Optimizing sampling methods for pollutant loads and trends in San Francisco Bay urban stormwater monitoring

Technical Report

Melwani, A.R., Lent, M., Greenfield, B.K., and McKee, L.

July 2010.

REPORT SUMMARY

The purpose of this document is to summarize efforts to evaluate the optimization of sampling methods for pollutant loads and trend monitoring at Guadalupe River (GR) and Zone 4 Line A (Z4LA). This report presents a technical evaluation of sampling methods, load estimators, and strategies for storm selection. The sampling optimization focused on Hg, PCBs, and suspended sediments (SS) since these are the high management priorities in San Francisco Bay. The information summarized here will facilitate further discussion to develop appropriate study designs to address MRP questions and priorities at these and future sites. The focus of this study was to evaluate sampling designs for obtaining annual loads estimates. The study included two components:

- Comparison of the accuracy and precision of a variety of stormwater monitoring designs and mathematical equations (estimators) for determining annual pollutant loads; and
- Determining the power and sample size needed to detect declining trends in Hg and PCBs in the next 10 - 40 years.

The MRP default design is the automated sampling of four random storms using a composite sample method. The estimated range in bias (- 50 – 13%) and standard error (4.3 – 6.5%) for the default MRP method was among the highest of the designs evaluated. Alternatives were explored such as increasing the number of samples and storms to six or 10 storms using a composite sampling method. Although sampling of 10 storms would provide better precision than four or six storms, a design with 10 storms would likely exceed budgetary limits. A six storm sampling strategy was simulated to include the first flush and largest storm. This design produced a similar range in bias (-16 – 31%) and standard error (1.4 – 3.6%) to the sampling of four storms (-13 – 57% and 2.2 – 5.0%, respectively). It is likely that the small improvement in precision with six storms would not warrant the extra on-going cost for this design, but inclusion of first flush and largest storms may warrant consideration. Automated sampling of two, four, or six storms using a discrete sampling method was also explored. The total number of samples was assumed to remain the same in each scenario, thus the range in bias (-7 – 4%) and standard error (0.1 – 1.4%) of these designs did not change. The best configuration was four storms (3 samples per storm).

The addition of turbidity was also explored using the turbidity surrogate regression estimator for the loads calculation method. This method produced the highest accuracy and least bias of all the alternative designs. To use regression on the turbidity surrogate records for estimating annual loads, at least 10 but ideally 16 samples per year should be collected at each site. Given results from the discrete among-storm evaluations, it is likely that scenarios that include first flush and one of the largest storms of the year would provide more robust loads estimates than random sampling alone when applying the turbidity surrogate method.

Power for detecting trends appeared to be possible with just 10 samples collected per year, based on a preliminary scenario in which the samples were randomly selected and did not confirm to any of the tested sampling designs.

ACKNOWLEDGEMENTS

The authors are grateful to The Regional Monitoring Program for Water Quality in San Francisco Bay (RMP), through its Sources, Pathways, and Loadings Workgroup (SPLWG) for funding this study. We thank Jim Griswold from Western Ecosystems Technology, Inc. (www.west-inc.com) for statistical guidance. The project work plan and draft report were guided by regular and timely input received from Arleen Feng (ACPWA / BASMAA), Richard Looker (Water Board), and Chris Sommers (EOA / BASMAA).

This report can be cited as:

Melwani, A., Lent, M., Greenfield, B., and McKee, L., 2010. Optimizing sampling methods for pollutant loads and trends in San Francisco Bay urban stormwater monitoring. A technical report for the Sources Pathways and Loading Workgroup of the Regional Monitoring Program for Water Quality. San Francisco Estuary Institute, Oakland, CA. Final Draft.

TABLE OF CONTENTS

INTRODUCTION

Worldwide, coastal ecosystems adjacent to large urban, industrial and agricultural centers are subject to contamination, toxicity, and subsequent demise of wildlife and fisheries (Lauenstein and Daskalakis 1998, Linkov et al. 2002, Trimble 2003, Newton and Mudge 2005). In response, environmental laws in many countries are being developed and implemented to slowdown or reverse the process of contamination and even restore lost ecosystem attributes. In many cases, estimation of ecosystem-scale mass loads emanating from sources is one of the first data requirements needed to develop a plan of action (e.g., Godfrey et al. 1995, Schiff and Bay 2003, Balcom et al. 2004). San Francisco Bay is one such ecosystem that has been highly impacted by a history of urban, industrial, agriculture and mining land uses spanning about 150 years (Flegal et al. 2007). Approximately seven million people currently live in the nine counties bordering the Bay, and runoff and contaminants from mining legacies, urban areas, and agriculture drain to the Bay from about 37% of California (McKee et al. 2006a, David et al. 2009). Today, mercury (Hg) and polychlorinated biphenyls (PCBs) are considered the greatest threat to human and wildlife uses of the Bay (Conaway et al. 2007, Davis et al. 2007, Flegal et al. 2007, Yee et al. in review). However, there are also concerns about a number of emerging contaminants (Oros et al. 2003, Hoenicke et al. 2007, She et al. 2008).

In San Francisco Bay, urban runoff is considered one of the largest controllable sources of pollutant discharge. Total maximum daily loads (TMDL) reports written by the local Regional Water Quality Control Board (Water Board), summarize current estimates of loads from the main sources and pathways (urban and industrial wastewater, urban stormwater, Central Valley rivers, atmospheric deposition). The TMDL reports also argue for studies linking loads and toxic effects to beneficial uses, and provide loads allocations for each source and pathway (SFRWQCB 2006, 2008). The allocations are particularly stringent for urban stormwater and allow for 82 kg of Hg and 2 kg of PCBs with the objective of improving water quality in the Bay to desirable standards in 20 years (2028 for Hg and 2030 for PCBs). These represent estimated reductions of 50% and 90% over the present load estimates of 160 kg of Hg and 20 kg of PCBs, respectively. However, these load estimates remain uncertain, since measurements have only been made in a few of over two hundred tributaries (SFEI 2010).

The Regional Monitoring Program for Water Quality in San Francisco Bay (RMP), through its Sources, Pathways, and Loadings Workgroup (SPLWG), has been conducting tributary loading studies for nine years. The focus has been to provide information on sediment and pollutant transport processes in urban watersheds around the Bay (McKee et al. 2004, McKee et al. 2005, McKee et al. 2006b, Davis et al. 2007, Oram et al. 2008, David et al. 2009, McKee and Gilbreath 2009). The primary objective of these studies has been to achieve precise and unbiased estimates of loads of particle-associated pollutants-of-concern (particularly, Hg, PCBs, and suspended sediments). Most of the sampling effort has been focused on three locations: Mallard Island on the Sacramento River; Guadalupe River in San Jose; and the Zone 4 Line A flood control channel in Hayward. At all three study locations, a turbidity surrogate methodology has been employed, as it has been reported to be an appropriate and cost-effective method for accurate and unbiased particulate loads calculation (Grayson 1996, Wall et al. 2005). The tributary loading studies have provided valuable information for the development of the

San Francisco Bay and Guadalupe River Hg TMDLs (Austin 2006, SFRWQCB 2006), and the Municipal Regional Stormwater NPDES Permit (MRP) (SFRWQCB 2009).

The TMDLs and the MRP call for the Bay Area Stormwater Management Agencies Association (BASMAA) to improve loads information. In response, the RMP developed a small tributaries loading strategy (STLS) to guide the development of loads information over the next five years and to ensure coordination between the RMP and BASMAA. The STLS and Provision C.8.e. of the MRP aim to answer the following management questions:

1. Identify which Bay tributaries (including stormwater conveyances) contribute most to Bay impairment from pollutants of concern,
2. Quantify annual loads or concentrations of pollutants of concern from tributaries to the Bay,
3. Quantify the decadal-scale loading or concentration trends of pollutants of concern from small tributaries to the Bay, and
4. Quantify the projected impacts of management actions (including control measures) on tributaries and identify where these management actions should be implemented to have the greatest beneficial impact.

All of these questions require some level of information on the concentrations and mass loads in tributaries but the focus here is on optimization of study design for questions 2 and 3.

There are a number of sampling methods and corresponding mathematical calculation methods available for developing mass loading information. The optimal balance of sampling frequency within and among storm events is important to achieve precise, accurate, and cost effective loads measurements. Several methods, such as random and time-interval based sampling designs, have already been evaluated in other studies and essentially rejected as ineffective methods for evaluating tools towards our management questions (Walling 1985, 1988, Leecaster et al. 2002, Ma et al. 2009), and thus need not be evaluated further. Other calculation methods, including flow-weighted means, have been tested previously in southern California (Leecaster et al. 2002, Ma et al. 2009), and additional methods (turbidity surrogate, simple means, and linear interpolation) were examined in this study.

The focus of this study was to evaluate sampling designs for obtaining annual loads estimates. The study included two components:

1. Comparison of the accuracy, precision, and cost of a variety of stormwater monitoring designs and mathematical equations (estimators) for determining annual pollutant loads; and
2. Determining the power and sample size needed to detect declining trends in Hg and PCBs in the next 10 - 40 years.

High quality loading data from local watersheds collected by the SPLWG and RMP provided a resource to evaluate potential future monitoring approaches. A variety of sampling and mathematical loads calculation methods were simulated by statistically subsampling the existing high temporal resolution empirical data sets. Combining empirical data with simulation methods to test and optimize loading measurements has been carried out in numerous studies before (e.g., Walling 1985, Walling 1988, Leecaster et al. 2002, Ma et al. 2009). This study focused on sampling optimization for Hg, PCBs, and suspended sediments (SS) since these are the high management priorities in San

Francisco Bay, and suspended sediment concentration and load is an important vector for transport of sediment-associated pollutants. However, the findings are likely relevant for other particulate substances in similar sized watersheds.

METHODS

Three years of urban runoff data from the Guadalupe River (GR) and Zone 4 Line A (Z4LA) monitoring stations in San Francisco Bay, California were statistically analyzed in this study. GR is located near San Jose, the largest city in the San Francisco Bay Area. Its watershed is the fourth largest in the Bay Area (about 500 km^2) and is a network of mostly natural channels that have been modified by impoundments and flood control engineering. The monitoring station operated by the SPLWG developed through collaboration with the US Geological Survey (USGS) (station number 11169025) is upstream from tidal influence, but resides downstream from five main reservoirs, the City of San Jose, and the majority of flood control channels. The typical flood hydrograph produced by heavy rainfall passes through this watershed over a period of about 12-24 hours but larger and later season floods may last for several days. The free flowing area downstream from reservoirs is 236 km^2, of which approximately 80% is urbanized landscape. In addition, this area also drains the Quicksilver County Park, formerly the New Almaden Mining District, which, since 1849, has produced 6% of the total Hg worldwide (McKee, L., unpublished data) and is a known source of Hg to San Francisco Bay.

In contrast, the monitoring station in Z4LA, located in Hayward, drains a relatively small 4.47 km^2 watershed of completely urbanized landscape with over 38% industrial land use. Historically, there was no creek draining this area of the Bay margin. The flood channel of Zone 4 Line A is entirely engineered with approximately one-third open to the air and two-thirds underground culverts and storm drains. The monitoring station resides approximately 1.7 km from the Bay and upstream from tidal influence. There are no reservoirs in this watershed and rain passes largely unabated through the network of flood channels in minutes to hours.

Though an excellent data set is also available for the Mallard Island on the Sacramento River for a range of trace contaminants (David et al. 2009), these data were not included in the present study. There were two primary reasons for this: 1. the system is not representative of typical small tributaries to be monitored in the MRP, due to extreme size of the watershed (covering two-thirds of the land area of California); 2. time and resource limitations precluded the detailed examination necessary to evaluate optimal sampling design for this system. Since the system is much larger than GR, Z4LA, and others watersheds previously studied (e.g., Leecaster et al. 2002, May and Sivakumar 2009), and has considerably higher water volume and longer flood wave travel times, an optimal sampling strategy for this system is likely quite different. Performing a similar exercise for larger watersheds such as represented by our sampling station on the Sacramento River at Mallard Island remains a data gap in the published literature.

Three water years (WY) of data that spanned a range in climatic conditions (and thus a range in wet weather discharge and pollutant concentrations) were chosen for each watershed. A water year begins October 1st each year and ends September 30th and is designated by the end date. At GR, during WYs 2003 – 2005, the peak discharge was 172 m^3/s on December 16th, 2002 in WY 2003 (Table 1). In comparison, WYs 2004 and 2005 had relatively lower maximum discharge and pollutant concentrations. At Z4LA, three water years of data were also available, although not all years have complete records due to permitting issues and upstream construction. During WYs 2007 - 2009, Z4LA wet season discharge varied from 4.7 – 6.7 m^3/s. WY 2008 exhibited relatively higher peak wet season discharge than the other two years, but this was predominated by many small storms.

"Best Estimate" Loads Calculations

Statistical simulations of sampling designs were compared to existing "best estimates". The "best estimate" of annual loads (based on wet season data) was determined for each watershed based on the mathematical combination of estimated pollutant concentration and discharge volume. Hg, PCBs, and suspended sediment (SS) loads were examined. At both study locations, a turbidity surrogate regression (TSR) methodology has been used. Specifically, turbidity was monitored at short time intervals (15 minutes or less) and a statistical regression developed with a subset of water samples analyzed for suspended sediment concentration. This turbidity: SSC regression was combined with the continuous turbidity measurements to generate a time-continuous SSC record.

Additional depth integrated water samples (10-40 samples per year) were collected manually during high-flow events (storms), and analyzed for trace contaminants. Clean hands protocols were used. All analytical results were certified by the RMP data management and quality assurance plan (Lowe et al. 1999). Subsequently, during well sampled floods, linear interpolation was used to estimate concentrations between data points which were then combined with short interval flow measurements to determine loads. During storm periods when no sampling was conducted or during dry weather flows, regression relationships were determined between turbidity and each pollutant and used to calculate time-continuous estimates of contaminant concentration (turbidity surrogate regression or TSR). These estimates were then combined with discharge measurements to calculate loads. These combined methods were used to determine the "best estimate" of annual pollutant concentrations and loads to these watersheds over the years. It has been acknowledged in previous reporting (McKee et al. 2006b) that under complex conditions (e.g., Guadalupe River in 2005) professional judgment was used to guide these calculations. For example, Hg loads in GR were often stratified based on the predominant source of runoff indicated on hydrographs, resulting in separate regression relationships for urban vs. non-urban signals. These professional-judgment-based turbidity surrogate load estimates were used as the best available load estimates, against which all sampling design scenarios were compared.

Loads Analysis

Sampling programs for watershed loads estimates are designed with two attributes in mind; the number of samples taken within a storm and the number of storms sampled during a year or wet season (Leecaster et al. 2002). Our analysis of the optimal sampling method was performed in two steps. First, within-storm load estimates were compared among sampling designs and calculation methods. Secondly, using the results of the optimal within-storm sampling strategy, scenarios among storms were examined. A number of prospective designs were considered, including variations on sample collection and loads calculation within individual storms (i.e., within-storm designs), and sampling designs across the wet season (i.e., among-storm designs). All designs were evaluated by Monte Carlo simulation. Bias and precision were calculated to evaluate all design scenarios. Bias was calculated as the median percentage difference between the expected estimate and the actual results. Precision was calculated as the variability in bias, measured as the standard deviation. Results were calculated and compiled using the statistical software package R (v. 2.10.1).

The within-storm design analysis considered four aspects of sampling design: 1. storm sampling basis (i.e., flow based vs. turbidity based); 2. sampling emphasis; 3. sample size; and 4. loads estimation method (Table 2). To examine the first aspect of the analysis, each year of data for the two watersheds was analyzed for the presence of sampled storms using a) flow thresholds and b) turbidity thresholds (Table 3a and 3b). The use of a pre-set threshold simulates protocols for an automated sampler, which will collect water samples when preprogrammed thresholds are surpassed that characterize flow and concentration during each storm event hydrograph. Since both flow and turbidity increase during high-flow events, both were evaluated for use as primary drivers for sampling design. To define and select a storm, thresholds were statistically established for the start, peak, and end of each storm hydrograph. Storm selection criteria differed between the two watersheds. For GR, flow events greater than 200 cfs, with peak flow greater than 736 cfs, were flagged as storms (Table 3a). For Z4LA, flow events greater than 5 cfs, with peak flow greater than 26 cfs, were flagged as storms. The storm selection criteria were chosen to achieve thorough coverage of storm flow events, without including baseflow events. For the purposes of this analysis, flow that did not meet these criteria was deemed baseflow. In contrast to flow thresholds, turbidity-based storm selection thresholds were similar between GR and Z4LA (Table 3b). For GR, turbidity measurements greater than 30 NTU, during storms with peak turbidity greater than 84 NTU, were flagged as storms. For Z4LA, the thresholds were 30 NTU and 89 NTU, respectively.

The second aspect of the within-storm design was sampling emphasis, which refers to relative frequency of sample collection in the rising vs. falling stage. Two approaches were considered: a) equal spacing of the samples across the rising and falling stages (i.e., 1:1 sampling emphasis) or b) rising-stage emphasis, where twice as many samples were spaced on the rising stage of the hydrograph relative to the falling stage (i.e., 2:1 sampling emphasis). The rationale for considering a rising stage emphasis is that suspended sediment pollutant loads are typically greater and more variable during the rising stage (McKee et al. 2006b).

The third aspect of within-storm design evaluated was the number of water samples collected per storm (i.e., sample size). For Hg and SS, four sample sizes were considered: 6, 12, 18, and 24 collections per storm. For PCBs, 6 and 12 collections per storm were considered; larger numbers of collection would be unfeasible due to the large sample volumes required for PCB lab analysis. The actual number of samples that could be evaluated for each scenario varied based on the size of each storm sampled.

The final aspect of within-storm design was the loads calculation method (Table 4, Equations 1 - 3). Methods for loads calculation will depend on the method used to integrate individual water collection events (Leecaster et al. 2002). Specifically, auto samplers may obtain discrete or composite samples. Discrete samples are small samples (referred to as "sips") taken by the auto sampler throughout the storm. Composite samples are the combined collection of many discrete sips that are used to represent conditions over an entire storm. The data generated from composite samples collected in this manner are often referred to as event mean concentrations (USEPA 2002, Ma et al. 2009). For the discrete sampling method, two loads calculation methods were examined: 1) linear interpolation; and 2) flow weighted mean (Table 2). For the composite sampling, only a simple mean method was used to estimate loads because the other estimators require discrete data. These three loads calculation methods were tested for all combinations of sampling basis (flow or turbidity), number of samples collected per storm, and sampling stage emphasis (Table 2). The resulting within-storm load estimates were compared to the "best estimate" loads to assess performance of the sampling design and loads calculation methods.

The among-storm design evaluations focused on number and type of storms sampled. Using the results from the optimum within-storm design, three strategies for sampling among storms were considered for their ability to estimate annual loads (Table 5). The first among-storm design (Design A in Table 5) sampled the first flush (i.e., the first storm of a wet season) plus a variable number of random storms. The second among-storm design (Design B in Table 5) sampled the first flush plus one of the three largest storms of the wet season and a variable number of random storms (Note we chose one of the three largest because although it is easy to define and then respond to a weather forecast for a large storm, we may also miss a storm that ends up larger than the forecast predicted - we can never know until the end of the season if we sampled the largest storm of the season or one of the three largest). The third among-storm design (Design C in Table 5) is the design written in the MRP and was evaluated using either 2 or 4 random storms. Designs A and B were evaluated for 2, 4, 6, and 10 storms (actual number depends on available data). To correspond with the MRP requirements, Design C evaluated 2 and 4 storms only. All results were extrapolated to an estimated annual load by dividing by the ratio of sampled storm flow volume vs. total wet season flow volume (Table 4, Equation 4). Note that WY 2008 at Z4LA was deemed inappropriate for this analysis since the sampling began later in the season, and thus, an assessment of first flush and largest storms was not possible.

The accuracy and precision for annual loads calculation using each sampling strategy was compared. Monte Carlo simulations were employed to obtain random storm subsamples under each design method. Each design was run 50 times for each number of storms (to allow for random selection of storms), and an annual load was calculated for each run. The optimum strategy was identified as the design with the median closest to the annual best estimate of load, and the lowest variability in estimated annual loads. Each year of data was analyzed separately to demonstrate performance under a variety of climatic conditions.

A parallel analysis was performed to examine performance of the turbidity surrogate regression method for annual pollutant loads estimation. First, Monte Carlo simulation was employed to examine the sample size requirements for developing a relationship of turbidity to Hg, PCBs, and SS. This was performed for each design within each year of data. The optimal sample size was examined by varying the number of samples in the TSR from 4 – 40 using Monte Carlo simulation (1000 simulations per design/year combination). Again, the actual sample sizes that could be simulated varied based on the empirical data sets. Each regression generated from a sub-sampled data set was converted into a continuous estimated pollutant concentration record by applying the regression to the continuous turbidity record. The continuous pollutant concentrations were then extrapolated to loads using the same methods as for the mass emission estimator. These annual loads based on sub-sampled data were compared against loads calculated using each water years' complete grab sample data set to determine how many grab samples were necessary to obtain precise and unbiased load estimates. Once the optimum number of TSR samples was identified for each pollutant, the average regression slope and intercept (from the 1000 runs at the optimal sample number) were applied to the continuous flow and turbidity records to calculate per-storm loads. Finally, the TSR-based storm loads were extrapolated to annual loads using the same three among storms sampling strategies (Table 5), employing 50 runs per design. The performance of the TSR in the among-storm designs was compared to the TSR loads using all samples collected in a year, and the "best estimate" of annual loads.

Trend Analysis

Provision C.8.e. of the MRP calls for testing for trends towards compliance with loads allocations (SFRWQCB 2009). To support that provision, the objective of the trends analysis performed here was to determine the power to observe declining trends in the ratio of SSC to Hg concentration or SSC: PCB concentration given the current mean slope and variability. This is consistent with the presentation of TMDL targets on SSC normalized basis (SFRWQCB 2006, 2008). Trends were examined for reductions in the estimated particle concentration [mass/unit mass] from its current value to a value of 0.2 mg Hg / kg suspended sediment (i.e., the SSC: Hg target) and 0.002 mg PCB / kg suspended sediment. These targets assume that urban suspended sediment loads in the Bay Area average 400 million kg annually (following Lewicke and McKee 2009). Note that for the Z4LA Hg trend analysis, initial regression results demonstrated that the current SSC: Hg slope estimates were below the 0.2 target (Appendix A). Therefore, in this analysis, the trend was examined for a target value of 0.05 mg Hg / kg suspended sediment (75% below 0.2 mg/kg).

Power to observe trends were evaluated at time periods of 10, 20, 25, and 40 years. The analysis examined the power to detect a decline in SSC: Hg and SSC: PCBs (at alpha = 0.05) based on the coefficient of variation (CV = s.e. / mean). Sample sizes in future years were assumed to be the same as current (approximately 12 to 20 PCB samples per year and 15 – 50 Hg samples per year) or reduced to 7 or 10 samples per year. The CV was adjusted for the n = 7 and n = 10 scenarios. Although intuitively, one might expect CV to diminish over time, since a trait of cleaner systems is lower concentration variation (Appendix B), in the absence of information to quantitatively predict the shape of such a trend for Bay Area watersheds and pollutants of interest, power was evaluated assuming that the CV would show a linear decline over the time period evaluated.

RESULTS

Optimal Sampling Designs for Estimating Annual Loads

Guadalupe River

Within-storm sampling design scenarios for Hg, PCBs, and suspended sediment (SS) generally indicated linear interpolation to be the most accurate estimator of loads per storm (Table 6). Complete results are tabulated in Appendix C.1. For all three pollutants, using either the flow-based or turbidity-based storm sampling methods, the accuracy (median bias) and precision (variability of bias) was higher at n = 12 than at n = 6. However, accuracy and precision for Hg did not notably improve from n = 12 to n = 18 or n = 24. Variability of bias generally decreased with increasing sample size. No improvement in accuracy or precision was evident using a rising stage (2:1) emphasis compared to samples evenly spaced over a storm hydrograph (1:1). Based on these findings, linear interpolation was used to characterize annual loads in the among-storm scenarios. To obtain a cohesive analysis, a single storm sampling method and sample size was used for each pollutant and site combination (Table 7).

Evaluation of the three among-storm sampling designs at GR (Table 5) indicated that the number of sampled storms strongly affected accuracy of estimated pollutant loads (Hg results in Figure 1; PCBs and SS results in Appendix E). Scenario results were generally similar for Hg, PCBs and SS loads, and for flow based vs. turbidity based storm sampling. The Hg flow based selection results are described in further detail here (Figure 1), while PCB and SS results, as well as all turbidity-based selection results, may be found in Appendices D.1 and E. The highest sample size of storms evaluated generally resulted in the lowest variability and bias in loads estimates (Figure 1). In WY 2004 and 2005, Design A (including first flush) and Design B (including first flush plus largest storm) demonstrated pronounced increases in accuracy and precision with each increase in storm sampling frequency (Figure 1b, 1c). Depending on the available data for simulations, either 6 or 10 storms were optimal for reducing bias. Design C (random storms only) consistently exhibited the least precision of the three designs at GR. However, Design C also exhibited less bias for 2 and 4 storms than the other designs.

Zone 4 Line A

Consistent with GR results, linear interpolation was the most accurate estimator of Hg, PCBs, and suspended sediment (SS) loads within individual storms (Table 6; Appendix C.2). Flow weighted mean performed particularly poorly for Z4LA. This may suggest that flow and concentrations were not as closely related in the storms analyzed (strong hysteresis). Accuracy and standard deviation of Z4LA load estimates were improved at the higher sample sizes, using either the flow-based or turbidity-based storm selection methods. For Hg and SS, the bias and precision were similar at n = 18 and n =

24. For PCBs, the highest sample size (n = 12) was optimal. Larger sample sizes were not evaluated due to the limitations on PCB sample volume in auto-samplers. Both the magnitude of bias and variability in bias generally decreased with increasing sample size, particularly for linear interpolation. Similar to GR, no change in accuracy or precision was evident using a rising stage (2:1) emphasis versus evenly spaced (1:1) emphasis. Based on these findings for Z4LA, the flow-based design with 18 samples per storm for Hg and SS, 12 samples per storm for PCBs, and linear interpolation was used to characterize annual loads in the among-storm scenarios (Table 7).

Simulation of the three strategies for sampling among storms indicated that 10 storms (the highest sample size) generally resulted in the lowest variability and bias in loads estimates (Figures 2, 3, and 4). Turbidity-based results are summarized in Appendix D.2. Designs A and B generally had the lowest variability for both Hg and PCB loads (Figures 2 and 3). In WY 2009, there was little difference between Designs A and B. In contrast, simulation of WY 2007 indicated more variable results among designs. For Hg, PCBs, and SS (Figures 2, 3, and 4), Design C with 4 storms sampled approximated the best estimate load as well as the other designs (i.e., similar accuracy) but was associated with much higher variability (i.e., lower precision).

Optimal Turbidity Surrogate Regression Designs for Estimating Annual Loads

Guadalupe River

Similar loads estimates could be obtained by the turbidity surrogate regression (TSR) method with significantly fewer samples than the full available sample size. Generally, all simulations indicated median loads that were similar to the best estimate load, reflecting the close relationships of these pollutants to turbidity. Simulations of the TSR showed that variability in the load estimates was markedly reduced at sample sizes of 7 or more (Figures 5, 6, and 7). For example, the median Hg load in WY 2004 at n = 7 was 12.8 ± 2 kg, compared to the best estimate load of 13.0 kg. Monte Carlo simulations of the TSR also indicated that 7 samples were needed to accurately estimate for PCBs and SS loads. For example, median PCB load estimates in WY 2003 and 2004 were 1.7 ± 0.4 kg and 1.1 ± 0.4 kg, respectively, compared to loads determined using all samples of 0.9 kg and 0.5 kg, respectively. In WY 2005, a limited pool of samples (n = 12) was available for PCB simulation, and thus annual loads exhibited wider variability. However, SS loads were well sampled in all years and thus were generally consistent. Based on the finding that 7 samples provided an adequate basis for TSR in most GR scenarios, among-storm sampling strategies were examined using this sample size.

Simulation of three among-storm sampling (Table 5) designs for annual loads estimation using TSR indicated that sampling 10 storms per year was optimal to approximate the best estimate loads (Figures 8 – 10). The error bounds for annual loads generally narrowed as the sample size increased, but there was considerable variability among years and pollutants. Simulation of Designs A and B most consistently produced the least bias estimates, but not in all cases. For example, WY 2004 results indicated similar Hg load estimates using either Design A or B, and wider variability for Design C. In contrast, estimated loads with Design C were more consistent in WY 2005 than either of the other designs. PCB and SS loads were less variable than Hg loads, but still indicated that 10 storms were required for minimum bias in loads.

As a final comparison, the bias and precision in sampling 10 storms using either Design A or B (Table 5) were compared between linear interpolation and TSR. For all three pollutants, linear interpolation provided more accurate and precise estimation of the best estimate load in WY 2005 (Table 8). A sampling strategy employing first flush and 10 total storms (Design A) with linear interpolation suggested relatively high accuracy for Hg and PCB loads of approximately 10%. Using TSR, PCB loads had very low accuracy, suggesting variability on the order of 50%. For SS, linear interpolation with Design B suggested the best design, which was estimated to have accuracy of approximately 1% under the WY 2005 scenario.

Zone 4 Line A

Simulations of the TSR for Z4LA supported the GR results that similar estimates of loads could be obtained with significantly less samples than the full available sample size (Figures 11, 12, 13). Simulations for Hg and PCBs indicated that 7 samples per year were needed to accurately estimate loads each year. Although the median load estimate did not vary greatly with increasing sample size, variability was significantly reduced with 7 or more samples. Simulation of these datasets indicated that the TSR was robust at all sample sizes in WY 2007 and 2008. Due to lower sample size of PCBs in WY 2009, there was greater variability in the load estimates. The accuracy of SS loads was relatively high at all sample sized evaluated due to the larger number of available sample points for simulation. However, the variability in loads spanned more than two orders of magnitude at sample sizes less than 10 (particularly in 2007). In summary, storm sampling strategies based on TSR as described below, were examined using 7 samples for Hg and PCBs, and 10 samples for SS.

Simulation of three among-storm sampling (Table 5) designs for annual loads estimation using TSR indicated that sampling the maximum number of storms each year was optimal for minimum bias and precise load estimates. For Hg loads, sampling of 10 storms per year using Design A or B achieved the least amount of variability and most accurate loads in 2007 (Figure 14). Simulations using 2009 data, indicated 4 – 6 storms using Design A would be sufficient, as the median load and variability did not vary greatly at greater sampling intensity or when one of the largest storms was included. Design C under predicted the best estimate loads in WY 2007, but attained reasonably close estimation of the best estimate load in WY 2009. PCB and SS loads were similar to Hg and best approximated loads at Z4LA by sampling of 10 storms (Figures 15 and 16).

Finally, TSR was compared to linear interpolation to evaluate bias in loads using the first flush designs when sampling 10 storms in WY 2009 (Table 5). Using either TSR or linear interpolation, Design A indicated better accuracy and precision relative to Design B. PCB loads were the most variable of the three pollutants in both methods, with an estimated bias of around 30% relative to the best estimate. However, estimated loads were very accurate for Hg (~ 1%) using either method and represented similar levels of precision (5%). Interestingly, SS loads were generally more accurate using TSR, but exhibited less precision than linear interpolation.

Trend Analysis

Guadalupe River

Trend analysis indicated that the power to detect trends in SSC: Hg using the current sampling intensity (Table 9) was generally greater than 90%. In WY 2004 and 2005, future sample sizes could be reduced to 7 samples per year from current sample sizes of 37 and 52, respectively, without loss of power to detect trends in the next 10 – 40 years (Table 10). Due to the lower SSC: Hg slope estimate (1.14) and weak regression relationship in 2003 (CV = 2.35; R^2 = 0.30), there was very low power to detect trends in that year.

Estimates of power to detect trends in the SSC: PCB relationship were also generally high (> 90%). Based on WY 2003 and 2004 data, future sample sizes could be reduced to 10 samples per year from current sample sizes of 21 and 19, respectively (Table 10). However, trend analysis performed with the lower SSC: PCB slope estimate measured in WY 2005 (0.12) indicated that relatively high power would not be achieved for a 10 year trend. Overall, the power analysis suggested that fewer sample sizes at GR would not inhibit the ability to detect declines in Hg and PCB concentrations.

Zone 4 Line A

The trend analysis for Z4LA indicated the power to detect trends in SSC: Hg using the current sampling intensity (Table 11) was generally greater than 90%. Scenarios run with the SSC: Hg slope estimate from 2009, indicated that future sample sizes could be reduced to 10 samples per year from a current sample sizes of 30 without loss of power for trends in 20, 25, and 40 years (Table 12).

Estimates of power to detect trends in the SSC: PCB relationship were very high (> 95%). In WY 2007 and 2008, future sample sizes could be reduced to 7 samples per year from current sample sizes of 15 and 14, respectively, for all trend scenarios evaluated (Table 12). However, at a sample size of 7 per year, 95% power would only be achieved in WY 2007 for 25 and 40 year trends. Overall, the power analysis suggested that lower sizes would also not inhibit the ability to detect declines in Hg and PCB concentrations at Z4LA.

SUMMARY

- The optimal within-storm design in GR and Z4LA evaluations was an equal-spacing, flow or turbidity-based sampling method, with the linear interpolation estimator.

- The optimal among-storm design was highly dependant on sample size. When small numbers of storms were simulated per year, sampling strategies that included first flush or largest storms per year (i.e., Designs A and B) exhibited substantial upward bias in estimated annual load. The first flush and large storm events generally have greater suspended sediment and pollutant concentrations than other storms; as a result, overemphasizing these events would result in overestimates of annual loads. Not surprisingly, the best estimates of annual loads were achieved in the largest sample sizes examined (10 storms per year).

- Designs that randomly sample storms throughout the year (i.e., Design C) without emphasizing first flush or large events have better accuracy at small sample sizes. However, these designs exhibit poor precision, with highly variable estimated loads.

- Evaluation of the turbidity-surrogate regression methods suggested that sampling frequency could be significantly reduced. For example, 10 storms sampled per year with one or two samples per storm were indicated.

- Results of the trend analysis indicated that power to detect long term trends in SSC: Hg concentrations and SSC: PCB concentrations should be high using a variety of sampling designs.

References

Austin, C. 2006. Guadalupe River Watershed Mercury Total Maximum Daily Load (TMDL) Project Report. San Francisco Bay Regional Water Quality Control Board, Oakland, CA.

Balcom, P. H., W. F. Fitzgerald, G. M. Vandal, C. H. Lamborg, K. R. Rolfllus, C. S. Langer, and C. R. Hammerschmidt. 2004. Mercury sources and cycling in the Connecticut River and Long Island Sound. Marine Chemistry **90**:53-74.

Conaway, C. H., J. R. M. Ross, R. Looker, R. P. Mason, and A. R. Flegal. 2007. Decadal mercury trends in San Francisco Estuary sediments. Environmental Research **105**:53-66.

David, N., L. J. McKee, F. J. Black, A. R. Flegal, C. H. Conaway, D. H. Schoellhamer, and N. K. Ganju. 2009. Mercury concentrations and loads in a large river system tributary to San Francisco Bay, California, USA. Environmental Toxicology and Chemistry **28**:2091-2100.

Davis, J. A., F. Hetzel, J. J. Oram, and L. J. McKee. 2007. Polychlorinated biphenyls (PCBs) in San Francisco Bay. Environmental Research **105**:67-86.

Flegal, A. R., C. H. Conaway, J. A. Davis, and M. S. Connor. 2007. Sources, transport, fate, and toxicity of pollutants in the San Francisco Bay Estuary. Environmental Research **105**:1-4.

Godfrey, J. T., G. D. Foster, and K. A. Lippa. 1995. Estimated annual loads of selected organic contaminants to Chesapeake, Bay via a major tribituary. Environmental Science and Technology **29**:2059-2064.

Grayson, R. B., Finlayson, B.L., Gippel, C.J., and Hart, T.B. 1996. The potential of field turbidity measurements for the computation of total phosphorus and suspended solids loads. Journal of Environmental Management **47**:257-267.

Hoenicke, R., D. Oros, J. Oram, and K. Taberski. 2007. Adapting an ambient monitoring program to the challenge of managing emerging pollutants in the San Francisco Estuary. . Environmental Research **105**:132-144.

Lauenstein, G. G., and K. D. Daskalakis. 1998. U.S. long-term coastal contaminant temporal trends determined from mollusk monitoring programs, 1965-1993. Marine Pollution Bulletin **37**:6-13.

Leecaster, M. K., K. C. Schiff, and L. L. Tiefenthaler. 2002. Assessment of efficient sampling designs for urban stormwater monitoring. Water Research **36**:1556-1564.

Lewicke, M., and L. J. McKee. 2009. Watershed specific and regional scale suspended sediment loads for Bay Area small tributaries. SFEI Contribution #566, San Francisco Estuary Institute, Oakland, CA.

Linkov, I., D. Burmistrov, J. Cura, and T. S. Bridges. 2002. Risk-based management of contaminated sediments: consideration of spatial and temporal patterns in exposure modeling. Environmental Science and Technology **36**:238-246.

Lowe, S., R. Hoenicke, J. Davis, and G. Scelfo. 1999. 1999 Quality Assurance Project Plan for the Regional Monitoring Program for Trace Substances. *in*. SFEI.

Ma, J.-S., J.-H. Kang, M. Kayhanian, and M. K. Stenstrom. 2009. Sampling issues in urban runoff monitoring programs: Composite versus grab. Journal of Environmental Engineering **135**:118 - 127

May, D., and M. Sivakumar. 2009. Optimum number of storms required to derive site mean concentrations at urban catchments. Urban Water Journal **6**:107-113.

McKee, L., J. Leatherbarrow, and R. Eads. 2004. Concentrations and loads of mercury, PCBs, and OC pesticides associated with suspended sediments in the Lower Guadalupe River, San Jose, California. 86, San Francisco Estuary Institute, Oakland, CA.

McKee, L., J. Leatherbarrow, and J. Oram. 2005. Concentrations and loads of mercury, PCBs, and OC pesticides in the lower Guadelupe River, San Jose, California: Water Years 2003 and 2004. San Francisco Estuary Institute, Oakland, CA.

McKee, L. J., N. K. Ganju, and D. H. Schoellhamer. 2006a. Estimates of suspended sediment entering San Francisco Bay from the Sacramento and San Joaquin Delta, San Francisco Bay, California. Journal of Hydrology **323**:335-352.

McKee, L. J., and A. Gilbreath. 2009. Concentrations and loads of trace contaminants in the Zone 4 Line A small tributary, Hayward, California: Water Year 2007. SFEI Contribution #563, San Francisco Estuary Institute, Oakland, CA.

McKee, L. J., J. Oram, J. Leatherbarrow, A. Bonnema, W. Heim, and M. Stephenson. 2006b. Concentrations and loads of mercury, PCBs, and PBDEs in the lower Guadalupe River, San Jose, California: Water Years 2003, 2004, and 2005. SFEI Contribution 424, San Francisco Estuary Institute, Oakland, CA.

Newton, A., and S. M. Mudge. 2005. Lagoon-sea exchanges, nutrient dynamics and water quality management of the Ria Formosa (Portugal). Estuarine and Coastal Shelf Science **62**:405-414.

Oram, J. J., L. J. McKee, C. E. Werme, M. S. Connor, and D. R. Oros. 2008. A mass budget of PBDEs in San Francisco Bay, California, USA. Environment International **34**:1137-1147.

Oros, D. R., W. M. Jarman, T. Lowe, N. David, S. Lowe, and J. A. Davis. 2003. Surveillance for previously unmonitored organic contaminants in the San Francisco Estuary Marine Pollution Bulleitin **46**:1102-1110.

Schiff, K., and S. Bay. 2003. Impacts of stormwater discharges on the nearshore benthic environment of Santa Monica Bay. Marine Environmental Research. Marine Environmental Research **56**:225-254.

SFEI. 2010. The Pulse of the Estuary: Monitoring and Managing Water Quality in the San Francisco Estuary. SFEI Contribution xxx, San Francisco Estuary Institute, Oakland, CA.

SFRWQCB. 2006. Mercury in San Francisco Bay Proposed Basin Plan Amendment and Staff Report for Revised Total Maximum Daily Load (TMDL) and Proposed Mercury Water Quality Objectives. . California Regional Water Quality Control Board San Francisco Bay Region, Oakland, CA.

SFRWQCB. 2008. Total Maximum Daily Load for PCBs in San Francisco Bay: Final Staff Report for Proposed Basin Plan Amendment

California Regional Water Quality Control Board San Francisco Bay Region, Oakland, CA.

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. California Regional Water Quality Control Board San Francisco Bay Region, Oakland, CA.

She, J., A. Holden, T. L. Adelsbach, M. Tanner, S. E. Schwarzbach, J. L. Yee, and K. Hooper. 2008. Concentrations and time trends of polybrominated diphenyl ethers (PBDEs) and polychlorinated biphenyls (PCBs) in aquatic bird eggs from San Francisco Bay, CA 2000–2003. Chemosphere **73**:S201-S209.

Trimble, S. W. 2003. Historical hydrographic and hydrologic changes in the San Diego creek watershed, Newport Bay, California. Journal Of Historical Geography **29**:422-444.

USEPA. 2002. Urban stormwater BMP performance monitoring. 821-B-02-001. , Office of Water, Washington, D.C. .

Wall, G. R., H. H. Ingleston, and S. Litten. 2005. Calculating mercury loading to the tidal Hudson River, New York, using rating curve and surrogate methodologies. Water, Air, and Soil Pollution **165**:233-248.

Walling, D. E., and Webb, B.W. 1985. Estimating the discharge of contaminants to coastal waters by rivers: some cautionary comments. Marine Pollution Bulletin **16**:488-492.

Walling, D. E., and Webb, B.W. 1988. Reliability of rating curve estimates of suspended sediment yield: some further comments. Pages 337-350 *in* Sediment budgets. IAHS, Porto Alegre.

Yee, D., L. J. McKee, and J. J. Oram. in review. A simple mass budget of methylmercury in San Francisco Bay, California. Environmental Toxicology and Chemistry.

Table 1. Summary of Guadalupe River and Zone 4 Line A data examined in this study.

Location	Water Years Examined	Peak Discharge (m^3/s)	Suspended Sediment FWMC (mg/L)	THg FWMC (ng/L)	TPCBs FWMC (ng/L)
Guadalupe River (GR)	2003	172	204	2190	55
	2004	124	191	329	26
	2005	112	79	140	45
Zone 4 Line A (Z4LA)	2007	6.3	212	48	27
	2008*	6.7	350	60	25
	2009	4.7	109	23	11

* Data for partial year that included 2 dry season months.

Table 2. Design options examined for sampling within storms.

Design Criteria	Design Options for Sampling Within Storms
1. Storm Sampling Basis	Flow-based or turbidity-based thresholds
2. Emphasis	Rising stage (2:1) or evenly spaced (1:1)
3. Max Sample Size (n) per Storm (actual n depends on storm size)	24*, 18*, 12, 6
4. Load Calculation Methods**	
Discrete Designs	LI – linear interpolation; FWM – flow-weighted mean
Composite Designs	SM – simple mean **

* Evaluated for Hg and suspended sediments only due to limitation on volumes required for PCB lab analysis

** Loads calculation methods differed for discrete vs. composite designs.

*** The other methods require discrete measurements

Table 3a. Flow-Based storm selection criteria by watershed.

Dataset	Flow Thresholds for Storm Selection (cfs)	Minimum Peak Flow for Storm Selection (cfs)
Guadalupe River	200	736
Zone 4 Line A	5	26

Table 3b. Turbidity-Based storm selection criteria by watershed.

Dataset	Turbidity Thresholds for Storm Selection (NTU)	Minimum Turbidity Peak for Storm Selection (NTU)
Guadalupe River	30	84
Zone 4 Line A	30	89

Table 4. Equations used to evaluate pollutant loads within- and among-storms.

Within-storm Estimators	Among-storm Ratio Estimator[1]
Simple Mean Equation 1: $L_{S.M.} = \left(\sum_{j=1}^{N} [x_j]/N\right) \left(\sum_{i=1}^{n} Q_i*\Delta t\right)$	
Linear Interpolation Equation 2: $L_{L.I.} = \sum_{i=1}^{n} [x_{i,int}]*Q_i*\Delta t$	Equation 4: $WY\_L_{L.I.} = \sum_{k=1}^{M} V_k * \sum_{k=1}^{m} L_k \,/\, \sum_{k=1}^{m} V_k$
Flow-weighted Mean Equation 3: $L_{F.W.} = \left(\sum_{j=1}^{N} [x_j]*Q_j\right)\left(\sum_{i=1}^{n} Q_i*\Delta t\right)/\left(\sum_{j=1}^{N} Q_j\right)$	

1 = Ratio Estimator used to calculate annual loads using optimal within-storm estimation method (i.e. Equations 1,2 or 3).

Where, L = estimate of mass loading for a storm; WY_L = estimate of annual mass emissions; Δt = time interval between discharge measurements; N = number of samples taken during storm; n = number of time intervals in storm (based on frequency of discharge measurements); $[x_j]$ = concentration of sample *j*; $[x_{i,int}]$ = $[x_j]$ interpolated to all *n* time intervals in storm; Q_i = discharge at time step *i*; Q_j = discharge at sampling event *j* V_k = discharge volume for storm *k*; m = number of storms sampled; and M = number of storms.

Table 5. Design options examined for sampling among storms.

Design Criteria	Design Options for Sampling Among Storms		
	A	B	C**
Which storms	First flush, and random N	First flush, largest storm*, and random N	Random N
Total Number Storms (N)	10, 6, 4, 2	10, 6, 4, 2	4, 2

* To account for selection uncertainty, the largest storm was selected randomly from the three highest total volume discharges per water year

** MRP design and sample sizes

Table 6. Comparison of bias in within-storm loads estimates for sample designs at Guadalupe River and Zone 4 Line A (sample emphasis = 1:1, flow-based criteria).

	Guadalupe River		
	Median Bias +/- St. Deviation		
Pollutant*	Simple Mean	Linear Interpolation	Flow-weighted Mean
Hg (n = 12)	-0.21 ± 0.18	-0.05 ± 0.25	-0.02 ± 0.33
PCBs (n = 12)	-0.04 ± 0.30	0.02 ± 0.28	0.15 ± 0.35
SS (n = 12)	-0.20 ± 0.24	-0.01 ± 0.24	0.07 ± 0.42
	Zone 4 Line A		
	Median Bias +/- St. Deviation		
Pollutant*	Simple Mean	Linear Interpolation	Flow-weighted Mean
Hg (n = 18)	-0.14 ± 0.23	0.01 ± 0.23	0.17 ± 0.38
PCBs (n = 12)	-0.26 ± 0.26	-0.11 ± 0.30	-0.09 ± 0.45
SS (n = 18)	-0.17 ± 0.27	0.003 ± 0.26	0.15 ± 0.44

* Number in parentheses designates number of samples evaluated per storm

Table 7. Within-storm design strategies used for among-storm analyses.

Site	Pollutant	Loads Estimation Method	Sampling Emphasis	Sample Size per Storm
Guadalupe River	Hg	Linear Interpolation	Evenly Spaced (1:1)	12
	PCBs			12
	SSC			12
Zone 4 Line A	Hg			18
	PCBs			12
	SSC			18

Table 8. Comparison of bias in median annual loads (+/- st. dev) resulting from turbidity-surrogate and linear interpolation in WY 2005 at GR and WY 2009 at Z4LA using two among storm sampling strategies (Design A and B in Table 5; N = 10 storms).

			Turbidity-surrogate		Linear Interpolation	
Watershed	Pollutant	Water Year	Design A	Design B	Design A	Design B
Guadalupe River (GR)	Hg	2005	-0.20 ± 0.25	0.09 ± 0.22	0.08 ± 0.05	0.09 ± 0.06
Guadalupe River (GR)	PCBs		0.54 ± 0.22	0.57 ± 0.20	0.11 ± 0.06	0.13 ± 0.07
Guadalupe River (GR)	SS		0.09 ± 0.23	0.17 ± 0.22	-0.04 ± 0.05	-0.01 ± 0.05
Zone 4 Line A (Z4LA)	Hg	2009	-0.02 ± 0.05	-0.01 ± 0.04	-0.01 ± 0.08	0.04 ± 0.08
Zone 4 Line A (Z4LA)	PCBs		0.26 ± 0.09	0.33 ± 0.10	0.26 ± 0.15	0.30 ± 0.14
Zone 4 Line A (Z4LA)	SS		0.13 ± 0.08	0.19 ± 0.09	0.08 ± 0.17	0.17 ± 0.15

Table 9. Data used to examine power for trend analysis at Guadalupe River.

Pollutant	Year	Current Sample Size	Mean Slope	S.D. Slope	95% C.I. Slope (lower, upper)	R^2
Hg	2003	25	1.14	13.4	-4.42, 6.70	0.30
	2004	37	1.44	0.73	1.20, 1.69	0.94
	2005	52	2.23	1.34	1.86, 2.61	0.93
PCBs	2003	21	0.06	0.05	0.03, 0.08	0.87
	2004	19	0.11	0.11	0.06, 0.16	0.85
	2005	12	0.12	0.18	0.00, 0.23	0.75

Table 10. Estimates of power to detect trends in the slope of SSC: Hg and SSC: PCBs at Guadalupe River.

		Hg			PCBs		
Year	Number of Years to Reach Target	Power for Current Sample Size*	Power for		Power for Current Sample Size**	Power for	
			n = 7	n = 10		n = 7	n = 10
2003	10	12	10	11	100	83	93
	20	13	11	11	100	98	100
	25	14	11	12	100	99	100
	40	17	12	13	100	100	100
2004	10	100	100	100	99	77	88
	20	100	100	100	100	96	99
	25	100	100	100	100	98	100
	40	100	100	100	100	100	100
2005	10	100	98	100	63	45	56
	20	100	100	100	87	69	82
	25	100	100	100	93	77	88
	40	100	100	100	99	91	97

* For 2003, n = 25; For 2004, n = 37; For 2005, n = 52
** For 2003, n = 21; For 2004, n = 19; For 2005, n = 12

Table 11. Data used to examine power for trend analysis at Zone 4 Line A.

Pollutant	Year	Current Sample Size	Mean Slope	S.D. Slope	95% C.I. Slope (lower, upper)	R^2
Hg	2007	30	0.13	0.11	0.09, 0.17	0.60
	2008	15	0.38	0.61	0.04, 0.72	0.31
	2009	21	0.08	0.05	0.05, 0.10	0.71
PCBs	2007	18	0.06	0.07	0.03, 0.10	0.82
	2008	15	0.16	0.13	0.09, 0.23	0.90
	2009	14	0.08	0.02	0.07, 0.10	0.98

Table 12. Estimates of power to detect trends in the slope of SSC: Hg and SSC: PCBs at Zone 4 Line A. Note that the SSC: Hg are currently below the target of 0.2 mg/kg, therefore the trend was examined for a target of 0.05 mg/kg, which is 75% below the original target of 0.2 mg/kg for urban stormwater (see Methods).

		Hg			PCBs		
Year	Number of Years to Reach Target	Power for Current Sample Size*	Power for n = 7	Power for n = 10	Power for Current Sample Size**	Power for n = 7	Power for n = 10
2007	10	98	55	68	95	66	79
	20	100	81	91	100	90	97
	25	100	88	96	100	95	99
	40	100	97	100	100	99	100
2008	10	62	38	48	100	91	97
	20	87	60	73	100	100	100
	25	93	68	81	100	100	100
	40	99	85	94	100	100	100
2009	10	74	38	48	100	100	100
	20	94	59	72	100	100	100
	25	98	67	80	100	100	100
	40	100	84	93	100	100	100

* For 2007, n = 30; For 2008, n = 15; For 2009, n = 21
** For 2007, n = 18; For 2008, n = 15; For 2009, n =14

Figure Captions

Figure 1. Comparison of annual Hg loads at Guadalupe River in 2003-2005 based on three designs for sampling among storms (Table 5). Loads were calculated using linear interpolation with flow-based storm selection criteria. Design A simulated sampling of the first flush only and a variable number of random storms. Design B simulated the first flush plus one of the three largest storms and a variable number of random storms. Design C only tested the random storm component. --------- = best estimate Hg load for year (96 kg, 13 kg, and 7 kg, respectively).

Figure 2. Comparison of annual Hg loads at Zone 4 Line A in 2007 and 2009 based on three designs for sampling among storms (Table 5). See Figure 1 caption and text for further information. --------- = best estimate Hg load for year (17 g and 11 g, respectively).

Figure 3. Comparison of annual PCB loads at Zone 4 Line A in 2007 and 2009 based on three designs for sampling among storms (Table 5). See also Figure 1 caption and text. --------- = best estimate PCB load for year (8 g and 5 g, respectively).

Figure 4. Comparison of annual SS loads at Zone 4 Line A in 2007 and 2009 based on three designs for sampling among storms (Table 5). See also Figure 1 caption and text. --------- = best estimate SS load for year (0.10 10^6 kg and 0.05 10^6 kg, respectively).

Figure 5. Results of Monte Carlo simulations to determine the optimum number of samples required to estimate Hg loads at Guadalupe River using the turbidity surrogate regression method. --------- = load determined using all samples collected in each year.

Figure 6. Results of Monte Carlo simulations to determine the optimum number of samples required to estimate PCB loads at Guadalupe River using the turbidity surrogate regression method. --------- = load determined using all samples collected in each year.

Figure 7. Results of Monte Carlo simulations to determine the optimum number of samples required to estimate suspended sediment (SS) loads at Guadalupe River using the turbidity surrogate regression method. --------- = load determined using all samples collected in each year.

Figure 8. Comparison of annual Hg loads at Guadalupe River in 2003-2005 based on three designs for sampling among storms (Table 5) using turbidity surrogate regression. --------- = Hg load from all storms sampled each year.

Figure 9. Comparison of annual PCB loads at Guadalupe River in 2003-2005 based on three designs for sampling among storms (Table 5) using turbidity surrogate regression. --------- = PCB load from all storms sampled each year.

Figure 10. Comparison of annual SS loads at Guadalupe River in 2003-2005 for three designs for sampling among storms (Table 5). --------- = SS load from all storms sampled each year.

Figure 11. Results of Monte Carlo simulations to determine the optimum number of samples required to estimate Hg loads at Zone 4 Line A using the turbidity surrogate regression method. --------- = load determined using all samples collected in each year.

Figure 12. Results of Monte Carlo simulations to determine the optimum number of samples required to estimate PCB loads at Zone 4 Line A using the turbidity surrogate regression method. --------- = load determined using all samples collected in each year.

Figure 13. Results of Monte Carlo simulations to determine the optimum number of samples required to estimate suspended sediment (SS) loads at Zone 4 Line A using turbidity surrogate regression. --------- = load determined using all samples collected in each year.

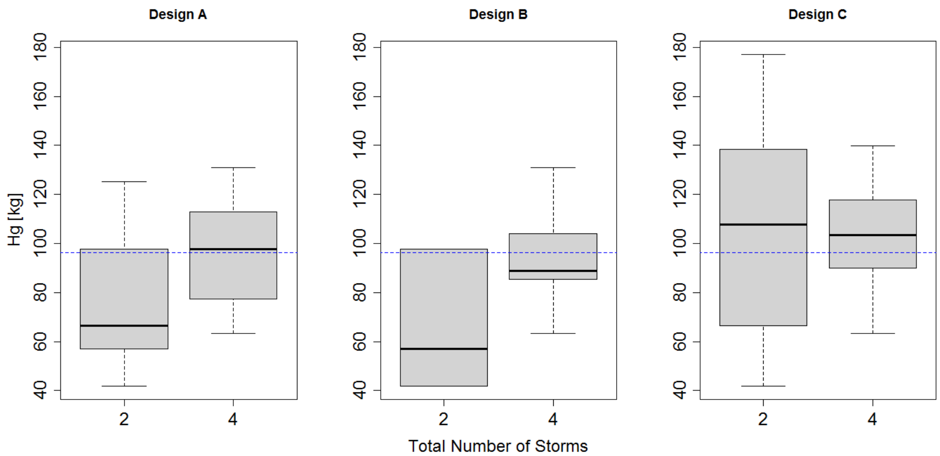
Figure 14. Comparison of annual Hg loads at Zone 4 Line A in 2007-2009 based on three designs for sampling among storms (Table 5) using turbidity surrogate regression. ---------- = Hg load from all storms sampled each year.

Figure 15. Comparison of annual PCB loads at Zone 4 Line A in 2007-2009 based on three designs for sampling among storms (Table 5) using turbidity surrogate regression. ---------- = PCB load from all storms sampled each year.

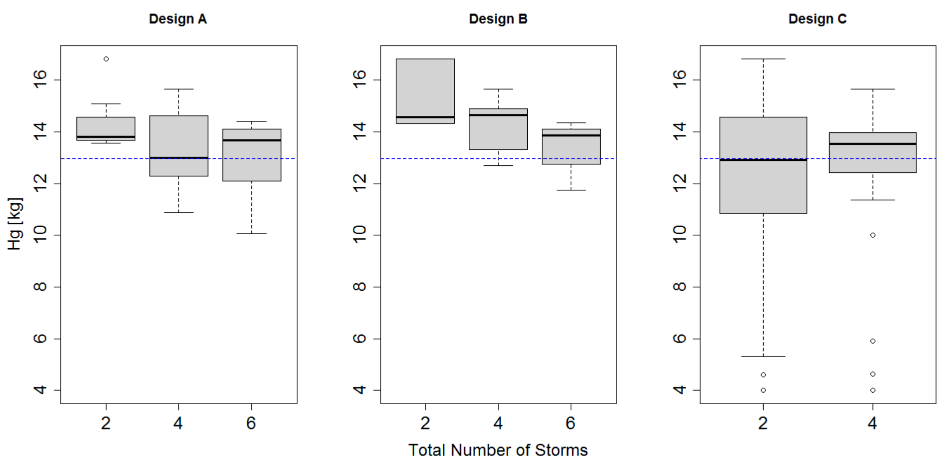
Figure 16. Comparison of annual SS loads at Zone 4 Line A in 2007-2009 based on three designs for sampling among storms (Table 5) using turbidity surrogate regression. ---------- = SS load from all storms sampled each year.

Figures

a) 2003



b) 2004



c) 2005

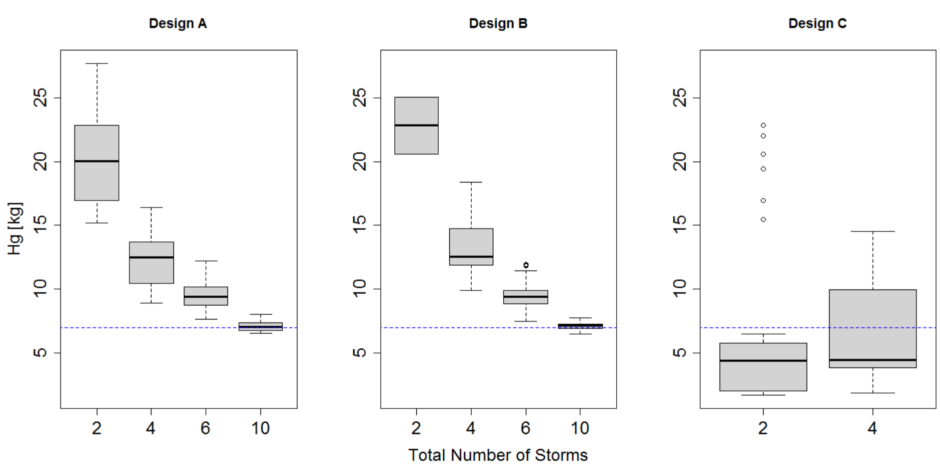
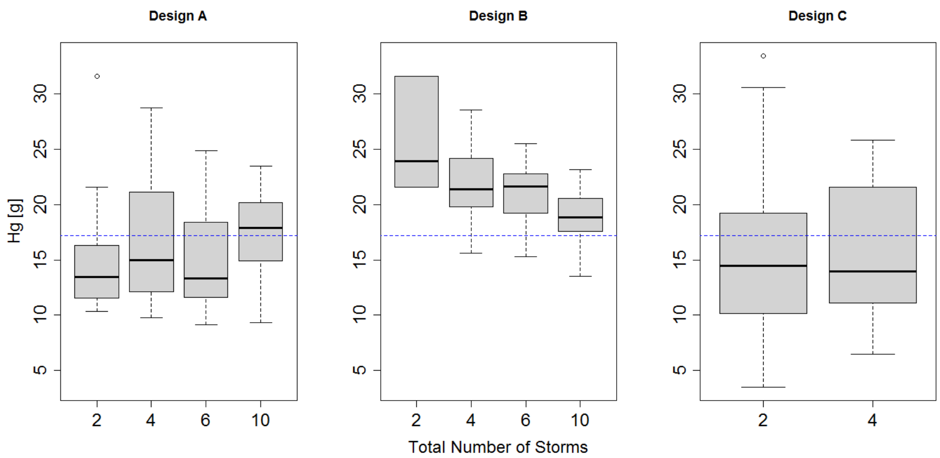


Figure 1a-c

a) 2007



b) 2009

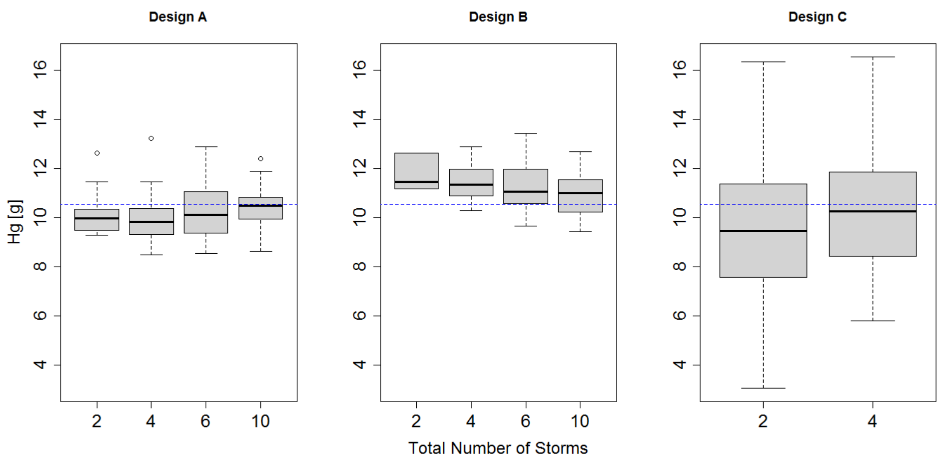
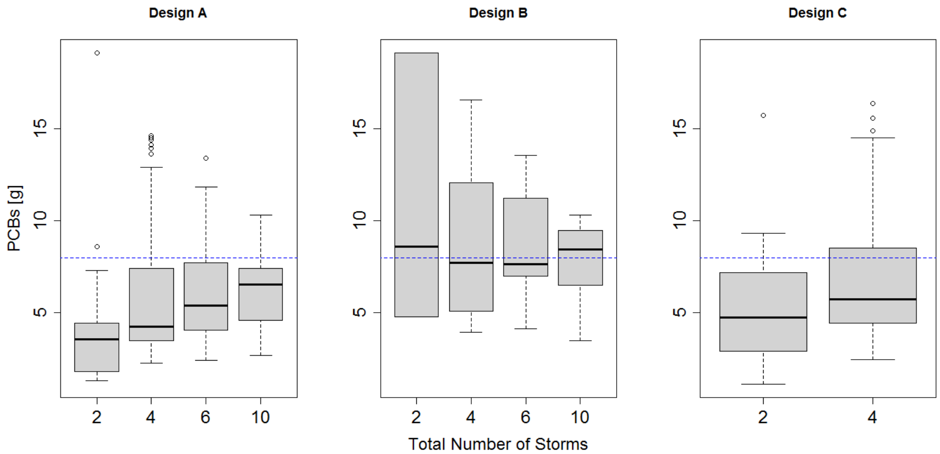


Figure 2a-b

a) 2007



b) 2009

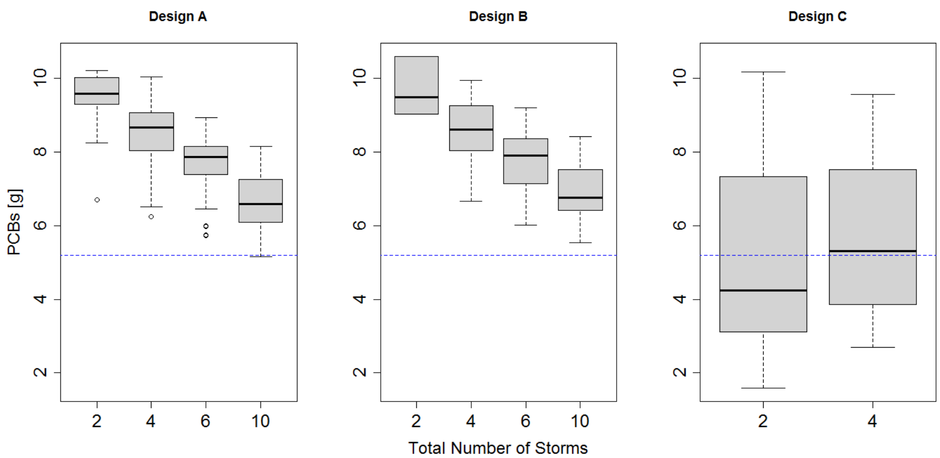
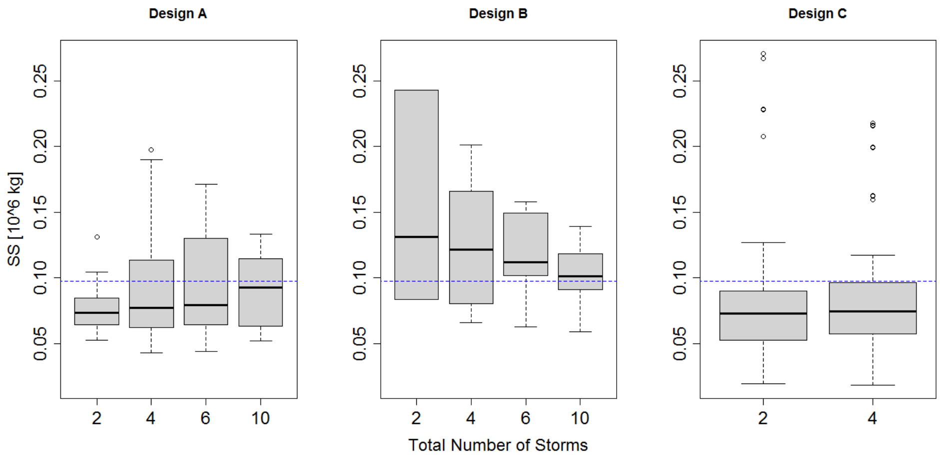


Figure 3a-b

a) 2007



b) 2009

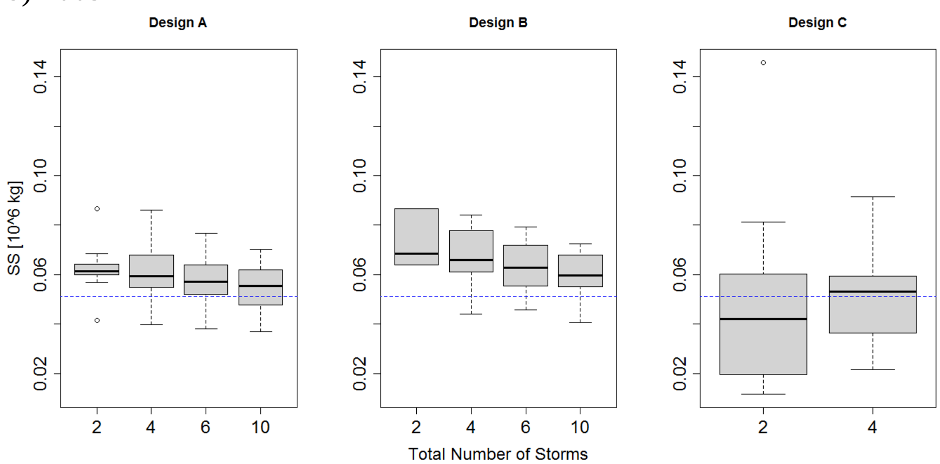


Figure 4a-b

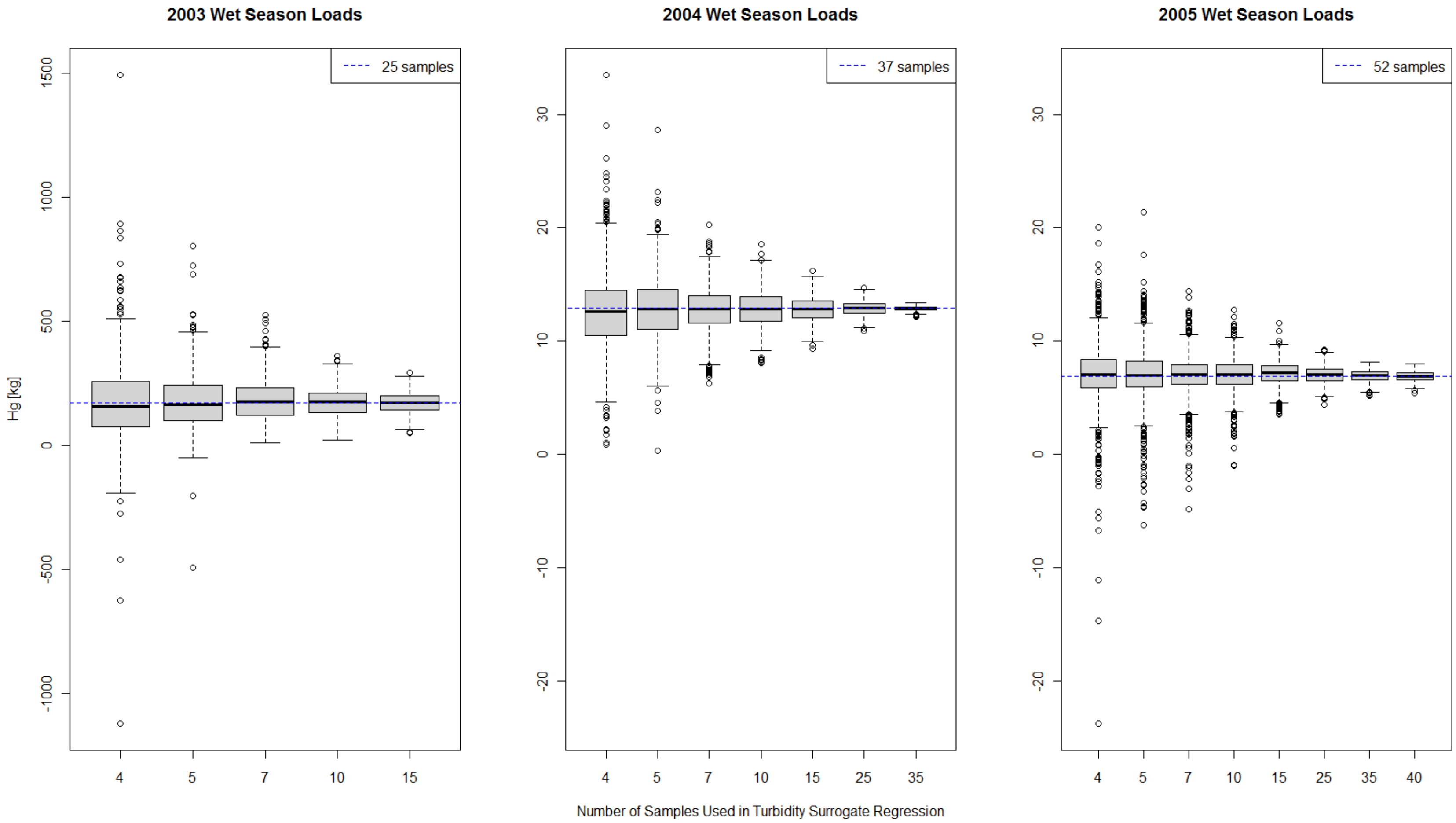

2003 Wet Season Loads
25 samples
Hg [kg]
1500
1000
500
0
-500
-1000
4
5
7
10
15
2004 Wet Season Loads
37 samples
30
20
10
0
-10
-20
4
5
7
10
15
25
35
Number of Samples Used in Turbidity Surrogate Regression
2005 Wet Season Loads
52 samples
30
20
10
0
-10
-20
4
5
7
10
15
25
35
40

Figure 5

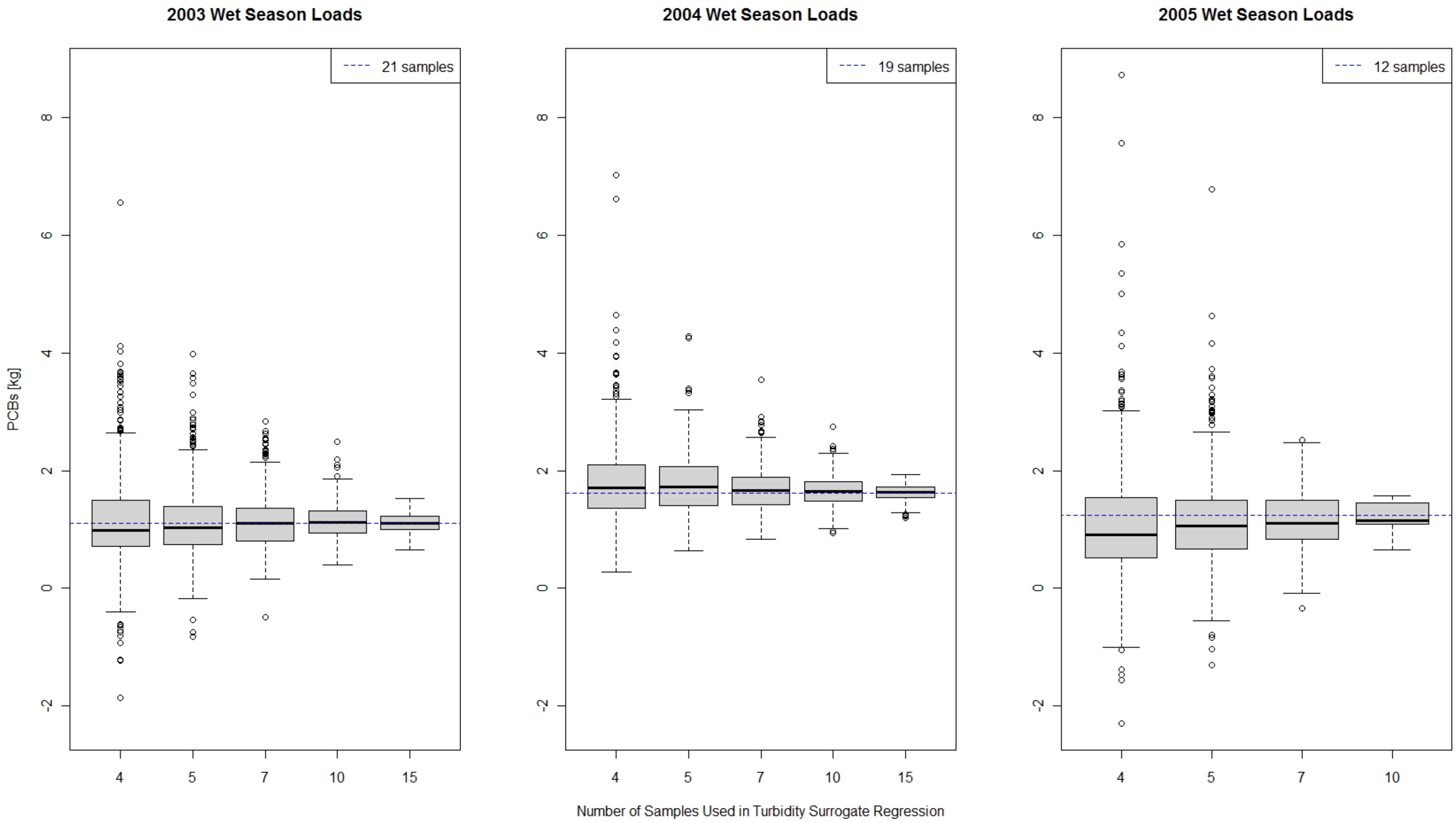

2003 Wet Season Loads
2004 Wet Season Loads
2005 Wet Season Loads
21 samples
19 samples
12 samples
PCBs [kg]
-2
0
2
4
6
8
4
5
7
10
15
Number of Samples Used in Turbidity Surrogate Regression

Figure 6

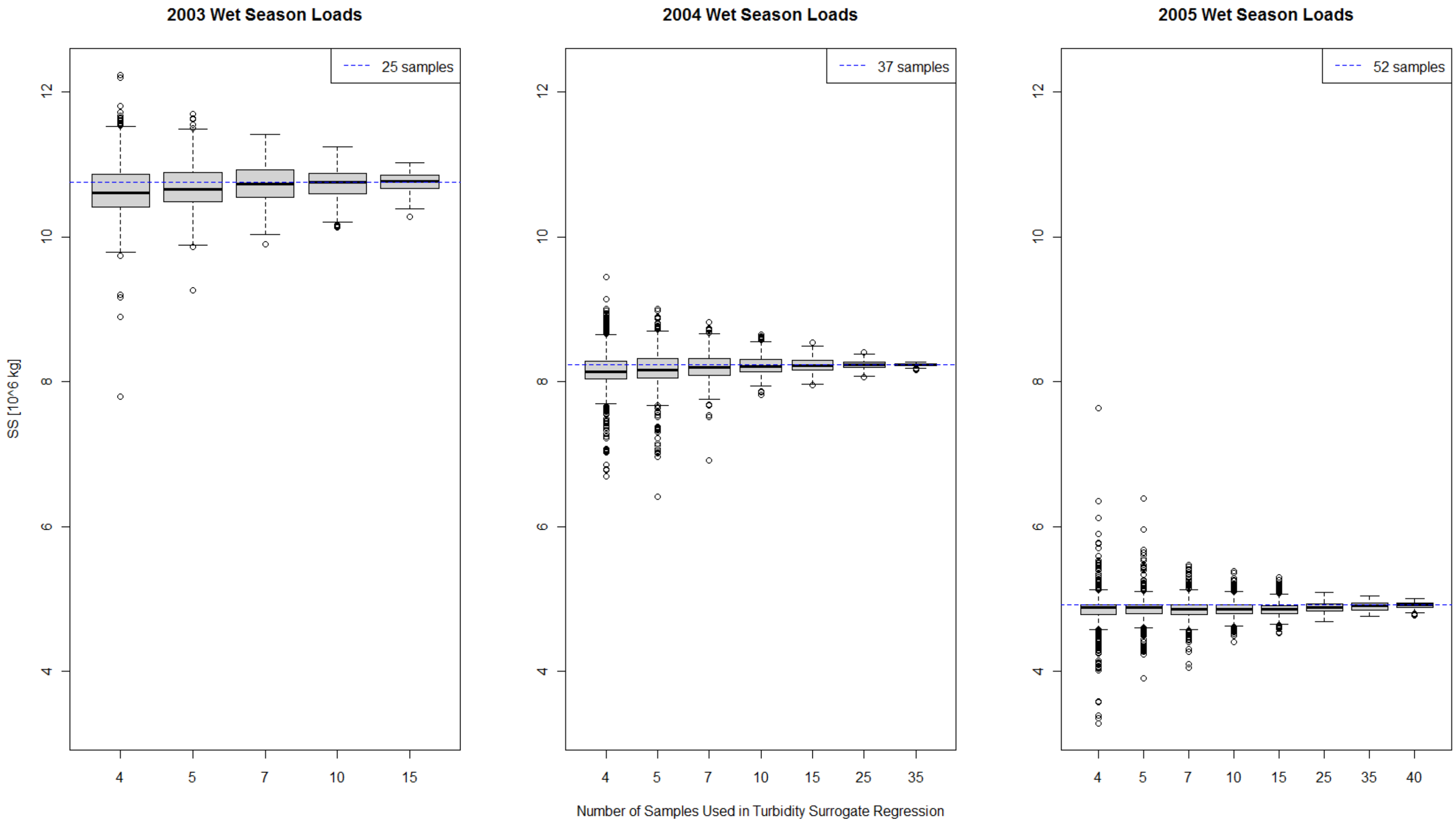
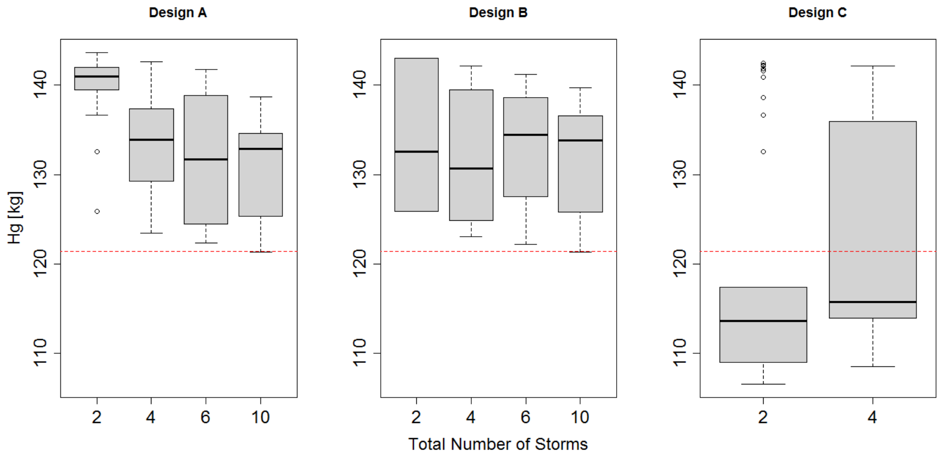
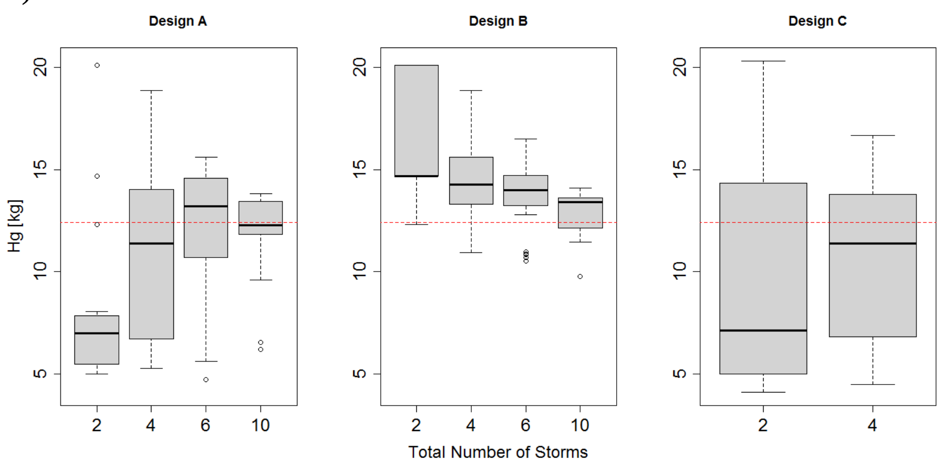

2003 Wet Season Loads
25 samples
2004 Wet Season Loads
37 samples
2005 Wet Season Loads
52 samples
SS [10^6 kg]
12
10
8
6
4
4
5
7
10
15
25
35
40
Number of Samples Used in Turbidity Surrogate Regression

Figure 7

a) 2003



b) 2004



c) 2005

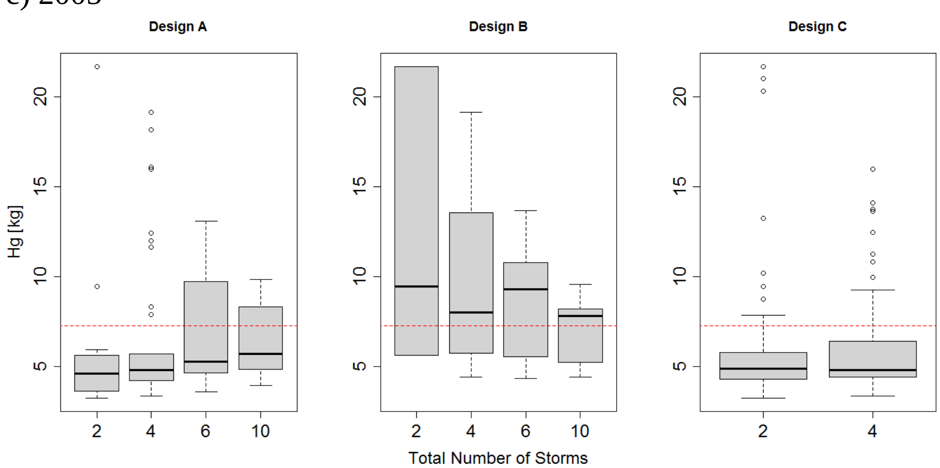
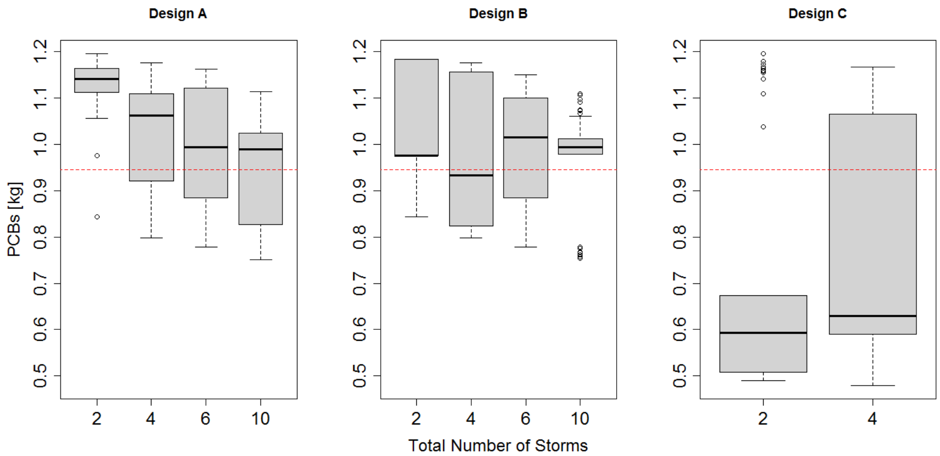
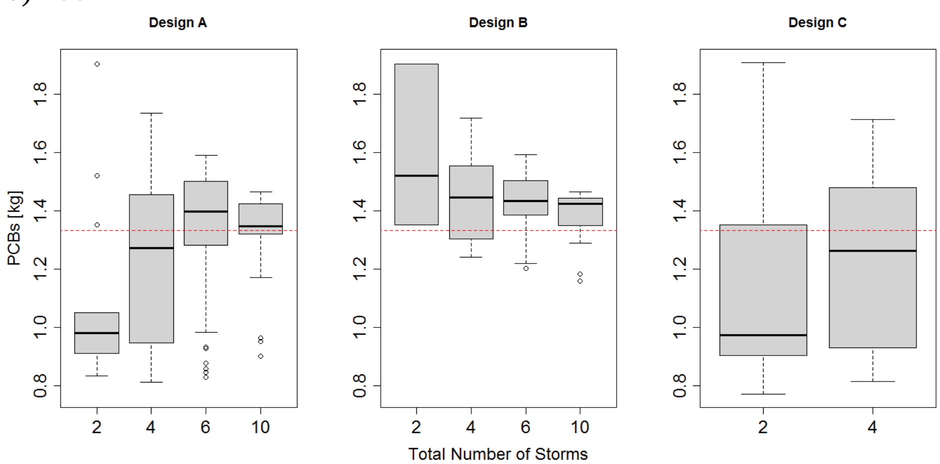


Figure 8a-c

a) 2003



b) 2004



c) 2005

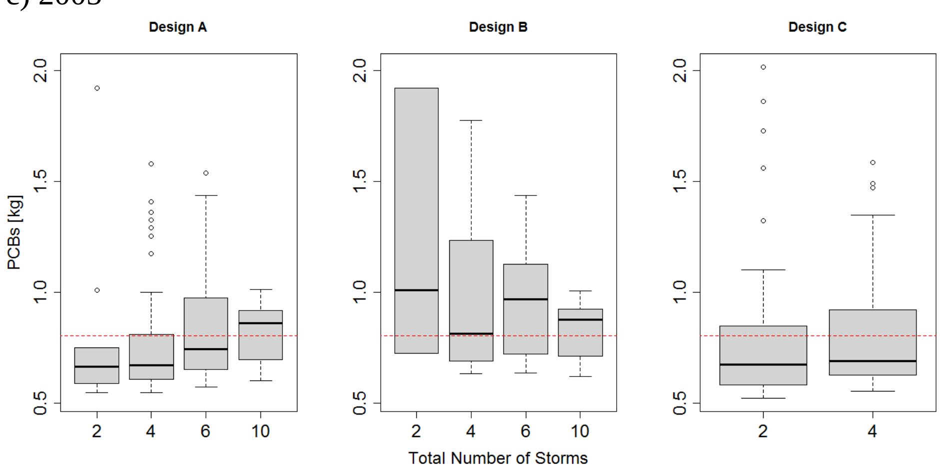
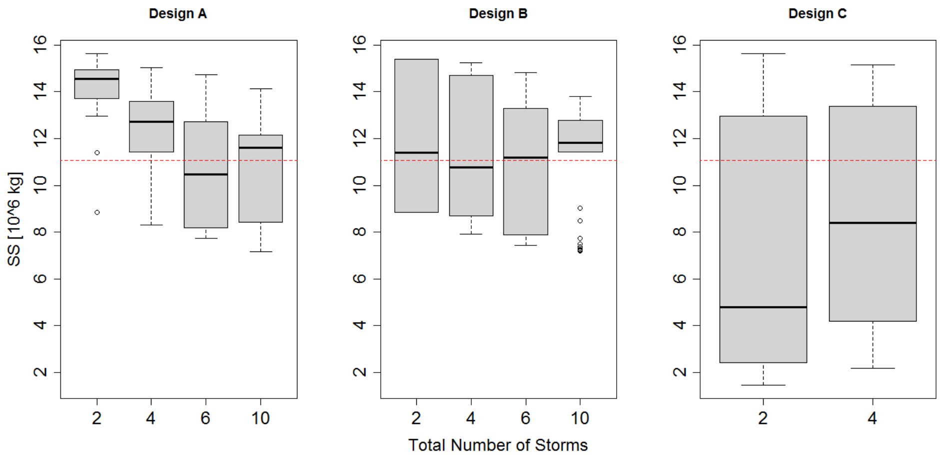
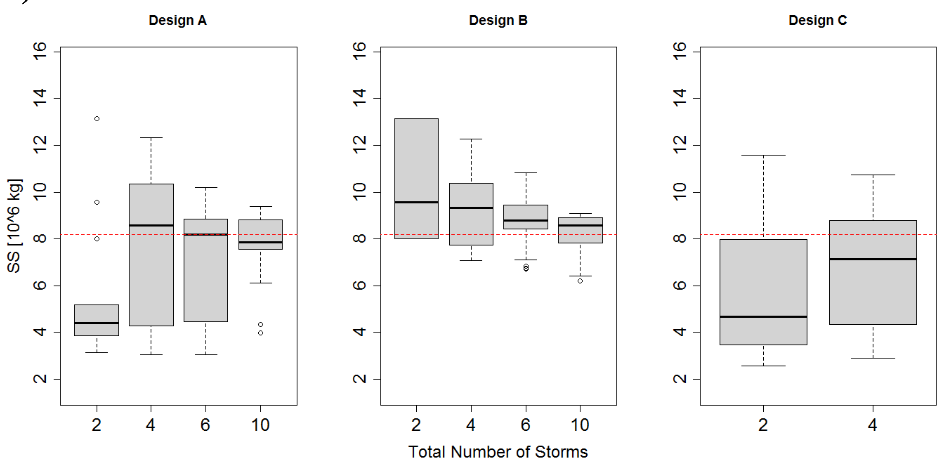


Figure 9a-c

a) 2003



b) 2004



c) 2005

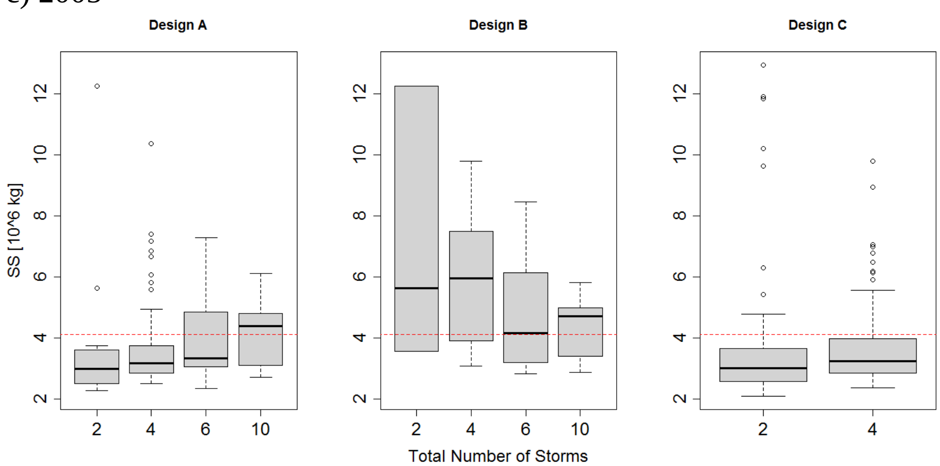


Figure 10a-c

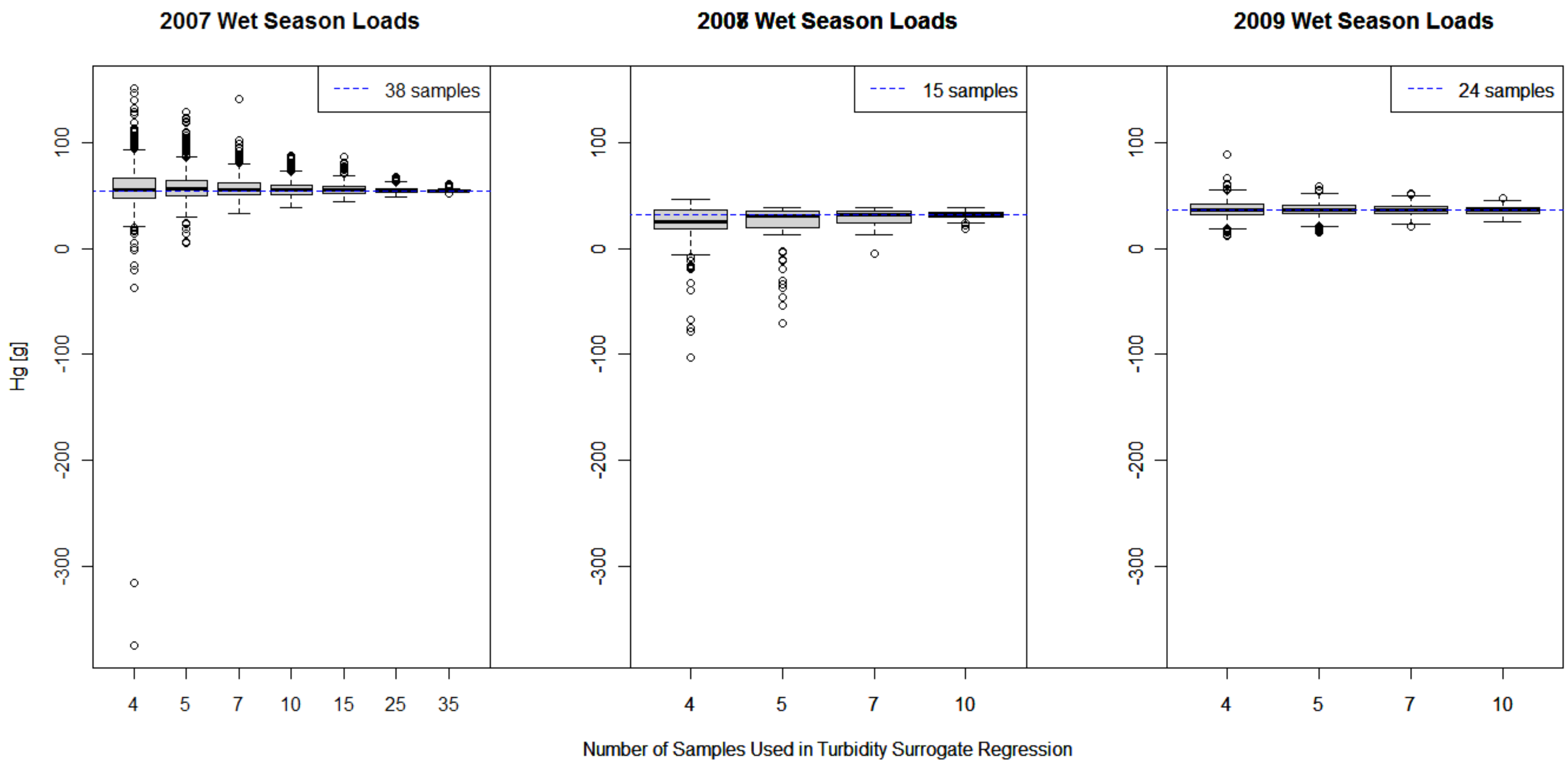

2007 Wet Season Loads
2008 Wet Season Loads
2009 Wet Season Loads
38 samples
15 samples
24 samples
Hg [g]
100
0
-100
-200
-300
4
5
7
10
15
25
35
Number of Samples Used in Turbidity Surrogate Regression

Figure 11

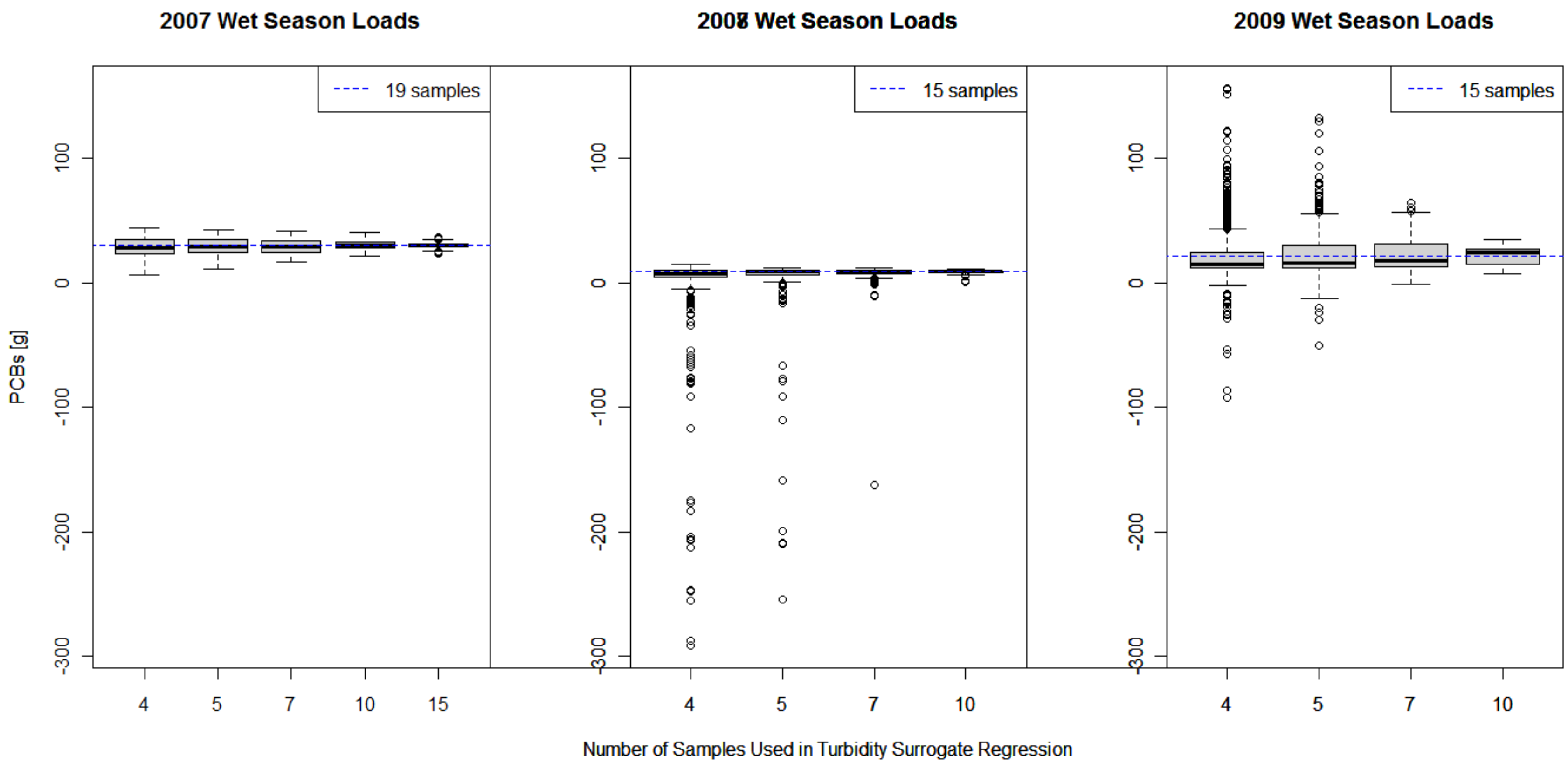

2007 Wet Season Loads
2008 Wet Season Loads
2009 Wet Season Loads
19 samples
15 samples
15 samples
PCBs [g]
100
0
-100
-200
-300
4
5
7
10
15
Number of Samples Used in Turbidity Surrogate Regression

Figure 12

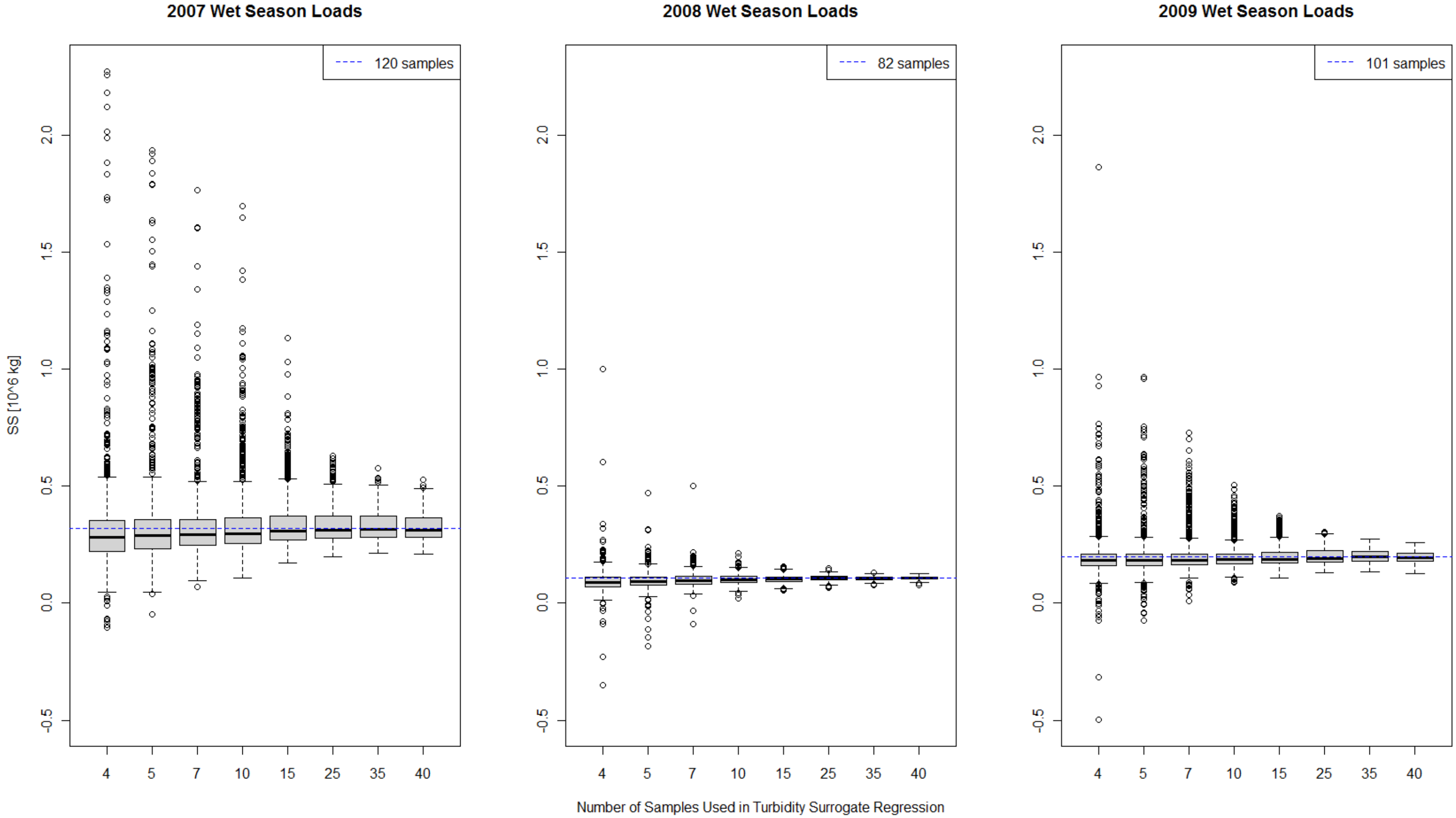
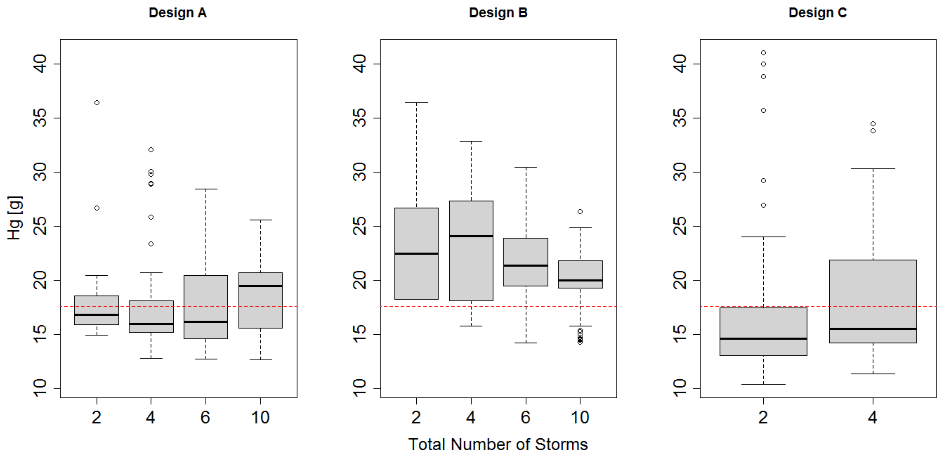

2007 Wet Season Loads
120 samples
2008 Wet Season Loads
82 samples
2009 Wet Season Loads
101 samples
SS [10^6 kg]
2.0
1.5
1.0
0.5
0.0
-0.5
4
5
7
10
15
25
35
40
Number of Samples Used in Turbidity Surrogate Regression

Figure 13

a) 2007



b) 2009

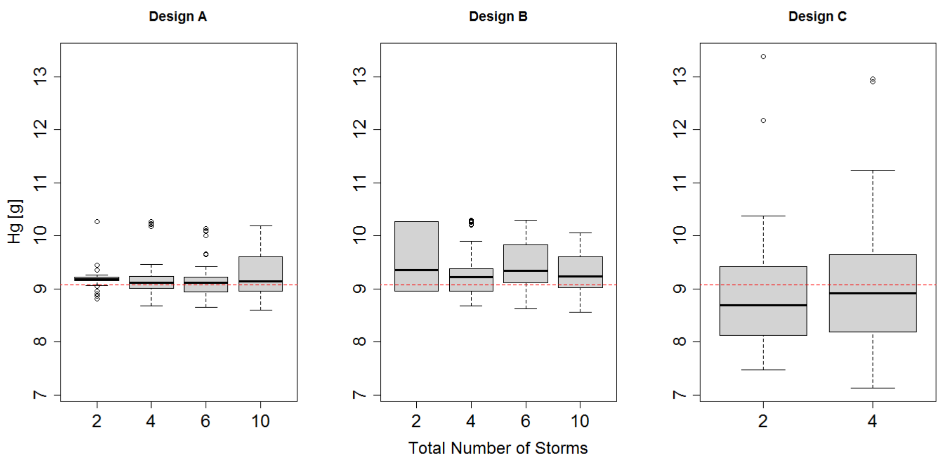
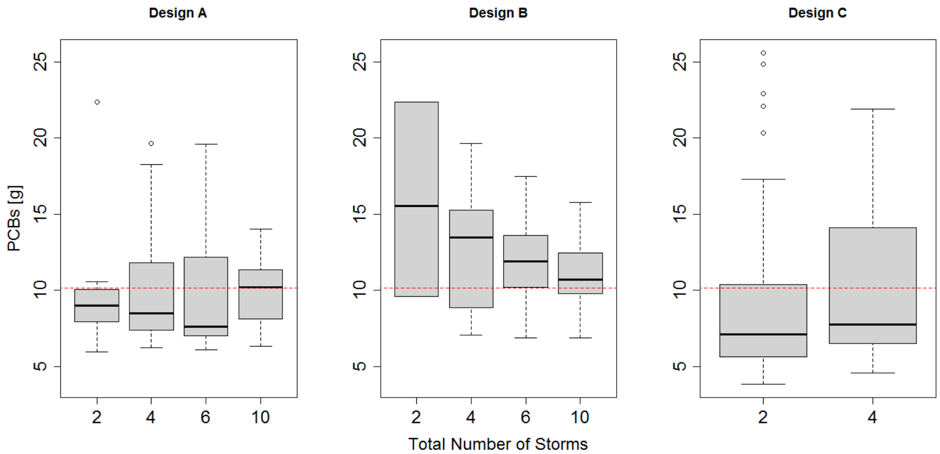


Figure 14a-b

a) 2007



b) 2009

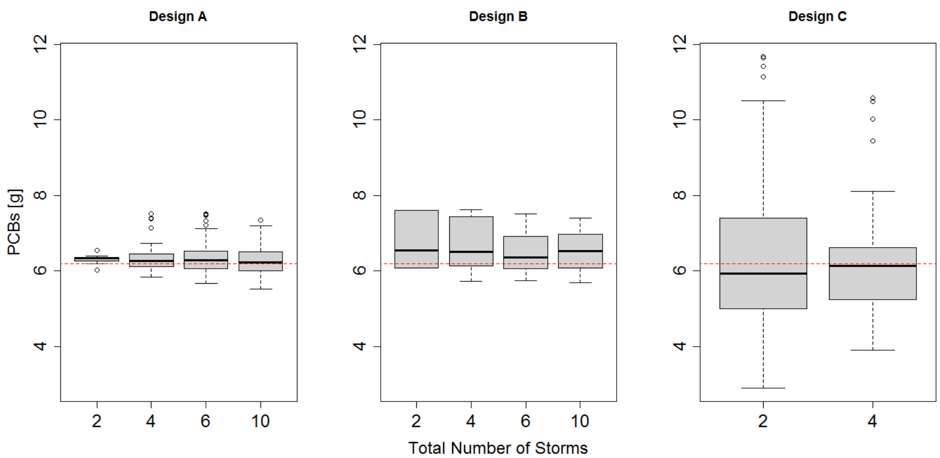
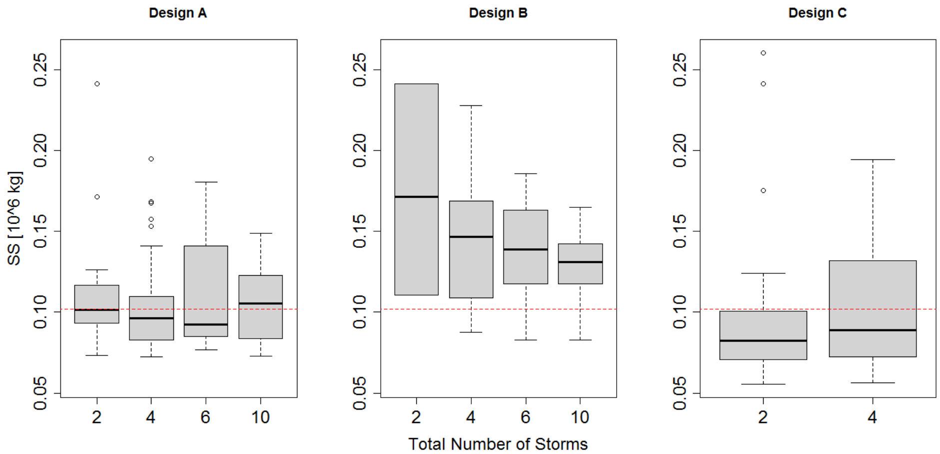


Figure 15a-b

a) 2007



b) 2009

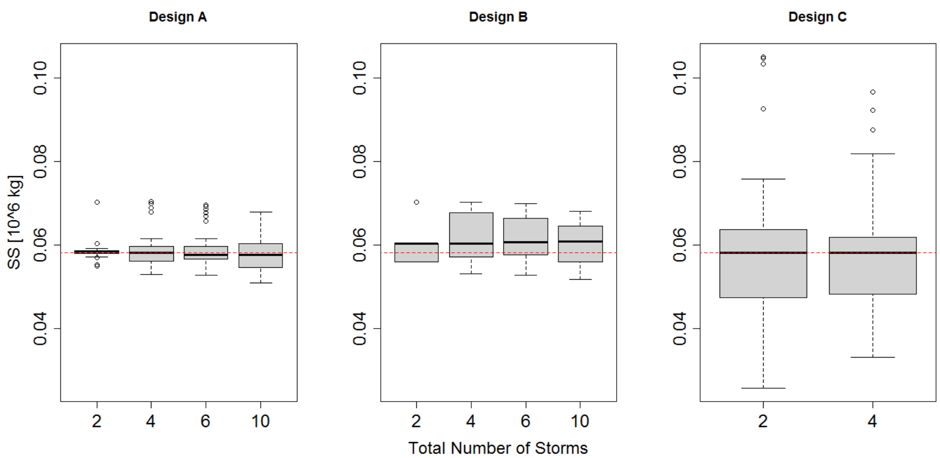


Figure 16a-b

Appendix A – SSC: Hg relationships at Zone 4 Line A

The mean slopes of SSC: Hg at Z4LA were below the target slopes for the trend analysis of 0.2 mg Hg / kg suspended sediment in two of the three years. Due to this situation, the Hg trend analysis examined power for trends assuming a target of 0.05 mg Hg / kg suspended sediment. The revised target was 75% below 0.2 mg/kg and was selected to ensure the trend could be examined for all years.

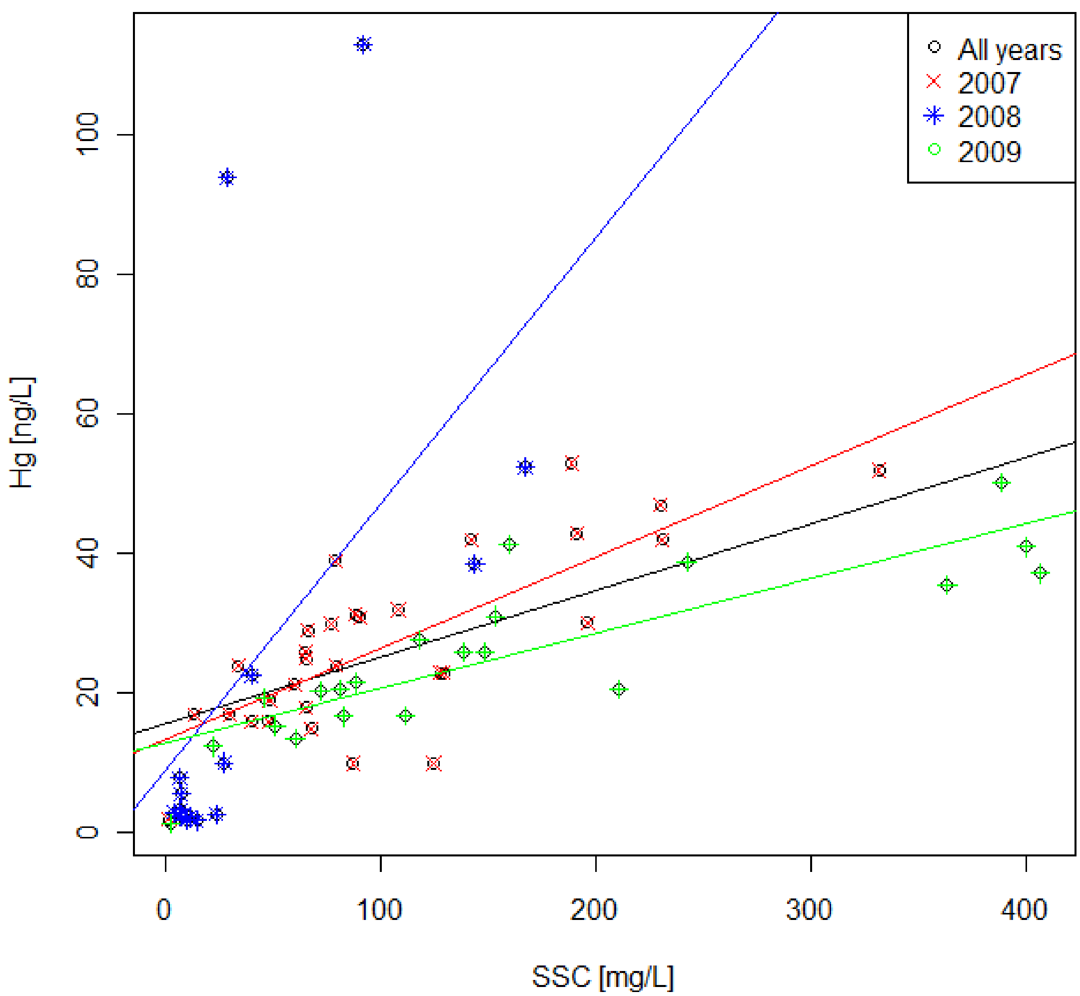
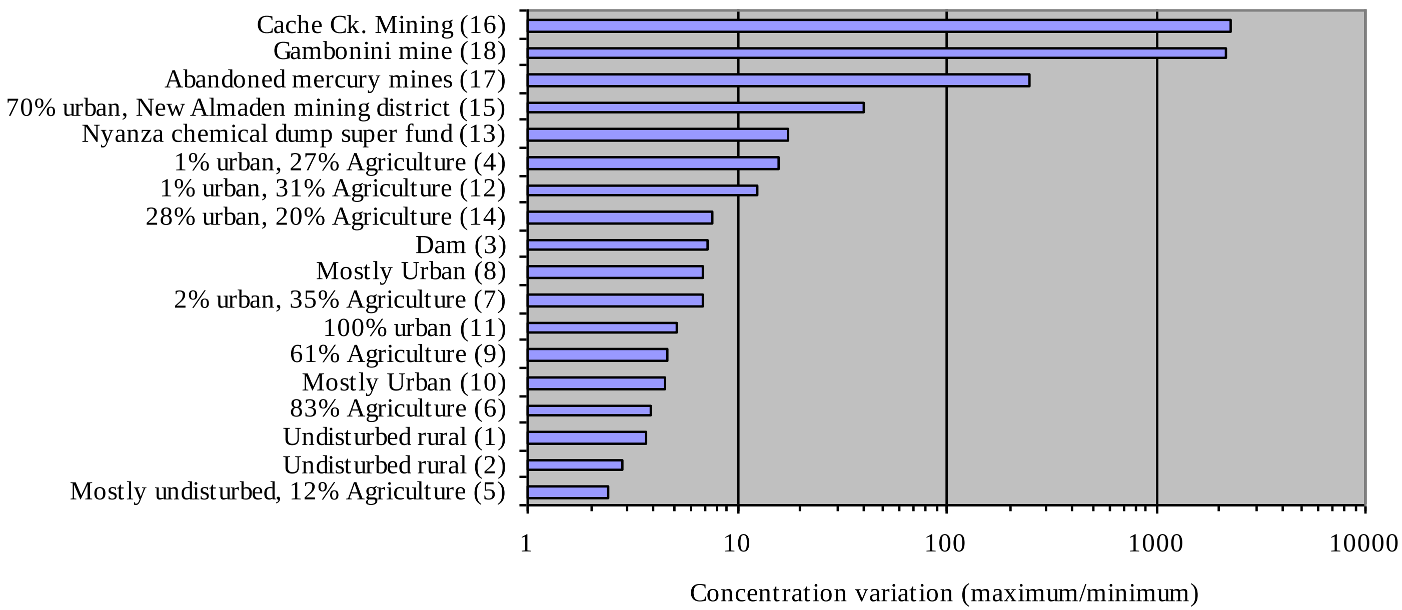


Figure A.1. SSC: Hg relationships at Zone 4 Line A. The mean slope estimate for 2007 – 2009 were 0.13, 0.38, and 0.08, respectively. The mean slope of all three years was 0.19.

Appendix B. Literature review of mercury variation by McKee et al. (2004)



	River (Source, see McKee et al. 2004)	Description	Min THg	Max THg	Max/min
5	Rappahannock R., Chesapeake (Lawson et al., 2001)	Mostly undisturbed, 12% Agriculture (5)	10.3	24.9	2.4
2	Site B1 Sudbury R., Massachusetts (Waldron et al., 2000)	Undisturbed rural (2)	1.9	5.4	2.8
1	Site B2 Sudbury R., Massachusetts (Waldron et al., 2000)	Undisturbed rural (1)	0.99	3.6	3.6
6	Choptank R., Chesapeake (Lawson et al., 2001)	83% Agriculture (6)	6.8	26.2	3.9
10	Anacostia R. NE. Branch (Mason and Sullivan, 1998)	Mostly Urban (10)	8.72	39.5	4.5
9	Susquehanna R., Chesapeake (Lawson et al., 2001)	61% Agriculture (9)	7	32.8	4.7
11	Herring Run R., Chesapeake (Lawson et al., 2001)	100% urban (11)	12.2	62.8	5.1
7	At Freeport, Sacramento Basin (Domagalski and Dileanis, 2000; Roth et al., 2001)	2% urban, 35% Agriculture (7)	4.2	29	6.9
8	Anacostia R. NW. Branch (Mason and Sullivan, 1998)	Mostly Urban (8)	4.45	30.8	6.9
3	Below Keswick Dam, Sacramento Basin (Domagalski and Dileanis, 2000; Roth et al., 2001)	Dam (3)	1.1	7.9	7.2
14	Potomac R., Chesapeake (Lawson et al., 2001)	28% urban, 20% Agriculture (14)	12.1	93.1	7.7
12	At Colusa, Sacramento Basin (Domagalski and Dileanis, 2000; Roth et al., 2001)	1% urban, 31% Agriculture (12)	6.5	81	12
4	Above Bend Bridge, Sacramento Basin (Domagalski and Dileanis, 2000; Roth et al., 2001)	1% urban, 27% Agriculture (4)	1.2	19	16
13	Site M1 Sudbury R., Massachusetts (Waldron et al., 2000)	Nyanza chemical dump super fund (13)	5.2	92	18
15	Guadalupe R., Bay Area (Leatherbarrow et al., 2002)	70% urban, New Almaden mining district (15)	18	730	41
16	Kuskakwim R. Basin, SW Alaska (Gray et al., 2000)	Abandoned mercury mines (17)	10	2500	250
17	Walker Ck. Marin County, California (Whyte and Kirchner, 2000)	Gambonini mine (18)	485	1040000	2144
15	Cache Ck., Sacramento Basin (Domagalski and Dileanis, 2000)	Cache Ck. Mining (16)	1	2250	2250

Appendix C –Within-storm Sampling Designs

Results are shown here for evaluation of within-storm sampling designs using flow and turbidity-based selection criteria. Bias here refers to the best estimate of loads per storm. Both flow and turbidity-based sampling criteria identified similar levels of accuracy (median load bias) and precision (standard error in load bias) in estimation of loads for the three pollutants (Hg, PCBs, suspended sediment).

C.1. Guadalupe River

Table C.1a. Summary of within-storm Hg loads (g) at Guadalupe River determined using three mass emission estimators. Flow-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	437	177	-50%	6%
Even	12		330	-21%	3%
Even	18		402	-14%	3%
Even	24		393	-10%	3%
Rising Stage	6		177	-50%	5%
Rising Stage	12		304	-25%	3%
Rising Stage	18		365	-15%	3%
Rising Stage	24		419	-10%	3%
			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	437	205	-40%	7%
Even	12		411	-5%	5%
Even	18		510	0%	4%
Even	24		429	3%	3%
Rising Stage	6		205	-40%	7%
Rising Stage	12		454	-3%	6%
Rising Stage	18		460	3%	4%
Rising Stage	24		520	2%	4%
			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	437	183	-47%	7%
Even	12		510	-2%	6%
Even	18		623	4%	5%
Even	24		533	10%	5%
Rising Stage	6		183	-45%	7%
Rising Stage	12		518	0%	7%
Rising Stage	18		479	9%	6%
Rising Stage	24		598	9%	6%

Table C.1b. Summary of within-storm Hg loads (g) at Guadalupe River determined using three mass emission estimators. Turbidity-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	285	149	-41%	4%
Even	12		241	-14%	3%
Even	18		251	-8%	-1%
Even	24		272	21%	3%
Rising Stage	6		149	-41%	4%
Rising Stage	12		237	-12%	4%
Rising Stage	18		277	-4%	4%
Rising Stage	24		289	0%	4%
			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	285	175	-28%	4%
Even	12		257	-4%	3%
Even	18		284	0%	1%
Even	24		275	-1%	2%
Rising Stage	6		175	-30%	6%
Rising Stage	12		270	7%	4%
Rising Stage	18		274	4%	3%
Rising Stage	24		299	2%	2%
			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	285	177	-30%	6%
Even	12		293	2%	4%
Even	18		321	12%	3%
Even	24		331	9%	3%
Rising Stage	6		178	-26%	7%
Rising Stage	12		307	13%	5%
Rising Stage	18		301	14%	5%
Rising Stage	24		324	17%	5%

Table C.1c. Summary of within-storm PCB loads (g) at Guadalupe River determined using three mass emission estimators. Flow-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	29	37	-18%	7%
Even	12		33	-4%	6%
Rising Stage	6		37	-15%	6%
Rising Stage	12		42	-1%	6%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	29	40	-9%	8%
Even	12		34	2%	5%
Rising Stage	6		43	-1%	7%
Rising Stage	12		52	5%	6%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	29	38	-13%	7%
Even	12		33	15%	7%
Rising Stage	6		39	1%	7%
Rising Stage	12		53	14%	7%

Table C.1d. Summary of within-storm PCB loads (g) at Guadalupe River determined using three mass emission estimators. Turbidity-based storm selection criteria.

			Simple Mean		
Emphasis	Sample Size	Median "Best Estimate" Load	Median Load	Median Bias	St. Error
Even	6	10	9.1	-10%	4%
Even	12		11.3	-2%	4%
Rising Stage	6		9.1	-8%	5%
Rising Stage	12		11.5	3%	6%

			Linear Interpolation		
Emphasis	Sample Size	Median "Best Estimate" Load	Median Load	Median Bias	St. Error
Even	6	10	9.3	-3%	4%
Even	12		11.3	-1%	4%
Rising Stage	6		9	-1%	7%
Rising Stage	12		13.5	12%	6%

			Flow-weighted Mean		
Emphasis	Sample Size	Median "Best Estimate" Load	Median Load	Median Bias	St. Error
Even	6	10	9.1	-3%	5%
Even	12		13.4	10%	6%
Rising Stage	6		9.1	-1%	8%
Rising Stage	12		14.9	19%	8%

Table C.1e. Summary of within-storm suspended sediment loads (kg) at Guadalupe River determined using three mass emission estimators. Flow-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	224	111	-50%	4%
Even	12		199	-20%	5%
Even	18		197	-10%	4%
Even	24		219	-10%	4%
Rising Stage	6		111	-46%	4%
Rising Stage	12		182	-17%	4%
Rising Stage	18		218	-8%	4%
Rising Stage	24		219	-1%	4%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	224	124	-40%	5%
Even	12		226	-1%	5%
Even	18		254	2%	4%
Even	24		245	2%	3%
Rising Stage	6		124	-34%	6%
Rising Stage	12		244	3%	6%
Rising Stage	18		258	6%	4%
Rising Stage	24		278	6%	4%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	224	114	-47%	6%
Even	12		270	7%	8%
Even	18		286	16%	6%
Even	24		280	19%	5%
Rising Stage	6		114	-38%	8%
Rising Stage	12		282	17%	7%
Rising Stage	18		281	22%	6%
Rising Stage	24		316	20%	5%

Table C.1f. Summary of within-storm suspended sediment loads (kg) at Guadalupe River determined using three mass emission estimators. Turbidity-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	70	47	-37%	4%
Even	12		57	-19%	3%
Even	18		63	-2%	3%
Even	24		64	-3%	3%
Rising Stage	6		47	-37%	1%
Rising Stage	12		68	-4%	4%
Rising Stage	18		68	3%	4%
Rising Stage	24		72	4%	4%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	70	50	-24%	5%
Even	12		68	-2%	3%
Even	18		72	2%	1%
Even	24		68	1%	2%
Rising Stage	6		50	-20%	6%
Rising Stage	12		78	18%	3%
Rising Stage	18		76	9%	3%
Rising Stage	24		77	5%	2%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	70	48	-28%	6%
Even	12		74	13%	4%
Even	18		78	22%	3%
Even	24		80	19%	3%
Rising Stage	6		48	-19%	7%
Rising Stage	12		86	31%	4%
Rising Stage	18		86	30%	3%
Rising Stage	24		89	27%	3%

C.2. Zone 4 Line A

Table C.2a. Summary of within-storm Hg loads (mg) at Zone 4 Line A determined using three mass emission estimators. Flow-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	290	142	-46%	3%
Even	12		180	-26%	3%
Even	18		248	-14%	3%
Even	24		258	-8%	2%
Rising Stage	6		143	-41%	3%
Rising Stage	12		275	-15%	4%
Rising Stage	18		272	-5%	3%
Rising Stage	24		295	1%	3%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	290	144	-44%	4%
Even	12		203	-11%	4%
Even	18		279	1%	3%
Even	24		278	3%	1%
Rising Stage	6		144	-38%	5%
Rising Stage	12		351	13%	4%
Rising Stage	18		311	9%	3%
Rising Stage	24		310	7%	2%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	290	146	-44%	4%
Even	12		225	-9%	6%
Even	18		351	17%	5%
Even	24		358	17%	3%
Rising Stage	6		153	-38%	6%
Rising Stage	12		401	28%	7%
Rising Stage	18		395	28%	5%
Rising Stage	24		388	28%	4%

Table C.2b. Summary of within-storm Hg loads (mg) at Zone 4 Line A determined using three mass emission estimators. Turbidity-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	152	93	-41%	3%
Even	12		113	-12%	4%
Even	18		138	-5%	3%
Even	24		136	-1%	3%
Rising Stage	6		99	-34%	3%
Rising Stage	12		127	-1%	4%
Rising Stage	18		147	5%	4%
Rising Stage	24		152	8%	4%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	152	108	-30%	4%
Even	12		143	3%	2%
Even	18		166	3%	1%
Even	24		158	3%	1%
Rising Stage	6		109	-22%	7%
Rising Stage	12		184	22%	4%
Rising Stage	18		177	12%	2%
Rising Stage	24		166	8%	2%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	152	95	-29%	5%
Even	12		157	13%	4%
Even	18		192	17%	4%
Even	24		187	18%	4%
Rising Stage	6		109	-22%	7%
Rising Stage	12		202	34%	5%
Rising Stage	18		191	33%	5%
Rising Stage	24		196	31%	4%

Table C.2c. Summary of within-storm PCB loads (mg) at Zone 4 Line A determined using three mass emission estimators. Flow-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	131	59	-47%	3%
Even	12		89	-26%	3%
Rising Stage	6		61	-42%	3%
Rising Stage	12		119	-15%	4%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	131	62	-44%	4%
Even	12		104	-11%	4%
Rising Stage	6		65	-40%	5%
Rising Stage	12		145	13%	4%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	131	61	-44%	4%
Even	12		113	-9%	6%
Rising Stage	6		63	-38%	6%
Rising Stage	12		180	27%	7%

Table C.2d. Summary of within-storm PCB loads (mg) at Zone 4 Line A determined using three mass emission estimators. Turbidity-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	67	39	-45%	3%
Even	12		54	-14%	4%
Rising Stage	6		44	-40%	3%
Rising Stage	12		58	-5%	5%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	67	44	-35%	4%
Even	12		63	3%	2%
Rising Stage	6		48	-22%	8%
Rising Stage	12		93	18%	4%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	67	47	-35%	5%
Even	12		68	14%	5%
Rising Stage	6		47	-22%	8%
Rising Stage	12		90	34%	6%

Table C.2e. Summary of within-storm suspended sediment loads (g) at Zone 4 Line A determined using three mass emission estimators. Flow-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	1244	536	-54%	4%
Even	12		705	-29%	4%
Even	18		1055	-17%	3%
Even	24		1171	-7%	2%
Rising Stage	6		536	-48%	4%
Rising Stage	12		1105	-15%	4%
Rising Stage	18		1163	-3%	3%
Rising Stage	24		1407	3%	3%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	1244	549	-51%	5%
Even	12		947	-11%	4%
Even	18		1150	0%	3%
Even	24		1272	3%	1%
Rising Stage	6		585	-46%	6%
Rising Stage	12		1381	15%	5%
Rising Stage	18		1430	11%	3%
Rising Stage	24		1443	8%	3%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	1244	551	-51%	6%
Even	12		1185	-10%	6%
Even	18		1337	15%	6%
Even	24		1592	18%	4%
Rising Stage	6		565	-43%	7%
Rising Stage	12		1739	33%	8%
Rising Stage	18		1594	35%	5%
Rising Stage	24		1836	32%	5%

Table C.2f. Summary of within-storm suspended sediment loads (g) at Zone 4 Line A determined using three mass emission estimators. Turbidity-based storm selection criteria.

			Simple Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	653	328	-43%	4%
Even	12		590	-11%	6%
Even	18		643	-6%	5%
Even	24		662	2%	6%
Rising Stage	6		342	-38%	4%
Rising Stage	12		673	-1%	7%
Rising Stage	18		791	8%	7%
Rising Stage	24		741	10%	7%

			Linear Interpolation		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	653	351	-35%	4%
Even	12		682	4%	4%
Even	18		653	4%	3%
Even	24		640	4%	3%
Rising Stage	6		351	-29%	7%
Rising Stage	12		956	19%	5%
Rising Stage	18		745	15%	3%
Rising Stage	24		655	9%	3%

			Flow-weighted Mean		
Emphasis	**Sample Size**	**Median "Best Estimate" Load**	**Median Load**	**Median Bias**	**St. Error**
Even	6	653	353	-34%	6%
Even	12		709	13%	7%
Even	18		768	22%	5%
Even	24		820	21%	6%
Rising Stage	6		353	-27%	7%
Rising Stage	12		918	37%	7%
Rising Stage	18		890	36%	7%
Rising Stage	24		811	34%	7%

Appendix D. Among-storm sampling designs (turbidity-based sampling)

Results are presented here for among-storm sampling designs with turbidity-based storm sampling. These results were generally more variable than the flow-based results presented in the main text of the report.

Figure Captions

Figure D.1a-c. Comparison of annual Hg loads at Guadalupe River in 2003-2005 based on three designs for sampling among storms (Table 5). Loads were calculated using linear interpolation with turbidity-based storm selection criteria. Design A simulated sampling of the first flush only and a variable number of random storms. Design B simulated the first flush plus one of the three largest storms and a variable number of random storms. Design C only tested the random storm component. --------- = best estimate Hg load for year

Figure D.2a-c. Comparison of annual PCB loads at Guadalupe River in 2003-2005 based on three designs for sampling among storms (Table 5). See Figure D.1 caption and text for further information. --------- = best estimate PCB load for year.

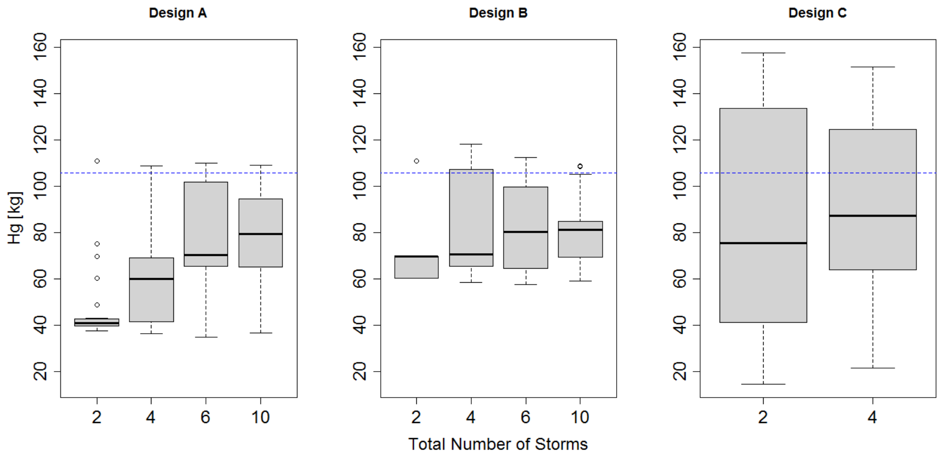
Figure D.3a-c. Comparison of annual SS loads at Guadalupe River in 2003-2005 based on three designs for sampling among storms (Table 5). See Figure D.1 caption and text for further information. --------- = best estimate SS load for year.

Figure D.4a-c. Comparison of annual Hg loads at Zone 4 Line A in 2007 and 2009 based on three designs for sampling among storms (Table 5). See Figure D.1 caption and text for further information. --------- = best estimate Hg load for year.

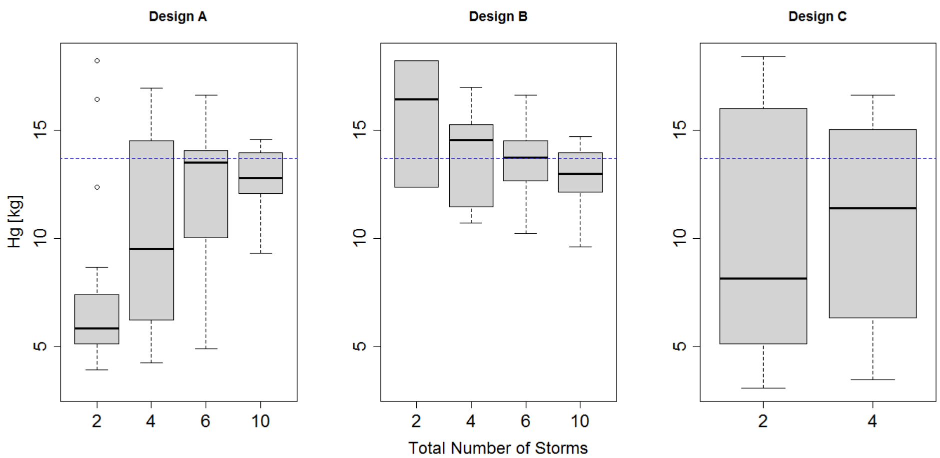
Figure D.5a-c. Comparison of annual PCB loads at Zone 4 Line A in 2007 and 2009 based on three designs for sampling among storms (Table 5). See Figure D.1 caption and text for further information. --------- = best estimate PCB load for year.

Figure D.6a-c. Comparison of annual SS loads at Zone 4 Line A in 2007 and 2009 based on three designs for sampling among storms (Table 5). See Figure D.1 caption and text for further information. --------- = best estimate SS load for year.

a) 2003



b) 2004



c) 2005

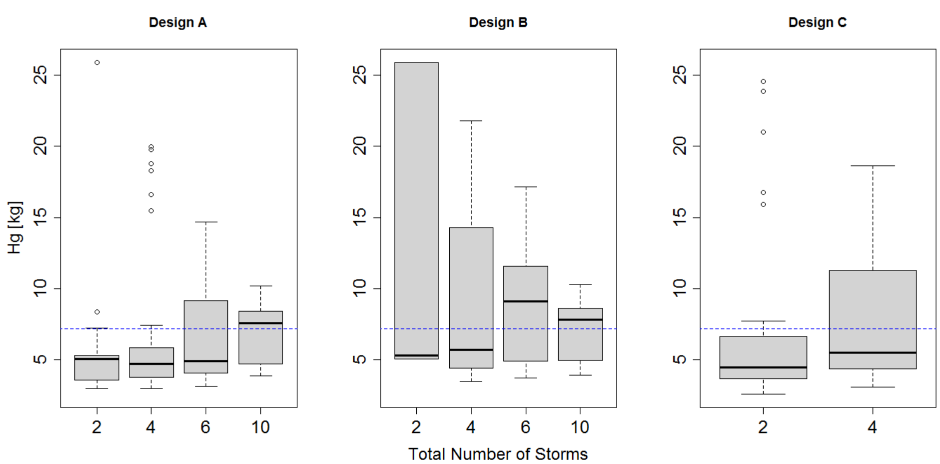
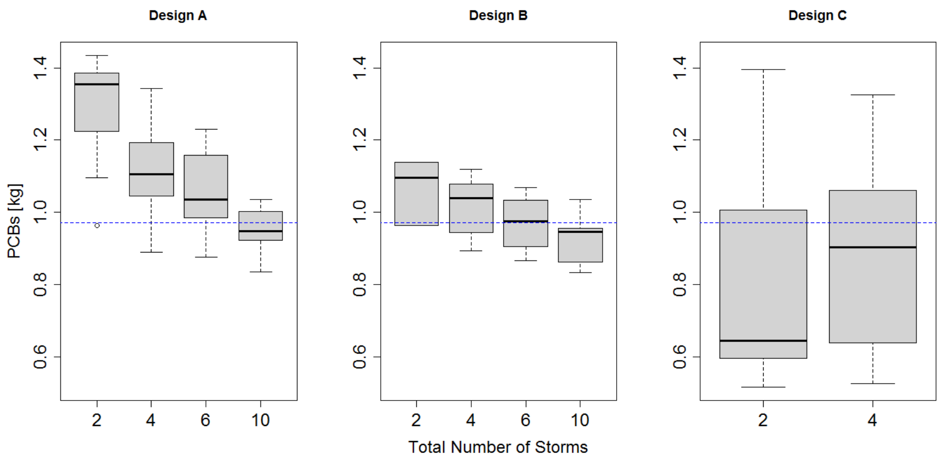
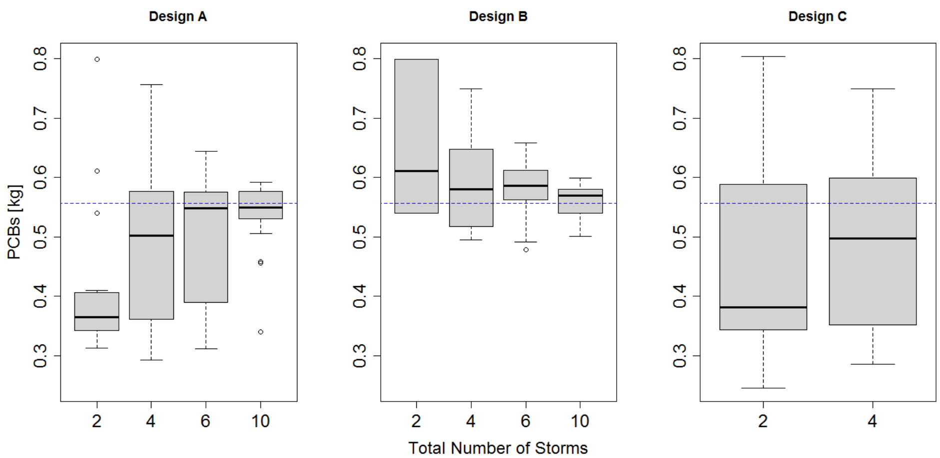


Figure D.1a-c

a) 2003



b) 2004



c) 2005

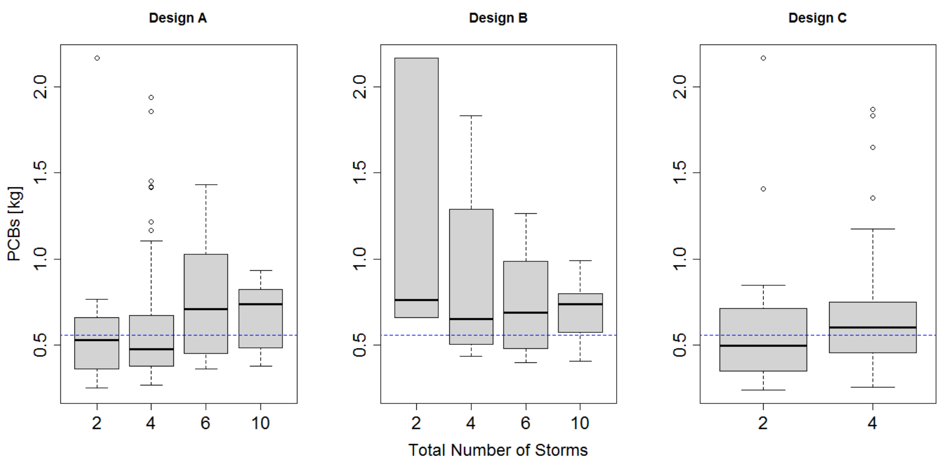
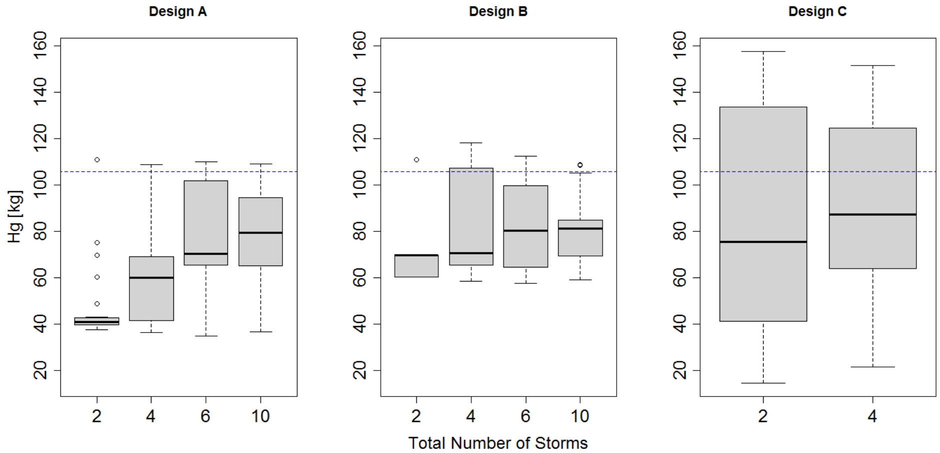
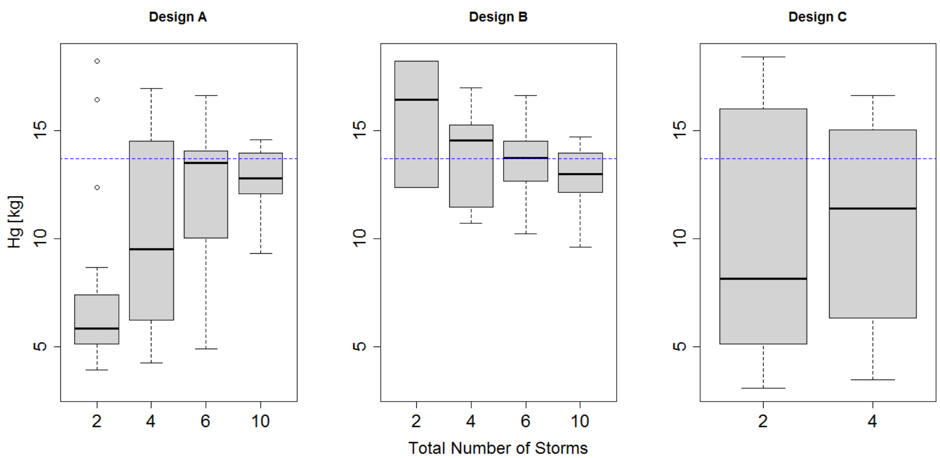


Figure D.2a-c

a) 2003



b) 2004



c) 2005

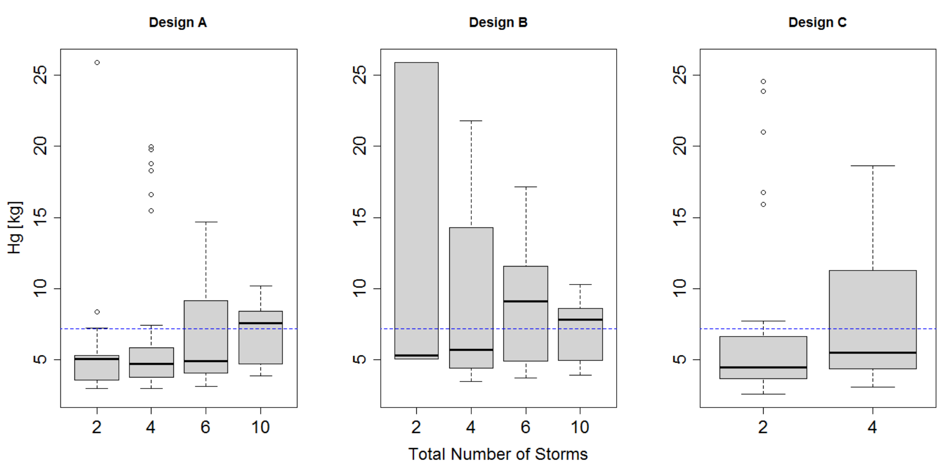
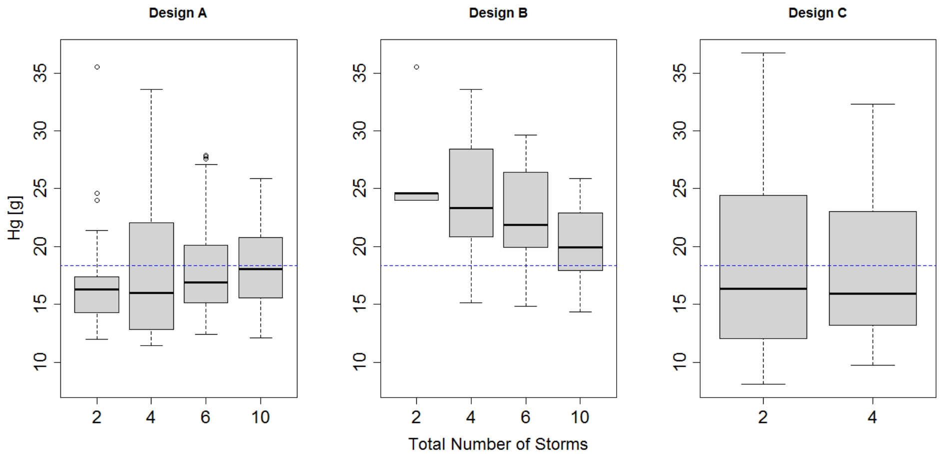


Figure D.3a-c

a) 2007



b) 2009

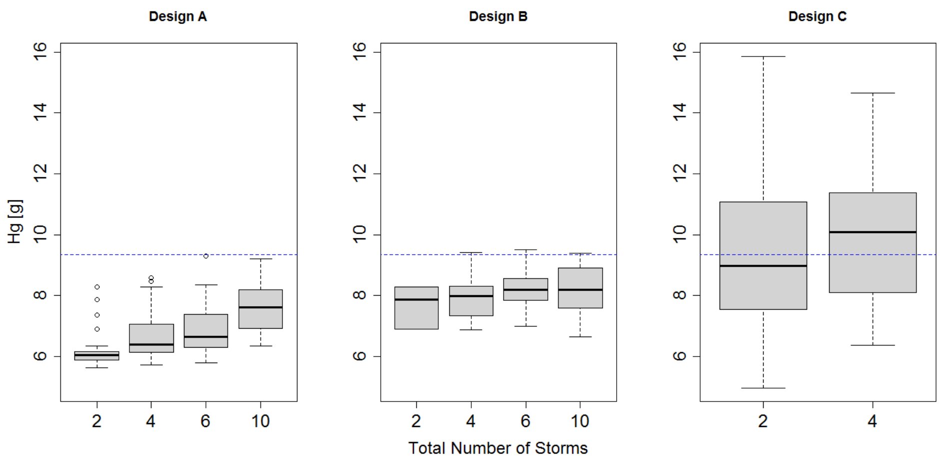
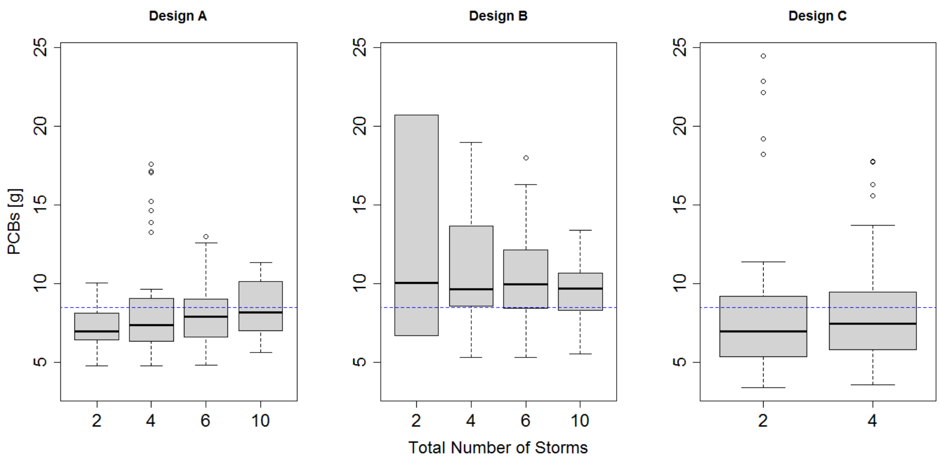


Figure D.4a-b

a) 2007



b) 2009

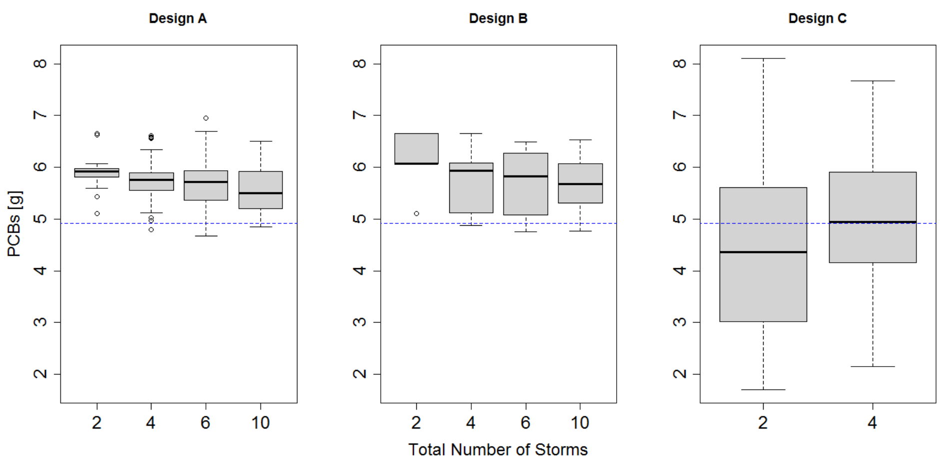
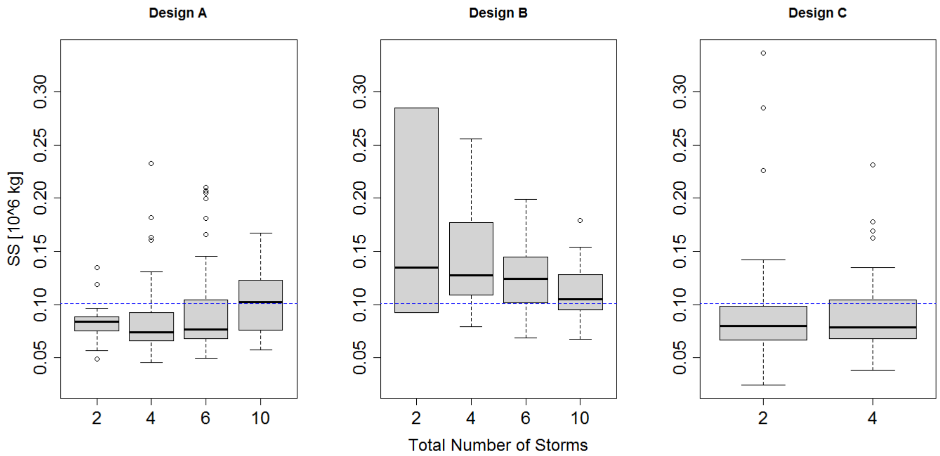


Figure D.5a-b

a) 2007



b) 2009

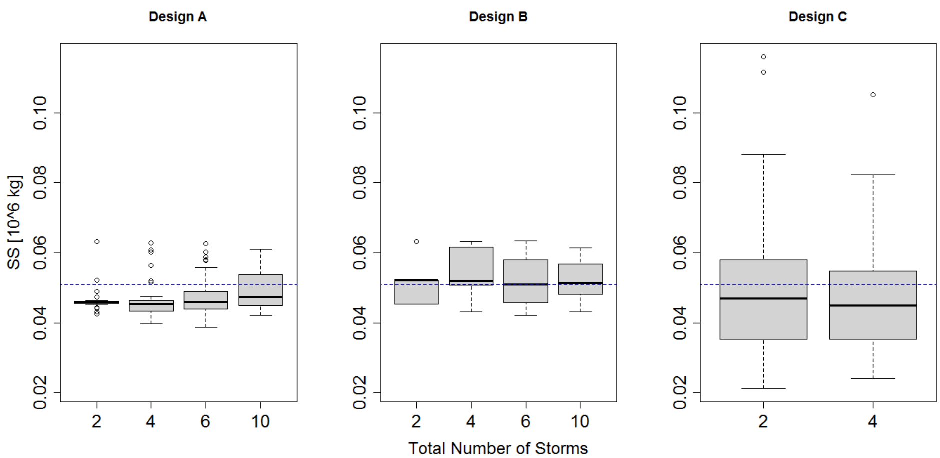


Figure D.6a-b

Appendix E. Comparison of annual PCB and SS loads at Guadalupe River in 2003 – 2005 based on three designs for sampling among storms.

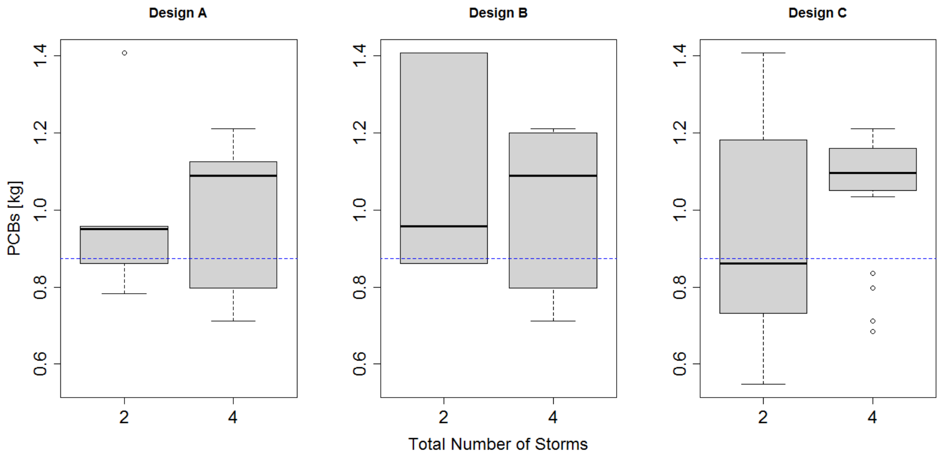
Results for estimation of annual PCB and SS loads at Guadalupe River mirrored that of Hg loads. Accuracy and precision were optimal at 6 or 10 storms samples per water year. Designs A and B performed the best and similarly well, with Design C exhibiting good accuracy, but poor precision.

Figure Captions

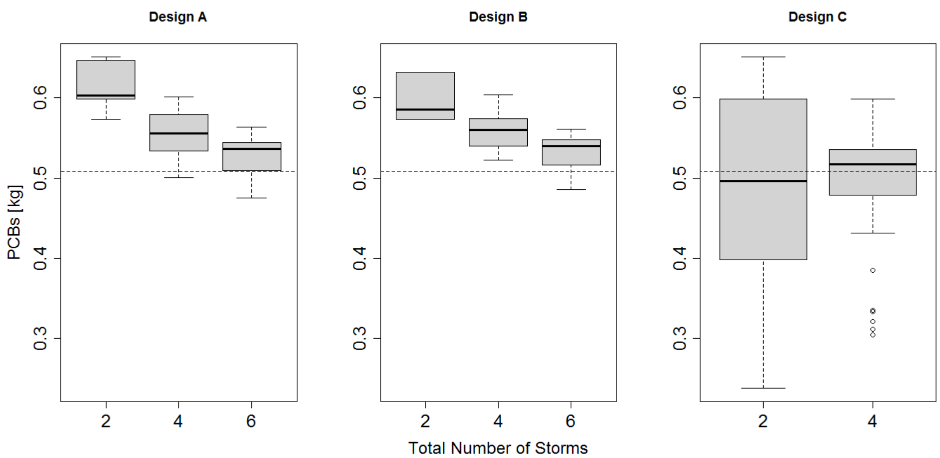
Figure E.1a-c. Comparison of annual PCB loads at Guadalupe River in 2003-2005 based on three designs for sampling among storms (Table 5). Loads were calculated using linear interpolation with flow-based storm selection criteria. In addition to the random component, Designs A and B simulated sampling of the first flush (A) and first flush plus one of the three largest storms (B). Design C only tested the random storm component. --------- = best estimate PCB load for year (0.9 kg, 0.5 kg, and 0.5 kg, respectively).

Figure E.2a-c. Comparison of annual SS loads at Guadalupe River in 2003-2005 based on three designs for sampling among storms (Table 5). Loads were calculated using linear interpolation with flow-based storm selection criteria. In addition to the random component, Designs A and B simulated sampling of the first flush (A) and first flush plus one of the three largest storms (B). Design C only tested the random storm component. --------- = best estimate SS load for year (10 x 10^6 kg, 8 x 10^6 kg, and 4 x 10^6 kg, respectively)

a) 2003



b) 2004



c) 2005

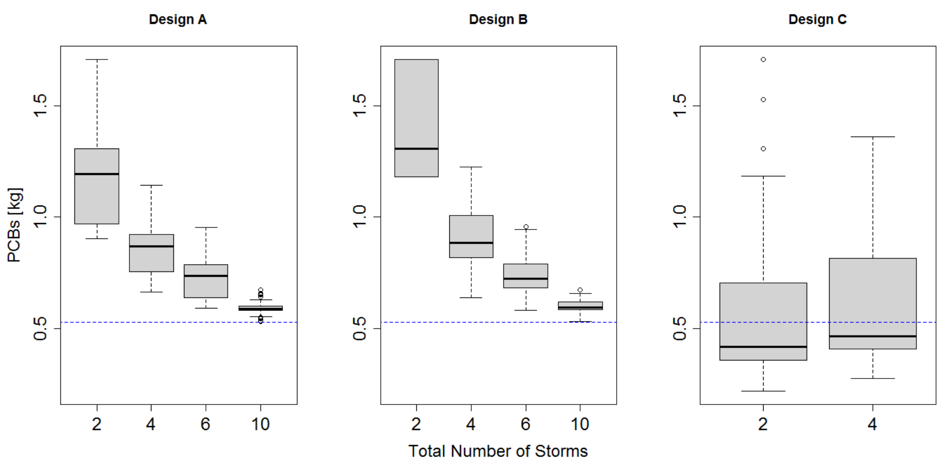
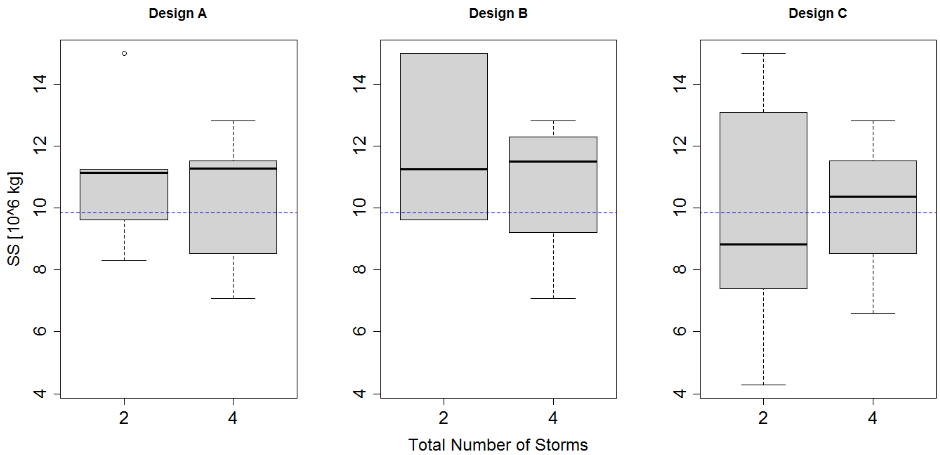
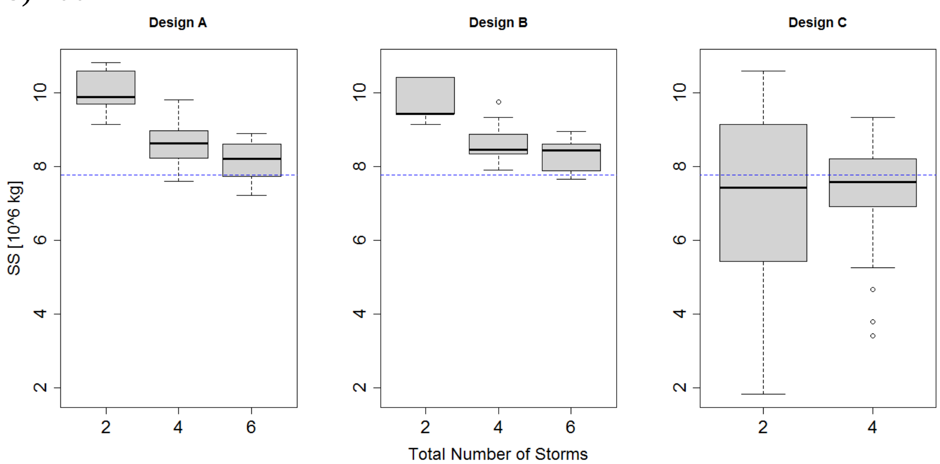


Figure 1a-c.

a) 2003



b) 2004



c) 2005

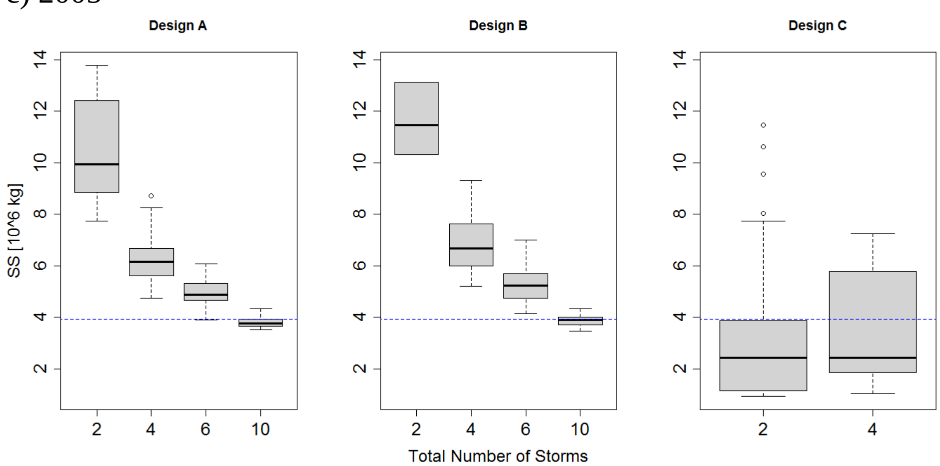


Figure 2a-c.