Appendix I-D

CONDITIONS ASSOCIATED WITH WATER COLUMN PCB CONCENTRATIONS:

THOMPSON ISLAND DAM 2009

Multivariable Analysis of Water Column PCB and Operational Data

March 5, 2010

Prepared for:

United States Environmental Protection Agency
Region 2
290 Broadway
New York, NY 10007-1866

Under Contract to:

Louis Berger Group, Inc.
412 Mt. Kemble Avenue,
Morristown, NJ 07960

.

KERN Statistical Services, Inc.
5175 NE River RD
Sauk Rapids, MN 56379

SUMMARY

It has been conjectured that water column concentrations can be "Predicted based only on PCB removal rate and river velocity". Based on this conclusion and application of such a model, GE broadly concludes (GE presentation to peer review panel *"Resuspension", Feb 15, 2010*) that:

1. EPA –proposed Phase 2 program cannot meet the resuspension performance standard
2. Drinking water standard will be exceeded frequently because of higher rates of dredging than in Phase 1.
3. Redeposition in non-dredge areas will compromise remedy benefits.
4. No practical means to reduce resuspension to standard.

PRIMARY FINDINGS

Subsequent to releasing these results, EPA has conducted a preliminary analysis of factors associated with water column PCB concentrations and has found that while water column PCB concentrations are indeed positively associated with mass of PCBs removed, a more careful analysis suggests that this relationship is due to a combination of several operational factors, some of which are readily manageable in ways that would logically be expected to reduce PCB releases associated with dredging operations.

Based on EPAs recent analysis, it can be concluded that the mechanisms associated with increased water column PCB concentrations are varied and likely, many and should not be simplified to a simple proportionality to mass removed, as suggested by GE. Mass removed is a surrogate for the net effect of all of the processes involved in dredging, and therefore correlates well with water column PCB concentrations. However, this does not preclude that of individual operational variables can be managed to reduce resuspension of PCBs.

Based on a multivariate analyses of the daily process and water column data, EPA finds that water column PCB concentrations are positively associated with several factors, all of which would be expected to influence release and resuspension of PCB contamination, including

1. Sediment removal (*i.e.*, bucket counts, volume removed, mass removed),
2. Flow rate,
3. Vessel traffic (primarily distance traveled by scows),
4. The number of CUs being backfilled in any given day,
5. The area and concentration of freshly disturbed sediments in CUs open to the water column each day,
6. Bucket fill-rate and other surrogates to sediment spillage,

Thirteen of the 28 process variables considered demonstrated statistically significant positive associations with water column PCB concentrations with squared Spearman Rank correlation coefficients ranging from approximately 0.2 to 0.4. EPA emphasizes that these levels of association are individually weak indicating that no single process can be identified as "*the source*" of resuspension, but rather a complex set of interactions among processes appears most likely to be "*causative*". Therefore it is expected that multiple variable models may be necessary to adequately explain variation in water column concentrations. It is further expected that controlling PCB resuspension may be accomplished through a combination of best management strategies applied to several stages in the sediment removal and disposal process—perhaps including application of multiple dredging and backfilling technologies.

Following is a description of the modeling approach used by EPA and the resulting relationships between operational variables and water column concentrations at TID.

OBJECTIVE

The primary objective of this analysis is to develop an empirical data-driven model describing water column concentrations of total PCB as a function of physical and operational variables associated with remedial activities in the Phase 1 project.

INTRODUCTION

Sediment transport investigations are often characterized by limited data, necessitating development of theoretical equations (e.g., models) to describe fate and transport of contamination from sediment to the water column. In contrast, the Phase 1 project is rich in data quantifying all aspects of the sediment removal process. These data are certification unit specific and available on a daily basis and can be associated with daily water column concentrations monitored at near- and far-field stations in Thompson Island Pool.

These rich data provide the basis to develop an empirical model of water column concentration as a function of measured data specific to the operations in the Thompson Island Pool in 2009. A combination of Factor Analysis (Seber, 1977) and Multiple Regression (Neter et al., 1996) is used to statistically identify groups of parameters most strongly associated with water column PCB concentrations. This empirical approach provides the opportunity to test hypotheses and assumptions that otherwise would remain untested in more typical situations where data are less rich—providing the means to evaluate the influence of various remediation processes on water column concentration.

METHODS

Data

Data were collected during the Phase 1 dredging project quantifying the primary aspects of operations throughout the dredging season. In addition, water column total PCB concentrations were measured daily, providing the potential to develop a retrospective model describing relationships between the mechanisms of the dredging process. The data include metrics quantifying, potential sources of PCBs associated with,

1. Flow and temperature conditions,
2. Debris removal,
3. Volume and mass removed,
4. Vessel traffic,
5. Efficiency of removal operations,
6. Resuspension of exposed PCB deposits in open CUs, and
7. Sediment disturbance associated with backfilling.

In all 28 variables were tested statistically for potential to predict water column PCB concentration at far field stations downstream of dredging operations. Data were summarized on a daily basis, so for most variables there were approximately 166 days (May 15th through October 27th) on which PCB concentration could be compared with dredging process variables. Some variables such as bucket fill rates were only available on the 127 days when active dredging occurred within the 166 day time period. Therefore the analyses were repeated for variables measured on all 166 days as well as for the subset of variables measured on just 127 days, primarily after June. Data measured on the smaller subset of days are more closely associated with <u>how</u> dredging was conducted and are therefore the more likely source of information on how dredging and other activities could be modified to reduce resuspension of PCBs to the water column.

Modeling Overview

This approach is used to develop a model of the form

$$C_{water} = C_{baseline} + K_1 C_{Source-1} + K_2 C_{Source-2} + \cdots + K_n C_{Source-n} \quad (1)$$

where the constants K_1, K_2, K_3,… K_n are loosely interpreted as "*net*" sediment to water partitioning coefficients for each source, where it is understood that source terms are based on surrogates for sediment removal processes.

The Phase 1 project is unique in the richness of data available to not only estimate these coefficients, but also to identify those combinations of measured processes that are most important for predicting water column concentrations. Multiple regression analysis is used to identify metrics contributing significantly to prediction of water column PCB concentrations.

Surrogates and Confounding

Metrics described here should be considered primarily as surrogates for physical processes of interest. For example, it is not clear that there is a partitioning coefficient between bucket counts and water column concentrations, however if one were to develop a mechanistic model relating sediment losses per bucket count, then there would be a "net" partitioning of PCBs relating sediment losses from bucket counts and water column concentrations. It is this net partitioning that is estimated by the coefficients of the regression model. Similarly, because the mass removed per day is derived from the bucket count, and other variables, it would also be expected that water column concentration would be correlated with the mass of material removed per day. In fact, mass removal is clearly a surrogate integrating all processes likely to cause PCB losses. Unfortunately this does not provide useful insight into how operational process can be modified to reduce PCB sources to the water column. Essentially any variables that are correlated to material disturbance and removal would be expected to correlate with water column concentration. Additionally some of those variables are also expected to be inter-correlated amongst one another, in statistical terms multi-collinear.

Cause and Effect

Because the data developed through this study are observational as opposed to based on a designed experiment and because many of the process variables are inter-correlated, one cannot infer cause and effect relationships directly. Particularly due to the surrogate nature of most of the metrics, it is important to consider relationships to be associative rather than causative unless other lines of evidence can be used to eliminate some plausible causative processes.   For example correlation between bucket counts and water column concentrations could lead one to conclude that use of larger buckets could reduce the daily bucket count per unit volume removed in efforts to reduce water column concentrations. However, if the actual acting mechanism is due to disturbances from scow traffic, which also would be expected to be associated with bucket counts, then modification to the bucket size in efforts to reduce bucket counts would be futile. <u>The data are observational and therefore their use in developing best management practices should take into account that individual metrics may be surrogates for what may be lurking un-quantified processes.</u>

Bivariate Correlation Analysis

The first step in the regression analysis was to analyze the pairing of each individual process variable with water column PCB concentration to test for a positive association. Bivariate relationships between process variables and water column PCB concentration are an indicator of at least a surrogate relationship that could also be a causative mechanism that should be considered for plausibility. These bivariate relationships were summarized by calculating Spearman Rank Correlation coefficients between water column PCB concentrations and each process variable of interest. Correlations were also calculated for water column concentrations lagged by 1 and 2 days respectively to determine if subsequent analyses should incorporate adjustments for travel time between the dredging areas and the far field monitoring station.

<u>Multiple Regression</u>

A fundamental assumption of multiple regression is that the predictor variables (*i.e.*, source terms) are statistically independent. In this situation it is clear that many of the source terms of interest are inter-correlated - more sediment volume removed requires more vessel traffic. Therefore some groups of source terms cannot be entered directly into multiple regression models without careful consideration of their inter relations. In the statistical literature the interrelated predictor variables are called multi-collinear. A great deal of effort has been devoted to the study of the effects of multi-collinearity and methods to mitigate the effects on interpretability of model coefficients as well as the predicted values.

Regression models are typically used for two purposes: 1) prediction of the response variable, and 2) testing and interpretation of the regression coefficients. For prediction one is primarily interested in estimating future water column concentrations under a set of conditions previously measured in the model fitting process. As long as the future conditions are within the range of the variables used to estimate the model coefficients, multi-collinarity generally does not adversely impact predictions. Conversely, multi-collinearity essentially precludes the use of models in efforts to differentiate causative relationships such as the importance of dredging relative vessel traffic more difficult. Careful model construction and evaluation of sub-models to deduce the plausibility of causative processes would be necessary.

Recognizing this limitation of multiple regression models to differentiate individual collinear factors, this analysis is designed to develop a predictive model and to qualitatively evaluate the relative importance of the factors associated with water column total PCB concentrations. From this analysis it is possible to identify groups of process variables that are collectively associated with water column PCB concentrations. Identification of such independent process variables suggests components of the dredging process that should be investigated for potential causative relationships.

<u>Factor Analysis</u>

A set of statistically independent predictor variables are derived from the collection of 28 inter-correlated variables. These independent variables are derived by applying a Factor analysis to the full collection of predictor variables, and deriving surrogates for the dredging processes that are statistically independent and can be entered jointly into multiple regression models. Factor analysis is similar to principal components analysis with the exception that the principal components are “rotated” through an orthogonal transformation that often results in component loadings that are more physically interpretable. It is recognized that there are no unique or optimal factor solutions, however, development of independent scores that are composed of interpretable groups of process variables is desirable for the purposes of developing a predictive model, as well as for interpretation of the relative importance of independent groups of process variables. It is fully recognized that because many process variables are inter-correlated, fully dissecting the relative importance of each process variable may not be possible, but to the extent that factor scores can identify independent groups of variables, the contribution of each group collectively can be distinguished through this approach.

Multiple Regression Analysis

The results of the factor analysis were used to transform each daily set of process variables into a linear combination of independent variables called factor scores. These factor scores have the advantage of being statistically independent and are therefore compatible with the assumptions of multiple regression. A multiple regression was used to identify those factors that were important to prediction of water column PCB concentration. Important factors were defined as those factors with regression coefficients that were significantly different for zero at the 5% level of statistical significance. The resulting model is suitable for use as a predictive model, and by inspection of the factor loadings can be used to identify independent combinations of process variables important to prediction of water column PCB concentrations. Effects due to variables nested within a common factor are difficult to distinguish without other lines of evidence.

The results of this analysis can also be used as a guide to the development of mechanistic models that are specific to individual process variables. In particular, when more process oriented variables are to be calibrated against water column data the interrelations found here through factor analysis should be respected and unless certain processes can be eliminated through other data and analysis, it would not be reasonable to assume that individual process variables can be eliminated purely through identification of other surrogates that are more strongly associated with water column concentrations.

For example, mass removed per day can be tracked relatively accurately, while losses from bucket lifts are much more difficult to measure directly. Therefore, the quality of mass removal data are expected to be much less variable and therefore more likely to correlate with water column concentration than sediment losses. Because of this difference in measurement quality among variables, mass removal per day would be the better apparent predictor of water column concentrations, but this would not eliminate the potential that loss percentages might be the more important process contributing to water column PCBs. The well known adage is that association does not imply causation, but conversely lack of apparent association also does not eliminate causation. Distinguishing the root causes of water column concentration will require extensive and careful multiple variable analyses combined with professional judgment and development of mechanistic models in order to develop sound best management practices. The analysis presented in this section is a first step in this direction, intended to provide an indication of the major groups of processes influencing water column PCB concentrations. Until more detailed sub-analyses are conducted it would be premature to eliminate any process variables from consideration for improvement and refinement.

RESULTS

Thirteen variables representing 5 groups of processes were identified that were associated with water column PCB concentrations at TID. Five variable groups (Factors) were identified that collectively explained 55% and 60% of the variation in water column PCB concentration in the 166 and 127 day models respectively. These factors represented volume and mass removed and efficiency, area of recently disturbed sediments in open certification units, vessel traffic and backfilling. Following is a summary of the results of the analysis.

<u>Bivariate Correlations</u>

Squared Spearman Rank correlation Coefficients for water column PCB concentration with each of the process variables are reported in Table 1. The analysis was repeated with process variables paired with one-day and two-day lagged PCB concentrations to evaluate the potential effects of travel time on the strength of correlation. These squared correlation coefficients represent the proportion of variation explained by the relationship between water column PCB concentration and each variable, analogous to an R^2 from a regression. The Spearman correlation coefficient is preferred because the assumption of linearity inherent in the Pearson's coefficient is relaxed. Results summarized in Table 1 show that:

1. Corrrelations between water column PCB concentrations and process variables are generally weak ranging from 2% for debris removal to 42% for volume and mass removal, indicating that no single variable could be expected to adequately explain the fluctuations in water column PCBs observed during the Phase 1 project.
2. Correlations for water column concentrations lagged by one day were less than those for concurrent measurements and two day lagged measurements produced still lower correlations.
    a. In contrast GE asserted that weekly averages were needed to counter the effects of travel time in their analysis of the water column PCB data.
    b. EPA views this as counterproductive given the lack of correlation between lagged water data and process variables.
    c. Weekly averaging would artificially reduce the power to detect subtle multiple variable relationships suppressing potentially important relationships between water column PCB concentration and operational variables.

3. Statistically significant positive associations were identified for most processes expected to disturb sediments
   a. Volume and mass removed (R^2=0.22 to 0.42)
   b. Dredging efficiency measures such as bucket fill rate and depth of cut (R2=0.08 to 0.22)
   c. Sources due to area of open CUs (R2=0.15 to 0.19)
   d. Debris removal (R2=0.02)
   e. Boat traffic (R2=0.11 to 0.26)
   f. Backfilling operations (Number of CUs being backfilled) (R2=0.09)
4. Weak statistical relationships may be indicative of surrogate relationships that are markers for important, but crudely quantified, sources of PCB resuspension.
5. Water column concentrations were negatively associated with flow at the Fort Edward Station, but the relationship was not statistically significant.

Multiple Variable Analyses

Because water column PCB concentrations were weakly associated with several operational variables, efforts were made to develop a multivariable model that would adequately explain water column PCB concentrations. Because several process variables were derived from basic measurements such as bucket counts it was expected that many process variables would be inter-correlated. In efforts to understand inter-relationships between process variables a factor analysis was conducted to identify a set of independent factors that would be both meaningfully interpretable, as well as providing inputs for a regression model predictive of water column PCB concentrations.

The factor analysis was conducted with only the predictor variables for the subset measured on all 166 days as well as the subset measured on just 127 of the 166 days. Resulting factor scores were used as predictors in a regression analysis to identify important factors for prediction of water column PCB concentrations.

Factor Analysis (127 day model)

Table 2 and Figure 1 show the factor loadings for each of the 28 process variables under consideration. The factor loadings are unitless and range from plus one to minus one and are considered meaningful when they exceed approximately 0.4 in magnitude. Loadings that are less than 0.4 in magnitude are within the opaque rectangular area. Cells in Tables 2 and 3 are shaded green to draw attention to loadings that exceed this nominal level.

There were 5 factors associated with total PCB concentration in the water column describing from 2% to 37% of the total 60% variance in water column PCB concentration explained by the regression model. Regression coefficients, standard errors, variance inflation factors and partial R^2 values are tabulated in Table 4.

Factor-1 includes loadings on bucket counts, mass removed, volume removed, residual Total PCB Concentration in Open CUs and the product of mass and removal efficiency (ME). This factor summarizes potential PCB sources from variables that are directly related to sediment removal, as well as efficiency of the removal process.

Factor-6 loads most heavily on the amount of backfilling being conducted and the product of flow and backfill (a surrogate for load from backfilling). This factor also has substantial negative loadings on bucket counts and temperature. This may reflect that bucket counts incidentally vary inversely with temperature and backfilling operations.

Factor-7 loads most heavily on concentration weighted surface area of open CUs and flow and concentration weighted surface area of open CUs. This factor has a clear signal exclusively related to the amount of open CUs at any point in time that is independent of volume and mass removal.

Factor-8 loads primarily on flow, and the product of flow and total vessel traffic. This factor is also independent of variables in Factor-1 describing removal metrics indicating that there may be an independent PCB source to the water column associated with vessel traffic. This variable is a crude measure of potential sources due to vessel traffic, because it does not account for either water depth or concentration of areas over which traffic occurs. It is expected that refinement of this variable will substantively improve its relative strength as a predictor of water column concentrations.

Factor-9 loads on boat distance which is a single metric that only accounts for distance traveled by vessels.

General Observation

These results suggest that resuspension of PCBs to the water column is associated with a combination of removal activities, backfilling activities, vessel traffic and the surface area and duration that disturbed residuals are exposed in open CUs. This suggests that best practices could be applied to one or several of these processes to reduce concentrations of PCBs in the water column.

Factor Analysis (166 Day Model)

The factor loadings for the 166 day model are summarized in Table 3 and Figure 2. Results were similar because the majority of data were common to both models. The model fit was slightly weaker with and adjusted R^2=55% as compared with 60% for the 127 day model. The five factors were qualitatively similar to those identified in the 127 day model representing variables associated with sediment removal (semi-partial R^2=28%) backfilling (semi-partial R^2=6%) concentration weighted surface area of open CUs (semi-partial R^2=4%) flow times vessel distance (semi-partial R^2=5%) and mass removed (semi-partial R^2=13%). Because performance data related to bucket filling rates were not included in the 166 day model, the separation of variables among factors was less obvious and general surrogates for overall activity such as mass and volume removal and boat traffic tended to group together in the first factor. This suggests that further refinements in the understanding of processes controlling fluxes of PCBs to the water column should focus on variables that characterize <u>how</u> dredging and other supporting operations are conducted as opposed to just on <u>how much</u> dredging is done.

Model Predictions

The fitted model results are plotted on Figure 3 showing that the modeled concentrations generally track day to day fluctuations in concentration in most months, including patterns observed in October that were not well described by GEs simpler model. Estimated regression coefficients, standard errors, partial R^2 and variance inflation factors are summarized in Tables 4 and 5. This suggests that GEs assertion that concentrations are driven exclusively by the amount of dredging may not be fully justified.

Also included in the plot are upper 95% prediction limits which are an added benefit of the regression approach to model development. It can be seen that the prediction intervals indeed capture at least 95% of observations and that when there are excursions above the prediction limits they are frequently tied to situations that may be well understood. For example excursions above or near the prediction limits occur in early August when dredging was halted due to exceedances of critical load thresholds. At these times the process variables are all simultaneously zero leading the predicted values to drop, whereas in these extreme conditions the corresponding reduction in PCB concentrations lagged the change in process operations.

Given that correlations were found to be strongest for water column concentrations paired with process variables measured on the same day (i.e. as opposed to lagged) it may seem somewhat contradictory that water column concentrations remained elevated for several days in August after operations were shut down. It is currently thought that this may be due to the fact that during this time, water was impounded in the East Rogers Island area and water column PCB concentrations were an order of magnitude higher than in the main flow of the river.

Because this water was impounded and isolated from the main flows of the river with approximately a 200 cfs discharge, these PCBs would influence TID water column concentrations as a relatively steady elevated concentration. Because of the slow discharge rate of this very high concentration water would require several days to flush out of the impounded area at a rate of 200cfs, therefore creating in the overall average concentration whereas the effects of day to day fluctuations were identified more immediately over and above the increase in base concentrations. Additional analyses will target the effects of separating these two sources of PCBs on the quality of model predictions.

CONCLUSIONS

This multiple variable analysis should be considered a first step in understanding the processes contributing PCBs to the water column. It is apparent that several processes may be contributing to the PCB loads to the TID far field stations and that there is potential to improve the dredging process while maintaining a high likelihood that the resuspension standard can be met in the Phase 2 project. The most likely factors contributing PCBs to the water column are not unexpected—mass and volume removal, vessel traffic, exposure of freshly disturbed residual sediments to active flows, processes associated with backfilling, and the extent to which dredge buckets may be overly full or dredging is hurried.

This analysis shows that a combination of processes are likely contributing measureable concentrations of PCBs to the water column which presents an opportunity to fine tune dredging operations in Phase 2.

This analysis stops short of development of a final model based solely on process variables, as opposed to factor scores, as this step involves a great deal of care and deliberation in selection of model variables and evaluation of the plausibility that resulting models might be reasonably expected to be causative. The model reported here is clearly associative, but does support the hypothesis that sources of PCBs to the water column are many and varied and that there are likely to be many opportunities to minimize PCB resuspension during the upcoming Phase 2 dredging project. Surrendering to the notion that *resuspension just happens* is probably not a reasonable response to the rich information that is available to further refine and optimize the dredging operation. EPA continues efforts along these lines to investigate factors identified in this analysis and their potential as causative agents as opposed to just surrogates. It is anticipated that these efforts will provide information necessary to develop operational management strategies.

REFERENCES

Neter, J., Kutner, M.H., Nachtsheim, C.J., and W. Wasserman. 1996. Applied Linear Statistical Models, 4th ed., Irwin. Chicago.

Seber, G.A.F. 1977. *Multivariate Observations.* Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons. New York, NY.

Table 1. Squared spearman rank correlation coefficients between water column total PCB concentrations lagged by 0, 1 and 2 days. Correlations are more often the strongest when based on concurrent measurements of water column PCBs and sediment disturbance and productivity factors.			
Variable	PCB_ngl	PCB_ngl_Lag1	PCB_ngl_Lag2
BargeDist	0.001	0.000	0.003
BargeV_D	0.005	0.004	0.001
BargeVel	0.003	0.002	0.001
BCntTotal	**0.217**	0.174	0.094
BoatDist	**0.338**	**0.315**	**0.189**
BoatV_D	**0.315**	**0.313**	**0.188**
BoatVel	**0.275**	**0.254**	**0.145**
Debris	**0.135**	**0.165**	**0.186**
DrdgDist	0.033	0.016	0.002
DrdgV_D	0.004	0.000	0.001
DrdgVel	0.001	0.002	0.010
Fill Rate	0.095	0.063	0.021
FlowFE	0.014	0.035	0.044
Load_Bfill	**0.091**	**0.094**	**0.094**
Load_CU_Area	0.191	0.190	0.193
Load_MassRem2	**0.379**	**0.316**	**0.191**
LoadBoats	**0.303**	**0.253**	0.136
MassRemTotal3	**0.417**	**0.323**	**0.185**
ME	**0.384**	**0.346**	**0.225**
SbDist	0.108	0.066	0.026
SbV_D	0.118	0.076	0.038
SbVel	0.022	0.007	0.000
ScowDist	**0.265**	**0.220**	**0.140**
ScowV_D	**0.257**	**0.225**	**0.144**
ScowVel	**0.178**	**0.125**	0.072
Temp_C	0.007	0.011	0.013
TotalBfill	**0.090**	**0.094**	**0.093**
tPCB_CU_AREA	**0.146**	**0.147**	**0.148**
VolRemTotal	**0.346**	**0.285**	**0.161**
Notes:			
1) Gray cells indicate when water current water column concentrations correlate more strongly than lag-1 measurements or when lag-1 measurements correlated more strongly than lag-2 measurements.			
2) Bold numbers indicate that correlations are significantly different from zero at the 5% level of significance.			

3) Number of days represents the number of paird observations for values measured concurrently. Sample sizes associated with one and two day lags are reduced by one or two respectively.

Table 2. Factor loadings for each variable for those factors found to be associated with water column Total PCB concentration in Thompson Island Pool from May to November in 2009, Hudson River. Loadings reange from minus 1 to plus 1 and values greater in magnitude than 0.4 (green shaded and bold) are thought to be meaningful. Based on 127 day model.

Variable	Factor1	Factor6	Factor7	Factor8	Factor9
BCntTotal	**0.66**	**-0.40**	0.19	-0.07	0.13
BargeDist	-0.09	0.22	-0.10	-0.04	0.08
BargeV_D	0.02	0.06	-0.04	-0.10	0.09
BargeVel	-0.10	0.22	-0.11	0.00	0.04
DrdgDist	0.04	0.08	0.02	0.01	0.11
DrdgV_D	0.03	0.00	-0.05	-0.03	0.04
DrdgVel	-0.11	0.03	-0.12	0.01	-0.08
Load_Bfill	-0.16	**0.94**	-0.03	0.16	0.06
FlowFE	-0.35	0.12	0.00	**0.86**	-0.19
Temp_C	0.18	**-0.72**	0.16	-0.11	0.09
Load_CU_Area	0.27	-0.09	**0.87**	0.24	0.05
MassRemTotal3	**0.92**	-0.11	0.10	-0.11	0.06
SbDist	0.10	0.05	0.02	0.02	0.12
SbV_D	0.09	0.13	0.02	-0.01	0.11
SbVel	-0.02	-0.03	-0.11	0.02	-0.13
ScowDist	0.34	-0.03	0.14	-0.03	0.20
ScowV_D	0.26	0.04	0.19	-0.02	0.10
ScowVel	0.28	-0.13	-0.15	0.04	-0.05
TotalBfill	-0.14	**0.93**	-0.05	0.03	0.09
VolRemTotal	**0.82**	-0.10	0.14	-0.04	0.08
tPCB_CU_AREA	**0.44**	-0.17	**0.78**	-0.18	0.13
TotalEfficiency	0.32	-0.10	0.16	0.09	0.03
ME	**0.92**	-0.11	0.10	-0.06	0.04
BoatDist	0.36	0.15	0.11	0.04	0.70
BoatVel	0.19	0.03	0.06	0.06	0.14
BoatV_D	0.36	0.09	0.14	0.03	0.68
LoadBoats	0.01	0.32	0.15	0.77	0.41
Semi-Partial R^2	37%	10%	2%	10%	2%
Factor Label	Volume/Mass Bucket Fill	Backfill and Flow Weighted Backfill	PCB/Flow Weighted CU Area	Flow Weighted Vessel Dist.	Vessel Distance/Velocity

Table 3. Factor loadings for each variable for those factors found to be associated with water column Total PCB concentration in Thompson Island Pool from May to November in 2009, Hudson River. Loadings reange from minus 1 to plus 1 and values greater in magnitude than 0.4 (green shaded and bold) are thought to be meaningful. Based on 166 day model.					
Variable	**Factor1**	**Factor4**	**Factor6**	**Factor7**	**Factor12**
BCntTotal	**0.84**	-0.23	0.08	-0.04	0.05
BargeDist	0.08	0.22	-0.12	-0.02	-0.01
BargeV_D	0.12	0.06	-0.04	-0.09	0.04
BargeVel	0.05	0.21	-0.12	0.02	-0.03
DrdgDist	0.18	0.09	0.00	0.01	-0.02
DrdgV_D	-0.02	0.00	-0.02	-0.04	0.08
DrdgVel	-0.16	0.00	-0.14	0.03	-0.07
Load_Bfill	0.06	**0.97**	-0.03	0.11	-0.03
FlowFE	**-0.45**	0.11	-0.03	**0.86**	0.02
Temp_C	0.32	**-0.65**	0.18	-0.12	-0.04
Load_CU_Area	0.32	-0.07	**0.88**	0.17	0.01
MassRemTotal3	**0.84**	-0.09	0.08	-0.12	**0.44**
SbDist	**0.42**	0.09	0.00	0.03	-0.01
SbV_D	0.36	0.16	0.03	-0.03	0.03
SbVel	0.18	0.00	-0.14	0.07	-0.01
ScowDist	**0.95**	0.03	0.08	-0.05	-0.09
ScowV_D	**0.91**	0.07	0.10	-0.02	-0.09
ScowVel	**0.91**	-0.03	-0.15	0.02	-0.14
TotalBfill	0.10	**0.94**	-0.05	-0.01	0.02
VolRemTotal	**0.87**	-0.04	0.08	-0.04	0.12
tPCB_CU_AREA	**0.49**	-0.14	**0.78**	-0.23	0.01
BoatDist	**0.85**	0.18	0.08	-0.02	-0.02
BoatVel	**0.61**	0.07	0.02	0.05	0.00
BoatV_D	**0.83**	0.14	0.12	-0.03	-0.01
LoadBoats	**0.59**	0.37	0.12	**0.60**	-0.12
tPCB_CU_AREA	**0.48**	-0.18	**0.76**	0.10	-0.23
TotalEfficiency	0.32	-0.09	0.16	**0.91**	0.06
ME	**0.91**	-0.10	0.10	0.23	-0.07
Semi-Partial R^2	28%	6%	4%	5%	13%
Factor Interpretation	Volume/Mass Bucket Fill	Flow Weighted Backfill	PCB/Flow Weighted CU Area	Flow Weighted Vessel Dist.	Mass Removed

Table 4. Coefficients, standard errors, tests of significance, squared semipartial correlation coefficients and variance inflation factors for regression of water column Total PCB concentration on factor scores. Analysis is based on the variables measured on 127 of the 166 day season.

Variable	Factor Interpretation	Coefficient Estimate	Standard Error	Students T-Statistic	Significance Level	Squared Semi-Partial Correlation	Variance Inflation Factor
Intercept	NA	237.17	6.31	37.57	<.0001	NA	NA
Factor1	Volume/Mass Bucket Fill	67.21	6.34	10.60	<.0001	37%	1.0
Factor6	Backfill and Flow Weighted Backfill	34.04	6.34	5.37	<.0001	10%	1.0
Factor7	PCB/Flow Weighted CU Area	16.36	6.34	2.58	0.0111	2%	1.0
Factor8	Flow Weighted Vessel Dist.	34.19	6.34	5.39	<.0001	10%	1.0
Factor9	Vessel Distance/Velocity	14.15	6.34	2.23	0.0275	2%	1.0

Table 5. Coefficients, standard errors, tests of significance, squared semipartial correlation coefficients and variance inflation factors for regression of water column Total PCB concentration on multivariate factor scores. Analysis is based on the variables measured on each of the 166 days of the season.

Variable		Coefficient Estimate	Standard Error	Students T-Statistic	Significance Level	Squared Semi-Partial Correlation	Variance Inflation Factor
	Intercepth	214.51	5.95	36.06	<.0001		0.00
Factor1	Volume/Mass Bucket Fill	59.67	5.97	10.00	<.0001	28%	1.00
Factor4	Backfill and Flow Weighted Backfill	27.52	5.97	4.61	<.0001	6%	1.00
Factor6	PCB/Flow Weighted CU Area	23.23	5.97	3.89	0.0001	4%	1.00
Factor7	Flow Weighted Vessel Dist.	24.16	5.97	4.05	<.0001	5%	1.00
Factor12	Mass Removed	41.01	5.97	6.87	<.0001	13%	1.00

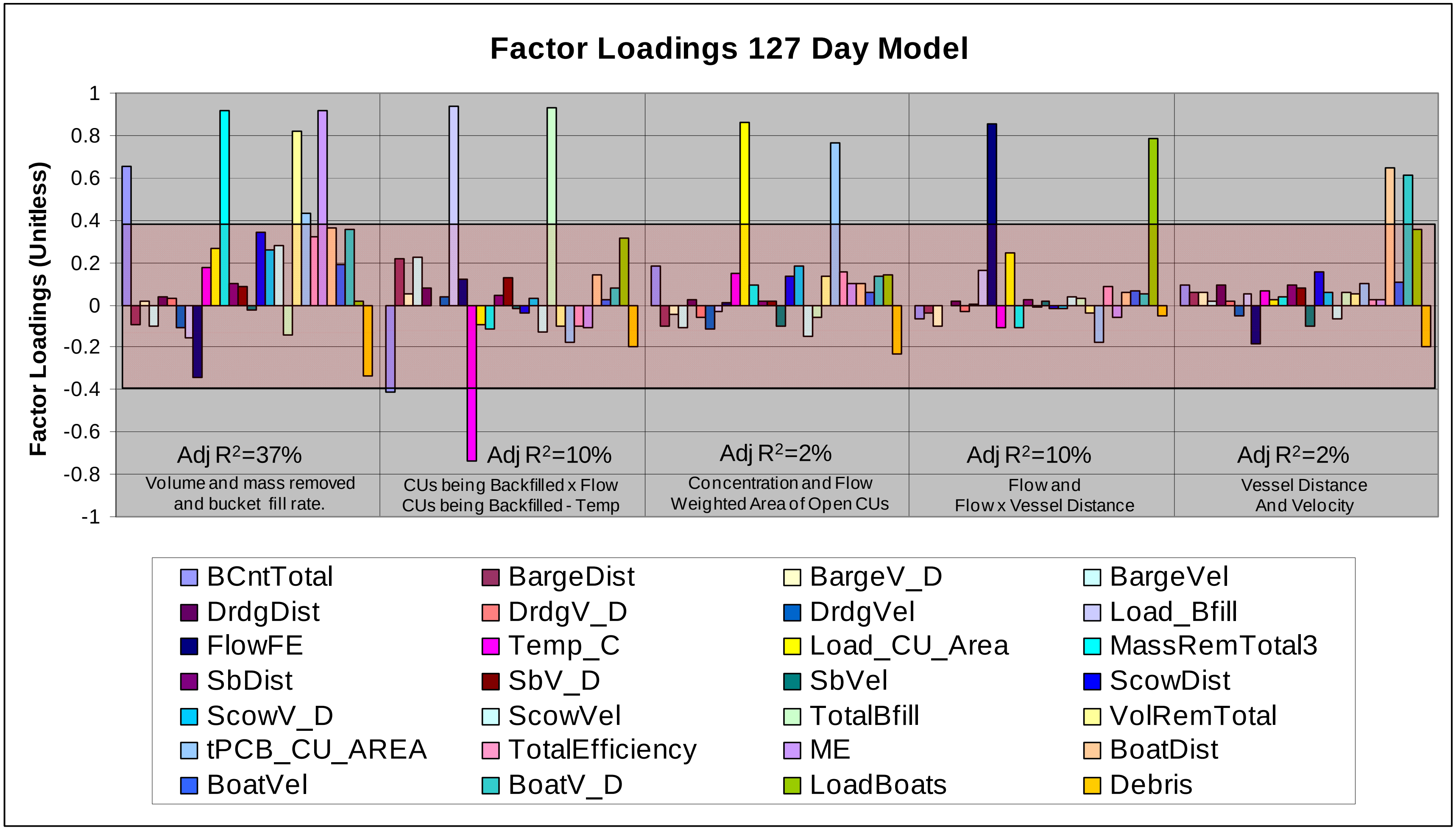

Figure 1. Factor loadings for six factors identified to be important factors for prediction of water column PCB concentrations. R^2 values represent the proportion of variance explained by each factor in multiple regression with water column PCB concentrations at far field stations in Thompson Island Pool. Loadings greater than roughly 0.4 in magnitude are considered meaningful.

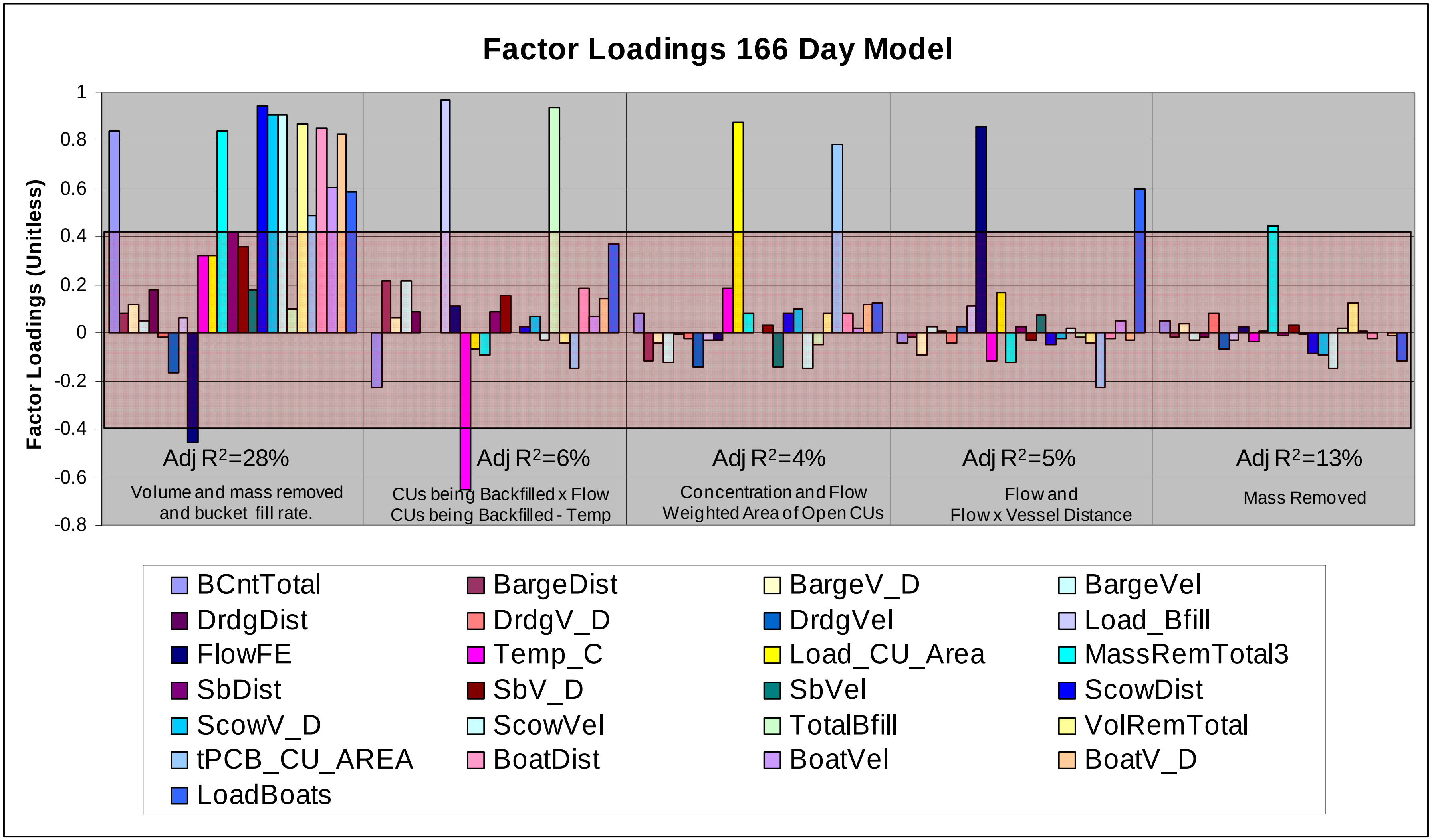

Figure 2. Factor loadings for six factors identified to be important factors for prediction of water column PCB concentrations. R^2 values represent the proportion of variance explained by each factor in multiple regression with water column PCB concentrations at far field stations in Thompson Island Pool. Loadings greater than than roughly 0.4 in magnitude are considered meaningful.

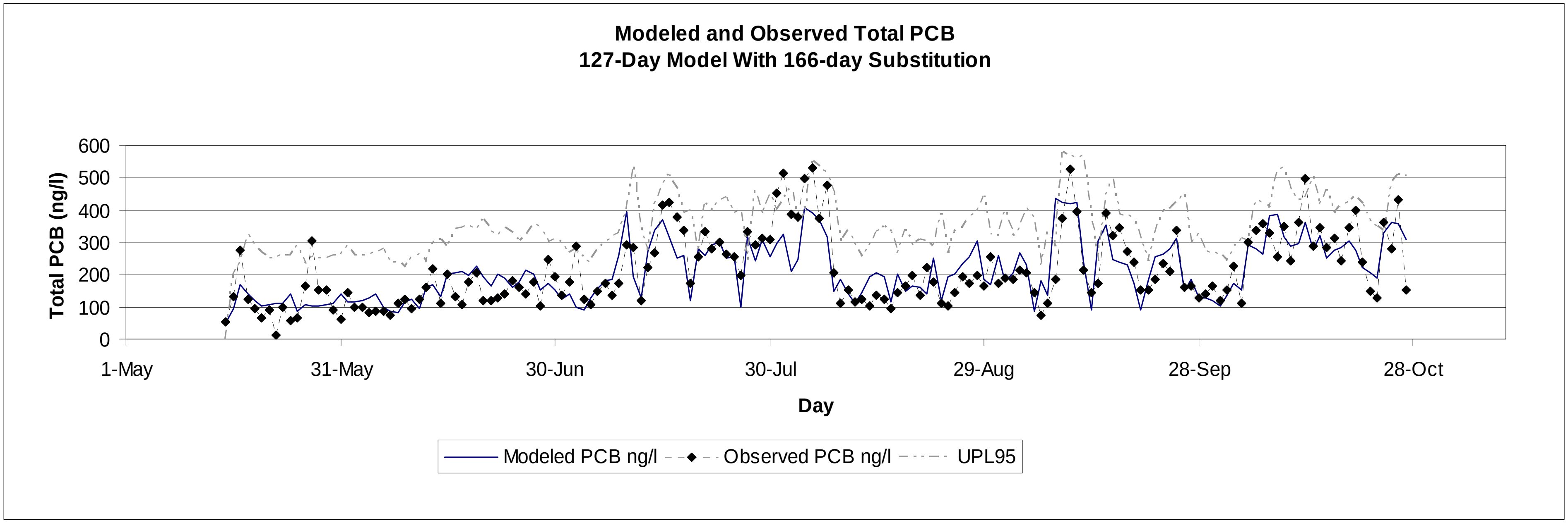

Figure 3. Observed and modeled values for water column PCB concentrations at far field station in Thompson Island Pool. The model is based on variables available on 127 of the 166 day season with modeled values from the 166 day model substituted on the remaining days—primarily in May and June.

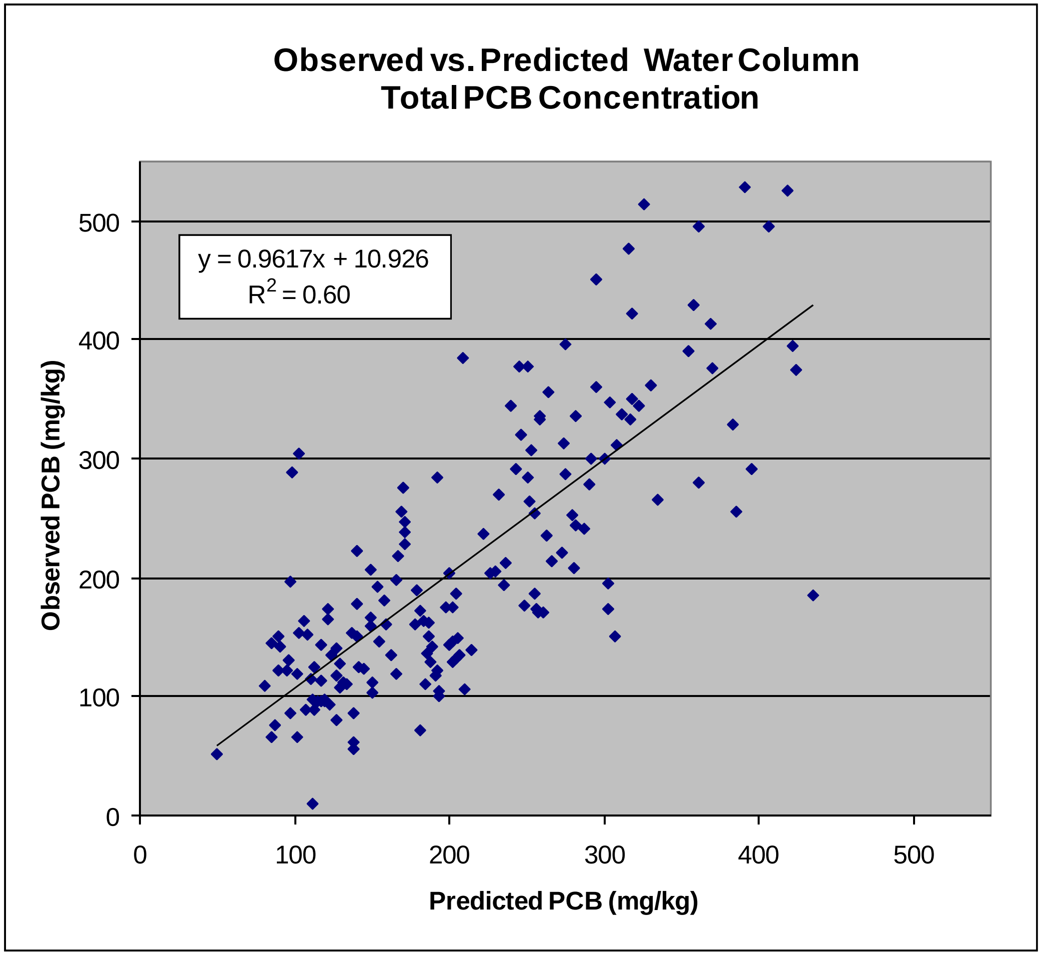

Figure 4. Observed water column Total PCB concentration plotted against modeled values for Thompson Island Pool based on the 127 day model with substitutions from the 166 day model for those days when predictor variables are missing—primarily Sundays, days when dredging was shut down and days prior to the onset of dredging.