Special Study Proposal: Advanced Data Analysis, Phase II

Summary: Reconnaissance data collected during single storms has provided good evidence to support enhanced management effort in watersheds with high PCB concentrations in water and on sediment particles. However, to date, such data has had more limited use for prioritization of management effort in watersheds exhibiting moderate or lower concentrations. This project proposes to enact the second phase of enhanced ranking and fingerprinting methods for the spatial prioritization and identification of watersheds, sub-watersheds and PCBs source areas. The outcome of this proposal will be a finalized stepwise methodology and application of that methodology to existing stormwater data sets to help prioritize areas for enhanced management or further sampling.

Estimated cost: $50k
Oversight Group: STLS/SPLWG
Proposed by: L McKee, A Gilbreath, J Hunt, J Wu, J Davis, and D Yee (SFEI)

Proposed Deliverables and Timeline

Task	Deliverable	2019											
		J	F	M	A	M	J	J	A	S	O	N	D
A	Complete refined methodology for identifying and ranking watersheds in relation to each other	!	!	!									
B	Complete refined methodology for identifying and ranking subwatersheds of importance within a larger watershed based on PCB fingerprinting	!	!	!									
C	Complete refined methodology for prioritizing source areas of importance within a larger watershed based on PCB fingerprinting	!	!	!									
D	For areas with no existing samples - develop and apply within a GIS, a stepwise statistically based method to predict new places in the landscape to look for PCBs	!	!	!									
E	Draft and final report					!!		!		!	!!		

! = STLS check in for review and course corrections
!! = SPLWG oversight and review

Background

During water years (WYs) 2011, 2015, 2016, 2017 and 2018, the RMP funded a stormwater characterization reconnaissance study aimed at locating high leverage watersheds and subwatersheds for management focus. Over these five years, a total of 81 sites have been characterized during at least one storm for PCB and Hg concentrations and particle ratios (the ratio of a pollutant concentrations measured in a stormwater sample to the suspended sediment concentration in that same sample) (McKee et al., 2012; Gilbreath et al., 2018 in SPLWG review; Gilbreath et al., in preparation). These data build upon and compliment a more detailed sampling program that was carried out at eight locations where samples were taken during multiple storms over a minimum of two years (WYs 2003-2010, 2012-2014) (Gilbreath et al., 2015; McKee et al., 2015).

While sampling a single storm has a relatively low cost, differing storm characteristics (intensity, duration, antecedent rainfall conditions) interplay with differing suspended sediment erosion and PCB source characteristics to confound comparisons between watersheds. Yet, to date, the primary method of interpretation has been one of simple ranking using two methods, directly measured PCB or Hg concentrations, and the ratio of concentrations of PCBs or Hg in the stormwater sample to the suspended sediment concentrations in the same stormwater. The most recent report (Gilbreath et al., 2018) ranked 79 sites from highest to lowest based on these two simple methods. Although this method coupled with other evidence including land use and source area characteristics, property surveys, and soil sampling has taken us a long way towards identifying a number of small industrial watersheds of management interest when concentrations are high, it has been less well suited to providing evidence and a rationale for how to prioritise management in watersheds exhibiting moderate or lower concentrations. However, it is recognised that there may be sources of interest in these watersheds also. There are a few main reasons for the failure to provide useful info across a broad range of concentration and watershed types:

1. Concentrations and particle ratios are indirect indicators of the real metric of interest (mass load)
2. The current ranking method has no adjustment for the dilution that occurs in mixed land use watersheds when stormwater and sediment derived from "cleaner" areas mixes with contaminated water from industrial areas and other PCB and Hg source properties. This dilution varies from watershed to watershed depending on the land use configuration but is is easy to conceive dilution factors as low as 2:1 and as high as 10:1.
3. There is no consideration of the PCB congener distribution (fingerprint) that may provide additional information about source types upstream or provide useful forensic evidence of the contribution of similarly fingerprinted water samples collected upstream in nested subwatersheds or similarly fingerprinted sediment and soil samples collected upstream adjacent to suspected source properties.

Recognizing these weaknesses, the Sources Pathways and Loading Workgroup of the RMP recommended an advanced data analysis project. The project aimed to mine and analyze all the existing stormwater data with the primary goals of developing an improved method for identifying and ranking watersheds of management interest for further screening or investigation, and to guide future sampling design. The RMP-funded project is underway in the first half of 2018. The outcomes of the 2018 project are expected to be:

1. The completion of version 1 of a new watershed intercomparison method for ranking based on mass for more accurate evaluation of relative contamination between sampling locations, completion of the documentation of how to apply the method and pilot demonstration of the method to a subset of sites (about a dozen). This new method will incorporate an uncertainty analysis as part of the methodology and include statistical methods such as Partial Least Square Regression and Random Forests Model to identify the most influential predictor variables for PCB concentrations and loads using a staged approach. First stage is to explore the relationship between the rainfall variables and dependent variables at each watershed. Second stage is to investigate the groups of

watersheds that were sampled during the same weather events to test the impact of landscape and source variables to help define critical factors of interest.

2. The development and refinement of PCB profile recognition methods (one based on aroclors and the other on homologs) and a pilot application in the nested sample locations Guadalupe River watershed to:
    a. relate stormwater data collected in a nested fashion to downstream stormwater samples (sub-catchment scale)
    b. relate sediment and soils data collected near source properties to downstream stormwater samples

This proposal is for a second year of funding to continue these promising data analysis techniques.

Study Objectives and Applicable RMP Management Questions

The main study objective is complete and apply improved methods for the spatial prioritization of areas for enhanced management interest. The methods are being developed and tested at a variety of scales relevant to management:

i) Watershed scale
ii) Nested subwatershed scale
iii) Source property scale

The outcomes are expected to be better support for decision making about how to spend limited resources on further investigations. Such investigations might include the initiation of enhanced sampling programs by BASMAA agencies to further characterize and identify source properties, and/or further stormwater reconnaissance characterisation sampling by the RMP or BASMAA to verify or refute the need for enhanced focus.

These objectives address management question (MQ) 1 & 3 primarily and 4 (secondarily).

MRP 2.0 Q1: Source Identification / Leverage: Which sources or source areas provide greatest opportunity for load reductions?
MRP 2.0 Q2: Impairment: Which source areas contribute most to impairment of Bay?
MRP 2.0 Q3: Management effectiveness: Provide support for planning future management actions or evaluate existing actions.
MRP 2.0 Q4: Loads: Assess POC loads, concentrations, or presence/absence.
MRP 2.0 Q5: Trends: What are the spatial and temporal trends in loads or concentrations?

Approach

Given the large variation in data available at each of the scales on management interest, improved methods for assisting management prioritization need to be developed at a variety of scales.

A. For watersheds where there are suspected sources but where there are no stormwater data in subwatershed (tributary locations) and no or limited data on soils and sediment concentration adjacent to potential source properties, data can be collected at the outlet of the watershed near the Bay margin and used (along with land use and source area maps and other local evidence to determine

importance relative to other watersheds. This watershed scale screening has been going on for years but the current simple ranking methods need refinement. The approach taken for refinement at this scale uses three types of evidence:

a. estimation of mass transport during a standard storm size. To do this, we have developed a mathematical adjustment methodology based on evaluation of data collected at the eight well sampled watersheds and made estimates of flow for each measured storm.
b. improved evaluation of land uses and source areas of interest for determination of yields (area normalization). To do this we used two methods - (1) the land use and source areas used in the RWSM and (2) we have explored statistical analysis of rainfall characteristics in relation to concentrations, particle ratios and loads to determine the most sensitive land uses and source areas for use in the area normalization step.
c. the use of congener patterns for recognizing watersheds of management interest (patterns that are atypical indicate follow up evaluation may be warranted)

In Phase I of the project (CY 2018) we developed beta versions of these methods for pilot set of watersheds (Guadalupe and nested sub-watersheds). In this, Phase II of the project, further refinement and testing on the larger data set will be completed.

B. For a smaller number of watersheds where stormwater samples have been collected in a nested fashion, a congener and homolog fingerprint methodology was piloted in CY 2018 for just one watershed (Guadalupe). In this, Phase II of development, the chosen method will be refined and applied to the remainder of the watersheds in the Bay Area where there are nested data available. The outcome will be an improved basis for prioritization of sub-catchments of greater management interest.

C. For a small group of source properties where sediment and soils data have been collected in adjacent public right of ways, a congener and homolog fingerprint methodology was piloted in CY 2018 for source properties in just one watershed (Guadalupe) using only recently collected (2015 and later) sediment and soils data. In this, Phase II of development, the chosen method will be refined and applied to the remainder of the watersheds in the Bay Area where there are nested data available. This may include exploring the use of older sediment and soils data for which there are congeners data available (the ~600 data data points collected prior to 2010). The outcome will be an improved basis for prioritization of sub-catchments of greater management interest.

D. For watersheds without stormwater data, pilot a new GIS methodology based on a statistical analysis of key factors that influence PCB concentrations and loads. The two options that will be explored are Partial Least Square Regression and Random Forest Regression. To reduce noise in the analysis, we plan to explore a subset of smaller watersheds (<6 sqkm), 70% of the data reported by Gilbreath et

al. (2018), and a group of watersheds that all have >50% impervious cover. Once, the key land use and source area factors have been statistically determined, patches with these factors with dimensions <1 sqkm will be ranked through GIS as places worth further management consideration.

E. The last task is to complete documentation that will include a stepwise series of methods for each management scale so that others can apply the methods themselves

Reporting

Outputs (Deliverables):
A technical report that provides the technical justification for the methods recommended and outlines in a stepwise fashion how to apply the methods to other Bay Area data sets. The draft will be presented in May and the final by November 2019.

Linkages to other STLS elements

Data collected in this project may provide enhanced methods for interpreting data that may be input for RAA, could be used to refine the calibration of the RWSM, and may provide useful context for development of the watershed loadings *Trends Strategy*.

References

Gilbreath, A.N., Hunt, J.A., Wu, J., Kim, P.S., and McKee, L.J., 2015. Pollutants of concern (POC) loads monitoring progress report, water years (WYs) 2012, 2013, and 2014. A technical report prepared for the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP), Sources, Pathways and Loadings Workgroup (SPLWG), Small Tributaries Loading Strategy (STLS). Contribution No. 741. San Francisco Estuary Institute, Richmond, California. http://www.sfei.org/sites/default/files/biblio_files/FINAL_WY2014_POC%20loads%20progress%20report.pdf

Gilbreath, A.N., Hunt, J.A., Yee, D., and McKee, L.J., 2018 (in SPLWG review). Pollutants of concern (POC) reconnaissance monitoring final progress report, water years (WYs) 2015 ad 2016. A technical report prepared for the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP), Sources, Pathways and Loadings Workgroup (SPLWG), Small Tributaries Loading Strategy (STLS). Contribution No. 840. San Francisco Estuary Institute, Richmond, California.

McKee, L.J., Gilbreath, A.N., Hunt, J.A., and Greenfield, B.K., 2012. Pollutants of concern (POC) loads monitoring data, Water Year (WY) 2011. A technical report prepared for the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP), Small Tributaries Loading Strategy (STLS). Contribution No. 680. San Francisco Estuary Institute, Richmond, California. http://www.sfei.org/documents/pollutants-concern-poc-loads-monitoring-data-water-year-wy-2011

McKee, L.J. Gilbreath, N., Hunt, J.A., Wu, J., and Yee, D., 2015. Sources, Pathways and Loadings: Multi-Year Synthesis with a Focus on PCBs and Hg. A technical report prepared for the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP), Sources, Pathways and Loadings Workgroup (SPLWG), Small Tributaries Loading Strategy (STLS). SFEI Contribution No. 773. San Francisco Estuary Institute, Richmond, CA. http://www.sfei.org/sites/default/files/biblio_files/MYSR%20Final%20Report.pdf

SFRWQCB, 2009. California Regional Water Quality Control Board San Francisco Bay Region Municipal Regional Stormwater NPDES Permit, Order R2-2009-0074, NPDES Permit No. CAS612008. Adopted October 14, 2009. 279pp. http://www.waterboards.ca.gov/sanfranciscobay/water_issues/programs/stormwater/Municipal/index.shtml

SFRWQCB, 2015. California Regional Water Quality Control Board San Francisco Bay Region Municipal Regional Stormwater NPDES Permit, Order No. R2-2015-0049, NPDES Permit No. CAS612008. November 19, 2015. 350pp. http://www.waterboards.ca.gov/sanfranciscobay/water_issues/programs/stormwater/Municipal/R2-2015-0049.pdf