THE BRAKE PAD PARTNERSHIP

Compilation of Technical Reviewers and Stakeholders Comments on Draft Report for Alameda Countywide Clean Water Program on Stormwater Quality Monitoring

February 22, 2005

Background

The Brake Pad Partnership is supporting the Alameda Countywide Clean Water Program's (ACCWP) collection and reporting of monitoring data on copper in stormwater runoff during the 2003-2004 wet season in Castro Valley Creek. The Brake Pad Partnership is working with Jim Scanlin of ACCWP who has collected flow-weighted and/or time-weighted samples from nine storm events from October through February 2004 (flow and time-weighted samples were collected for seven events). All samples were analyzed for total copper using EPA Method 160.2 and total suspended solids using EPA Method 200.8. The data obtained adds a ninth year of data to Alameda County's copper monitoring data set that was begun in 1989. The entire data set is being used to calibrate and validate the watershed model.

The Brake Pad Partnership Steering Committee is seeking an independent expert review of the Draft Report for Stormwater Quality Monitoring to ensure that the approach and results of this element of the Partnership's work are technically sound and to help build in-depth understanding of and confidence in the technical studies on the part of the Steering Committee and the stakeholder communities.

Charge

With the aim of meeting these objectives, the Steering Committee has developed the following questions on which it is seeking specific comments from the reviewers:

1. What is your assessment of the quality of the methods used and resultant data?

2. Previously collected monitoring data from the same watershed will be used to calibrate and/or validate the Brake Pad Partnership's watershed model. Do you recommend that these data also be used for this purpose?

3. Based on your experience with similar monitoring efforts, what advice do you have regarding the interpretation and use of these data?

Comments Received

Comments of Arthur J. Horowitz, U.S. Geological Survey (February 2, 2005).

The current report needs some editorial review. The author needs to read through the document and supply the missing word(s). For example, on page 8, first full paragraph, fourth line "transported the container". Data is a plural, not singular. The tables that use abbreviations need explanations.

In addition to the limited number of editorial problems, I have several technical concerns as well.

For someone like myself, who received a copy of this report 'cold', without any information on the history associated with the development of this project, a clearer set of objectives/purposes would be appropriate. The current objectives section really doesn't say very much. Is this supposed to be a basic data collection exercise, to simply determine annual fluxes and, in turn, long-term trends, or are there additional objectives such as evaluation of remediation efforts, effects of land-use changes, meeting regulatory requirements, etc. There appears to be a concurrent modeling effort, but there is no indication of the purpose for the model. Also, why is the monitoring limited only to TSS and Cu? Why aren't other trace elements being monitored?

It's obvious that the author is concerned about how well the ISCO sampler collects representative samples. Further, there appears to be sufficient evidence that the suspended sediment in the creek contains sand-sized particles. Typically, sand-sized material is not evenly distributed in fluvial cross-sections. Further, as the author states, a pumping sampler can generate a size-fraction bias in the material it collects. Hence, since the ISCO is a pumping point sampler, there is some concern as to whether or not the samples it collects are, in fact, representative of the material in transport. Lacking any direct comparisons with manually collected representative samples (e.g., depth- and width-integrated isokinetic samples collected using either EDI or EWI) leaves doubt about the utility of the data for calculating annual fluxes, or simply concentrations relative to any existing regulatory limits.

Based on an examination of the graphs provided in the text, there appears, not surprisingly, to be relationships between flow, TSS, and total Cu. Why haven't these relationships been explored and evaluated/discussed. It may be possible to reduce the sampling and/or analytical load through the use of surrogates (e.g., discharge for TSS).

As currently constituted, the program is a stormwater monitoring program. As such, I have several concerns about the current sampling scheme, above and beyond the relatively simple issue of 'representativeness' of the ISCO samples. These include:

Hysteresis Issues: If the sampler reacts to changes in gage height, and the suspended sediment peak leads the hydrographic peak, then it is possible the first ISCO storm sample(s) will miss the 'first flush' of material coming off the floodplains or from the creek bed because it will pass the site before sufficient gage height is reached to trigger the sampler.

Antecedent Dry Period: Individuals working with storm sampling are very much aware that the magnitude of the TSS as well as the associated chemical levels can depend on the length of time between storms. Typically, the longer the antecedent dry period, the higher the levels of TSS and associated chemical constituents. What procedural/programmatic controls have been developed to deal with this issue? In regulatory situations, storms occurring within a 72 hour period or less from the previous event, are not sampled?
Temporal Resolution: The objectives section indicates a desire to generate annual fluxes, is this the maximum temporal resolution required? If fluxes for shorter intervals are required (e.g., daily for TMDL development), then more frequent measures of TSS and Cu may be required.

I have never been terribly fond of EPA analytical procedures for solid-phase materials. I much prefer total analysis because there is no ambiguity about the amount of a particular chemical constituent present. In the case of this particular study, has there been any evaluation of whether or not the primary source(s) of Cu are quantifiable using the current analytical procedure. My presumption, from the title of the report/study is that the presumptive major source of the Cu is highway/roadway runoff of brake dust coming from automobile traffic. If that is the case, has any attempt been made to ensure that all brake-associated Cu is quantified by the current analytical procedure? If not, the current analytical procedure may require reconsideration.

The current analytical procedure has a fairly substantial range of recoveries (75 to 125%). That's almost a factor of two, and I presume, is based on either determinations on replicate samples and/or reference materials. If it's the latter, how well does the standard matrix reflect actual sample matrices? If it doesn't, you need to find a different set of standards. Further, can current programmatic goals be achieved with results that could be ±100%? If not, then once again the analytical procedure may have to be altered.

Comments of Robert Holmes, Central Valley Regional Water Quality Control Board (February 3, 2005).

1. What is your assessment of the quality of the methods used and resultant data?

   The use of the EPA methods identified in the report may be used to generally meet the objectives of the study. Objectives were identified as "collect data for model development and analyses of long-term copper trends". Details of the model expectations were not provided. To meet the second part of the objective (long term analyses) would depend upon consistency in the approach over time (consistent sampling methods, analyses methods, QA/QC, etc). Such a plan for long-term consistency was not identified in the report.

   When monitoring concentrations and loads of contaminants in surface waters it is common to investigate such data under different hydrologic conditions to characterize the timing, magnitude, frequency, and duration of the contaminant. For example, if the interest is in storm water loads of contaminants – goals would typically include: to examine loads prior to storm events (baseflow conditions – background), to examine individual components of the hydrograph (rising limb, peak, decline) and to examine loads for periods of time after the hydrograph has declined. Also, goals typically include examination of such conditions with respect to rainfall events and non-rain events, and under different water years. Water year and the amount of rain received prior to and during sampling events have large influences on the transport of contaminants in aquatic systems. Such data are valuable for calibration and development of aquatic contaminant transport models. Also, calculations of aquatic loads are typically used for identification of sources of contaminants in a watershed. Concentrations are important for knowing aquatic toxicological significance. How will your sampling design (location of sites, sampling frequency, sampling duration, etc.) allow you to meet objectives? This is not clear in report.

Overall, the report is somewhat vague and simplistic. More information would allow for more critical analyses of the approach/methods and the potential for the study design and data collected to meet project goals and objectives. It is suggested that the goals and objectives be more clearly identified as a first step. What questions are the data intended to answer? What will the model be used for? Who will use the model? What information is needed from the model? Also, the term data is plural. Example from page 2 "rainfall data was collected" should be "were" collected. Example from page 4 " stream flow data was" should be stream flow data "were". Same comment applies throughout the report.

2. Previously collected monitoring data from the same watershed will be used to calibrate and/or validate the Brake Pad Partnership's watershed model. Do you recommend that these data also be used for this purpose?

   Development of models typically relies on the use of data from different studies. If the data from other studies meet your criteria for the model they may be used. Data acceptability criteria requirements of the model, to be collected as part of this study, and from other studies were not outlined in the report.

3. Based on your experience with similar monitoring efforts, what advice do you have regarding the interpretation and use of these data?

   Every study has strengths and limitations. The answer to this question depends on the intended use(s) of the data and the questions being asked. For example you may be able to report what was found in your study on days you studied. However, it is unclear to me based upon available information that – the copper that may be observed in storm events is due to brake pads (such background information would be useful in the report), that your results are typical of storm events of specific magnitudes and durations, that your results adequately reflect different hydrological conditions and water years, etc.

Comments of Jim Carleton, EPA Watershed Modeler (February 5, 2005).

General Impressions: The draft document provides a brief description of the recent BPP-sponsored water quality sampling conducted in Castro Valley Creek. The parts of the document that I found to be most informative and potentially helpful were the "water quality sample collection" subsections of the Methods and Results sections, which go into some detail about the sampling aspects of the project. However, as a presumably authoritative record on the BPP sponsored monitoring, I find this draft report to be lacking in much of the kind of detail one might hope to find in such a document.

Specifics:

1. The Introduction section states that the report provides "a summary of the rainfall, stream flow, and stormwater quality data collected…". The report does describe the water quality data in some detail, but where are the rainfall and stream flow data? Besides the average storm volumes listed in Table 1, and the streamflow data shown graphically in the figures in Appendix D, this information is apparently missing from the report. An unnamed table in the appendices, which appears to be a somewhat garbled printout of a spreadsheet, may contain some of this information. Unfortunately this table is impossible for the reader to interpret in its present form. I might be able to use the precipitation data to drive HSPF if it were provided. If not intended for this purpose, why were these data collected?

2. Regarding stream flow data, has the project involved the collection of any additional information beyond that which is already being collected continuously (and at no cost to the BPP) by the USGS at their stream gage on Castro Valley Creek? If so, this should be documented. If not, the study authors should not take credit for data collection they had no part in.

3. Where are appendices B and C?

4. No description of the analytical methods employed for copper and TSS is given, besides a listing of the numbers for the EPA methods used. It would be nice to at least know what sort of instrument was used to analyze the samples for copper.

5. Some interpretive synthesis and discussion of the monitoring data could be provided. The authors have undoubtedly looked at the data in more detail than anyone else, and may have gained insights into trends that the reader might not see, or otherwise have observations that would be worthwhile to include in the write-up.

Comments of Roger James, Water Resources Management (February 19, 2005).

My review of the subject report identified a number of issues that should be addressed before the results can be used in the development of a model of the discharge of copper from streams into San Francisco Bay. These include: study approach, representative sampling, analytical methods and sample management, data reporting and analysis and QA.

Study Approach

The reliance on total suspended sediment and copper data from a few storm events during a single year in development of a model is questioned. Considerable more data including bed sediment data collected over representative water years and storm events will be needed. The Steering Committee is strongly encouraged to carefully review Arthur J. Horowitz's, September 1995 IAHS Special Publication No. 4 "The Use of Suspended Sediment and Associated Trace Elements in Water Quality Studies". That publication identifies many issues with the study approach and raises significant questions about use of the data collected during the study. Foremost of those issues is the role that bed sediments play in estimating short and long-term temporal changes of suspended sediment concentrations and trace elements.

Horowitz notes the old adage about fluvial transport which says that "90% of the transport can take place during 10% of the time". My experience at Santa Clara Valley Water District was that the episodic high energy events generate the bedloads that must be considered in any modeling efforts. These events transport sediments deposited in storm drains (enclosed pipes), natural streams and open channels during the low flow events that are mobilized during high energy events. These represent the most significant source of suspended sediment and trace element loadings to San Francisco Bay.

The events monitored did not include the high energy events that would mobilize these sediments. The study was conducted during a period when rainfall was about 80% of the long-term annual average, runoff 76% of the long-term average annual flow, but perhaps most important the maximum daily flow that occurred on February 25, 2004 was only 117 cfs which is significantly below historic levels.

Representative Sampling

The report concludes that the single point sample is representative of the entire flow in the creek based on a qualitative assessment. There are multiple methods for collection of suspended and bedload sediments and analysis of those samples for particle size distribution during the study could have confirmed whether representative samples were being collected. Absent this type of data and based on multiple studies conducted by USGS it is not possible to conclude that representative samples were obtained.

Water quality monitoring studies have been performed using automatic samplers that found when suspended particles exceed 63µm (very fine sand) in runoff the sampler is not capable of obtaining representative samples. I have recently completed a review and summary of 17 different studies that have measured pollutants associated with various ranges of particle size of street dirt and runoff. Many of the studies contain data on copper concentrations and loadings by particle sizes. The largest loadings are associated with particles >63µm and most of the studies reported particles much greater than 150µm. One study conducted in the Bay Area on street dirt found that over 80% of the copper was associated with particles >177µm.

The particles >63µm travel as bedload and were unlikely sampled by the study. The USGS in cooperation with the Federal Highway Administration has published the "The National Highway Runoff Data and Methodology Synthesis", publication no. FHWA-EP-03-054, 2003 documenting these issues associated with monitoring and analyzing runoff.

<u>Analytical Methods and Sample Management</u>

The draft report indicates that EPA Method 160.2 was used to analyze for total suspended solids. USGS in a 2000 Policy, in multiple studies and discussed in the above referenced FHWA report have found that the TSS method of analysis significantly underestimates suspended sediment concentrations. USGS adopted the SSC method of analysis (ASTM 3977-97) as the reliable method for determining the concentration of suspended solids in urban runoff. USGS indicates that use of the EPA method can result in solids loadings that are erroneous by several orders of magnitude.

I understand that ToxScan in 2003 may have been using a modification of the EPA method that could mitigate this problem where they were processing the entire sample rather than a 100ml sub-sample as specified in EPA Method 160.2. I recommend that the exact sub-sampling and analytical procedures that were used by the study be documented before use of the TSS data.

The difference in suspended solids concentration between the ASTM and EPA methods is derived from the sub-sampling techniques (pouring and pipetting) used in the EPA method versus the processing of the entire sample used in the ASTM method. It is note worthy that the sub-sampling technique used for TSS analysis is also commonly used in obtaining sub-samples for analysis of other parameters. When this is used for heavy metals it is likely to result in higher concentration of metals than in the original sample.

A review and analysis of the discrete and composite TSS data where they overlap for some of the events monitored suggest that the laboratory was experiencing sub-sampling issues. Data from the December 14, 2003 is an example.

<u>Data Reporting and Analysis and QA</u>

The draft report would be significantly improved by performing an analysis of the collected data and comparison with historic data including:

- Statistical analysis of the relationships between suspended solids, copper, runoff rates, rainfall depths that correspond to the catchment's time of concentration, etc.
- Comparison of historic water quality data to data collected during the study and discussion and explanation of differences.

- Comparison of rainfall data to historic data and where possible include the above mentioned data.
- Comparison of runoff data to historic data.

The value of the report would also benefit from adding the following information:

- Include or cite the reference to the USGS methods for collecting stream flow data.
- Include or cite the reference to the methods used for collection of rainfall data by Alameda Flood Control District.
- More rainfall event data – start, stop intensity, antecedent dry conditions, maximum 10 minute rainfall depth, maximum depth of rain that occurred during an event that corresponds to the catchments Time of Concentration.
- A better description of the actual sampling equipment layout including lift and length of tubing from the sample intake strainer to the sampler. This will allow an analysis of the sampler's capability to deliver larger particles and whether sample line purging was adequate.
- Rainfall data for a station nearest to the USGS gauging station to see if there is significant difference at that point from the referenced stations.

Comments of Bill Selbig, Hydrologist, U.S. Geological Survey (February 25, 2005).

Thank you for the opportunity to review the subject report. I found the methods employed by the primary researcher to be similar to those used by typical USGS investigations. I was encouraged to hear there were several years of historical data in the Castro Valley basin. I sincerely hope the efforts put forth on this project will go a long way to mitigate source-area contributions of copper to San Francisco Bay. Below are several comments that highlight the major issues I would like to see addressed before final publication. There were several smaller talking points, but I refrained from including those in this response as it did not seem appropriate. Perhaps we can cover those issues during the upcoming conference call.

Page 1, Objective: The author mentions the data will be used to develop a model but does not describe which model. Is a model going to be developed from scratch? If so, will the results be compared to a model that has been used for numerous hydrology and water-quality studies such as HSPF, SLAMM, etc.? Is it a coupled hydrologic and water-quality model? It would be helpful to learn more about what the model data requirements are and how it uses that data to calculate something useful for water resources managers. Any model is really only as good as the data that goes into it. There is no mention of calibration and verification of the model with available historical data. This would go a long way to validate the model used in this study.

Page 1, Geographical Setting, 2nd paragraph: The stream setting is described but there is no mention of geomorphic conditions, in particular streambed composition. Wisconsin typically has very soft, silty streambeds which may be very different than a typical streambed in California. This is important when developing a sampling protocol as to whether bedload is or is not an important contributor to storm load. If the streambed composition were mostly sand and gravel then a large part of the sediment load is likely in the form of saltation and bedload.

Page 3, Stream Flow: A reference to the USGS standard methods for stream gaging should be provided.

Page 3, Water Quality Sample Collection, 2nd paragraph: The sample threshold seems to be set a bit too high above baseflow conditions. It is important to capture storm flow and associated contaminants that stem from the "first flush" effect. Setting a sample threshold too high will miss the first flush and will generally produce a negative bias in pollutant concentration whether it be time or flow-composite. This is especially true for rapid responding streams most often associated with urban drainages.

Similarly, the volume of water required to trigger a flow-composite sample appears to set to high. It is crucial to adequately represent the rising limb, peak, and recession limb of a hydrograph. By setting the sample volume to high, it is likely that the rising limb, the portion of the hydrograph typically associated with the greatest contaminant concentrations, is underrepresented. The author had a 24-bottle configuration in the ISCO sampler yet only collected a maximum of 12 samples per each event. A smaller sample volume threshold would more adequately represent the true event mean concentration.

Page 3, Water Quality Sample Collection, 3rd paragraph: There is mention of several years worth of historical data at the sampling site but no information is given as to how the current year's data compares. If it is beyond the scope of this report to go into those details then at least a reference to any published materials or the principal investigator would be appropriate.

Page 3, Water Quality sample Analysis: Total suspended solids is no longer the preferred test to measure the concentration of suspended solids in urban runoff. A recent USGS publication has determined suspended sediment as a more adequate test. In fact, when comparing suspended sediment concentrations to total suspended solids, the total suspended solids are underestimated by up to an order of magnitude. This is especially true for urban runoff containing particles greater than 63um.

There does not appear to be any mention of load computations. Although concentration data is an important component to a water-quality study, the effects of concentrations are more adequately represented in terms of a total event load. If the concern of the investigating party is to determine the impact of copper from brake wear debris on water quality in the South San Francisco Bay, then concentrations are only a part of the story. Previous studies have demonstrated an inverse relationship with particle size and concentration of sediment-associated trace metals. The configuration of the autosampler was preferential to smaller (<63um) particles. These particles tend to have the highest concentrations but also have the least amount of mass and therefore overall load to the Bay. Larger particles play an important role in determination of average concentrations of copper that might be used in the model. Since the autosampler demonstrated a bias against larger particles, little is known as to how much moved past the sampling station as bedload. Furthermore, previous studies suggest the type of sampler, the time of sampling, and the horizontal and vertical location of the sampling device can affect concentrations. Furthermore, there is no discussion as to the relevance of copper concentration and what it means to San Francisco Bay. Why is copper a problem in the first place? This discussion could be placed in the introduction but it may also reappear in this section. Copper is often bound to sediment very tightly such that it is not readily available for uptake by aquatic species. If the concern is toxicity, then a more complete discussion of copper concentration on a range particle size is warranted.

If copper and suspended sediment loads were computed then I would express caution when dealing with time-weighted samples. When examining the spacing of the time-weighted samples on many of the storm hydrographs, several of the samples were taken during what appears to be baseflow conditions. These should not be included. Similarly, some of the flow-weighted samples are spaced a fair distance apart and appear to cover two rather than a single event. The 11/14/03 event appears to have only a single sample for each rise in the hydrograph.

If copper and suspended sediment loads were not computed then there should be a discussion on the relationship of concentration and discharge. If the model uses only average concentration as its input variable it would be worthwhile to investigate whether copper and/or suspended sediment have any correlation with discharge. A simple regression could verify or dispute a relationship. It would also give the modeler some sense of variability and error associated with the concentrations used in the model.

Page 4, Rainfall, 1st paragraph: More attention should be given to rainfall data. A more complete statistical analysis of the data would provide useful information. For example, what were the intensities of each sampled event? Why was an average used instead of Thiesen polygons for an area-weighted average? How did the basin respond in terms of runoff percentage on a unit-area basis. All of the gages used in this study were in the headwaters of the basin. Large differences in rainfall can occur within a very small radius. A rain gage at the USGS monitoring station would have been useful. Was one available?

Page 4, Rainfall, 2nd paragraph: I understand the project budget limited the number of samples collected during the study period. However, if the data is to be used in a model that relies on average conditions then it is better to have concentration data for a variety of hydrologic conditions as well including baseflow vs stormflow, high-intensity vs low-intensity storms, and bedload vs suspended load.

Page 4, Stream Flow: The author provides the total flow for the 2004 water year and its comparison to the long term annual average but does not mention how the total flow during the study period compares to the long term average for the same period. Were the months sampled normal, above normal, or below normal? We are told how the flow measured for the year but perhaps the summer months were very dry and the winter months were very wet.

Page 7, Water Quality Sample Collection, 3rd paragraph: A quantitative assessment of the precision of the sampling method could be conducted if loads were computed. Since there were in fact co-located samplers, assuming the time-weighted and flow-weighted samplers were in the same location, the concentration results from each sampler could be used to compute an event load.

The author assumes the stream segment is well mixed where the samplers are located. This may be true but more information about the monitoring station and sampler configuration would be helpful. How far from and how high above the stream is the sampler located? As part of the quality assurance protocol several Equal Width Increment (EWI) samples should have been taken during both baseflow and stormflow conditions. This would help reduce some of the horizontal and vertical variability in sediment and sediment-associated trace element concentrations and provide a direct comparison to autosampler concentrations.

Page 8, Water Quality Sample Collection, 4th paragraph: More information on blank sample collection would be useful. It is unclear whether the blank sample taken after cleaning and subsequent blank samples actually tested each component of the sampling process. The sample tube is only one part of the process. A better test of the sampler integrity would include the tube, sampler head, collection bottle, Milli-Q, etc…essentially everything the sample water comes into contact with during the sample collection process. Also, there is not any discussion on the grab sample that was taken as a QA sample (11/08/03). It is interesting how the copper and zinc concentrations of this grab sample are elevated yet the TSS concentration is very low. This could be explained by how the grab sample was acquired.

Page 9, Water Quality Sample Analysis, Completeness: It is unclear what the author means by “the number of samples that ideally would have been collected and analyzed”. I assume the samples collected would cover 90% of the total storm volume based on the length of the storm. This is an appropriate metric to evaluate sample completeness but there is no discussion as to what defines a storm period.