this been decided? What role does diffusion due to bioturbation play in this region?

Response: The decision was made to use two layers for TMDL development. Sediment was considered to be “clean” if the top 20+ centimeters of sediment were in compliance with the TMDL targets. Since the TMDL model considers vertical diffusion, the assumption is reasonably conservative. No special consideration was made for bioturbation, which was not quantifiable based on available data at the time of analysis.

(46). The sentences that read, “TPCB concentrations also reduce over time; however, the slope of decrease appears to be more level. Chlordane shows a significant trend of increase with time because the incoming stormwater has high concentrations of chlordane under the baseline condition. In all cases, the bed toxicity shows detectable response to watershed loading.” (Page 36, 1st full paragraph, 3rd and 4th sentences)

Comment: This is all quite vague: "more level", "significant tend of increase", "high concentrations", "detectable response". These descriptions need to be quantitative.

Response: See comment III.B.3., above. Qualitative description is often necessary to help assess quantitative modeling results.

(47). The sentences that read, “The results show that under baseline loading conditions, the TPAH and TPCB concentrations decrease sharply over time during the first three years. The decrease of TPAH and TPCB shows a step-shape pattern, which is caused by the addition of bed layer at the time point when sufficient sediment is deposited to the bed. (Page 36, 2nd full paragraph, 2nd and 3rd sentences)

Comment: Delete "point."

Response: The recommended change was incorporated into the report.

(48). The sentences that read, “Unlike TPAH and TPCB, the bed chlordane level shows a trend of increase due to higher incoming concentrations. The lindane concentrations seem pretty stable over the three years of simulation, indicating that the incoming loading has minimum impact on the bed toxicity of lindane.” (Page 36, 2nd full paragraph, last 2 sentences of paragraph)

Comment: More vague statements: “trend of increase due to higher incoming concentrations,” “seem pretty stable,” and “has minimum impact.”

Response: See comment III.B.3., above. Qualitative description is often necessary to help assess quantitative modeling results.

(49). The sentences that read, "The TPAH concentration in the bed sediment shows a slight upward trend during the first three years, while TPCB tends to decrease with time. As with Paleta Creek and Switzer Creek, the chlordane levels tend to significantly increase over time due to the high concentration inflows." (Page 36, 3rd full paragraph, 2nd and 3rd sentences)

Comment: All vague, undefined: "slight upward trend," "tends to decrease," significantly increase," and high concentration inflows."

Response: See comment III.B.3., above. Qualitative description is often necessary to help assess quantitative modeling results.

(50). The sentence that reads, "As shown in the temporal response examples in Section 6.6, the sediment toxicity in the impaired shoreline areas tend to have a detectable relationship to the watershed loading." (Page 36, 4th full paragraph, 1st sentence)

Comment: How is this defined?

Response: Section 6.6 clearly shows that the sediment toxics concentration is related to watershed loading; therefore, the report states that bed sediment toxicity has a detectable relationship to the watershed loading.

(51). The sentence that reads, "Under existing conditions, chlordane has very low concentrations in the bed at Downtown Anchorage, therefore, the TMDL allocation for watershed loading can include a specification of a suspended solids concentration equal to the numeric target for bed sediment concentration, and the bed sediment toxicity will not further worsen." (Page 37, 1st paragraph)

Highlight emphasis appears on the words "very low concentrations" with no embedded comment. Prof. Mehta was contacted to clarify whether he had a comment. His reply was that the highlight was made for his own reference and should have been deleted. There is no comment.

(52). The sentence that reads, "Based on scenario 2, the watershed suspended solids organics concentrations will be increased to levels between existing and numeric target values for bed sediments." (Page 37, last paragraph, 2nd sentence)

Comment: Please explain what this means. There is little discussion on the role of organics in this report. This statement seemingly appears out of nowhere. Do you mean petroleum products?

Response: For the purpose of this report, organics refers to the TMDL pollutants being evaluated in the study. The organic constituents evaluated were PAHs, PCBs, Chlordane, and Lindane.

(53). Regarding the figures in Appendix A of this document.

Comment: This looks good, but what are typical current velocity magnitudes? Provide descriptions in the text since the reader is more likely to relate to currents than to water levels. How good are the predicted current values?

Response: Current data were not available during model development, so no comparisons can be made.

(54). Regarding the figures in Appendix B of this document.

Comment: Indicate what you mean by “Index” on the x axes.

Response: The “Index” is described in the report, Section 5.1.2, as “Note in these figures, the “index” on the horizontal axis represents the five statistics, from 1 to 5 (representing minimum, 25th percentile, median, 75th percentile, and maximum).”

(55). Regarding the figures in Appendix C of this document.

Comment: I think it would help if 95% confidence bands were included.

Response: This is a good suggestion; however, adding confidence bands will not change the judgment on the model performance.

d. Appendix E –

The following specific comments were provided as embedded comments by the reviewer in the report titled, “Watershed Modeling for Chollas, Switzer and Paleta Creek Watersheds for Simulation of Loadings to San Diego Bay,” dated March 3, 2011:

(1). The bullet that reads, “Non-road sediment in the watershed consisted of 5% sand, 40% clay, and 55% silt.” (Page 2, section 2.2, 3rd bullet)

Comment: Have sources of information been cited some place?

Response: Sediment fractionation was based on best professional judgment and similar studies conducted in regional watersheds.

(2). The sentence that reads, “Additionally, rain gages were co-deployed by Mactec, Inc. at select water quality sampling stations to provide a more accurate measurement of the non-homogeneous rainfall in the watershed due to orographic enhancement and storm patterns.” (Page 4, last sentence)

Comment: Please define the term “orographic.”

Response: An orographic enhancement or effect is when rainfall increases as a result of increasing elevation, such as when an air mass is lifted over mountain ranges.

(3). The sentences that read, “The monitoring methods and results are presented in detail in the previous reports. A discussion of the water quality monitoring results from the 2006 study can be found in the 2007 watershed modeling report. The sampling locations in the 2006 study were located at the bottom of North Chollas, South Chollas, Paleta and Switzer Creeks …” (Page 6, 2nd paragraph)

Comment: Revise font and text placement.

Response: There appears to have been an error with the PDF conversion software. Bookmark coding that was placed on references to figures within the text caused the conversion software to insert the figure images at those reference points in the text, splitting the paragraph across several pages. The entire content of the paragraph is present and will be corrected for the final document.

(4). The sentences that read, “…Figure 2) and focused on characterizing the loads from the watersheds to the Bay. The results of the water quality sampling in the second study are addressed in detail in the City of San Diego’s storm drain characterization studies (City of San Diego, 2010 a, City of San Diego 2010b). The second study was more extensive and monitored sites to characterize storm water quality at the catchment …” (Page 7)

Comment: Revise font and text placement.

Response: See response to comment III.C.1.d.(3), above.

(5). Regarding the parenthesis at top of Page 8.

Comment: Delete parenthesis.

Response: See response to comment III.C.1.d.(3), above.

(6). Regarding appearance of Figure 1 on Page 8.

Comment: Delete figure and caption.

Response: See response to comment III.C.1.d.(3), above. This duplicate figure appears to have been inserted by the PDF conversion software at the reference to figure 1 in the text.

(7). The sentence that reads, "…Figure 2) and land use (Figure 4) scales. The land use distribution of the twelve smaller land use sites (Figure 5 and Figure 6) characterized the predominant land uses within the three watersheds." (Page 9)

Comment: Revise figure caption.

Response: See response to comment III.C.1.d.(3), above. This is the end of the paragraph that was started on Page 6 and not a figure caption. It should also be noted that the figure numbers were changed by the conversion software as well. The figures referenced in this partial sentence should be Figures 1, 3, 4 and 5, respectively.

(8). Regarding Figure 4 on Page 10.

Comment: Reference to Figure 4 missing?

Response: See response to comment III.C.1.d.(3) and III.C.1.d.(7), above.

(9). Regarding Figure 5 on Page 11.

Comment: Please cite this figure in the text.

Response: See response to comment III.C.1.d.(3) and III.C.1.d.(7), above.

(10). Regarding Figure 6 on Page 12.

Comment: Please cite this figure in the text.

Response: See response to comment III.C.1.d.(3) and III.C.1.d.(7), above.

(11). The sentence that reads, “Total suspended sediment (TSS) event mean concentrations (EMCs) showed variability within a land use type, across land use categories, and between storms (Figure 7); however, the EMCs were typically within the 95th percentile flow weighted confidence intervals.” (Page 12, 1st paragraph, 2nd sentence)

Comment: Include or cross-reference an exact definition.

Response: The calculation for the flow-weighted 95th percentile confidence intervals was included in the report. Confidence intervals provide an assessment of the variability of data.

(12). The sentences that read, “Significant improvements were included in the current LSPC models for the Chollas, Paleta, and Switzer Creek watersheds as compared to the previous study.” and “A significant difference with the previous effort was that storm drains were explicitly modeled as a separate land use category and thus a source of sediments in the watershed, as described in Section 4.2.” (Page 16, 1st paragraph, 1st and 4th sentences)

Comment: Do the two highlighted clauses refer to the same aspects of the two efforts? If so avoid repetition.

Response: These references point to different aspects of the two modeling applications.

(13). The sentence that reads, “However, because LSPC is a lumped model it does not simulate each individual storm drain in the watershed, therefore, it was not possible to develop a set of shear/scour parameters that represent the deposition of sediments in the storm drains.” (Page 16, 5th paragraph, 3rd sentence)

Comment: Recommend that the sentence be edited as follows: “… watershed. Therefore, … “

Response: This recommended change has been made to the report.

(14). The sentence that reads, “The runoff volume from the surface land uses was converted to “rain” for inclusion in the storm drain model and the water quality was processed to be included as a mass point source.” (Page 17, 1st sentence)

Comment: The clause, “… and the water quality was processed to be included as a mass point source,” does not read well.

Response: The sentence has been revised.

(15). The sentence that reads, "To allow sediments to accumulate on the storm drain land uses and still have dry weather runoff, flows below a threshold were routed directly to the storm drain channels (which would represent flows in the lower part of the physical storm drain and sediments accumulating on either side of those flows)." (Page 17, 2nd sentence)

Comment: Rephrase this sentence as well. Rewrite sentence.

Response: The sentence has been revised.

(16). Regarding units on the y axes on graphs in Figures 11 through 21 on Pages 18 through 27.

Comment: Discharge should be reported in cubic meters per second (only). Volume in cubic meters (only).

Response: Updated figures have been incorporated into the final report.

(17). Regarding the units on the y axis in Figure 22 on Page 28.

Comment: What is ug/L? Do you mean mg/L?

Response: The units in the figure are micro grams per liter (μg/L). The letter "u" was used by the author as a short-hand for the symbol micro (μ).

(18). The sentences that read, "This shows the importance of accurately representing the watershed hydrology to mobilize and transport sediment via storm water runoff. Furthermore, including storm drains as a separate land use category where sediment was allowed to build up and be transported in storm water improved the model prediction and provided an additional realistic source of sediments in the watershed." (Page 30, 3rd and 4th sentences)

Comment: It would help if you could itemize the specific changes that led to improved prediction.

Response: The discussion contained in Section 4.2 provides the reader with a description of the changes that were made to improve the model's capability to predict.

(19). The sentences that read, "A reason for this could have been a better representation of storm hydrology or a more targeted and better designed sampling protocol. Also the current dataset was more robust with 22 site-events compared to 12 site events in the 2006 sampling.

This second sampling likely is more representative of current water quality and storm water conditions from these watersheds and provides increased reliability for use in calculating TMDLs for the listed toxic pollutants." (Page 31, 2nd paragraph, 2nd, 3rd, and 4th sentences)

Comment: These reasons should be itemized before this section (see previous comment).

Response: The discussion contained in Section 4.2 provides the reader with a description of the changes that were made to improve the model's capability to predict.

(20). Regarding Sections 5 and 6.

Comment: It is necessary to go over Sections 5 and 6 to improve the text (even though there is nothing "wrong" grammatically).

Response: These sections have been revised.

(21). Regarding the following reference, "Longanathan, B., K. Irvine, K. Kannan, V. Pragatheeswaran, and K. Sajwan. 1997. Distribution of Selected PCB Congeners in the Babcock Street Sewer District: A Multimedia Approach to Identify PCB Sources in the Combined Sewer Overflows (CSOs) Discharging to the Buffalo River, New York. Archives of Environmental Contamination and Toxicology, 33, 130-140." (Page 32)

Comment: Change to: Loganathan.

Response: This correction has been incorporated in the document.

e. Appendix F – Comments relating to converting fps units to SI units, correcting a capitalization error, and rounding values presented in Table F-11 PAHs and PCBs in bulk sediments.

Response: All embedded recommendations were incorporated into the document with the exception of rounding values presented in Table F-11. The values were presented exactly has they were presented in the referenced report and there was no reference in the text that the values would be rounded. No changes were made to the Table F-11.

2. ***Comment***: 2). In Appendix D, EFDC modeling, although well done (the investigators do deserve credit), suffers from the paucity of data. In hydrodynamics the absence of current velocity data to compare with is a major drawback because comparison with tides alone can be inadequate, especially given that the comparison is made on a visual basis and does not insure that currents will be reproduced within a satisfactory error margin. (Note that the turbulent shear stress is proportional to velocity squared!)

Response: San Diego Water Board agrees with the reviewer's comment; however, data limitation does not allow for such comparisons to be made.

3. ***Comment***: 3). The main drawback in Appendix D is in the area of selecting the input parameters for fine sediment transport. The state of suspended and bottom sediment has not been described. (a) What is the mineral composition of the sediment? (b) Is the fine sediment flocculated (in suspension and in the bed)? (c) Does the suspended matter, which can be expected to be finer than bed sediment, contain any organic matter of biogenic origin (other than petroleum products)? (If so it would be lighter than mineral sediment and move more easily), (d) What are the distributions of sediment fractions in suspension versus sediment in the bed in the mouth area? (e) What are the size and density distributions of aggregated and unaggregated particles? (d) What is the composition and state of density stratification of the top ~1 m of the bed? (f) Is any fluid mud ever present, such as during or immediately after a storm? (g) Have any dredging data ever utilized to calibrate the sediment model? (h) How would that "long-term" calibration stand up against short-term SSC-based calibration?

   Response: The reviewer provided very good comments about details in the sediment transport modeling; however, for the current study there is no data available to quantify these factors such as mineral composition, flocculation, etc. In the future, the current model could be refined for more reliable predictions if more systematic data became available.

4. ***Comment***: 4). Neither erosion nor deposition rate data appear to be available (How reliable are the settling velocity values taken from an earlier investigation?). In general this would make the problem of modeling fine sediment transport indeterminate, since more than one combination of settling velocity and erosion rate may produce the same or similar SSC time-series. In the present case this dilemma is mitigated by the fact that the mouths are mainly depositional environments. However, one would expect that during and immediately following a major storm some fluid mud would generated due to the hindered settling nature of the fine fraction of suspended matter as it settles close to the bed. Some erosion of the bed or entrainment (or horizontal transport) of fluid mud may occur. Although it appears that the investigators have modeled reasonable scenarios in the mouth areas, there is practically no backup evidence available to support or negate the choices made for model inputs.

   Response: The San Diego Water Board agrees with the reviewer; however, data availability prevents any further refinement of the model.

5. ***Comment***: 5). The wave field has not been considered. I expect this may not be a poor assumption, but in any case no evidence has been supplied to support it. Given the right directions the bay does seem to have fetches that would generate short period waves during severe sea storms. However, in the absence of data we do not know if this is so. Sea storms have not been addressed in any detail.

   Response: As previously stated, data is not available to address the influence of sea storms. A significant level of effort would be needed to provide wave modeling and is likely not necessary for the purpose of this study.

6. ***Comment***: 6). In order to produce Appendix D that can withstand technical challenges it will be necessary to address the above issues. In its present form this appendix does not meet technical standards of acceptance. <u>A re-review of the revised Appendix D will be essential</u>.

   Response: In accordance with the law that governs the peer review process (Health & Safety Code 57004) the San Diego Water Board is only allowed to communicate with peer reviewers regarding clarifications. To be considered an independent review, there can be no interaction that could be perceived as collaboration in writing the documents. In complying with this requirement, San Diego Water Board cannot re-submit the documents requested for additional review.

   Reply: Professor Mehta was notified of the above response and he replied with the following comment.

   A re-review would have allowed me to go over authors' responses. The report needs some strengthening of explanations for sediment transport model application. In any event I trust the authors, who seem to be good at numerical modeling applications but perhaps do not have an extensive background in fine sediment transport, will be able to provide satisfactory answers to my queries and concerns.

7. ***Comment***: 7). It is conceivable that several questions raised in this review regarding sediment transport may not be easy to address fully. The correct technical approach would be to collect the necessary data. One could envisage a single roaming instrumented tower assembly installed at two spots in one of the mouths, e.g. Chollas, one at a location along the upstream flow cross section where the creek enters the basin and another along the outer cross-section of the basin. At each spot the tower could collect data for, say, 1.5 months. Water level (tides and waves), current velocity, and SSC time-series would be collected. In addition, the bottom 1 m of the bed would be cored for stratigraphic data and assessment of the presence of fluid mud. Separately from this the settling velocity of particles must be measured, e.g. using an imaging procedure similar to one used in South San Francisco Bay by ERDC.

Response: The San Diego Water Board agrees with the reviewer and appreciates his comments.

Visual Hypothesis Testing with Confidence Intervals

Robert W. Smith, EcoAnalysis Inc., Ojai, CA

ABSTRACT

A simple visual method of comparing means given normal 95% confidence intervals is proposed. Monte Carlo simulations were used to derive a visual rule of thumb that can be used to test the null hypothesis of equal means, given the confidence intervals of the means, and the sample sizes (or degrees of freedom). The type-1 error rate of the test is approximately .05. The rule involves accepting the null hypothesis when a confidence interval for either mean overlaps the other mean, and rejecting the null hypothesis otherwise. In some cases with moderate to high sample sizes and/or similar-sized confidence intervals, the original confidence intervals need to be extended before the null hypothesis can be rejected by the rule. The rule is most easily applied when the means and confidence intervals are displayed with auxiliary information that includes the degrees of freedom and potential extensions to the confidence interval bounds. To display results in this manner, a SAS macro utilizing the annotate facility with the SAS GPLOT procedure is provided.

INTRODUCTION

The relationship between a confidence interval and a hypothesis test is that the confidence interval contains all the values of the population mean that could serve as the null hypothesis value (for equality of means) and the null hypothesis would not be rejected at the nominal type-1 error rate (Nelson 1990). Rather than rely solely on p values from statistical tests, there are practical advantages of using confidence intervals for hypothesis testing (Wonnacott 1987, Nelson 1990). For most people, confidence intervals are easier to understand. Confidence intervals also provide additional useful information since they include a point estimate of the mean, and the width of the interval gives an idea of the precision of the mean estimation.

The most direct manner of comparing pairs of means with confidence intervals is to the compute confidence interval for the *difference* between each pair of estimated means. If the confidence interval covers a value of zero, then the null hypothesis is accepted at the type-1 error rate of *100-p* (Gardner and Altman 1989, Hsu and Peruggia 1994, Lo 1994), where p is the percent coverage of the confidence intervals. With this approach the visual advantage of the confidence interval of the mean is lost. The individual means and their uncertainty will be obscured. Also, if several estimated means are to be compared, there will be $n(n-1)/2$ separate confidence intervals of the differences to display (n=the number of means compared). This can be a very large number of confidence intervals.

Another common approach is to reject the null hypothesis when the 95% confidence intervals of the means do not overlap. Barr (1969), Nelson (1989), and Lo (1994) show that using the overlap criterion with 95% confidence intervals of the mean will be associated with an actual type-1 error rate as low as .005 and relatively low power. Thus this approach represents a relatively insensitive test.

Most other graphical techniques for comparing means with statistical intervals involve intervals that are specialized for hypothesis testing. For example, with *uncertainty or comparison intervals* (Gabriel 1978, Andrews et al. 1980, Hochbert et al. 1982, Hochbert and Tamhane 1987, Lenth 1988, Hsu and Peruggia 1994, Goldstein and Healy 1995), the interval bounds are computed so that the null hypothesis of equal means is rejected (at a nominal type-1 comparisonwise or experimentwise error rate) when the uncertainty intervals of a pair of means do not overlap. The uncertainty interval methods can be exact when the samples sizes are balanced and standard errors for the means are homogeneous. Otherwise, the interval sizes will usually be rough approximations since different sized intervals would be required for comparisons of different pairs of means (when more than two means are to be compared). Hsu and Peruggia (1994) discuss approximation methods for unbalanced designs, and Hochberg (1976) and Lenth (1988) give approximations for unequal standard errors of the means.

My objectives in developing the technique presented in this paper were to provide a method where

1. hypothesis tests on the equality of means could rapidly be made with a simple rule of thumb and without computations,
2. the comparisonwise type-1 error of the tests will approximate a known nominal type-1 error rate whether the design is balanced or unbalanced and whether the standard errors of the means are homogeneous or heterogeneous, and
3. a commonly displayed statistical interval is used directly in the hypothesis test. The 95% (or 90%) confidence interval of the mean was chosen for this purpose. This objective precluded using uncertainty interval methods.

Browne (1979) provides an approach for using 95% confidence intervals of the mean to test for the equality of means at a nominal .05 type-1 error level. The method is based on a D statistic, which for each pair of means is the ratio of the difference in the two upper confidence bounds to the length of the smaller of the two confidence intervals. A table is provided with D values for several sets of sample sizes and different M ratios. An M ratio is the ratio of the larger interval to the smaller interval for a pair of means being compared. The D values in the table are minimum D values for which the means will be significantly different. From the patterns of the D values in the table, the author provided a few rules of thumb that covered some but not all situations. The D values in Browne's table (Browne 1979) were directly computed using the definition of the t statistic for the difference in two means. I use a simulation approach in an attempt to obtain a more general and simpler rule of thumb.

In this paper, I describe a visual test using confidence intervals to test the null hypothesis of equal means with known approximate nominal type-1 error rates. The test is visual in that it requires no statistical computations, and can be directly applied to results presenting confidence intervals. Only the degrees of freedom and the approximate ratio of the lengths of the respective confidence intervals are needed. I propose a simple rule of thumb to accomplish this for confidence intervals based on normal distributions. My main focus is on two-tailed tests with a nominal type-1 error rate of .05 (α=.05). I present an example plot that includes useful auxiliary information for easily applying the proposed method, along with a SAS macro that can be used to produce similar plots.

METHODS

Rule of thumb for hypothesis testing

As a starting point for developing a rule of thumb for hypothesis testing with confidence intervals, I propose the following rule of thumb with two-tailed 95% confidence intervals.

Rule of thumb: *Accept the null hypothesis of equal means for two estimated means if* ***either*** *mean is within the 95% confidence interval of the other mean. Otherwise, reject the null hypothesis.*

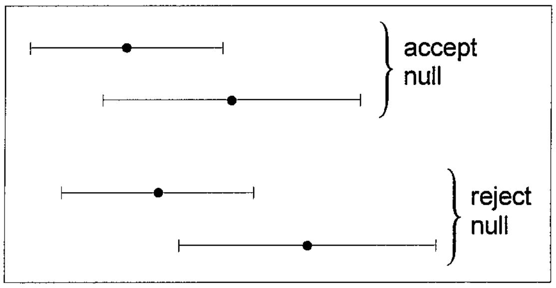


Figure 1. Rule of thumb. The null hypothesis is accepted if either confidence interval includes the mean of the other.

The rule is illustrated in Figure 1. To test type-1 error levels associated with this rule, I performed Monte Carlo simulations for two normal populations with equal means. Many simulation scenarios with different sample sizes and population variances were used to approximate a wide variety of situations. Simulations were performed in SAS. Each simulation scenario involved randomly drawing 100,000 samples from each of the two normal populations, using the RANNOR function. For each sample, the population sample mean and confidence limits were computed. The upper (*U*) and lower (*L*) bounds of two-tailed confidence limits were computed as

$$L = \bar{x} - t_{.975,df}\ s/\sqrt{n} \qquad (1)$$

and

$$U = \bar{x} + t_{.975,df}\ s/\sqrt{n}\,, \qquad (2)$$

where $\bar{x}$, *s*, and *df* are the estimated mean, standard deviation, and degrees of freedom, respectively, for a sampled population. In the simulations, *df=n-1*, where *n* is the sample size. The $t_{.975,df}$ is the value for the 97.5th percentile of the cumulative *t* distribution with *df* degrees of freedom.

Preliminary simulations showed that the actual type-1 error rates varied with the degrees of freedom and the ratio of the lengths of the confidence intervals being compared. The ratio of confidence intervals lengths (*R*) for a simulation scenario is computed as

$$R = \frac{Max(\bar{L}_1, \bar{L}_2)}{Min(\bar{L}_1, \bar{L}_2)}\,, \qquad (3)$$

where $\bar{L}_1$ and $\bar{L}_2$ are the average estimated lengths of the confidence intervals for the two populations over all simulations. *R* is simply the ratio of the expected value of the larger confidence interval to the expected value of the smaller confidence interval. Variation in *R* is produced by using different population variances and sample sizes in the different simulation scenarios.

Figure 2 summarizes the pattern of actual type-1 error rates, which are the proportion of the simulations (within a scenario) for which the null hypothesis of equal means is rejected using the rule of thumb. Separate plots are shown for different ranges of *R* that produced similar results. In these, the simulation type-1 error rates are shown as a function of the degrees of freedom for the two population samples. Lines are drawn to delimit blocks of sample-size combinations where the actual type-1 error levels are roughly similar. Type-1 error rates within the block labeled "*A=0*" are close to .05 or, in a few cases involving minimal degrees of freedom, the rates get low as .01. Most of the type-1 error rates in the block labeled "*A=.3*" are over twice the .05 rate, and the type-1 error rates in the blocks labeled "*A=.1*" or "*A=.2*" are moderately above the .05 rate. The results in Figure 2 indicate that in most cases where the *R* is relatively high (Figure 2c) or the sample size of at least one of the populations is relatively low (Figures 2a and 2b), the rule generally works well if one is interested in a nominal type-1 error rate of .05.

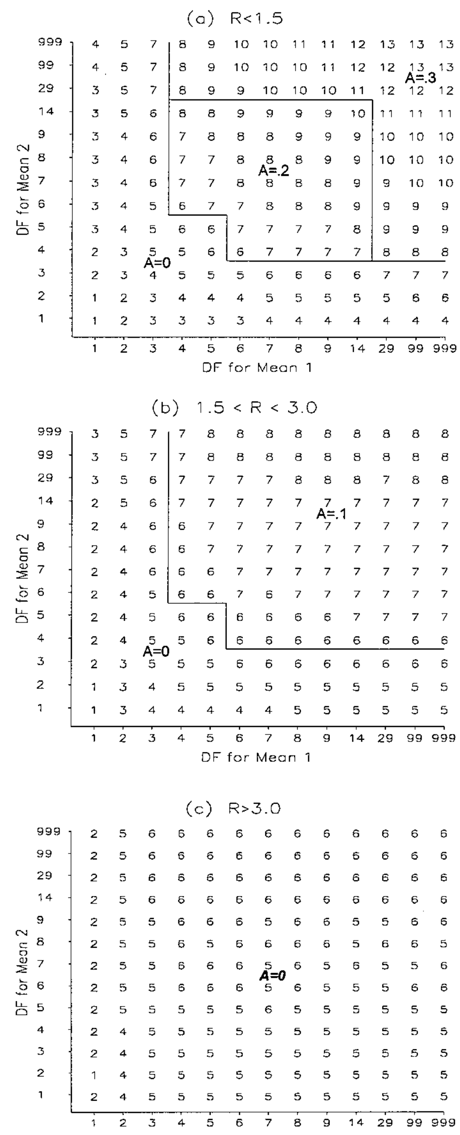


Figure 2. Actual simulation type-1 error rates using rule of thumb for multiple simulation scenarios involving different degrees of freedom and R ratios. The two-tailed error rates are shown as 100(error rate), rounded to the nearest digit. For example, an error rate of .051 would appear as a "5", and an error rate of .066 would appear as a "7". The rates are blocked into regions with generally similar error rates. The indicated values for A are used to modify the confidence intervals to improve the type-1 error rate. Where R>1, the expected confidence interval length for mean 1 is greater than the expected confidence interval length for mean 2. For (a) R ≤1.5, (b) R =1.5 to 3, and (c) R >3.

Application of the rule of thumb to extended confidence intervals. The more common situation where rule of thumb fails involves lower R values and/or higher degrees of freedom (*A>0* in Figure 2). To correct for this, I propose extending the interval bounds to produce type-1 error rates closer to .05 in cases where the initial type-1 error rates are too high. The rule of thumb is then applied to the extended confidence bounds. For a pair of means being compared, the rules for interval modification are as follows.

1. Make no modifications to the interval bounds if the null hypothesis has been accepted by the rule, since extension of the bounds cannot cause rejection of the null hypothesis.
2. Make no modifications if the confidence intervals for the two means do not overlap. As noted in the introduction, a test using interval overlap will generally be associated with very low type-1 error, so we will not reverse the decision to reject the null hypothesis in cases of non-overlap.
3. Make no modifications for situations corresponding to areas labeled "A=0" in Figure 2. The type-1 error in these areas are already for the most part satisfactory.
4. Make no modifications if it appears that as much as a 30% extension of the half intervals will not change the decision on rejection of the null hypothesis. As shown below, the maximum proposed extension of the half intervals is 30% (A=.3), so if this amount of extension will not change our decision, there is no need to go through the process of determining the best modification for the specific situation.
5. If rule 4 above is not applied, make modifications for situations corresponding to areas labeled "A=.1", "A=.2" or "A=.3" in Figure 2.

The modification of the confidence intervals, which involves a visual extension of the interval lengths for the means being compared is now described. The confidence interval length ratio for a pair of samples is

$$R_{est} = \frac{Max(\widehat{L}_1, \widehat{L}_2)}{Min(\widehat{L}_1, \widehat{L}_2)}\,, \qquad (4)$$

where $\widehat{L}_1$ and $\widehat{L}_2$ are the estimated confidence interval lengths for the first and second samples, respectively. The R_{est} value for a pair of means is needed to determine the value of the adjustment factor *A* (see below).

A separate potential modification is required for each *pair* of means to be compared. The modified upper and lower confidence interval bounds for each mean being compared are

$$U_{\text{mod}} = U + A(U - \bar{x})\,, \qquad (5)$$

and

$$L_{mod} = L - A(\bar{x} - L), \qquad (6)$$

where *L*, *U*, and $\bar{x}$ are defined in (1) and (2). The value for *A* is taken from Figure 2, and depends on the R_{est} and the degrees of freedom associated with the two estimated means. If $R_{est} \leq 1.5$ use Figure 2a, or if $1.5 < R_{est} < 3.0$ use Figure 2b. When $R_{est} > 3$, no adjustment is needed since *A=0* for all degrees of freedom (see Figure 2c). Note that the adjustment factor *A* is applied to one half the length of the confidence intervals ($\bar{x} - L$ or $U - \bar{x}$) rather than to the full length of the confidence intervals (*U-L*) . I could just as easily halved the size of the *A* values and used the entire confidence interval length. However I chose to work with the half size interval because it is easier to visually apply the modification to the shorter length of the half interval.

Figure 3 shows the type-1 error rates when the rule of thumb is applied with extended intervals where necessary. The simulation type-1 error rates are now closer to .05. The error rates when the larger confidence interval (for mean 1) is associated with 1 degree of freedom are still low. To keep things simple, I made no attempt to correct for this situation since it would entail reducing the interval sizes. It is probably not a bad idea to accept a conservative test when the degrees of freedom are so minimal for one of the means.

(a) R<1.5 with Rule of Thumb

DF for Mean 2 \ DF for Mean 1	1	2	3	4	5	6	7	8	9	14	29	99	999
999	4	5	7	4	4	4	4	5	5	5	5	5	6
99	4	5	7	4	4	4	4	4	5	5	5	5	6
29	3	5	7	4	4	4	4	4	4	5	5	5	5
14	3	5	6	4	5	5	5	5	5	6	5	5	5
9	3	4	6	4	4	5	5	5	5	5	4	5	5
8	3	4	6	4	4	5	5	5	5	5	4	5	5
7	3	4	6	4	4	4	4	5	5	5	4	4	4
6	3	4	5	4	4	4	4	4	4	5	4	4	4
5	3	4	5	6	6	4	4	4	4	5	4	4	4
4	2	3	5	5	6	4	4	4	4	4	4	4	4
3	2	3	4	5	5	5	6	6	6	6	7	7	7
2	1	2	3	4	4	4	5	5	5	5	5	6	6
1	1	2	3	3	3	3	4	4	4	4	4	4	4

(b) 1.5 < R < 3.0 with Rule of Thumb

DF for Mean 2 \ DF for Mean 1	1	2	3	4	5	6	7	8	9	14	29	99	999
999	3	5	7	4	4	4	5	5	5	5	5	5	6
99	3	5	7	4	4	4	5	5	5	5	5	5	6
29	3	5	6	4	4	4	4	5	5	5	5	5	5
14	2	5	6	4	5	5	5	5	5	5	5	5	5
9	2	4	6	4	4	5	5	5	5	5	5	5	5
8	2	4	6	4	4	4	5	5	5	5	5	5	5
7	2	4	6	4	4	4	4	5	5	5	5	5	5
6	2	4	5	4	4	4	4	4	5	5	5	5	5
5	2	4	5	6	6	4	4	4	4	5	4	5	5
4	2	4	5	5	6	4	4	4	4	4	4	4	4
3	2	3	5	5	5	6	6	6	6	6	6	6	6
2	1	3	4	5	5	5	5	5	5	5	5	5	5
1	1	3	4	4	4	4	4	5	5	5	5	5	5

(c) R>3.0 with Rule of Thumb

DF for Mean 2 \ DF for Mean 1	1	2	3	4	5	6	7	8	9	14	29	99	999
999	2	5	6	5	5	5	5	5	5	6	6	6	6
99	2	5	6	5	5	5	5	5	5	5	6	6	6
29	2	5	6	4	5	5	5	5	5	5	5	6	6
14	2	5	6	5	5	5	5	5	5	5	5	6	6
9	2	5	5	5	5	5	5	5	5	5	5	5	5
8	2	5	5	4	5	5	5	5	5	5	5	5	5
7	2	5	5	5	5	5	5	5	5	5	5	5	5
6	2	5	5	5	5	5	5	5	5	5	5	5	5
5	2	5	5	5	5	5	5	5	5	5	5	5	5
4	2	4	5	5	5	5	5	5	5	5	5	5	5
3	2	4	5	5	5	5	5	5	5	5	5	5	5
2	1	4	5	5	5	5	5	5	5	5	5	5	5
1	2	4	5	5	5	5	5	5	5	5	5	5	5

Figure 3. Simulation type-1 error rates using rule of thumb for simulation scenarios with (a) R ≤1.5, (b) R =1.5 to 3, and (c) R>3. See Figure 2 for explanation of values in plots.

The confidence interval modification is illustrated with an example. The confidence interval for one sample mean is *11±5*, the confidence interval for a second mean is *17±2*, and the respective degrees of freedom are 7 and 39. Using (4), (5), and.(6), the modified lower confidence interval bound for the first mean is L_{mod} *= 6-.1(11-6)=5.5*, and U_{mod} *= 16+.1(16-11)=16.5.* Here *A=.1*, since R_{est}*=10/4=2.5* and thus *A* is taken from Figure 2b for degrees of freedom of *7* and *39*. For mean 2, L_{mod} *= 15-.1(17-15)=14.8,* and U_{mod} *= 19+.1(19-17)=19.2.*

Given the rules for modification above and the patterns in Figure 2, a summary of the rule of thumb as modified to apply to the extended intervals is given in Figure 4. In practice, the application of the rule of thumb need not be too complicated nor require repeated reference to Figure 2. In most cases, it should be easy to visually determine the approximate value for R_{est}, and with some experience, extensions of the intervals corresponding to the different values of *A* should become simple. When the data are displayed as in the example shown below (Figure 5), the application of the rule should be even more simplified.

Incidentally, a one-tailed comparison of means with approximately a .05 type-1 error rate can be obtained with the exactly the same rule of thumb by using 90% instead of 95% confidence intervals.

Use the rule of thumb directly if:

no overlap of confidence intervals, or
null hypothesis accepted with rule of thumb, or
high R_{est} ($R_{est} > 3$), or
low degrees of freedom (*Max(df)<6* or *Min(df)<4*)

Otherwise extend the intervals as follows before applying the rule of thumb:

If $R_{est} > 1.5$ let $A=.1$,
otherwise
if *Max(df)<14,* let $A=.2$
otherwise let $A=.3$.

Figure 4. Rule of thumb extended to apply to modified intervals if necessary. Max(df) is the maximum degrees of freedom for the two sample means being compared, and Min(df) is the minimum degrees of freedom for the two means being compared.

RESULTS

Example

Figure 5 shows a hypothetical example plot showing some estimated means and confidence intervals. To aid in applying the proposed rule of thumb, I have included the associated degrees of freedom (df=) and three dots corresponding to interval adjustments for A=.1, A=.2, and A=.3 (in case modified intervals are needed). Table 1 summarizes the rules applied and acceptance or rejection of the null hypothesis for each pair of treatments.

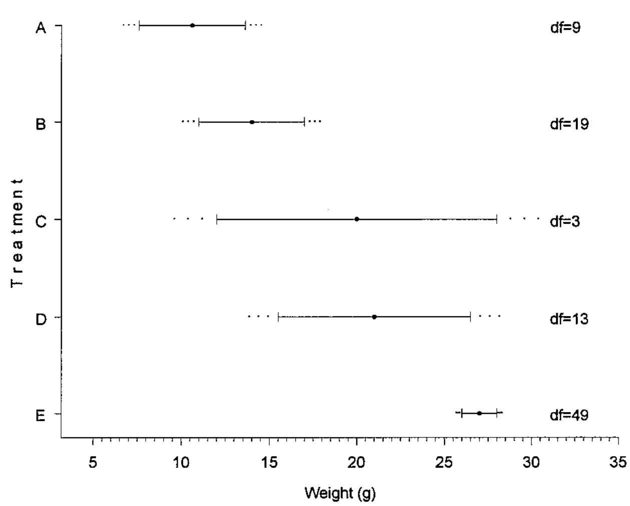


Figure 5. Hypothetical example for body weights of test organisms exposed to five different treatments (A-E). The five treatment sample means and their estimated confidence intervals, along with the associated degrees of freedom (df=) are shown. The three dots outside the interval bounds correspond to modifications of the interval bounds corresponding to A=.1, A=.2, and A=.3, respectively.

Table 1. Results of hypothesis tests for all pairs of means in Figure 5 using the rule of thumb. An "M" in the lower diagonal indicates that modified intervals were used in the decision. In the upper diagonal, "A" indicates acceptance and "R" indicates rejection of the null hypothesis for the treatment pair using the rule of thumb.

TREATMENT	A	B	C	D	E
A	-	A	R	R	R
B	M	-	A	R	R
C			-	A	A
D		M		-	R
E					-

Given the display in Figure 5 with the degrees of freedom and the dots for potential adjustments, it is a simple matter to determine which comparisons will require using an adjusted interval. For the example data, only the A-B and B-D comparisons require modified intervals before application of rule of thumb. All other comparisons are directly handled by the rule because they meet at least one of the criteria at the top of Figure 4.

Macro BPLOT

The SAS macro BPLOT, which produces output similar to that shown in Figure 5, is presented in Appendix A. Macro BPLOT uses the GPLOT procedure and the Annotate facility of SAS/GRAPH®.

CONCLUSION

A simple visual rule of thumb is presented for comparing means using 95% confidence intervals. For the most part, the method controls the type-1 error for a 2-tailed test of the null hypothesis of equal means at approximately .05. A SAS macro is included for presenting means and confidence intervals in a manner that facilitates applying the rule of thumb.

ACKNOWLEDGMENTS

Dr. Allan Stewart-Oaten provided some comments that were very useful in refining my approach.

TRADEMARKS

REFERENCES

Andrews, H.P., R.D. Snee, and M.H. Sarner. 1980. Graphical display of means. The American Statistician 34 (4): 195-199.

Barr, D.R. 1969. Using confidence intervals to test hypotheses. Journal of Quality Technology 1(4): 256-258.

Browne, R.H. 1979. On visual assessment of the significance of a mean difference. Biometrics 35: 657-665.

Gabriel, K.R. 1978. A simple method of comparison of means. Journal of the American Statistical Association 73 (364): 724-729.

Gardner,M.J. and D.G. Altman. 1989. Statistics with Confidence: Confidence intervals and statistical guidelines. Published by British Medical Journal, London.

Goldstein, H. and M.J.R. Healy. 1995. The graphical presentation of a collection of means. Journal of the Royal Statistical Society A 158 Part 1: 175-177.

Hochberg. Y. 1976. A modification of the T-method of multiple comparisons for a one-way layout with unequal variances. Journal of the American Statistical Association 71: 200-203.

Hochberg, Y. and Tamhane, A.C. 1987. Multiple Comparison Procedures. John Wiley and Sons. New York.

Hochberg, Y. G. Weiss, and S. Hart. On graphical procedures for multiple comparisons. Journal of the American Statistical Association 77 (380): 767-772.

Hsu, J.C. and M. Peruggia. 1994. Graphical representations of Tukey's multiple comparison method. Journal of Computational and Graphical Statistics 3 (2): 143-161.

Lenth, R.V. 1988. Graphics for multiple comparisons. ASA Proceedings of the Statistical Graphics Section. 70-75. The American Statistical Association.

Lo, N.C.H. 1994. Level of significance and power of two commonly used procedures for comparing mean values based on confidence intervals. CalCOFI rep. 35: 246-253.

Nelson, L.S. 1989. Evaluating overlapping confidence intervals. Journal of Quality Technology 21(2): 140-141.

Nelson, L.S. 1990. Comments on significance tests and confidence intervals. Journal of Quality Technology 22(4): 328-330.

Wonnacott, T. 1987. Confidence intervals or hypothesis tests? Journal of Applied Statistics 14 (3): 195-201.

APPENDIX A. THE BPLOT MACRO.

```
%macro bplot(
 dsn=,        /* Name of sas dataset with plot data */
 yvar=,       /* Name of variable with y data values (A
                 separate mean with confidence
                 intervals plotted for each value of
                 the yvar= variable */
 yformat=,    /* Format for y variable (yvar=) */
 yorder=,     /* If yorder= (blank) the order on the Y
                 axis will be in ascending order (top
                 to bottom) of the xmean= variable.
                 If yorder=variable name(s), then the
                 order on the Y axis will be in
                 ascending order of the values of the
                 yorder= variable(s) */
 xmean=,      /* Name of variable with x mean values */
 xlci=,       /* Name of variable with x lower
                 confidence limit */
 xuci=,       /* Name of variable with x upper
                 confidence limit */
 hsize=7,     /* Horizontal size of plot area */
 vsize=10,    /* Vertical size of plot area */
 units=in,    /* Units for hsize and vsize */
 xtick=,      /* Set xtick= tick values for x axis */
 df=,         /* Name of variable with degrees of
                 freedom to be plotted */
 dfsize=.6,   /* Size of df text */
 dotprope=.1 .2 .3, /* Proportion expansions for each
                        half ci */
 dotsizee=.6, /* Size of the ci expansion dot(s) */
 dotsizem=.6, /* Size of the mean dot */
 cisize=1,    /* Size of vertical part of the
                 confidence interval bars */
 lnsize=1,    /* Size of horizontal line for confidence
                 interval bars */
 title=,      /* Main title for plot */
 titlsiz=1,   /* Size for title text */
 xtitle=,     /* X-axis title */
 xtitlsiz=1,  /* Set xtitlsiz = size for x axis title*/
 ytitle=,     /* Y-axis title */
 ytitlsiz=1,  /* Set ytitlsiz = size for y axis title*/
 rotytitl=Y,  /* Set rotytitl=Y to rotate the y title
                 90 degrees */
 setgopt=Y,   /* If setgopt=Y, GOPTIONS statement
                 used */
 device=LJIVPS,  /* Device parm for GOPTIONS
                    statement*/
 tardev=LJIVPS,  /* targetdevice parm for GOPTIONS
                    statement */
 gaccess=,       /* gaccess parm for GOPTIONS
                    statement */
 goutmode=,      /* goutmode parm for GOPTIONS
                    statement */
 bfont=SWISS,    /* Name of font for vertical bars at
                    ci bounds
                    Note: SIMPLEX works best with cgm
                    device file to be imported into
                    WORD */
 titlfont=SWISSB, /* Font for plot main title */
 texfont=SWISS);  /* Font for text in plot */

%local nprop ngapp dflab setgopt ndot word srtv nform
mxlen i ndot dot1
  dot2 dot3 maxx minx
  c1 c2 c3 c4 c5 c6 c7 c8 c9 c10 c11 c12 c13 c14 c15
c16 c17 c18 c19 c20
  f1 f2 f3 f4 f5 f6 f7 f8 f9 f10 f11 f12 f13 f14 f15
f16 f17 f18 f19 f20;

%*---- SOME DEFAULTS;
%let nprop=.10;%let ngapp=.03;%let dflab=%quote(df=);
%let title=%quote(&title);
%*---- PARSE DOT PROPORTIONS;
%let ndot=0;
%do %until(&word= );
 %let word=%qscan(&dotprope,%eval(&ndot+1),%str( ));
 %if &word ne  %then %do;%let ndot=%eval(&ndot+1);%let
dot&ndot=&word;%end;
%end;

%*---- GOPTIONS STATEMENT;
%if %upcase(&setgopt)=Y %then
%do;
GOPTIONS RESET=ALL HSIZE=&hsize &units VSIZE=&vsize
&units
%if &device ne  %then %do; DEVICE=&device %end;
%if &tardev ne  %then %do; TARGETDEVICE=&tardev %end;
%if &goutmode ne  %then %do; GOUTMODE=&goutmode %end;
%if &gaccess ne  %then %do; GACCESS="&gaccess" %end;
;RUN;
%end;

%*---- SET UP DATA FOR Y AXIS ORDER;
%if &yorder ne  %then
%let srtv=&yorder; %else %let srtv=&xmean;

PROC SORT DATA=&dsn OUT=SS_; BY DESCENDING &srtv;

DATA SSS_(KEEP=&yvar &xmean &xlci &xuci &df _YCD);
 LENGTH _YCD $ 4 _IV $ 4 _YFORM $ 16;
 RETAIN _I _MAXLEN 0;
 SET SS_(KEEP=&yvar &xmean &xlci &xuci &df) END=EOF;
 IF &xmean=. THEN DELETE;
 ELSE
 DO;
  _I=_I+1;
  _YCD=PUT(_I,Z4.);
  IF &xlci=. THEN &xlci=&xmean;
  IF &xuci=. THEN &xuci=&xmean;
  %if &yformat=  %then %do; _YFORM=&yvar; %end;
  %else %do; _YFORM=PUT(&yvar,&yformat); %end;
  _MAXLEN=MAX(_MAXLEN,LENGTH(_YFORM));
  _IV=LEFT(PUT(_I,4.));
  CALL SYMPUT('c' || _IV,_YCD);
  CALL SYMPUT('f' || _IV,_YFORM);
 END;
 IF EOF THEN
 DO;
  CALL SYMPUT('nform',LEFT(PUT(_I,4.)));
  CALL SYMPUT('mxlen',LEFT(PUT(_MAXLEN,2.)));
 END;
RUN;

PROC FORMAT;
 VALUE $YFORM
  %do i=1 %to &nform;
```

```
  %if %length(&&f&i)>&mxlen %then %do;
"&&c&i"="%substr(&&f&i,1,&mxlen)" %end;
  %else %do; "&&c&i"="&&f&i" %end;
  %end;  ;

DATA PLOTDS_;
 SET SSS_;
 FORMAT _YCD YFORM.;
 RENAME &xuci=_XUCI &xlci=_XLCI &xmean=_XMEAN;

%*---- ANNOTATE DATA SET WITH LINES, N, AND DOTS;
%if &df&dotprope ne  %then
%do;
DATA NPOS_(KEEP=_MAXX _MINX _MAXXF);
 RETAIN _MAXX _MAXXF -999999 _MINX 999999;
 SET PLOTDS_ END=_EOF;
 %if &ndot>0 %then
 %do;
 _X=_XUCI+(_XUCI-_XMEAN)*&&dot&ndot;
 _XMIN=_XLCI-(_XMEAN-_XLCI)*&&dot&ndot;
 %end;
 %else %do;  _X=_XUCI;_XMIN=_XLCI; %end;
 _MAXX=MAX(_MAXX,_X);
 _MINX=MIN(_MINX,_XMIN);
 IF _EOF THEN
 DO;
  %if &df ne  %then %do;
_MAXXF=_MAXX+(&nprop+&ngapp)*(_MAXX-_MINX); %end;
  %else %do;  _MAXXF=_MAXX; %end;
  OUTPUT;
  CALL SYMPUT('maxx',LEFT(PUT(_MAXX,BEST15.)));
  CALL SYMPUT('minx',LEFT(PUT(_MINX,BEST15.)));
  STOP;
 END;
RUN;
%end;

DATA ANNO_;
  LENGTH FUNCTION $5 POSITION $ 1 TEXT $ 16 STYLE $ 8;
  RETAIN XSYS YSYS '2';
  SET PLOTDS_ END=_EOF;
  FUNCTION='MOVE'; *--  move to beginning of CI;
  YC=_YCD; X=_XLCI; OUTPUT; *-- draw horizontal line;
  FUNCTION='DRAW'; SIZE=&lnsize; YC=_YCD; X=_XUCI;
  OUTPUT;
%*-- dots;
  %if &dotprope ne  %then
  %do;
  POSITION='B';SIZE=&dotsizee;TEXT='.';YC=_YCD;
  %do i=1 %to &ndot;
  FUNCTION='MOVE'; X=_XUCI+(_XUCI-_XMEAN)*&&dot&i;
  OUTPUT;FUNCTION='LABEL';OUTPUT;
  X=_XLCI-(_XMEAN-_XLCI)*&&dot&i; FUNCTION='MOVE';
  OUTPUT; FUNCTION='LABEL'; OUTPUT;
  %end;
  %end;
%*-- df values;
  %if &df ne  %then
  %do;
  POSITION='6';STYLE="&texfont";SIZE=&dfsize;
  %if &dflab ne  %then %do;TEXT="&dflab" ||
LEFT(PUT(&df,6.));%end;
  %else %do;TEXT=LEFT(PUT(&df,6.));%end;
  YC=_YCD;_XADD=&ngapp*(&maxx-
&minx);FUNCTION='MOVE';X=&maxx+_XADD;OUTPUT;
  FUNCTION='LABEL';OUTPUT;
  %end;

%*---- ADD MIN AND MAX POSITIONS TO PLOT DATA;
DATA PLOTDS1_;
 %if &df&dotprope ne  %then %do; MERGE PLOTDS_ NPOS_;
%end;
 %else %do; SET PLOTDS_; %end;
 IF _XLCI=_XUCI THEN DO;_XLCI=.;_XUCI=.;END;

%*---- PLOTTING;
SYMBOL1 VALUE=dot HEIGHT=&dotsizem COLOR=BLACK;
SYMBOL2 VALUE='l' FONT=&bfont HEIGHT=&cisize
COLOR=BLACK;
SYMBOL3 VALUE=dot HEIGHT=.00001 COLOR=WHITE;
AXIS1 LABEL=  (F=&titlfont HEIGHT=&xtitlsiz "&xtitle")
      OFFSET= (5 PCT,0 PCT) VALUE= (FONT=&texfont)
      %if &xtick ne  %then %do; ORDER=(&xtick) %end;  ;
AXIS2 LABEL= (F=&titlfont HEIGHT=&ytitlsiz
             %if &rotytitl=Y %then %do; ANGLE=90 %end;
"&ytitle")
      OFFSET= (5 PCT, 0 PCT) VALUE=  (FONT=&texfont);

PROC GPLOT DATA=PLOTDS1_;
 PLOT _YCD*_XMEAN=1 _YCD*_XLCI=2 _YCD*_XUCI=2
       %if &df&dotprope ne  %then %do;_YCD*_MINX=3
_YCD*_MAXXF=3 %end;
       / OVERLAY HAXIS=AXIS1 VAXIS=AXIS2
ANNOTATE=ANNO_;
 %if &title ne  %then %do;TITLE FONT=&titlfont
HEIGHT=&titlsiz "&title";%end;
RUN; QUIT; RUN;
%mend bplot;
```

Contacting the Author

Robert W. Smith
EcoAnalysis Inc.
221 E. Matilija St. # A
Ojai, CA 93023
(805) 646-1461
bsmith@ecoanalysis.com