RMP
REGIONAL MONITORING PROGRAM FOR WATER QUALITY IN SAN FRANCISCO BAY

sfei.org/rmp

Statistical Methods Development and Sampling Design Optimization to Support Trends Analysis for Loads of Polychlorinated Biphenyls from the Guadalupe River in San Jose, California, USA

Prepared by

Aroon Melwani, Applied Marine Sciences

Don Yee, Lester McKee, Alicia Gilbreath, Philip Trowbridge, Jay Davis, San Francisco Estuary Institute

CONTRIBUTION NO. 876 / September 2018

Citation:

Melwani, A.R., Yee, D., McKee, L., Gilbreath, A., Trowbridge, P., and Davis, J.A. (2018). Statistical Methods Development and Sampling Design Optimization to Support Trends Analysis for Loads of Polychlorinated Biphenyls from the Guadalupe River in San Jose, California, USA, Final Report. Applied Marine Sciences: Livermore, CA

Table of Contents

Executive Summary

The Regional Monitoring Program for Water Quality in San Francisco Bay (RMP), with guidance from its Small Tributaries Loading Strategy (STLS) Team, has been conducting small tributary loading studies at several sites in the Bay Area since 2003. A current priority for the STLS is developing a strategy for measuring trends in pollutant loadings from small tributaries to the Bay.

This technical report presents the work to develop a statistical model for trends in loads of polychlorinated biphenyls (PCBs), and estimate the power for proposed monitoring designs as a basis for detecting trends for the Guadalupe River watershed (San Jose, California). The statistical approach builds upon the turbidity surrogate methodology that has been employed in the STLS since 2003. A novel, two-stage statistical modeling approach was used to incorporate the significant turbidity-PCB relationships that exist, and evaluate climatic, seasonal, and inter-annual factors as additional potential drivers of PCB loads. The longest-running time series of tributary monitoring by the STLS on the Guadalupe River was selected as the case study for developing and testing the statistical approach. As a result of this effort, a multiple linear regression model for detecting trends in PCB loads in the Guadalupe River was developed.

There are two main findings from the modeling effort. Firstly, the statistical models did not find a significant linear inter-annual trend in current PCB loads for the period 2003-2014 after accounting for climatic variability. Secondly, simulations to estimate the power for detecting significant trends in PCB loads, employing a grab-based design using the proposed statistical methods would have sufficient (> 80%) power to detect 25% or greater trends over a 20-year period. A parallel analysis to simulate a composite-based sampling design indicated that this approach was less sensitive to trends, and could only detect larger trends (>75%) over 20 years. Overall, the simulations indicated that the higher sample size and relatively low standard error of load estimates of the grab-based methodology holds promise for detecting trends of management interest in the future. Finally, a cost estimate for grabs and composite designs was included as an appendix using the following assumptions: grab sampling in 10 years over a 20 year period with 4 storm events/year and 4 grab samples/storm; composite sampling in 12 years over a 20 year period with 2 storm events/year.

To date, the statistical models developed in this project have only been parameterized for the Guadalupe River, which may be subject to refinement and testing, as future PCB monitoring data are collected in the watershed. A similar approach, may also be effective if developed for other tributaries, but the combination of variables used to develop the models will likely be different. Therefore, the modeling techniques developed here could be applied to other watersheds in the future, to determine if the approach, if not the specific model parameters and coefficients, is generally applicable.

Introduction

The Regional Monitoring Program for Water Quality in San Francisco Bay (RMP), through its Small Tributary Loading Strategy (STLS) Team has conducted small tributary loading studies at several sites in the Bay Area since 2003 (David et al. 2009; Davis et al. 2007; McKee et al. 2004; McKee et al. 2005; McKee and Gilbreath 2009; McKee et al. 2006; Oram et al. 2008). The primary objective of these studies has been to determine annual loads of particle-associated pollutants of concern (particularly, Hg, PCBs, and suspended sediments). The longest running time series of loads collected in the RMP exists for Guadalupe River in San Jose (2003-2014; McKee et al. 2017), where a turbidity surrogate methodology has been employed, as it has been demonstrated to be an appropriate and cost-effective method for 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; SFRWQCB 2015).

The San Francisco Bay Regional Water Quality Control Board (Water Board) has developed Total Maximum Daily Load (TMDL) requirements for major pathways of Hg and PCB loading to San Francisco Bay. Stormwater load allocations are particularly stringent, allowing for annual discharge of 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. These targets represent estimated reductions of 50% and 90% from the present load estimates of 160 kg of Hg and 20 kg of PCBs, respectively.

To meet these goals, the TMDL reports and the MRP call for improved loads information to answer four key management questions:

1. Identify which San Francisco Bay tributaries contribute most to impairment from pollutants of concern (PCBs, Hg, suspended sediments),
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.

A current priority for the STLS is developing a strategy for measuring trends in pollutant loading from small tributaries to the Bay. STLS monitoring data have demonstrated a high-level of precision for quantifying annual loads. However, recent analyses conducted to evaluate intra- and inter-annual variability, suggested an apparent lack of power for detecting trends, unless unrealistic sampling effort is performed (Melwani 2016). Upon detailed review of these results, the STLS Team concluded the apparent lack of power might not be simply due to insufficient sampling effort or the consequence of the sampling approaches used. Variability due to physical climatic processes affecting inter-annual patterns in loads needs to be examined.

In this study, a statistical modeling procedure was developed to test whether the temporal pattern in loads of a sediment-associated pollutant of concern (i.e., PCBs) may be estimated by incorporating climatic surrogates for various fluvial processes in combination with measured turbidity. The basis for including climatic processes into a statistical modeling framework was to improve the predictive power for annual loads estimation. However, employing such an approach for detecting trends in loads has not previously been investigated for PCBs or other particle-associated contaminants. Therefore, this report describes a new approach to testing for trends in small tributary loads, and subsequently employs power analyses to examine sampling designs that could be implemented in the STLS monitoring program to collect data to support implementation of the statistical framework. This work focused on PCBs in the Guadalupe River as a case study.

Methods

Background

The Guadalupe River is located near San Jose, the largest city in the San Francisco Bay Area. Its watershed is the fourth largest in the Bay Area, representing about 500 km^2. A station in the lower watershed has been monitored with funding by the RMP and Santa Clara Valley Urban Runoff Pollution Prevention Program (SCVURPP), in collaboration with the USGS (station number 11169025) since October 2002. The station is located upstream from tidal influence, but resides downstream from five main reservoirs, the City of San Jose, and the majority of flood control channels. The continuous flowing area downstream from reservoirs is estimated to be 236 km^2, of which approximately 80% is urbanized landscape. The typical flood hydrograph produced by heavy rainfall will pass through the watershed over a period of ~ 12-24 hours, but larger and later season (February – April) floods may last for several days.

Data Used

Eight years of urban runoff data collected from the Guadalupe River during a 12-year span (2003-2014) were statistically analyzed in this project. The eight water years (WY; October 1-September 30) spanned a range in climatic conditions, and thus a range in wet weather discharge and pollutant concentrations. Four WYs during the period were not monitored: 2007-2009, and 2011. Data collected during the 12-year span has shown that sediment loads are almost exclusively transported during high flow events, and not during base flow. The peak discharge during this period was 6071 cfs in WY 2003. In comparison, during WYs 2005 and 2012-2014, the maximum discharge did not exceed 1000 cfs. Under base flow conditions, peak discharge was usually under 500 cfs.

Over the course of the eight years, continuous monitoring of turbidity and discharge was performed, in addition to the manual collection and analysis of grab samples for pollutant and suspended sediment concentrations at several times during the course of storms.

These data were the basis for the statistical modeling procedures that follow. Further descriptive summaries of these data have been described previously (Melwani 2016).

Statistical Modeling Approach

Three key statistical analysis steps were employed to develop and test a statistical model for trend analysis in the Guadalupe River watershed. All statistical and graphical analyses were conducted in R Studio (running R v. 3.3.1). The following section describes the details of each step in the analysis process.

Step 1

A generalized linear modeling approach was used to examine variation in natural log PCB concentrations. Several iterations of the model evaluation process were conducted to test the climatic and temporal variables that provided the most robust predictions. Forty-four linear models consisting of 26 predictor variables were examined. Table 1 summarizes the general types of predictors and parameters considered. It should be noted that the initial list of predictor variables tested was more extensive to test support for varying antecedence and rainfall lag times, and methods for incorporating antecedence patterns based on best professional judgment by the STLS Trends Team (Appendix A). The footnote to Table 1 specifies the final set of model parameters employed in the model selection framework.

The general equation of the regression model was:

$$ln(PCBs) = B_0 + B_1 ln(Turbidity) + B_2 X_i + B_3 T_i + \epsilon$$

Where, ln(PCBs) is PCB concentration in natural log units, B_0 – B_3 are fitted coefficients of the model, ln(Turbidity) is turbidity in natural log units, X_i values are additional raw or transformed predictor variables, T_i values are time variables representing month (categorical), season (categorical), and/or water year (continuous), and ε is the unexplained variation (residual error).

Model parameters were estimated by least squares multiple regression following a Maximum Likelihood approach. In all candidate models, turbidity was treated as the primary continuous predictor variable in the default model – Model 1. In examining subsequent sets of models, the analysis sought to incorporate factors related to discharge, rainfall patterns, and hysteresis signals of the watershed. All candidate models were evaluated independently. Both linear and polynomial functions were considered, as were the interaction between predictor variables over water year. Temporal trend was evaluated using water year (centered on the mean of the dataset). Rainfall variables with values of < 0.02” were substituted with 0.02” to facilitate logarithmic transformation. Selection of the ‘best’ model was made based on the AICc statistic, the AICc weights, statistical significance of model terms, and the minimum number of significant terms.

Following selection of the top model(s), diagnostic tests were performed on the model residual values (i.e., the unexplained variance). Tests included checking for parametric assumptions of normality (Shapiro-Wilk test) and homogeneity of variance (Goldfeld-

Quandt test). Graphical assessments of patterns in residuals were conducted to search for violations of independence plotted against predicted values, covariates (e.g., discharge and SSC), and temporal variables (i.e., month, water year). To ensure that all systematic temporal variability was captured in this model, linear regression of residuals against water year (centered on the mean) and month was performed. If any systematic temporal variability remained unexplained at this stage, it would invalidate the subsequent trend analysis conducted on predictions from the model. Summary statistics of the models included evaluation of adjusted and predictive R^2 values and the PRESS statistic. The predictive R^2 and PRESS statistics are cross validation measures commonly used to gauge the quality of models with regards to making predictions. Due to the complexity of statistical testing of time series with irregularly sampled data, autocorrelation bias could not be assessed with these data.

The last set of analyses to examine the concentration model outputs was to conduct two cross-validation (CV) experiments. Cross validation is a routine model validation technique for assessing how well the results of a statistical analysis will generalize to an independent data set. The first method used k-fold CV, where all the samples were randomly partitioned into 'k' sample sets (referred to as folds) of approximately equal size. Subsequently, the selected model was fit using all the samples except the first subset and the prediction error of the model calculated against the held-out samples. This procedure was repeated for all sample sets, and the model performance calculated as the mean error among the different sets. Secondly, the Leave-One-Out CV method was employed due to concerns that limited sample sizes may bias the k-fold method. In this case, the selected model was fit using all the samples except a single sample, and the prediction error of the model calculated in a similar fashion as described above.

Step 2

The second step in the analytical procedure was to apply the concentration model from Step 1 to a long-term continuous record of turbidity and precipitation data to estimate storm event loads. In the case of Guadalupe River, this consisted of 15-min observations spanning October 2002 to April 2014.

To identify the data reflecting storm events, the presence of storms was assessed based on flow and precipitation criteria. All selected storm events required a peak flow of at least 500 cfs. The start of an event was defined by flow rising above 150 cfs (or cumulative precipitation >0.1" if baseline flow was already >150 cfs), and the end by flow dropping below 150 cfs, or no rainfall at either CSJ or LP stations for 2 and 7 hours, respectively (enough time for surface runoff from either area to reach the monitored station). For the purposes of this analysis, flow that did not meet these criteria was deemed base flow.

Following the storm selection criteria, predicted PCB concentrations and their respective standard errors comprising each set of flagged storms were converted to instantaneous loads. As the predictions were generated on a natural log-scale, both the predicted values and the standard errors were back-transformed for the loads calculation. The back-transformation included a bias correction factor (BCF) following the method of Newman (1993). Specifically, the residuals of the selected model were used to calculate the mean-square-error (MSE), and subsequently the BCF = MSE/2. The re-transformed values were then converted to loads by multiplying concentration by discharge, and the instantaneous loads summed for all predictions comprising an event to determine 'event loads'. Seventy-five storms were identified through this method, and reflect the data employed in the trend model analysis in Step 3.

Two calculations were performed to propagate the standard errors generated from the concentration predictions into the event load derivation. First, the standard errors in predicted concentrations was added to the error in discharge measurements (assumed to equal 10%) to determine the error in instantaneous PCB loads, according to the 'simple rule of product and ratios' from Kirchner (2001):

$$\frac{s_{\bar{z}}}{\bar{z}} = \sqrt{\left(\frac{s_{\bar{x}}}{\bar{x}}\right)^2 + \left(\frac{s_{\bar{y}}}{\bar{y}}\right)^2 + \left(\frac{s_{\bar{q}}}{\bar{q}}\right)^2 + \left(\frac{s_{\bar{w}}}{\bar{w}}\right)^2 + \cdots}$$

Next, the instantaneous load error estimates were summed for all estimates comprising the same storm to derive the error in event loads, following the 'simple rule for sums' from Kirchner (2001):

$$s_{\bar{z}} = \sqrt{(s_{\bar{x}})^2 + (s_{\bar{y}})^2 + (s_{\bar{q}})^2 + (s_{\bar{w}})^2 + \cdots}$$

In the following step, the PCB event loads resulting from the procedure above were used as the dependent variable in the statistical trend test. Both the event loads and corresponding standard errors were then employed in the power analysis procedure that followed.

Step 3

The third step in the modeling procedure involved testing support for a suite of regression models that best describe the relationship between PCB event loads over time (referred to as the trends model). Here, the goal of the model testing was to determine the significance of inter-annual trends in event loads after accounting for predictor variables associated with discharge and other sources of event variability (e.g., seasonality). Variables considered for the development of the trends model are listed in Table 5. Notably, turbidity and discharge were discounted from consideration in this step, due to concerns about their utility in models applied to future years, as they may be response variables directly impacted by management actions (e.g., reductions in sediment load or peak discharge through BMPs). Similar to Step 1, the evaluation of models followed a Maximum Likelihood Approach, with the AIC statistic, AICc weights, statistical significance, and minimum number of parameters forming the basis for model selection.

The general equation of the regression model was:

$$ln(PCB_{Loads}) = A_0 + A_1 ln(X_i) + A_2 T_i + e$$

Where, $\ln(PCB_{Loads})$ is PCB event load in natural log units, A_0 – A_2 are fitted coefficients of the model, Xi values are alternative raw or transformed predictor variables, Ti values are time variables representing month (categorical), season (categorical), and/or water year (continuous), and ε is the unexplained variation (residual error).

For the model development and testing, the mean estimate of PCB loads for each storm was the dependent variable used in the models. For selected models, statistical and graphical residual analysis and cross validation experiments (k-fold and LOOCV) to assess errors in the model was performed using the same general methods described for Step 1.

Composite Sampling Strategy

A key assumption of the modeling framework described above, is that a grab-based methodology will be used to characterize loads in the future, and high-resolution flow and turbidity data are available. To evaluate other sampling options, a parallel analysis was performed to simulate an alternative sampling strategy involving the collection of event composite samples. However, as composite samples have yet to be collected in the Guadalupe River, a simulation approach was employed to generate 'synthetic' composite loads for use in a trend model evaluation. The details of the analysis to generate synthetic loads are described in Appendix F in this report.

Development of a composite-based trend model took the same general form as the grab-based load regression model described previously in Step 3. Here, the dependent variable was loads estimated from synthetic composites by a linear interpolation of grab concentrations rather than the loads estimated by the regression model. Additionally, there are two major differences to the grab-based loads method: 1. Storms with a single collected grab were not considered in composite model calibration, and 2. Only loads for storms with more than one collected grab were estimated for development of the composite trend model (vs. the grab-based method, for which loads for ALL storms in sampled water years are extrapolated from the turbidity and climate factor relationships derived from the subset of sampled storms).

Power Analysis

Following selection of trend models for the grab- or composite-based methods, Monte Carlo simulations were developed to test the statistical power for determining declines in loads with the models, assuming varying levels of trend over 20 years. Here, 'Power' is defined as the probability of detecting a trend of a certain magnitude during a specified monitoring period (years), where a Type I error rate (α) is set at 5%[1]. Assuming this definition, a power analysis procedure was developed to test the likelihood of detecting significant inter-annual trends based on future scenarios of storm sampling. The goal of the analysis was to determine the statistical power associated with detecting significant inter-annual trends using the selected trend model based on several sampling designs and magnitudes of PCB change (Appendix B).

[1] A Type I error rate of 5% is conventional practice but is biased toward avoiding "false positives". Type I error rates of 10 - 20% might be more appropriate for environmental trend applications where the risk of false positives should be balanced with the risk of false negatives. For this study, we used the conventional 5% error rate as a starting point. Increasing the Type I error rate would have the effect of increasing the power to detect trends.

Sampling design simulations were developed by assuming that 1) future concentrations and loads will be sampled in the same manner as the current data, and 2) the distribution of storm events (number, timing. and magnitude) will be similar in the future. Subsequently, geometric trends of 10% to 90% decline over 20 years were imposed to the data by year. In the grab-based designs, PCB concentrations for the sequence of future years were simulated based off the original set of grab samples. For the composite-based sampling strategy, a similar approach was used except simulations were based on the calculated composite loads using the linear interpolation approach described in Appendix F.

To simulate grab/composite data for 20-years, concentrations/loads were repeated from the 12-year dataset, maintaining their original sequence including missing years of data. This methodology resulted in 56 storms over 20 years for grab-based simulations and 32 storms over 20 years for composite-based simulations. To simulate sampling at lower levels of effort to the previously sampled storms (e.g., 75% of 56 storms) a random selection of those storms were sub-sampled from the 20 year pool of simulated storms (250 trials per level of effort) and propagated through Steps 1 and 2 in the model approach to generate estimated event loads and standard errors. Subsequently, Monte Carlo simulations (400 runs) of the trends model (Step 3) were conducted using the estimated event loads and associated standard errors in loads. In the composite approach, since no model error in loads estimation was calculated, two standard error estimates were tested to bridge the expected range in composite load error; 5% and 30%. The estimated statistical power for each sampling design was calculated as the mean proportion of MC simulations (250 sample sets x 400 runs) where a significant water-year term was detected with the corresponding trend model.

Results and Discussion

Data Evaluation

Figure 1 summarizes the PCB concentrations, turbidity, and particle-ratio concentrations determined from grab storm samples collected at Guadalupe River used in statistical modeling. PCB concentrations and turbidity were most variable during the 'wet' water years (particularly, 2003 and 2005), as well as during 2012, and lower during the 'dry' water years, such as in 2006, 2010, and 2013. This is illustrated in the PCB data normalized to suspended sediment (i.e., particle ratio concentration), where mean concentrations appear relatively consistent over time, with few events in the time series having more variable particle ratio concentrations.

Table 1. Predictors Tested in PCB Concentration Models

Parameter	Transformation	Type	Guadalupe River
Turbidity	*Natural-Log*	*Continuous variable*	*Default predictor in all models*
Accumulated Rainfall Year-To-Date	*Natural-Log*	*Continuous variable*	Upper, middle, lower watershed USGS gauges

Rainfall Intensity	*Natural-Log*	*Continuous variable*	Upper, middle, lower watershed USGS gauges[1]
Season	N/A	Categorical variable	Early or Late[2]
Month	N/A	Categorical variable	October to March
Water Year	Raw	*Continuous* variable; mean centered	2003-2006, 2009-2010, 2012-2014

[1] 30 min-7 hr lagged rainfall with 30 min-2 hr antecedence for each gauge; [2] Early/Late Season = season-to-date rainfall less than/greater than 50% mean annual rainfall

Concentration Model

The first step in our model evaluation was to construct linear models to predict PCB concentrations as a function of turbidity, accumulated rainfall, rainfall intensity, seasonal, intra-annual, or inter-annual variability (Table 1). Several climatic variables related to discharge and precipitation indicated linear relationships (Figure 2). It should be acknowledged that a discharge parameter was discounted from consideration in the final tested models because it is a component of the loads calculations in Step 2. However, as precipitation and discharge are correlated, this influence on PCBs has to a certain extent been accounted for in the tested parameters. All models included turbidity as the primary predictor variable.

Forty-four models were examined for Guadalupe River. Summary statistics for the top five models in addition to the turbidity model are shown in Table 2. None of the top models differed greatly in the amount of explained variation (R^2 = 0.781 – 0.811). However they all represent a significant improvement in R^2 relative to the Default model that related concentration to turbidity alone (R^2 = 0.519). Two models were selected where all model terms were statistically significant; Model 42 and Model 21 (highlighted in gray on Table 2). These two models consisted of five (Model 21) and six (Model 42) statistically significant predictor variables each; turbidity, rainfall indicators for the upper and lower watershed, and a temporal term for season. The only difference was the inclusion of a polynomial term for one of the rainfall indicators in Model 42. These two models were further scrutinized in residuals analysis prior to final model selection.

Diagnostic plots of the residuals of Model 42 and 21 suggest that both models conform to assumptions of normality and homoskedacity (Appendix C). The normal Q-Q plots correspond to a relatively straight line (suggesting normality) and there did not appear to be major outliers based on the Cook's distance plot. The inferences were supported through statistical tests for normality (Shapiro-Wilk test) and homogeneity of variances (not presented), and plots of residuals vs. fitted values and fitted values vs. standardized residuals (i.e., residuals divided by its standard error). The graphical assessment indicated residuals that span the range and did not point to highly skewed values or non-random errors in the predicted values.

Figure 1. Guadalupe River PCB Concentrations, Turbidity, and Particle Ratio Concentrations by Water Year. Widths of boxes are proportional to the square-root of the number of observations in each group

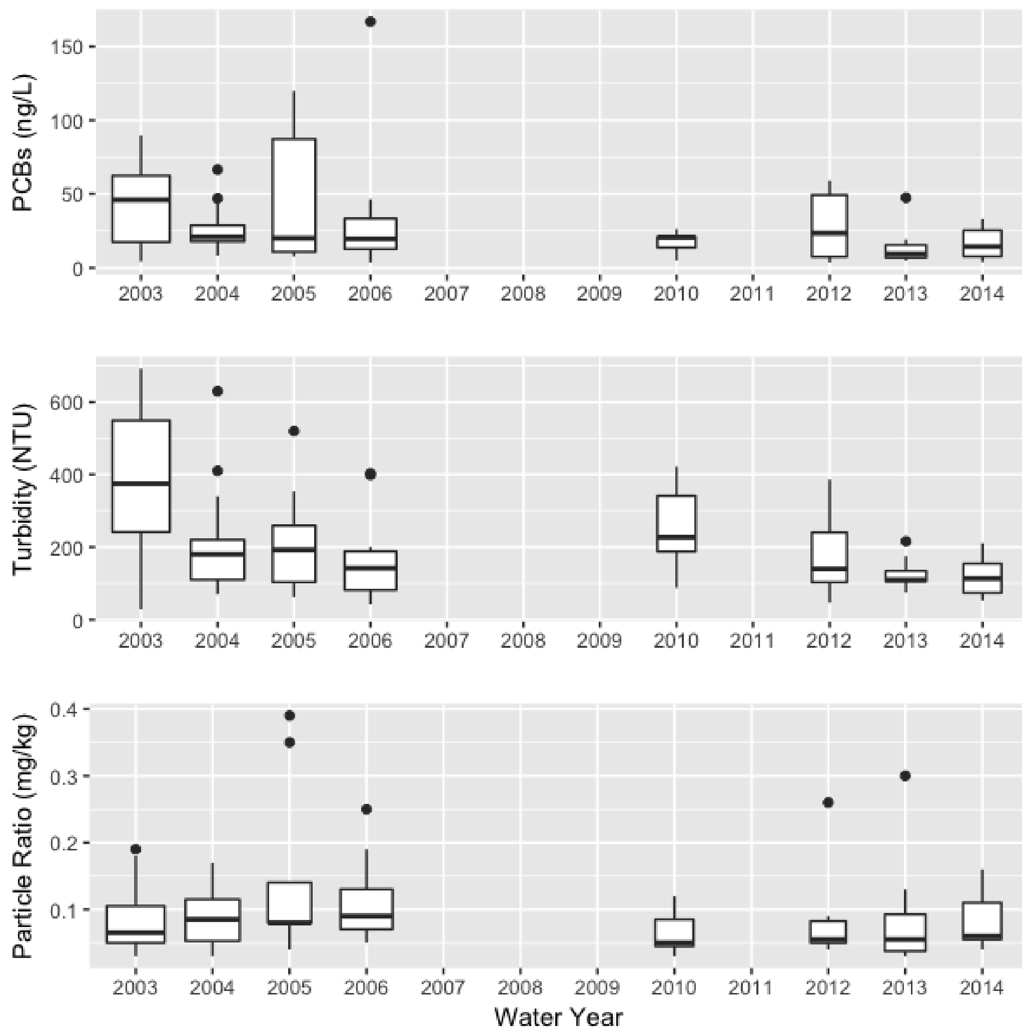


Table 2. Summary AIC Statistics for the highest ranked Guadalupe River PCB Models.
All terms were significant ($p < 0.05$) in each model except for those terms shown in *italics*

Model	Parameters	K	AICc	Delta AICc	AICc Wt	Cum .Wt	Adj. R^2
Model 42	Turbidity + 30minLag:2hrSum[CSJ] + 30minLag:2hrSum[CSJ]^2 + 7hrLag:7hrSum[LP] + Season[a]	7	88.12	0	0.96	0.96	0.811
Model 20	Turbidity + 30minLag:2hrSum[CSJ] + *4hrLag:4hrSum[ALM]* + 7hrLag:7hrSum LP] + Season[a]	7	96.09	7.97	0.02	0.98	0.791
Model 21	Turbidity + 30minLag:2hrSum[CSJ] + 7hrLag:7hrSum[LP] + Season[a]	6	97.72	9.60	0.01	0.99	0.783
Model 43	Turbidity + 30minLag:2hrSum[CSJ] + 7hrLag:7hrSum[LP] + *7hrLag:7hrSum[LP]^2* + Season[a]	7	98.16	10.04	0.01	0.99	0.786
Model 23	Turbidity + 30minLag: 2hrSum [CSJ] + 7hrLag: 7hrSum[LP] + *Season[1]+ Turbidity* Season[a]*	7	99.75	11.63	0	1	0.781
Model 1	Turbidity	3	157.85	69.72	0	1	0.519

[a] Jan/Feb vs. all other months (categorical; 2 levels)

Figure 2. Relationships Between Climatic Variables and PCB Concentrations at Guadalupe River

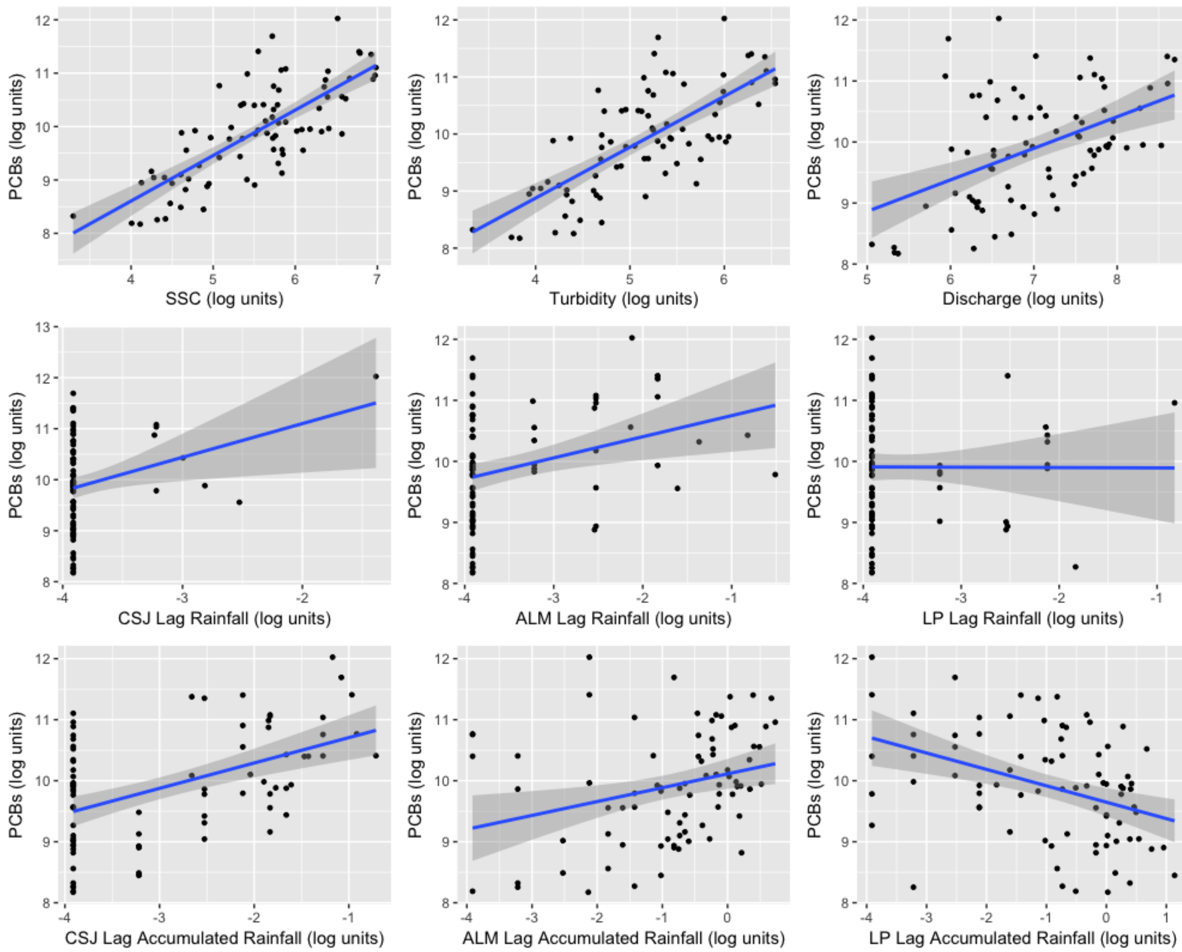


Table 3. Summary Statistics For Model 21 Relative to the Default Model

Model	Model Parameters	K	Adj. R^2	Pred. R^2	RMSE	PRESS
21	Turbidity + 30minLag:2hrSum[CSJ] + 7hrLag:7hrSum[LP] + Season	7	0.78	0.78	0.41	15.2
1	Turbidity	3	0.52	0.51	0.62	32.5

Table 4. Coefficients for Proposed Guadalupe River PCB Concentration Model (Model 21)

Model Parameters	Estimate	S.E.	t-value	p-value
Intercept	6.681	0.387	17.27	< 0.0001
Turbidity	0.802	0.066	12.23	< 0.0001
30minLag:2hrSum[CSJ]	0.311	0.049	6.32	< 0.0001
7hrLag:7hrSum[LP]	-0.143	0.040	-3.59	0.0006
Season	-0.422	0.103	-4.09	0.0001

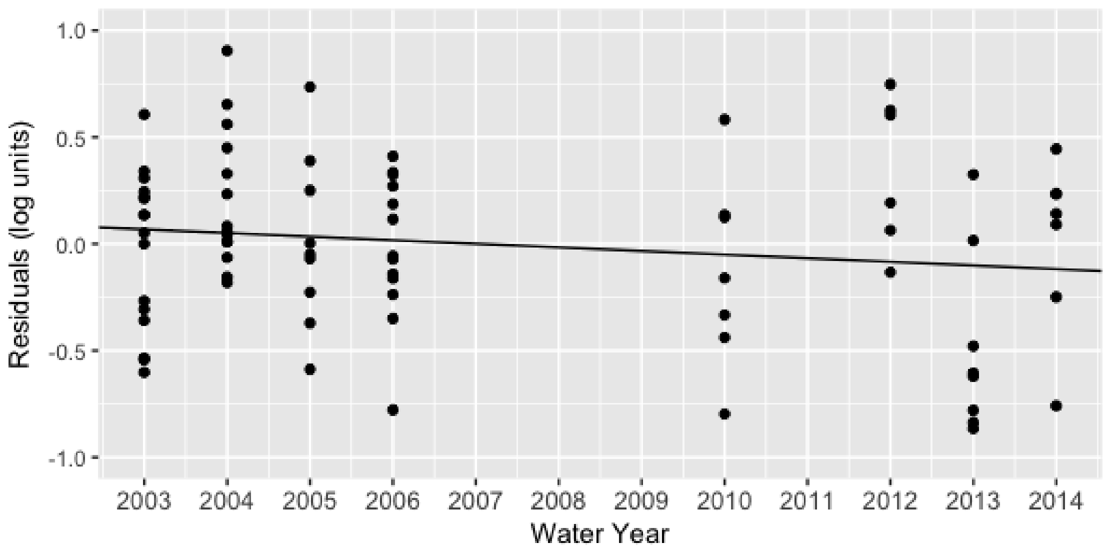
Given the similarity in the results of Models 21 and 42, the simpler model formulation of Model 21 (without polynomial terms) was chosen to proceed with. Summary statistics of the selected model are shown in Table 3. Clearly, Model 21 exhibits relatively higher model R^2 and predictive R^2, and lower RMSE and PRESS statistics relative to Model 1 (the Default model).

Residuals were plotted against water years and months to examine temporal structure in the unexplained variance. If any systematic temporal trend in the residuals vs. time were observed, it would invalidate the subsequent test using the trend model. The residuals against water year were relatively well distributed around zero during the earlier period of sampling and 2014, but deviated in 2012 and 2013 (Figure 3). Although the residuals for samples collected during these years appeared higher and lower, respectively, relative to other years, this apparent difference was not statistically significant when tested by regression (not presented). Similarly, November events tended to have lower residuals than other months, but this difference was not statistically different.

To further explore whether model error patterns can be explained by event-scale variation not already captured in the selected predictors, the residuals were plotted against two intermediary climatic response variables (specifically, observed discharge and suspended sediment concentration). These graphical analyses indicate the model residuals are randomly distributed with respect to these response variables (Figure 4). None of the linear regressions of residuals against flow or SSC were significant (not presented).

Figure 3. Guadalupe River PCB Concentration Model 21 Residuals Plotted Against Water Year (A) and Month (B). Points reflect residual values, and the regression line resulting from a linear model of residuals against water year or month. Both regression slopes were not significantly different from zero.

A)



B)

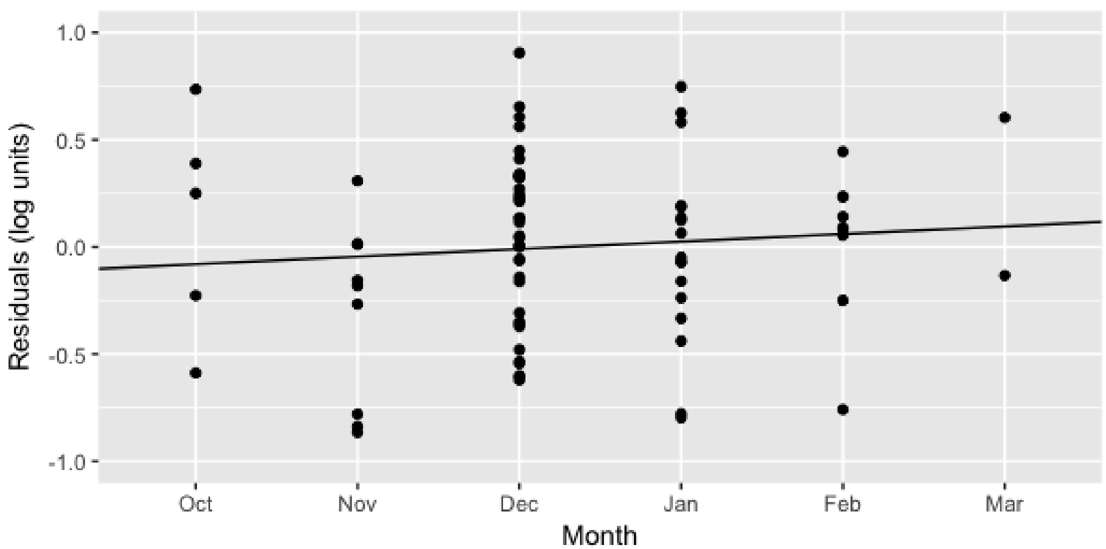
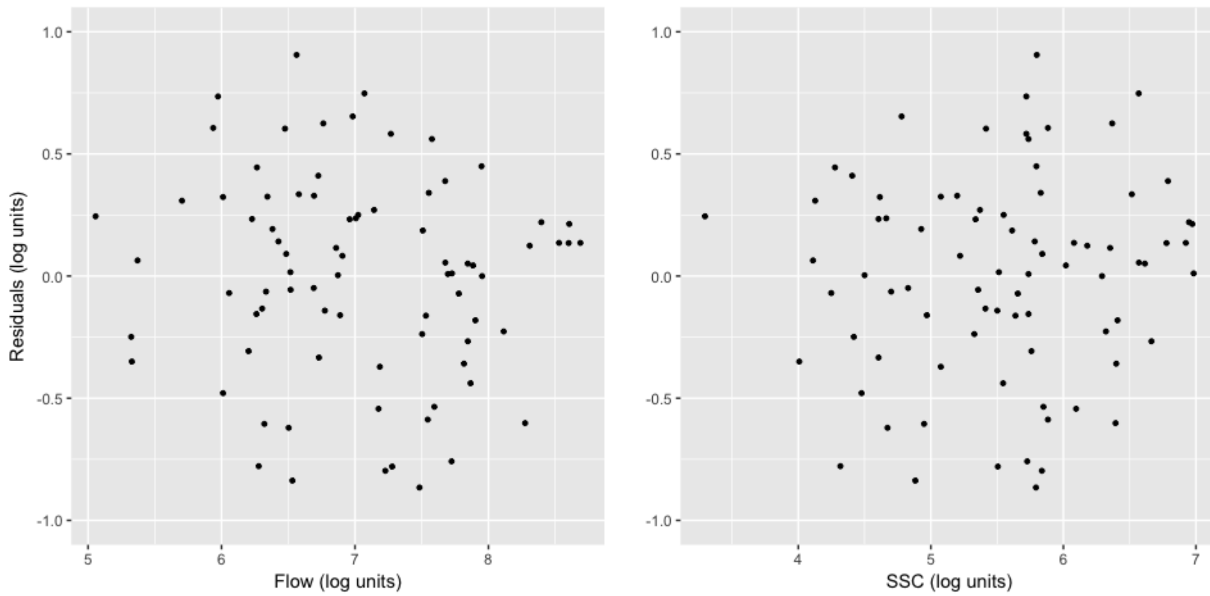


Figure 4. Guadalupe River PCB Concentration Model 43 Residuals Plotted Against Flow and Suspended Sediment Concentrations.



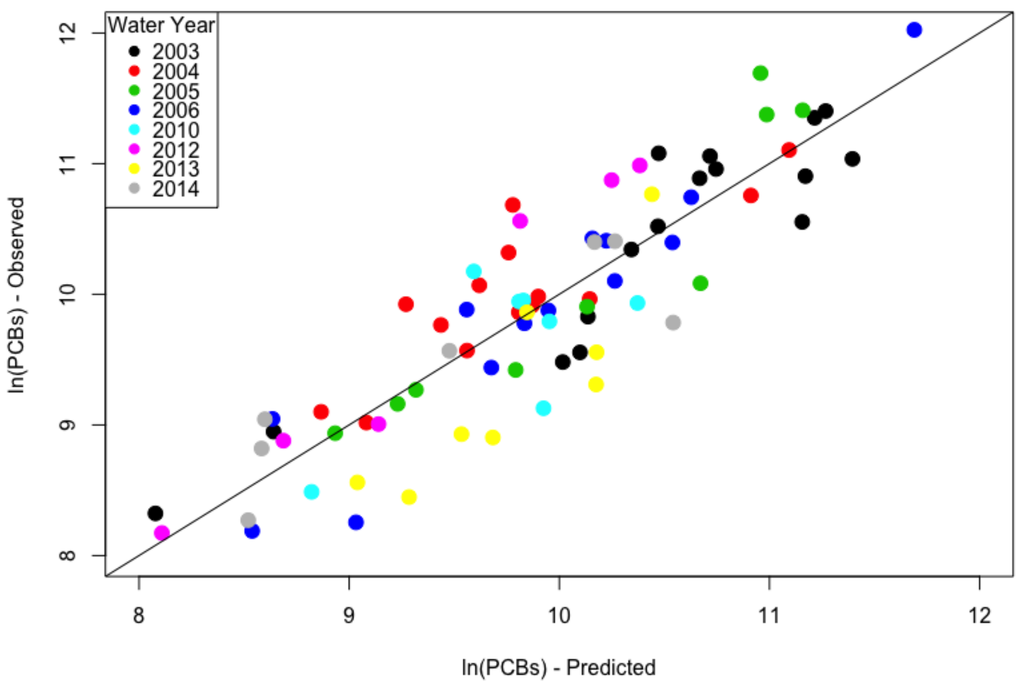
Next, cross-validation methods were conducted to examine estimates of model error and the robustness for making predictions with an independent dataset. Two common methods were applied; Leave-One-Out (LOOCV) and k-fold cross validation (with k = 5). The cross validation results of both methods on the two selected models were similar (Appendix D). Had there been a large difference in RMSE or R^2 it may have been concluded that the models were overfitting the data, which could lead to more spurious predictions.

Finally, the predictions from Model 21 were plotted against the observed PCB concentrations. Figure 5 shows the majority of Guadalupe River data are predicted reasonably well by the model, with few samples showing large deviations from observed values.

Storm Selection

The next step in the analysis was to use the selected concentration model to predict PCB concentrations from wet season turbidity and rainfall data across the entire period of sampling in the watershed (2003-2014). Note that since the grab sample dataset that formed the basis of the model only contained samples from October to March to generate the seasonal variable in the selected model, only concentrations for those months in the dataset could be predicted. Therefore, predictions for data records in April and May of each water year was not performed due to concerns over extrapolation beyond the range in the data (un-sampled months or years).

Fig 5. Guadalupe River PCB Concentration Model 21. Relationship between predicted and observed grab samples. Predicted concentrations were back-transformed to ng/L



Subsequently, storm selection criteria based on peak flow, minimum flow, and precipitation were applied to the continuous record to determine 'storm events' to use in the development of the trends model. Figure 6 summarizes the peak flow for the selected storms (n = 75). Generally, the discharge during events in any single year had a median around the selected minimum peak flow of 500 cfs. However, 2003 was associated with much more variable and higher storm flows, while the more recent years were drier, with many storms exhibiting flows that were generally less than 500 cfs, (except at peak flow). These features of the dataset further support the need for the trend model to explain climatic response patterns in PCB loads.

Loads Calculation

The 15-minute predicted concentrations at Guadalupe River associated with these 75 storms were converted into loads by combining the modeled concentrations with measured flow obtained from the USGS gauge (11169025) on the lower Guadalupe River. Figure 7 shows the subsequent propagation of the prediction errors from predicted (15-minute interval) PCB concentrations to event loads from this step in the analysis. The storms that exhibited less precision were those with predicted event loads greater than 100 g per event. For example, the highest PCB load was estimated for a storm on 12/19/03 that exhibited an event load of 461.3 ± 30.4 g (Storm # 3). Generally, individual storms with event loads less than 50 g exhibited errors of approximately 5% of the event load estimate.

Trends Model

The final step in the statistical analysis procedure was to evaluate models that test for significant temporal trend in event loads, while accounting for the relative influence of climatic parameters (e.g., related to antecedence, accumulated rainfall and individual storm lagged rainfall indicators, Table 5). Twenty-five predictor variables were selected for testing, which yielded 44 regression models that were examined for AIC statistics and statistical significance, in the same manner as in Step 1. Additional complexities to the model structure were also considered such as using polynomial terms and interactions.

Table 5. PCB Trends Model Parameters Explored

Category	Transformation	Type	Parameter
Accumulated Rainfall	Natural-log	Continuous variable	YTD at City San Jose [CSJ] or Loma Prieta [LP] rain gauges
Lag Rainfall	Square-root	Continuous variable	CSJ 30min Lag or LP 7hr Lag
Max Rainfall	Raw	Continuous variable	CSJ or LP 1 to 6 hr max
Antecedence	Square-root	Continuous variable	CSJ or LP 1-day, 7-day, 14-day
Storm Flow Duration	Natural-log	Continuous variable	Length of storm (days)
Inter-annual	Raw	Continuous variable; mean centered	Water Year (2003-2014)
Season	N/A	Categorical variable; Jan/Feb vs. other months	Month (Oct-Mar)

Figure 6. Peak discharge (natural log of peak in cfs units) for storms selected for the trends model.

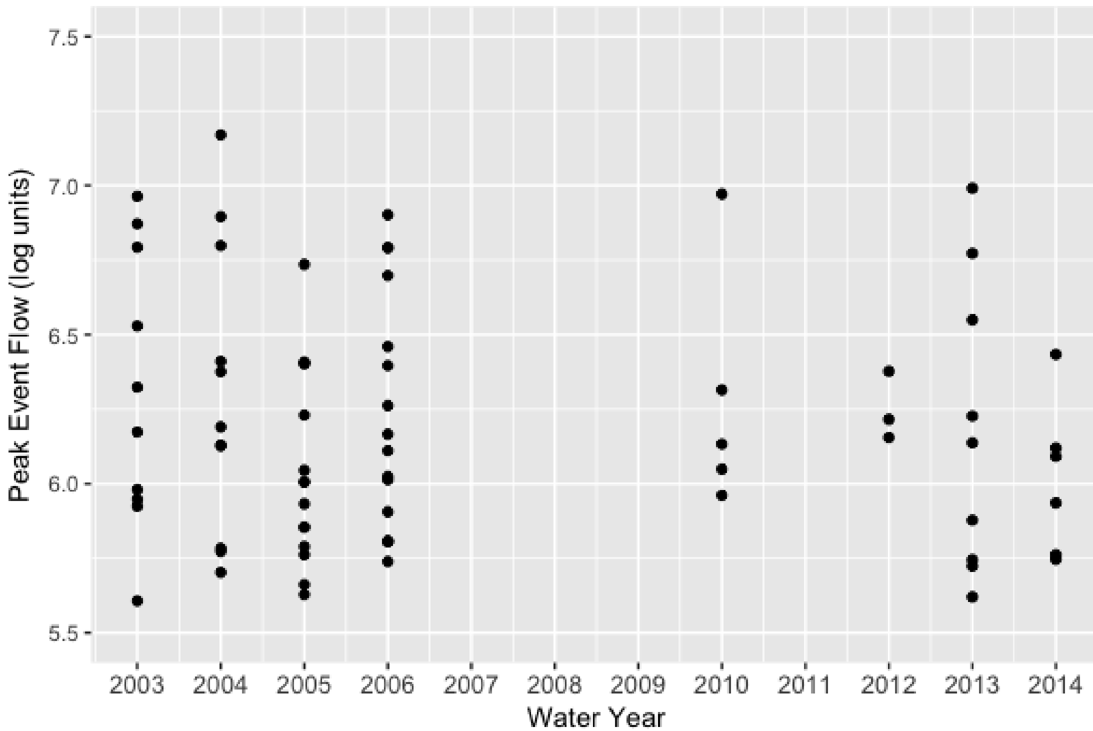


Figure 7. PCB Loads (+/- standard error) for Storm Events in Guadalupe River (N = 75; 2003-2014). Estimated loads and standard errors were determined from predictions and errors of the concentration model, and assume a 10% standard error in discharge.

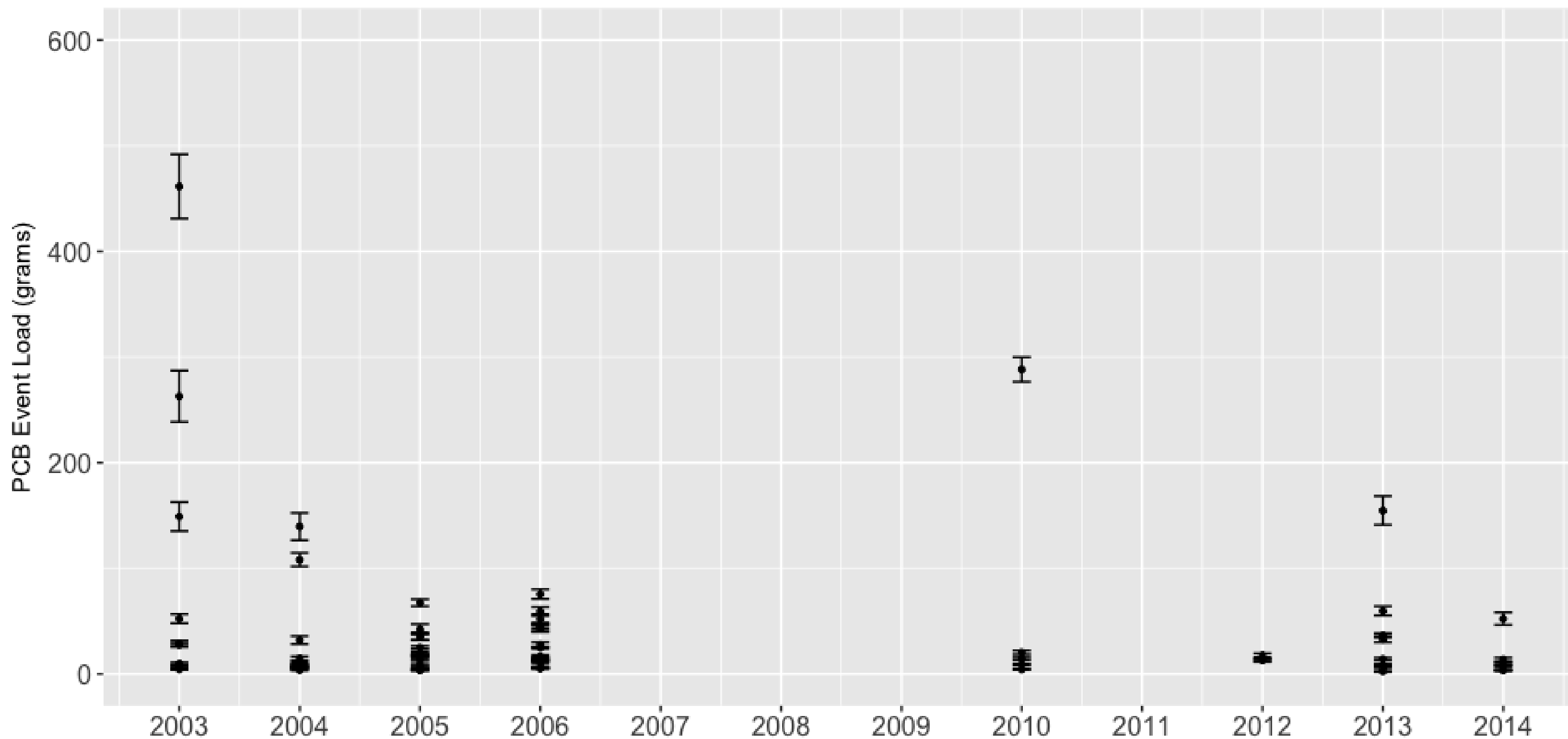
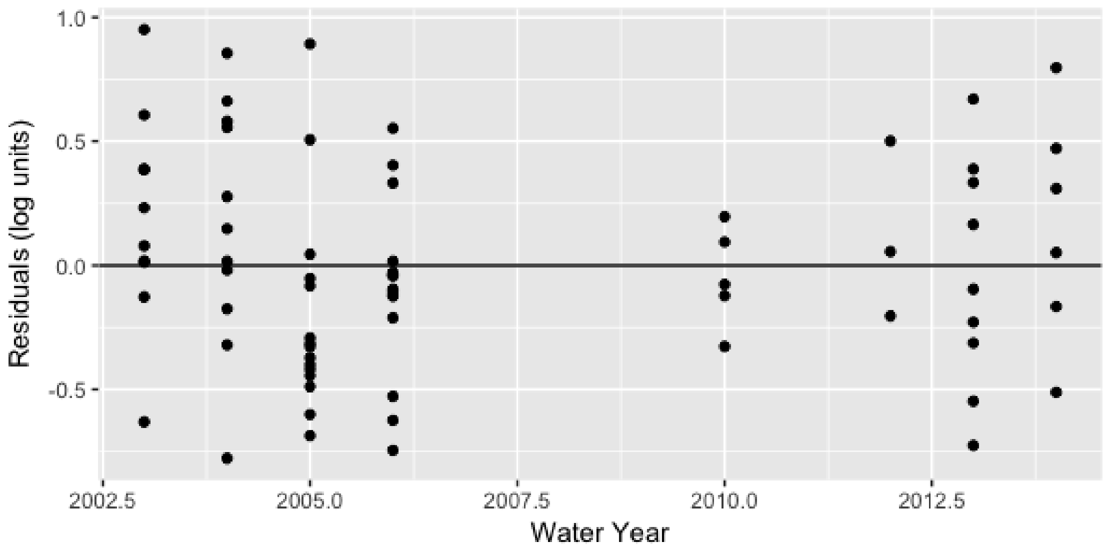


Table 7. Guadalupe PCB Trend Modeling AIC Statistics. Only for Model 30 were all terms significant ($p < 0.05$). Not significant model terms are shown in *italics*

Model	Parameters	K	AICc	Delta AICc	AICc Wt	Cum. Wt	Adj. R^2	Pred. R^2	RMSE	PRESS
Model 34	30minLag:2hrSum[CSJ] + 7hrLag:7hrSum[LP] + 3hrMax Rainfall[LP] + StormFlowDuration[CSJ] + Season + *WaterYear*	8	100.4	0.00	0.36	0.36	0.86	0.84	0.42	16.1
Model 35	30minLag:2hrSum[CSJ] + 7hrLag:7hrSum[LP] + 3hrMax Rainfall[LP] + StormFlowDuration[CSJ] + *StormFlowDuration[CSJ]^2* + Season + *WaterYear*	9	100.6	0.26	0.32	0.68	0.86	0.84	0.41	16.0
Model 36	30minLag:2hrSum[CSJ] + *30minLag:2hrSum[CSJ]^2* + 7hrLag:7hrSum[LP] + 3hrMax Rainfall[LP] + StormFlowDuration[CSJ] + Season + *WaterYear*	9	100.7	0.29	0.32	1.00	0.86	0.84	0.41	16.2
Model 32	30minLag:2hrSum[CSJ] + *30minLag:2hrSum[CSJ]^2* + 7hrLag:7hrSum[LP] + 3hrMax Rainfall[LP] + Season	7	124.0	23.6	0.00	1.00	0.80	0.77	0.50	23.0
Model 30	30minLag: 2hrSum [CSJ] + 7hrLag: 7hrSum [LP] + 3hr Max Rainfall [LP] + Season	6	125.3	25.0	0.00	1.00	0.77	0.80	0.50	23.0

Figure 8. Guadalupe River Trends Model 34 Residuals Plotted Against A) Water Year and B) Month. Points are residuals, and the regression line resulting from a linear model of residuals against water year or month. Both regression slopes were not significantly different from zero.

A)



B)

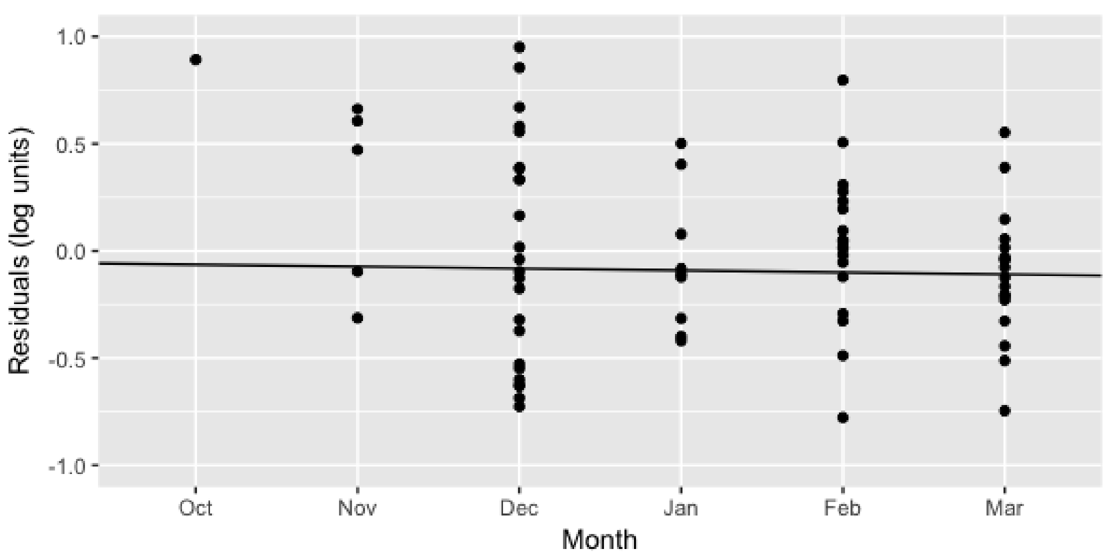
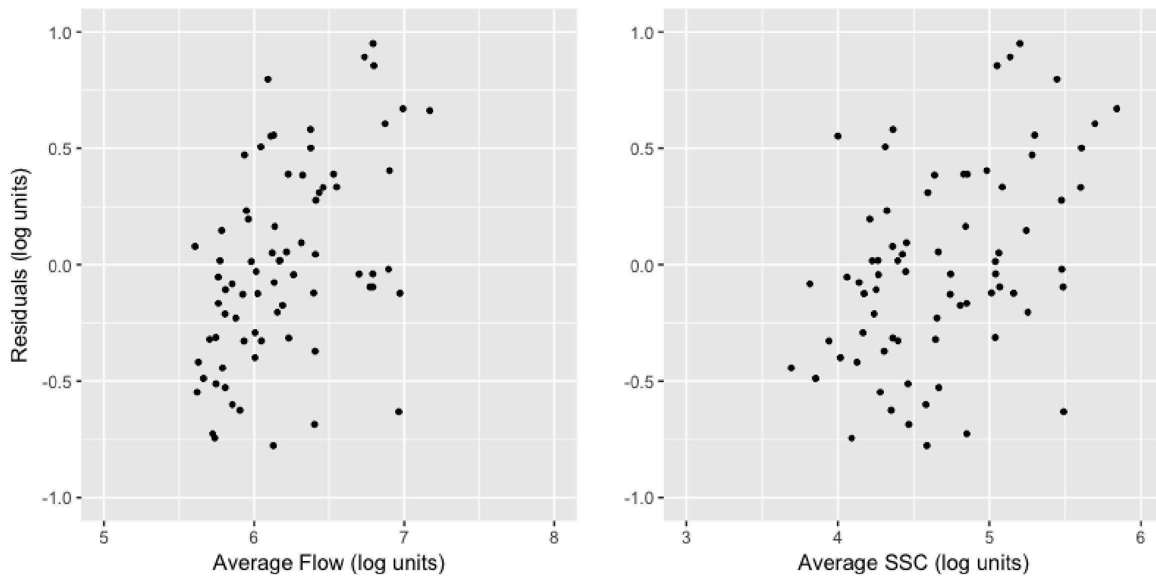


Figure 9. Guadalupe River Trends Model Residuals Plotted Against Flow and SSC for each corresponding storm employed in the trend model



Statistical testing of 44 generalized linear models indicated three models with the highest likelihood of 'best' describing PCB event loads; Model 34, 35, and 36 (Table 7). They all consist of at least two continuous climatic factors from the lower and upper watershed gauges (San Jose and Loma Prieta, respectively), a categorical variable for season, and continuous inter-annual trend. Models 35 and 36 differed from Model 34 by the addition of polynomial variables (not significant in either case). In all cases the water year term was also not statistically significant (Table 9). Following the same rationale of the concentration model, the simplest model with the highest number of statistically significant model terms was selected. In this case, Model 34 fit these criteria and was also the highest ranked based on AIC statistics, and thus was selected as the proposed trends model for Guadalupe River. In Figure 8, it is evident there is no remaining seasonal or interannual trend to the residuals over time. Model residuals also passed formal testing for normality and homoscedasticity (Appendix E).

Table 8. Coefficients of Proposed Trend Model for Guadalupe River (Model 34)

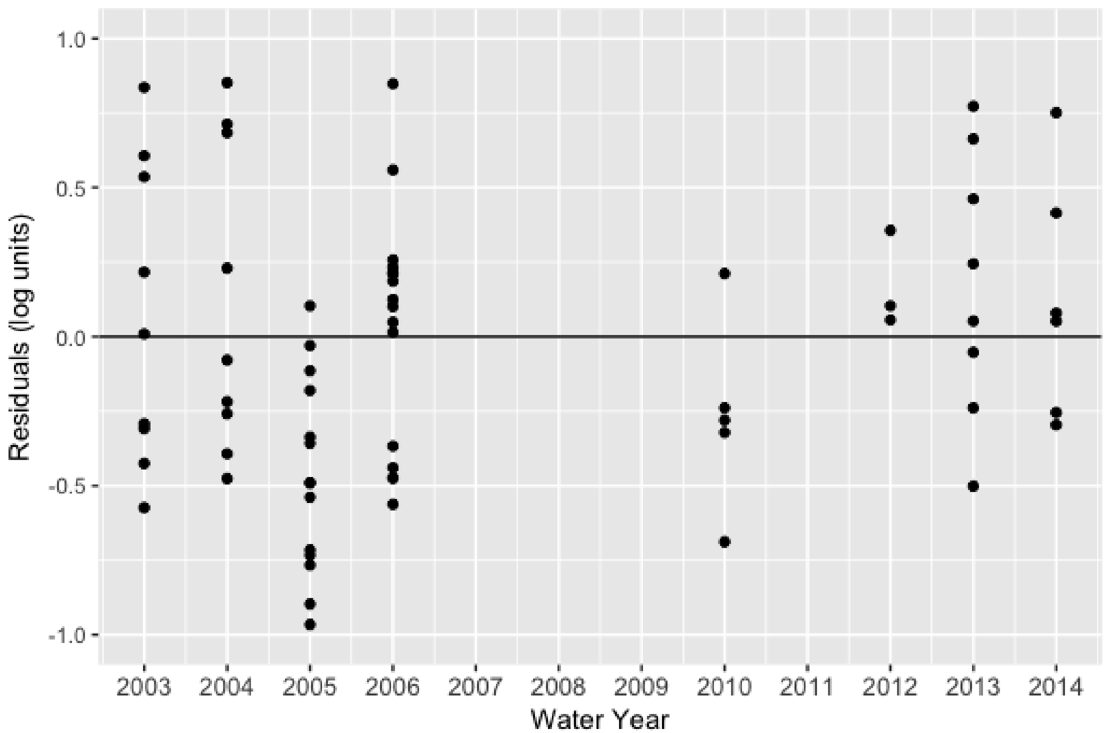
Model Parameters	Estimate	S.E.	t-value	p-value
Intercept	2.11	0.18	11.9	< 0.0001
30minLag:2hrSum[CSJ]	1.23	0.14	8.50	< 0.0001
StormFlowDuration[CSJ]	0.69	0.12	5.76	<0.0001
7hrLag:7hrSum[LP]	-0.79	0.24	-3.33	0.001
3hrMaxRainfall[LP]	1.63	0.25	6.48	<0.0001
Season	-0.67	0.11	-6.24	< 0.0001
Water Year	0.02	0.01	1.13	0.260

Regression of the model residuals to discharge and SSC (Figure 9) suggest that Model 34 has largely captured the variation at event-scales, as both the positive and negative residuals span the range in flow and SSC conditions. At the highest flow and SSC events the residuals indicate some heterogeneity, exhibited by more positive residuals. This would reflect an under-prediction of event loads under those conditions. Linear regressions of peak SSC (slope = 0.26, R2 = 0.17, p = 0.0002) and peak flow (slope = 0.30, R2 = 0.17, p = 0.0003) against the residuals were both significant. These apparent differences are not manifested directly in the predicted PCB loads, however, as there was no statistical relationship (slope = 0.002, R2 < 0.01, p = 0.96) between predicted loads and the residual error from the trend model.

Two cross-validation experiments were conducted to examine estimates of model error. As in Step 1 of the analyses, two common methods of CV were applied; Leave-One-Out (LOOCV) and k-fold cross validation (with k = 5). As shown in Table 7, the RMSE estimated during the development of the model was 0.43. Based on both the LOOCV method and k-fold CV, the RMSE was estimated at 0.47. There were also only minor differences in R^2, with LOOCV indicating R^2 = 0.814 and k-fold CV of 0.859. These estimates were very similar to the R^2 and adjusted R^2 for the model, 0.850 and 0.835, respectively. The lack of large differences in RMSE and R^2 in cross validation relative to the model, suggests that the model is not over-fitting the data.

To finally demonstrate the current slope of trend in PCB loads, the predicted event loads were summed to determine the annual storm loads and associated standard errors for yearly estimates. Figure 10 illustrates that despite relatively high PCB loads during the earlier years of monitoring (Figure 7), there is a lack of trend in the residuals of the trend model. This means that that any apparent trend in the loads data over time can be accounted for by climatic factors.

Figure 10. Regression of water year vs. residuals of proposed Guadalupe River trend model (all variables except water year included). Slope of 0.00 confirms the lack of linear trend in loads over time.



Power Analysis

Simulations were conducted to test the power for detecting trends in PCB loads assuming either grab-based or composite-based sampling designs. Grab simulations were conducted based on sampling between 18 (33% of past effort) and 56 storms (100% of past effort) over 20 years, with an average of 3 grabs collected on each event. In comparison, the composite simulations tested sampling a single composite per storm between 16 to 32 storms over 20 years, or 50% to 100% of past sampling effort. For each of the scenarios, a given annual rate (geometric) change of interest (ranging from 25% to 90% net decline in 20 years) was imposed on the past sequence of empirical data, and the grab or composite load calculation methodology re-applied to the hypothetical empirical data, subsampled at different levels of effort.

For the grab sampling designs, power curves indicate that most trends of interest would exhibit high statistical power. Inter-annual trends of 25% or more would be detectable with > 80% power when sampling 28 storms (50% effort) over 20 years (Figure 11). This design would equate to sampling at least 2 storms in 13 out of 20 years. During those 13 years, 4-6 grab samples are assumed to be collected during each storm as performed in the current sampling methods. In all grab-based sampling designs (i.e., 33-100% effort) the power to detect imposed trends of 90% decline over 20 years (the TMDL target) would be 100%.

For the composite scenarios, power curves indicate that only larger trends of interest would be detectable with high power. Sampling of 24 storms (75% effort) in 20 years would be sufficient to detect trends of 75% or more (Figure 12). Similar to the grab design, this suggests approximately 2 storms per monitoring year, assuming 13 of 20 years are sampled. In the composite-based sampling designs, the power to detect trends of 90% over 20 years (the TMDL target) would be 100% if at least 16 storms are sampled and the assumed error in loads is 5%. Under the higher variance scenario (30%), only the design for 100% effort (32 events in 20 years) would attain 80% power.

Finally, to test the power of the grab-based methods under more variable conditions (e.g., with concentrations less reliably predicted by turbidity and climatic factors), the sampling designs were re-tested assuming an inflated variance in the sample data. Assuming random error was increased by +/- 20%, the size of trend that could be detected with > 80% power at either 50% or 100% effort was approximately the same as with the original analysis (Figure 13). However, if random variability increased to +/- 50%, the size of detectable trend with > 80% power and 50% sample effort would increase to 50% or greater declines.

Figure 11. Power Curves for Grab Sampling Design Scenarios

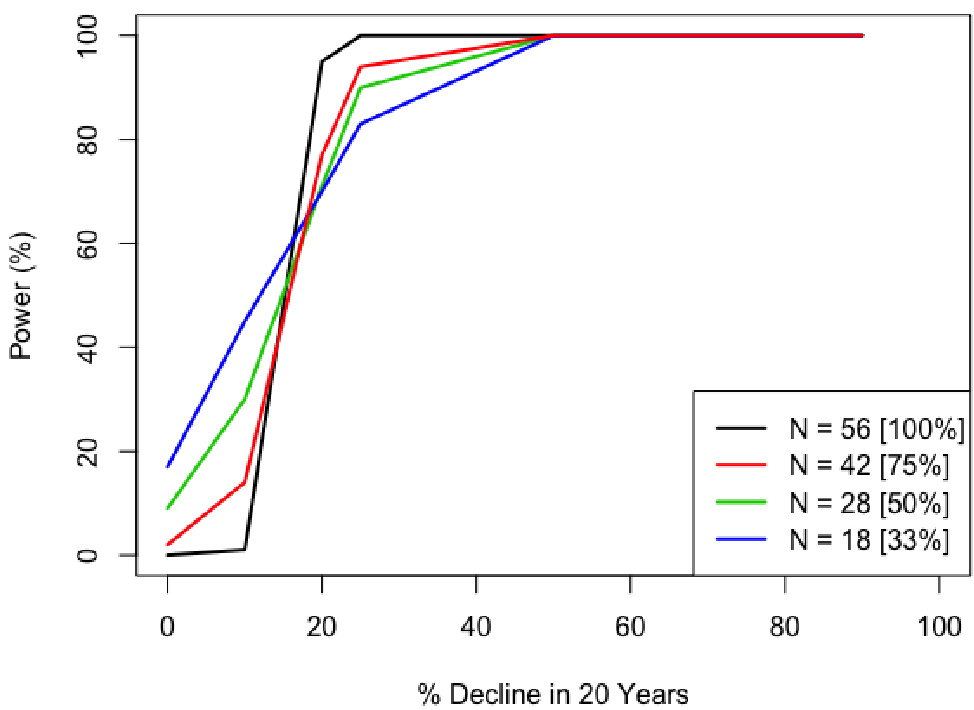


Figure 12. Power Curves for Composite Sampling Design Scenarios. Scenarios with solid lines include a 5% standard error in loads; dotted lines include a 30% standard error in loads.

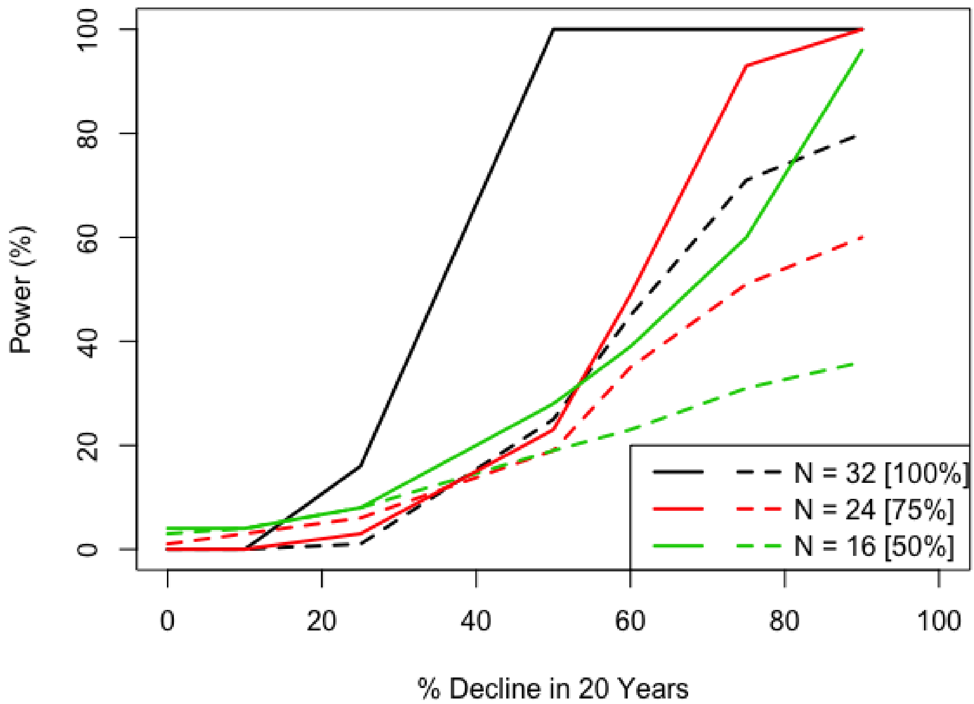
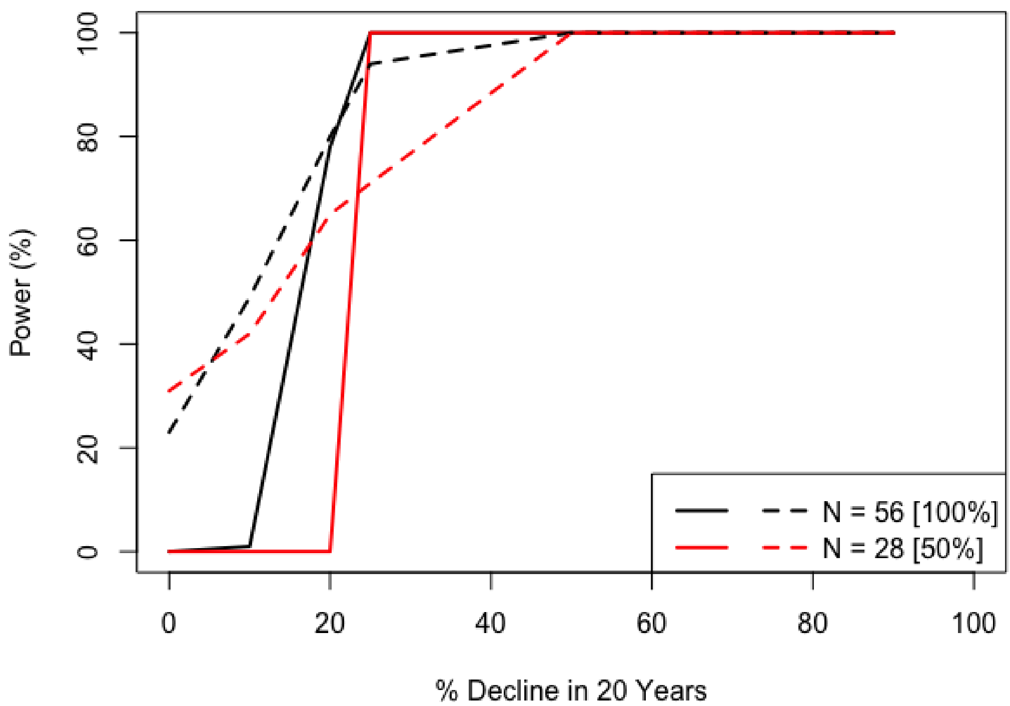


Figure 13. Power Curves for Grab Sampling Design Scenarios Including an Inflated Random Error. Scenarios with solid lines include +/-20% random error, while dotted lines include +/-50% random error.



Conclusions

The statistical analysis approach developed in this project has shown the potential for minimizing error associated with estimates of PCB loads by accounting for major sources of variability in climatic response. Applying the statistical model to the Guadalupe River data supports earlier reports that document a lack of significant inter-annual trend in PCB loads (McKee et al. 2017). The power analysis results indicated that there is a relatively low chance of concluding that there is a significant trend with the modeling approach under the null hypothesis (lack of a trend).

The key feature of the trend model procedure is that it tests for a change in loads after accounting for climatic variability. An apparent decrease in loads in the Guadalupe River from the early, wet years of the dataset to the more recent, drier years was found to not constitute a significant trend after explicitly accounting for rainfall amount and timing in the watershed. This feature of the model is critical to the conclusion of whether a statistical trend may exist in the future. For example, if there were a trend in PCB loads in the future caused solely by changes in climate, the trend model would continue to support the result of 'no trend'. However, if climatic factors followed a new pattern of more frequent rainfall, which cannot be accounted for solely using the existing trend model factors, the conclusion from the trend test may be different. As a result, future scenarios may arise where PCB loads decline due to management changes, but overall, loads are actually increasing due to increased rainfall patterns. Thus, the influence of shifts in climatic patterns on the trend model will warrant consideration in future analyses.

Another important aspect of the statistical approach was the use of storm selection criteria to select storms for testing with the trend model. This approach increased the sample size for modeling by encompassing all storms. However, by definition the method assumes that the conditions of all selected storms in the dataset behave in the same manner as the sampled storms. This assumption is likely an oversimplification of the behavior of loads in the watershed.

The power analysis work further showed that the grab-based model approach exhibited higher power for detection of trend over composite methods. The power simulations showed that a 20-year monitoring program where at least two storms are sampled in 65% (13 of 20 years) of the time series would have high statistical power to detect a 25% or greater declining trend in PCB loads. The simulations indicate that the higher sample size and relatively low standard error of load estimates of the grab-based methodology holds promise for detecting trends of management interest in the future. Such an approach could be useful to verify relatively small future changes at Guadalupe River over time, or at a select number of other watersheds where such methods can be applied.

In the event that high frequency turbidity data are unavailable in the future, and load estimates are based on event-composite samples, it is likely that a 20-year program will lack the statistical power to detect smaller changes in loads over time. Using the composite-based trend model presented in this study, it appears that declines of 75% or more over 20 years would be detectable, unless a greater frequency of storms are sampled during 20 years than were assessed here (i.e., focusing on capturing more storms with fewer samples or only a single composite per storm).

Finally, the current modeling framework is based on the empirical response of the watershed to past climate drivers. Since the distribution patterns of climate drivers, and watershed characteristics (current land use distribution, imperviousness, soil and stream bank erodibility, among other factors) may change over time, it is important that future efforts to employ the model framework consider the potential for future changes due to uncontrollable driving factors, as opposed to response factors that may be directly or indirectly managed or mitigated (e.g., watershed efforts to reduce or increase suspended sediment supply from various areas or sources). Additionally, the statistical approaches presented have the potential to optimize future parameter estimates for improved annual loads estimation, and detection of future trends in loads. Yet, given the scope of the study, it would be premature to generalize the trends model for Guadalupe River to other Bay Area watersheds. Therefore, future efforts will also need to examine the utility of designing similar trend models and supporting power analysis scenarios for other individual watersheds. The case study for Guadalupe River has shown that by employing high frequency turbidity and concentration data, there is the potential to separate climatic responses from inter-annual trends.

The final step towards selecting a preliminary monitoring design for the Guadalupe River (or other target STLS watersheds) was to evaluate the cost-effectiveness of the various sampling approaches at levels of sampling intensity indicated by the power analysis. Appendix G presents cost scenarios for both grab-based and composite-based methods; assuming grab sampling in 10 years over a 20 year period with 4 storm events/year and 4 grab samples/storm; or composite sampling in 12 years over a 20 year period with 2 storm events/year.

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, 150.

David, N., L. J. McKee, F. J. Black, A. R. Flegal, C. H. Conaway, D. H. Schoellhamer & 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(10):2091-2100.

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

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.

Kirchner, J. W., 2001. Chapter 5: Uncertainty Analysis and Error Propagation Data Analysis Toolkit. University of California at Berkeley.

McKee, L., J. Leatherbarrow & R. Eads, 2004. Concentrations and loads of mercury, PCBs, and OC pesticides associated with suspended sediments in the Lower Guadalupe River, San Jose, California SFEI Contribution. San Francisco Estuary Institute, Oakland, CA, 1-83.

McKee, L., J. Leatherbarrow & 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 A Technical Report of the Regional Watershed Program: SFEI Contribution 409. San Francisco Estuary Institute, Oakland, CA, 72 pp.

McKee, L. J., A. Bonnema, N. David, J. A. Davis, A. Franz, R. Grace, B. K. Greenfield, A. N. Gilbreath, C. Grosso, W. A. Heim, J. A. Hunt, J. E. Leatherbarrow, S. Lowe, S. A. Pearce, J. R. M. Ross & D. Yee, 2017. Long-term variation in concentrations and mass loads in a semi-arid watershed influenced by historic mercury mining and urban pollutant sources. Science of The Total Environment 605-606(Supplement C):482-497 doi:https://doi.org/10.1016/j.scitotenv.2017.04.203.

McKee, L. J. & A. Gilbreath, 2009. Concentrations and loads of trace contaminants in the Zone 4 Line A small tributary, Hayward, California: Water Year 2007. San Francisco Estuary Institute, Oakland, CA, 57 pp.

McKee, L. J., J. Oram, J. Leatherbarrow, A. Bonnema, W. Heim & M. Stephenson, 2006. Concentrations and loads of mercury, PCBs, and PBDEs in the lower Guadalupe River, San Jose, California: Water Years 2003, 2004, and 2005. San Francisco Estuary Institute, Oakland, CA, 47pp + Appendix A and B.

Melwani, A. R., Yee, D., Gilbreath, A., McKee, L.M., 2016. Technical Appendix to the Small Tributaries Trend Design. San Francisco Estuary Institute.

Newman, M. C., 1993. Regression analysis of log-transformed data: Statistical bias and its correction. doi:10.1002/etc.5620120618.

Oram, J. J., L. J. McKee, C. E. Werme, M. S. Connor & D. R. Oros, 2008. A mass budget of PBDEs in San Francisco Bay, California, USA. Environment International 34:1137-47.

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, 116.

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.

SFRWQCB, 2015. San Francisco Regional Water Quality Control Board Municipal Regional Stormwater NPDES Permit. Order R2-2015-0049, NPDES Permit No. CAS612008. 152 pp plus appendices.

Wall, G. R., H. H. Ingleston & 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.

Appendices

Appendix A. Steps in developing predictors for modeling PCB concentrations from limited storm water sampling

1) Literature review – turbidity has been used as a surrogate for:
   * Mercury (Quémerais et al. 1999, Wall et al. 2005, David et al. 2009, Ruzycki et al. 2011, Riscassi and Scanlon 2013, David, et al. 2015, McKee and Gilbreath 2015)
   * PCBs (David et al. 2015, Gilbreath and McKee 2015)

2) Parameter development and selection - Stratified the data into subsets:
   * Lag times,
   * Rainfall intensity
   * Timing (early/later, rising/falling)

3) Also considered conceptual models in relation to:
   * Source homogeneity
   * Wetting and drying

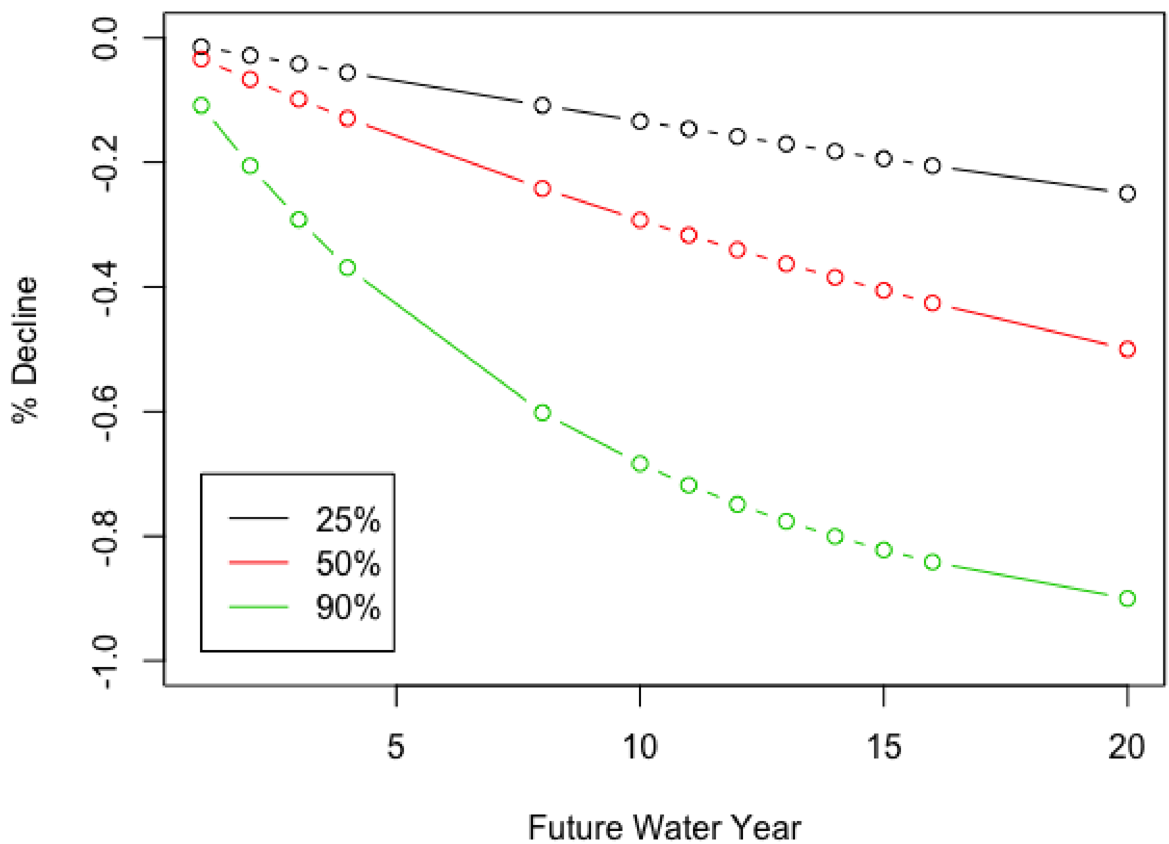
4) Sought terms that would impact processes
   * Washout/exhaustion effect on source
   * Short term- rising falling stage
   * Early or late wet season (antecedent rain)

5) Types of factors
   * Site Observations
     i) Define hydrograph peak to categorize rise/fall
   * Calculated values
     i) Cumulative precipitation threshold for early vs. late season
     ii) Event cumulative precipitation (event antecedent rain, e.g., hours to days)

6) For Guadalupe River, also attempted to convert rainfall into a % urban source surrogate
   * #1 = concurrent rainfall at 3 gauges, ratio urban/total
   * #2 = 0.5hr lag of 2hr sum urban, 4 & 4 mid, 7 & 7 upper watershed
   * #3 = #2 but average per ¼ hour (/8, /16, /28 of #2 components)
   * #3.5 = #3 with area and runoff coeff weights for low/mid/up
   * #3.6 = #3.5 but 1, 2, 4 hr avgs (same lags)
   * #3.7 = like 3.6, but 2, 3, 4 hr avgs, 0.5, 3, 5 hr lags for low/mid/up
   * Denominators of #3 to #3.7 are time averaged rainfalls with lags

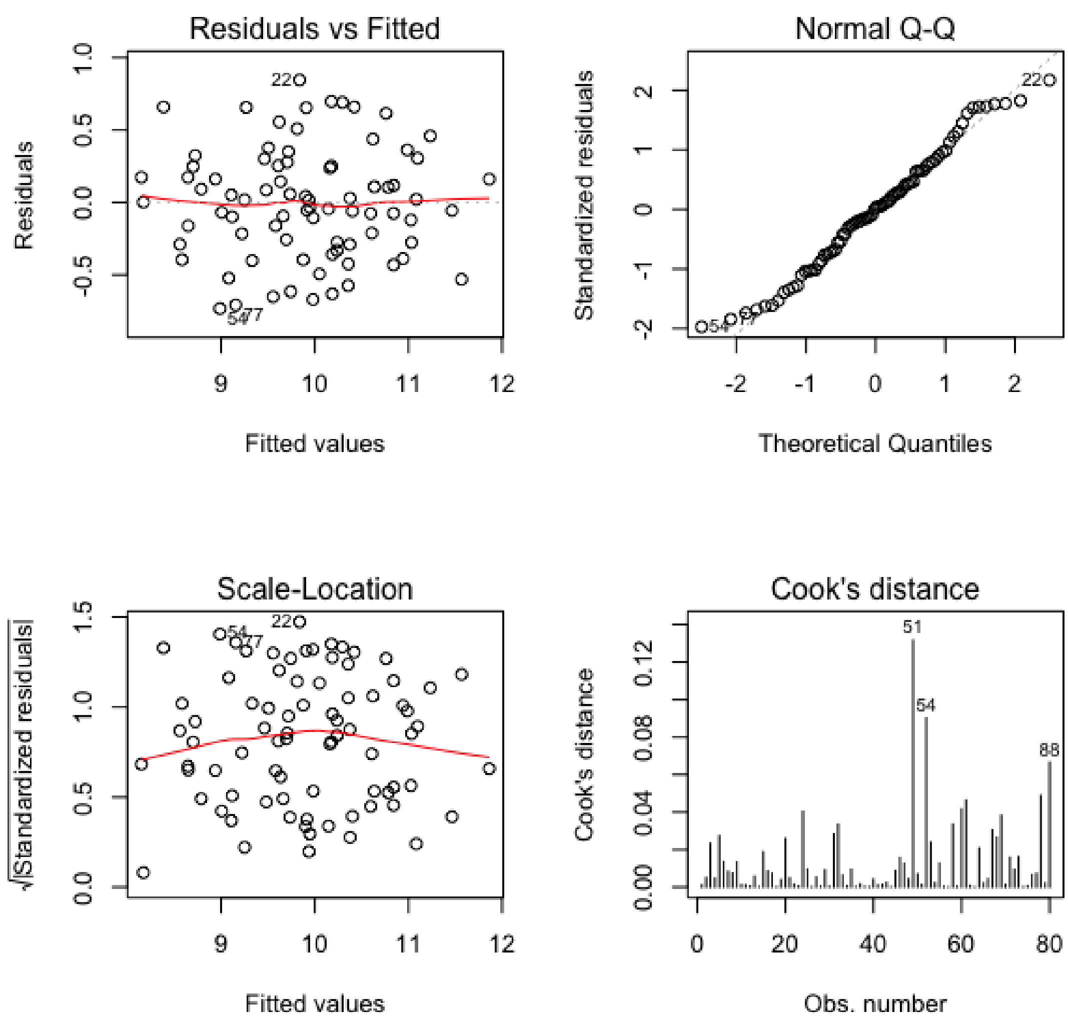
Appendix B. Example representation of geometric declining trends tested in power analyses.

Points represent years when sample data were simulated based on the pattern of sampled years in the current Guadalupe River dataset



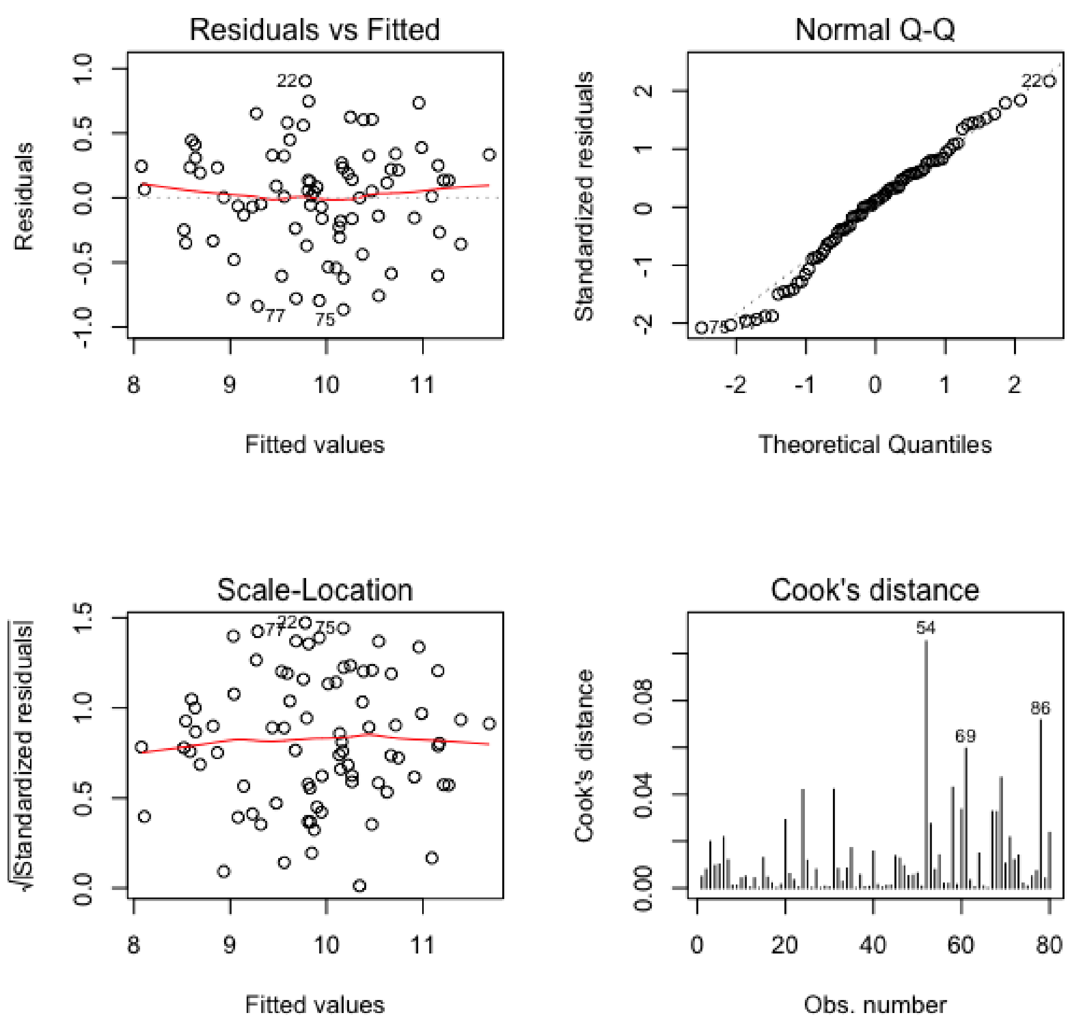
Appendix C. Guadalupe River PCB Concentration Model Residuals

Diagnostic Plots of Residuals – Model 42



Appendix C. (continued)

Diagnostic Plots of Residuals – Model 21

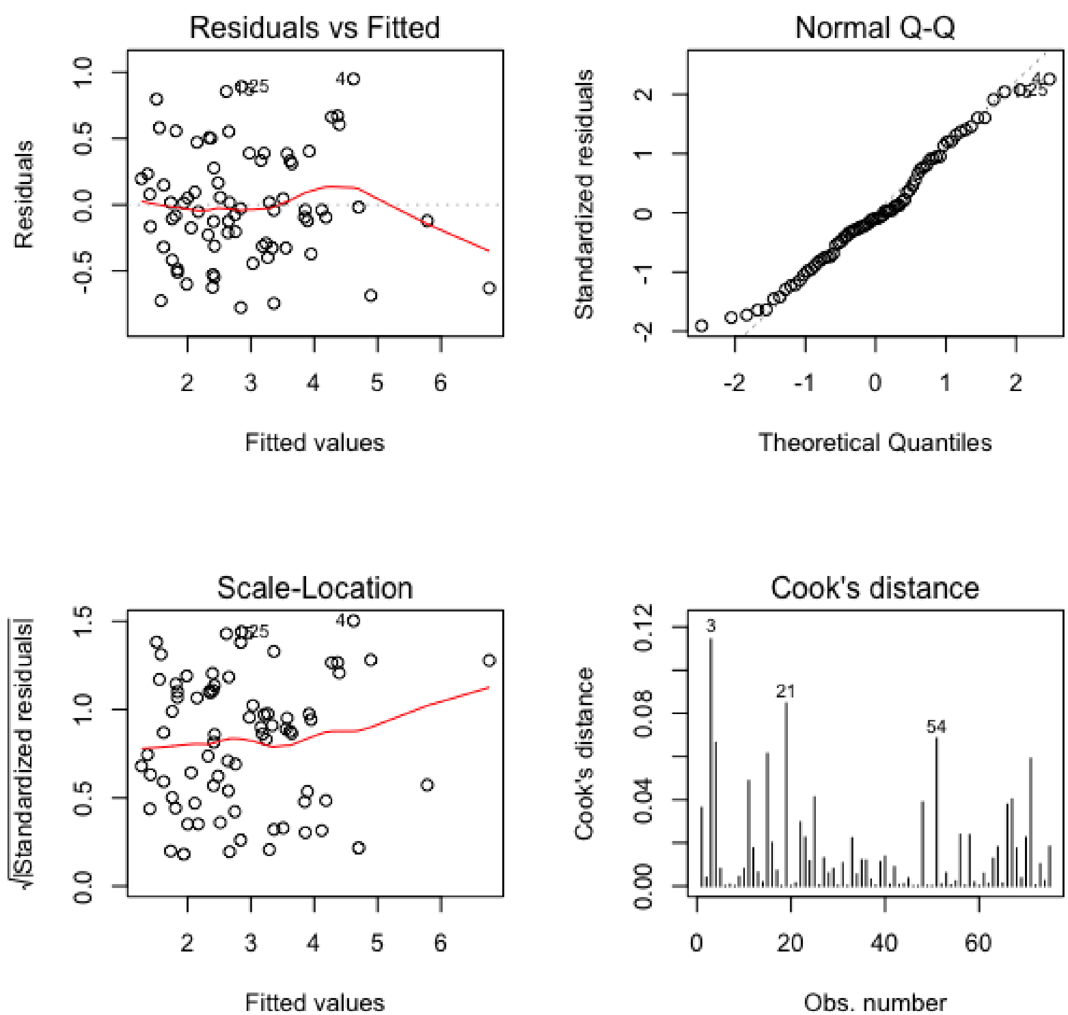


Appendix D. PCB Concentration Model Cross Validation Results

CV Method	Model	RMSE Model	R2 Model	RMSE CV	R2 CV
LOOCV	42	0.381	0.811	0.411	0.794
K-Fold				0.432	0.812
LOOCV	21	0.410	0.783	0.436	0.768
K-Fold				0.428	0.780

Appendix E. Guadalupe River Trends Model 34

Diagnostic Plots of Residuals



Appendix F. Composite Scenarios

Simulations of Composite Sampling Strategy

Approach:

A key assumption of employing the statistical modeling approach developed in this project is the use of multiple grab samples to characterize the PCB particle-ratio concentration of storms. However, it is feasible that the STLS program may switch to a composite-based sampling approach in the future to lower analytical costs. Therefore, to evaluate a trend procedure in a composite-based sampling program, an alternative analytical approach was evaluated.

In Step 1, composite samples were simulated, as prior composite data for PCBs have not been collected from the GR. Therefore, linear interpolation functions were generated between each pair of grab samples that have been collected. All storms were included where at minimum of two samples were collected (N = 19).

In Step 2, the interpolation functions were applied to entire storm periods (using the storm selection criteria described in the Methods of this report), to determine the instantaneous PCB concentrations at 15 min time-steps. In order to extend the interpolation to the beginning and end of the storm window, the first and last grab sample concentration was assumed constant for each 15-minute time point (Figure F1). This step provided a consistent manner for selecting storm events for testing trend models. Subsequently, the PCB concentrations were converted to event loads, corresponding to the same storm periods as for the grab based approach.

In Step 3, the estimated composite loads were employed in a trend model evaluation to test support for event-scale model parameters, same as for the grab-based methodology. The difference being that the number of storms used to test models was limited to just "sampled" storms (N = 19), not the entire pool of storms selected (N = 75).

Upon consideration of AIC model statistics and results of residuals analysis, a proposed composite-based trend model was employed in a power analysis procedure. The goal of this analysis was to assess the statistical power for detecting trends under a composite strategy, where high-resolution turbidity data and grab sample concentrations are not feasible. The power analysis methods for both grab-based and composite-based sampling are described in the Methods.

Results:

Figure F1. Example of the linear interpolation method to generate composite data (Steps 1-2). Red circles are the sampled grabs, and black circles are 15-minute sample intervals. Grab concentrations were interpolated between each pair of grab samples, and the interpolation function applied to every 15-minutes between sets of grabs. The interpolation method was used to estimate PCB concentrations every 15-minutes.

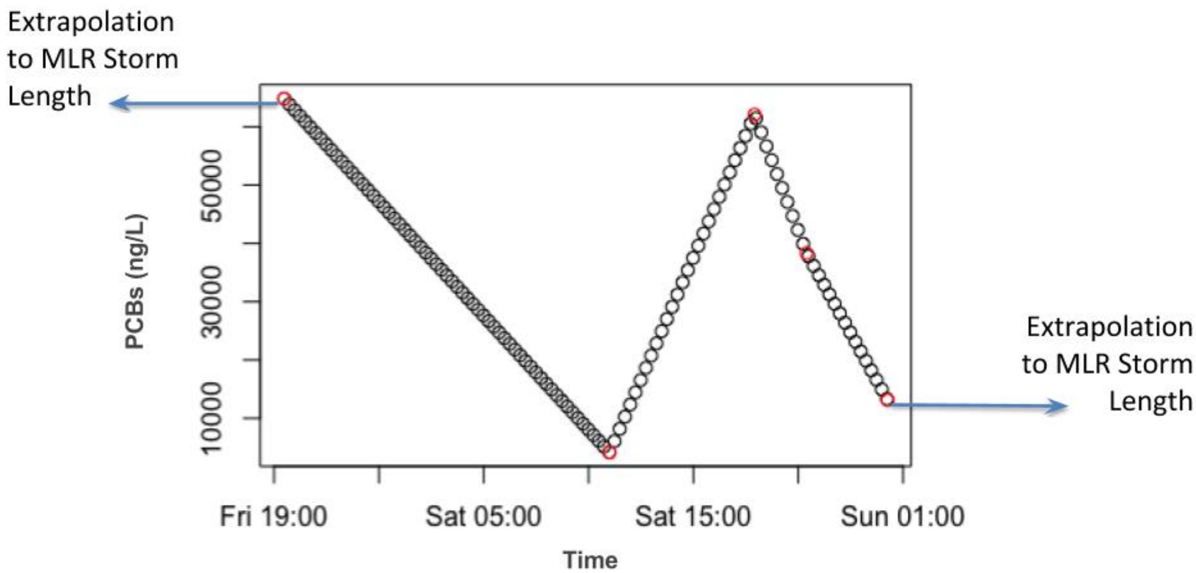


Figure F2. PCB event loads calculated by the grab-based regression (MLR Model) vs. composite method (N = 19). There was a good relationship between the loads estimated from the two methods. Only one storm (12/19/02) is underestimated by the composite method relative to the regression model. These results suggest the linear interpolation functions are a relatively good fit to the storm loads.

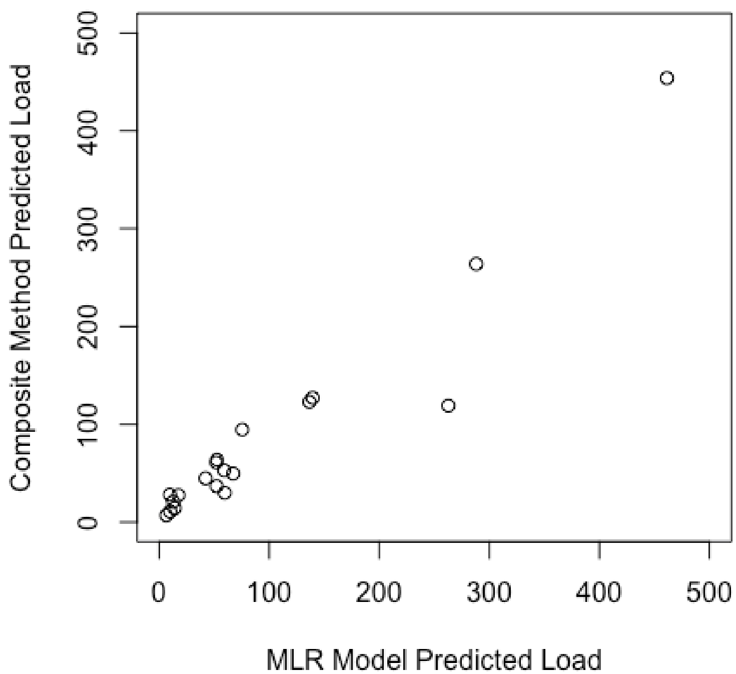


Figure F3. Composite Trend Model 1 Residuals. The regression line is the slope of the water year term (-0.05, p=0.20). This slope illustrates the shallow non-significant decline to the composite loads.

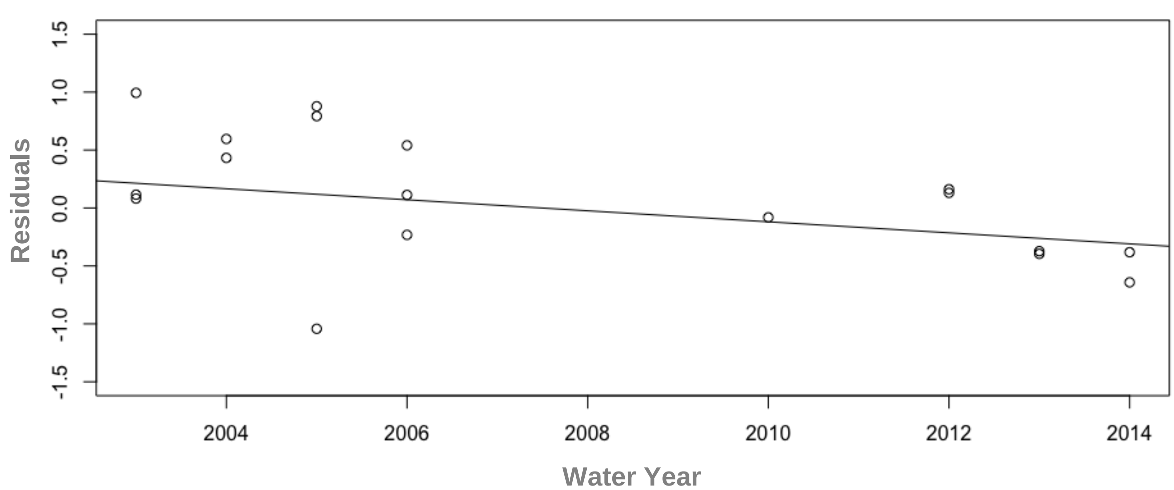


Table F1. "Best" Trends Models for the Composite Data. These were the only models where at least one significant model term was found. In both cases, a non-significant declining slope was also indicated.

Model	Model Parameters	Adj. R^2	RMSE	YEAR T-value (p-value)	Pred. R^2	PRESS Statistic
1	StormDuration[CSJ] + WaterYear	0.59	0.62	-0.05 ($p = 0.202$)	0.51	9.8
2	30minLag:2hrSum[CSJ] + WaterYear	0.54	0.62	-0.09 ($p = 0.056$)	0.46	10.8

Appendix G. Trends 20 Year Estimated Monitoring Budget

Table G.1. Sampling scenario and assumptions for Guadalupe River trends budget estimate

Trend decline detected	Sampling scenario	Sampling methodology	Watershed	Assumptions
25%	10 sampling years over a 20 year period; sampling at least 4 events/year; 4 grab samples/storm	Grab	Guadalupe River	1. Set up turbidity and flow monitoring equipment 2. Set up data telemetry 3. Sample 4 medium duration storm events/year for 10 out of 20 years-focused on urban hydrograph 4. Sampling only PCBs & SSC 5. Assumes full DM every year 6. Assumes full reporting every 5 years & SPL PPT any year monitoring occurs
75%	sample 24 storms over 20 years which equates to 2 storms per year for 12 out of 20 years	Composite	Guadalupe River	1. Set up turbidity and flow monitoring equipment 2. Set up data telemetry 3. Sample 2 medium duration storm events/year - focused on urban hydrograph 4. Assumes full DM every year 6. Assumes full reporting every 5 years & SPL PPT any year monitoring occurs

Table G.2. Budget estimates for field, laboratory and grab and composite sampling over 20 year period at Guadalupe River. Note this is a ballpark budget estimate with uncertainty in each line item. A more accurate and robust budget will be developed should the project be considered for funding via the RMP or other program.

Item	Grab	Composite
Labor 20 year budget	$947,000	$974,000
4 full reporting efforts	$140,000	$140,000
PCB analysis	$144,000	$22,000
SSC analysis	$18,000	$3,000
ODCs	$31,000	$27,000
20 year total	$1,280,000	$1,165,000
Total Average/Year	$64,000	$58,000