RMP Special Study Proposal: Advanced Data Analysis, Phase III

Summary: Reconnaissance data collected during single storms have provided good evidence to support enhanced management efforts in watersheds with high PCB concentrations in water and on sediment particles. However, to date, such data have had limited value for prioritizing management efforts in watersheds exhibiting moderate or lower concentrations, yet these watersheds likely contain patches with elevated concentrations. This project proposes to analyze additional available data (likely all the remaining sites) using enhanced ranking and congener fingerprinting methods. The outcome of this project will be a completed application of these new methodologies to existing stormwater datasets 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)
Time Sensitive: No.

Proposed Deliverables and Timeline

Deliverable	**Due Date**
Compilation of data	01/2020
Analysis of data and presentation of draft results	02/2020
Draft report for STLS review	03/2020
Draft report for SPLWG review	04/2020
Presentation of results to SPLWG	05/2020
Address review comments and finalize report	08/2020

Background

During water years (WYs) 2011, 2015-2018, the RMP funded efforts to characterize a total of nearly 80 sites 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., 2016; Gilbreath et al., 2017; Gilbreath et al., 2018, Gilbreath et al., 2019). These data build upon and complement 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) (McKee et al., 2015). Water concentrations and particle ratios from all sites have been ranked from high to low to help identify watersheds for management consideration (Gilbreath et al., 2019). In addition, well over 1000 samples from soils and sediment in storm drains and around potential source properties have now been collected and are also being used to identify focal areas for management action (Yee and McKee, 2010; Geosyntec and EOA, 2017).

When concentrations are high, RMP samples, coupled with other evidence—including land use and source area characteristics, records review, reconnaissance surveys, and facility inspections—have identified a number of small industrial watersheds and properties within them of management interest. However, evidence and a rationale for management prioritization in watersheds and on potential source properties exhibiting moderate or lower concentrations are still lacking, thus catalyzing the funding of an advanced data analysis project in 2018 and 2019.

In 2018, three new techniques were developed and applied as a pilot study to support decision making. The first technique used data from seven well-sampled watersheds, and stormwater concentration data collected from 15 locations in the Guadalupe River watershed along with rainfall data, RWSM watershed specific runoff coefficients, empirical storm rainfall-load relationships and land-use data to estimate storm based yields for standard sized storms. The pilot study results from all 22 locations provide new evidence on areas to consider for management. For example, East Gish Road Storm Drain increased five places in priority based on concentration and nine places based on particle ratio. Charcot Avenue Storm Drain dropped nine places in priority based on concentration and seven places based on particle ratio. East Gish Road Storm Drain, an area of just 0.438 km^2 (0.19% of the Guadalupe watershed area) had an estimated “standard” load of 0.36 g or 2.6% of the total Guadalupe River load and the relative yield of 4.3-fold that of Guadalupe River as a whole. These results were not predicted from the field-based concentrations and particle ratios, which were less at East Gish Road Storm Drain than those measured for Guadalupe River watershed as a whole.

The second technique used congeners representative of the four most commonly used Aroclors as indicators to estimate relative contributions of source areas in watersheds to the concentrations and congener profile seen at the outlet. In the Pulgas Pump Station watershed, the results indicated that outflow was uniquely dominated by Aroclors 1242 and 1260 due to two distinct source areas in the watershed. If eliminated, it was estimated that the load exported could be reduced by 50% or more. For the Coyote Creek watershed, the similarity in congener profiles for the highest concentration sediment samples and the stormwater samples suggest that the important source areas in the watershed have been identified and that reduction of loading from an area at the south end of the Charcot Avenue Storm Drain watershed would yield the greatest reduction in export at the Coyote Creek station. For the Guadalupe River watershed, the concentrations and congener profiles in stormwater and sediment from the Guadalupe River watershed indicate the presence of one source area that is likely a substantial contributor to PCB export from the watershed, and that all of the significant source areas may not yet have been identified.

Overall, the results from the two methods provide additional supporting evidence for management focus and that reducing loads in sub-areas with higher yields or specific congener profiles within the larger watersheds may help speed the recovery of the nearby receiving waters and eventually the Bay as a whole. In addition, a third method was trialed using a combination of partial least squares regression (PLSR) and random forest (RF) routines to explore the relationships of PCB water concentrations, particle ratios, and area yields with a large array of candidate climatic and landscape variables. The most influential variables identified largely matched our expectations (e.g., storm load correlated to total event rainfall, or yield [load per area] inversely correlated to percent open space), but some results were ambiguous (e.g., positive or negative correlations to antecedent rainfall, depending on watershed). At this time, we were unable to identify a combination of variables that point to new sites of interest or alternative methods for accounting for other factors (e.g., for normalizing ranking metrics). Further exploration is expected with 2019 funding.

With method development complete (2018 funding), $50k of funding in 2019 is being used to further refine these pilot applications and test them in a greater number of watersheds. The 2020 funding would support the continuation of these applications to the remaining areas of interest.

Study Objectives and Applicable RMP Management Questions

The study will provide information essential to understanding the relative importance of watersheds, subwatersheds, and source areas for management consideration in relation to reducing loading-based impacts of PCBs in San Francisco Bay. The objectives of the project and how the information will be used are shown in Table 1 relative to the RMP's high-level management questions.

Table 1. Study objectives and questions relevant to RMP management questions.

Management Question	Study Objective	Example Information Application
Q1: What are the loads or concentrations of Pollutants of Concern (POCs) from small tributaries to the Bay?	Use yield and congener-based analysis methods to identify watersheds, subwatersheds, and source areas for management consideration.	The project will provide new insights into relative loads for about 100 watershed locations.
Q2: Which are the "high-leverage" small tributaries that contribute or potentially contribute most to Bay impairment by POCs?		The project will identify high yielding areas and areas with unique congener patterns in relation to general concentrations found in Bay sediment and fish tissue.
Q3: How are loads or concentrations of POCs from small tributaries changing on a decadal scale?		
Q4: Which sources or watershed source areas provide the greatest opportunities for reductions of POCs in urban stormwater runoff?		The study will compare yields and congener profiles for individual source areas to those of other sources area within subwatersheds and watersheds to identify which may be of higher interest.

Q5: What are the measured and projected impacts of management action(s) on loads or concentrations of POCs from small tributaries, and what management action(s) should be implemented in the region to have the greatest impact?		

Approach

We will continue to apply these promising data analysis techniques to watershed data collected in 2019 by the RMP (and data provided by BASMAA over the past five years).

Method 1. Yields-based Comparisons

Yields for watersheds where there are suitable concentration data collected during storms will be estimated using the following steps:

1. Using storm-based rainfall data and estimates of runoff from the RWSM, storm volume will be estimated
2. The storm volume will be combined with estimated event mean concentration data to estimate storm loads
3. Storm loads will be adjusted to a standard sized storm
4. Standard storm yields will be estimated by dividing the loads by the source areas of interest in the given watershed or subwatershed
5. These will then be used to compare sites and estimate relative leverage to identify areas for management consideration

Method 2. Congener Profile-based Comparisons

The relative contributions of four different Aroclor mixtures (1242, 1248, 1254, and 1260) in stormwater and sediment will be estimated using the following steps.

1. Determine the percent contributions of four relatively unique indicator congener sets for each Aroclor that are a major contributor to the overall sum of PCBs (Aroclor 1242: PCBs 18, 28, 31, 33; Aroclor 1248: 44, 49, 66, 70; Aroclor 1254: PCBs 487, 101, 110, 118; Aroclor 1260: PCBs 149, 170, 180, 187).
2. For each sediment, soil, or water sample, compute the index as the sum of the percent contributions of the indicator congeners for each Aroclor.
3. Standardize the index for each Aroclor as a percentage of the sum of the four indices.
4. The data for the Aroclor indices are then binned into the following categories
    a. greater than or equal to 40% of the sum of Aroclor indices (primary contributor);
    b. greater than or equal to 20% and less than 40% of the sum of Aroclor indices (secondary contributor); and
    c. less than 20% of the sum of Aroclor indices (minor contributor).
5. Compare the profiles found in individual sediment and water samples in upstream areas of a watershed with those observed at the outlet further down to

make inferences about relative contributions.

Method 3. Statistical Data Exploration
For the PLSR, the initial analysis examined each watershed individually, using only climatic variables. Continued work would expand to:

1. Examine all watersheds at once with land-use factors
2. Test transformations (e.g., log, since PLSR expects linearity)
3. Incorporate interactions (match/normalize lag and integration periods to watershed size)

For the Random Forest Approach, it could continue by:

4. Trying other Random Forest routines/packages
5. Examining other ranking metrics
6. Using bigger runs (more stable outcomes)
7. Testing subsets of watersheds or adding new watersheds (to test representativeness, impacts of permutations)
8. Deepening the review of outputs to better understand ranks/responses (e.g., negative vs positive correlations for different watersheds of some variables)

Budget

The following budget represents estimated costs for this proposed special study (Table 2). At this time, it is not known which method or methods will be applied. Results from 2019 work are not yet available.

Table 2. Proposed Budget.

Expense	Estimated Hours	Estimated Cost
Project Staff	280	$36,400
Senior Management Review	40	$8,000
Project/Contract Management		
Data Technical Services		
GIS Services	40	$5,600
Grand Total	360	$50,000

<u>Budget Justification</u>
Experience from work completed in 2018.

References

Geosyntec and EOA, 2017. Clean Watersheds for a Clean Bay (CW4CB), Final Report. http://basmaa.org/Portals/0/documents/Project%20Reports/Final_Report/CW4CB%20Project%20Report_No_Appendices_.pdf

Gilbreath, A.N., Hunt, J.A., Yee, D., and McKee, L.J., 2017. Pollutants of concern reconnaissance monitoring final progress report, water years 2015 and 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. 817. San Francisco Estuary Institute, Richmond, CA. http://www.sfei.org/sites/default/files/biblio_files/SFEI%20WY2016%20POC%20monitorng%20report%20FINAL%20June2017.pdf

Gilbreath, A.N., Wu, J., Hunt, J.A., and McKee, L.J., 2018. Pollutants of concern reconnaissance monitoring final progress report, water years 2015, 2016, and 2017. A technical report prepared for the Regional Monitoring Program for Water Quality in San Francisco Bay (RMP). Contribution No. 840. San Francisco Estuary Institute, Richmond, CA. https://www.sfei.org/documents/pollutants-concern-reconnaissance-monitoring-water-years-2015-2016-and-2017

Gilbreath, A.N., Hunt, J.A., Yee, D., and McKee, L.J., 2019 (in final revisions). Pollutants of concern (POC) reconnaissance monitoring final progress report, water years (WYs) 2015, 2016, 2017 and 2018. 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. 942. San Francisco Estuary Institute, Richmond, CA.

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, CA. 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). Contribution No. 773. San Francisco Estuary Institute, Richmond, CA. http://www.sfei.org/sites/default/files/biblio_files/MYSR%20Final%20Report.pdf

McKee, L.J., Gilbreath, A.N., Yee, D., and Hunt, J.A., 2016. Pollutants of concern (POC) reconnaissance monitoring draft progress report, water year (WY) 2015. 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. 787. San Francisco Estuary Institute, Richmond, CA. http://www.sfei.org/documents/pollutants-concern-poc-reconnaissance-monitoring-final-progress-report-water-year-wy-2015

Yee, D., and McKee, L.J., 2010. Task 3.5: Concentrations of PCBs and Hg in soils, sediments and water in the urbanized Bay Area: Implications for best management. A technical report of the Watershed Program. Contribution No. 608. San Francisco Estuary Institute, Oakland, CA. http://www.sfei.org/documents/concentrations-pcbs-and-hg-soils-sediments-and-water-urbanized-bay-area-implications-best