

South River Statistical Modeling of Bank Stabilization Project

John W. Green, Ph.D., Ph.D.

Principal Consultant: Biostatistics

DuPont Applied Statistics Group

Models for South River

- **Reasons for revision**
 - Bulk sediment measurements were included before which greatly increased the variability of sediment mercury values, reducing sediment model responsiveness, possibly impacting surface water and fish models
- **Models revisions:**
 - Bulk sediment excluded
 - New surface water, sediment, & fish data collected, included
 - All calculations and modeling started over from beginning
 - Time period covered: 2001-2014
- **Surface water and fish tissue THg and MeHg and sediment THg models have been completed.**
 - Both more sensitive to bank stabilization project than before

Pre-Model Calculations

- **For each fish, surface water and sediment sample, the following calculations were made**
 - **Land use adjacent to nearest river bank**
 - **Total mercury (& other chemicals) within 1, 2, and 3 RRM upstream from each of the sources: sediment, riverbank (left, right, total), floodplain, trib, pore water**
 - **Rainfall & discharge rate within last 1-10 days**
 - **Environmental factors within 1-3 RRM upstream**

Fish Pre-Model Calculations

- **In addition to the above:**
 - **Dietary Thg and MeHg within 1, 2, and 3 RRM**
 - **Length of fish**
 - **Found earlier to be correlated with mercury bioaccumulation**
 - **Maximum discharge rate 6-24 months prior to sample**
 - **Previous work suggesting a time lag of up to 3 years between storms and fish mercury levels**
 - **Half-life of methyl mercury in fish tissue estimated 1- to 3-years**
- **In all models, interactions among these terms were included and seasonal adjustments were allowed**

General Methodology

- **Stepwise regression on large number of factors**
- **Data standardized prior to modeling**
- **Modified to keep variance inflation factors reasonable**
 - VIF generally below 5, always below 10
 - Eliminates collinearity, confounding
- **Some scientific judgment used when choosing which terms to use, which to eliminate, after stepwise**
- **Validity checks**
 - PCA, residual analysis
 - Multiple regression models, all pointing to same general conclusions (except fish MeHg)

Role of Statistical Models

- **Statistical models are driven by data**
 - Terms considered for model suggested by scientific understanding
 - Model terms are evaluated by their degree of association with the response being modeled and their effect on model stability
- **Statistical models can be good because**
 - No theoretical mechanistic model is force fit to the data
 - Statistical modeling attempts to evaluate all data for relevance
- **Statistical models can be bad because**
 - There may be no framework by which to explain the associations
- **A good statistical model offers credible, quantifiable associations and accounts for a substantial percentage of the variability observed**
- **Ideally, mechanistic and statistical models will agree, at least in broad outline, when both exist**

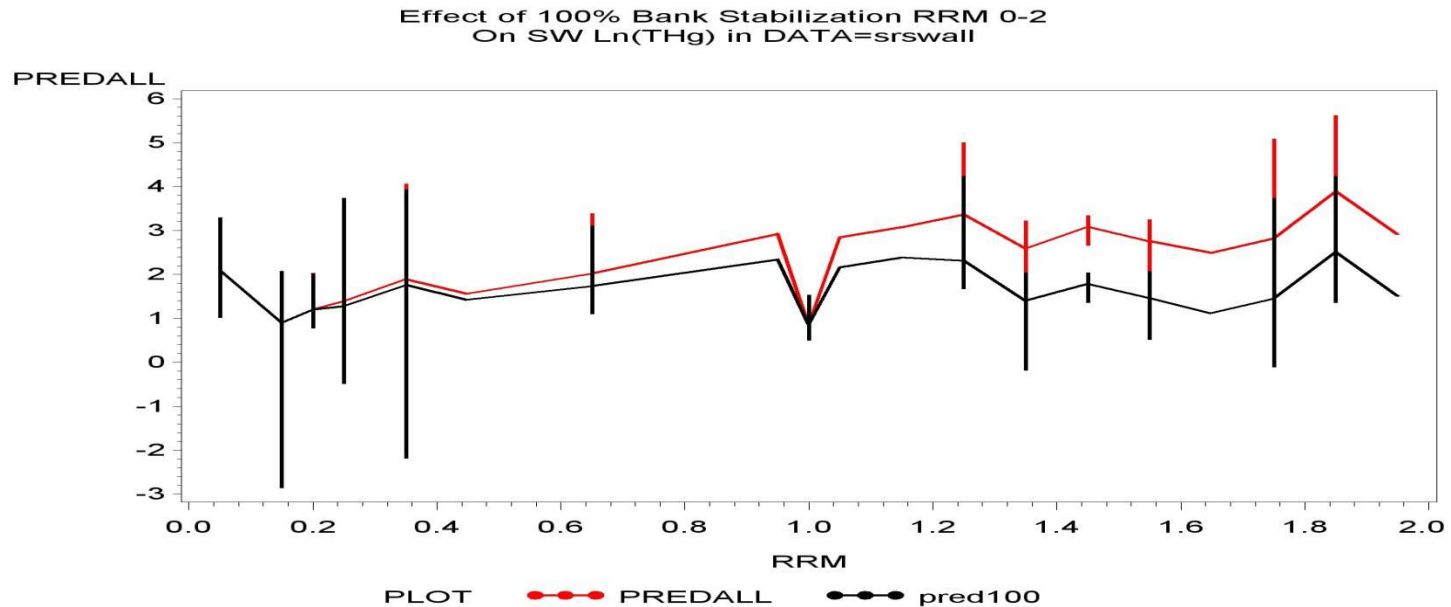
Basic Conclusions from Models

- Riverbank mercury loading extremely important predictor of SW (THg & MeHg), sediment (THg), and fish tissue mercury (THg & MeHg) levels
 - By itself (e.g., thru normal erosion) and associated with high discharge rates and rainfall
- No other factor *consistently* predicts greatly reduced Hg
- Floodplain Hg, associated with rainfall, also important
- Adjacent land use and trib Hg, river character (gradient, floodplain area) also important for sediment
 - These & sediment values, associated w/rainfall, important for SW
- Dietary items important for fish tissue Hg

Predicting Effect of Bank Stabilization

- **Start with model fit using all available data**
 - Estimate SW Hg as a function of RRM, month, etc.
- **Modify the bank Hg loading rate and bank erosion rates to reflect X% reduction**
 - X=100, 75, 50, 25
- **Apply same model equation to modified input**
 - Not as simple as it may sound: cumulative effects and interactions and standardizations all have to be modified but consistent with previous data
 - Estimate SW Hg at same location, time, etc. & compare

Surface Water THg Values



Generally, there is a reduction in SW THg in the region of the bank stabilization. The plot is on a natural log scale. The actual decrease at RRM 2 is predicted to be **15.4 ng/L**, from 23.5 to 8.1 (**65% REDUCTION**), with approximate 95% Conf. Bounds=(10,21).

Prediction assumes total removal (Pred100) of bank Hg loading to the river in 0-2 RRM stretch compared to no stabilization (PredAll).

SW THg Reduction with Partial Bank Stabilization in RRM 0-2

% Reduction	RRM	Predall	Predpct	Diff	Diff%	LCBADJ	UCBADJ
25	2.35	29	21	9	30	16	27
25	7.15	60	42	18	30	31	56
50	2.35	29	14	15	50	11	19
50	7.15	60	29	30	51	22	39
75	2.35	29	10	19	65	8	13
75	7.15	60	21	39	66	15	27
100	2.35	29	7	22	75	5	9
100	7.15	60	14	45	76	11	19

Predall is predicted SW THg value (ng/L) with no bank stabilization.

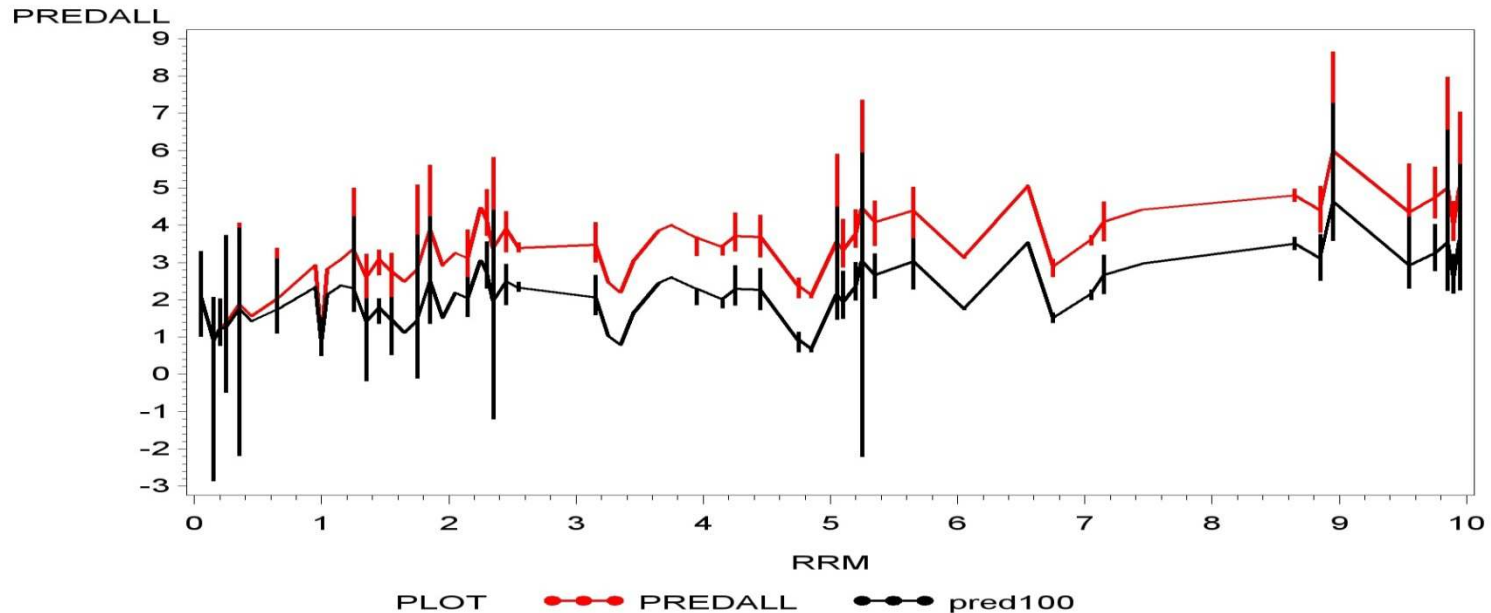
PredPct is predicted SW THg with bank stabilization.

LCBADJ and UCBADJ =95% confidence bounds on PredPct

%Reduction is the percent reduction of bank Hg entering river from RRM 0-2.

Predicted Impact on SW THg Continues Down River

Effect of 100% Bank Stabilization RRM 0-2
On SW Ln(THg) in DATA=srswall



Generally, there is a reduction in SW THg in the region of the bank stabilization. The plot is on a natural log scale. The actual decrease at RRM 2 is predicted to be **15.4 ng/L**, from 23.5 to 8.1 (**65% REDUCTION**), with approximate 95% Conf. Bounds=(10,21).

Plot assumes total removal (Pred100) of bank Hg loading to the river in 0-2 RRM stretch compared to no stabilization (PredAll).

Testing to Determine Whether a Decrease in SW THg has Occurred

- There are variations associated with seasons, land use, flooding frequency, mercury in the flood plain, storms, rainfall, discharge rate.
- To test for a decrease, sample in the
 - same location (RRM & distance from center of river)
 - same season
 - similar discharge rates (if possible)
 - non-storm times
 - non-rainfall times

How Many SW Samples?

- SW THg is log-normally distributed
 - Compute standard deviation from model for Ln(THg): Use $S=1.4^*$ for RRM 0-2
 - $$N = \frac{2*(1.96*S)^2}{Diff^2}$$
 - Where Diff= is the expected difference on Ln-scale
 - Example: At RRM=2.35, the predicted values are Ln(29)=3.4, Ln(7)=2.0
 - Diff=1.4, so N=11
 - This is the number of samples required to be 95% confident that the sample means (before and after bank stabilization effects are realized) will be *significantly* different.
- * From regression, but this value of S takes in many sources of variation. S=1.4 for sediment. Table of standard deviations for each measurement provided below.**

Reducing Sample Sizes

- Sample size = 11 takes into account variation in sampling month, land use, etc.
 - *This may be higher than necessary*, as it captures variation across years, discharge rates, rainfall, etc.. The more variables that are controlled, the smaller the required sample sizes.
- A reduction in sample size might suffice if sampling is restricted to periods where variability in results is low, e.g., July

Recommendation for SW Sampling

- **Collect 11 SW samples around RRM 2 during a one month period to establish a baseline.**
- **Compute standard deviation, S , of $\ln(\text{THg})$ values**
- **Compare this to value used in calculation (1.4).**
 - **If $S \gg 1.4$ or $S \ll 1.4$, revise sampling size for subsequent years**
 - **In any case, compare sample means at same month under similar conditions (discharge rate, etc.) across years**

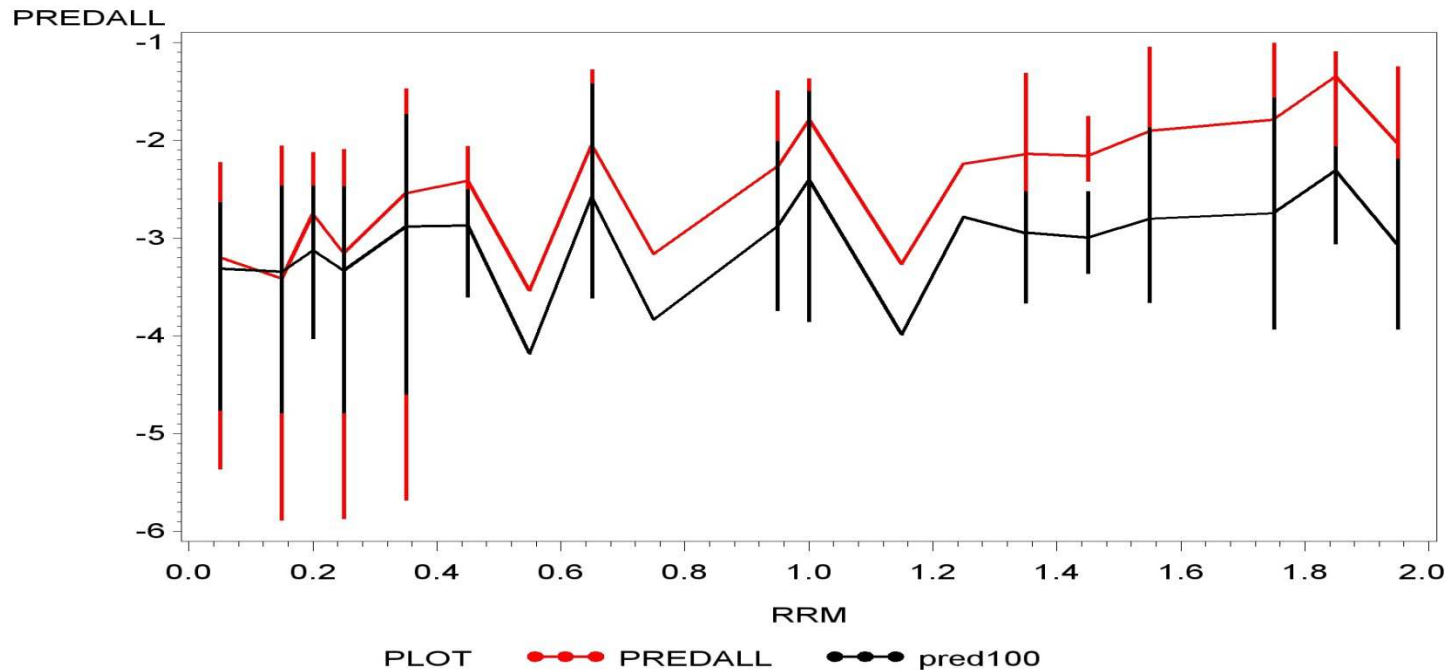
Standard Deviations to Use for Sample Size Calculations

Response	SWTHg	SWMeHg	SEDTHg	FishTHg	FishMeHg
STD	1.4	1.2	1.25	0.67	0.74

These values are from the regression models fit to the natural logarithms of the indicated responses.

Surface Water MeHg Values

Effect of 100% Bank Stabilization RRM 0-2
On SW Ln(MeHg) in DATA=srswall



Generally, there is a reduction in SW MeHg in the region of the bank stabilization. The plot is on a natural log scale. The predicted decrease at RRM 2 is approximately **22 ng/L**, from 29 to 7 (**75% REDUCTION**).

Prediction assumes total removal (Pred100) of bank Hg loading to the river in 0-2 RRM stretch compared to no stabilization (PredAll).

SW MeHg Reduction with Partial Bank Stabilization in RRM 0-2

% Reduction	RRM	Predall	Predpct	Diff	Diff%	LCBADJ	UCBADJ
25	2.35	0.22	0.14	0.08	35	0.11	0.18
25	7.15	0.43	0.19	0.24	56	0.15	0.24
50	2.35	0.22	0.10	0.12	54	0.08	0.13
50	7.15	0.43	0.11	0.32	74	0.09	0.14
75	2.35	0.22	0.07	0.14	67	0.06	0.09
75	7.15	0.43	0.07	0.36	84	0.06	0.09
100	2.35	0.22	0.05	0.16	76	0.04	0.07
100	7.15	0.43	0.04	0.39	90	0.03	0.05

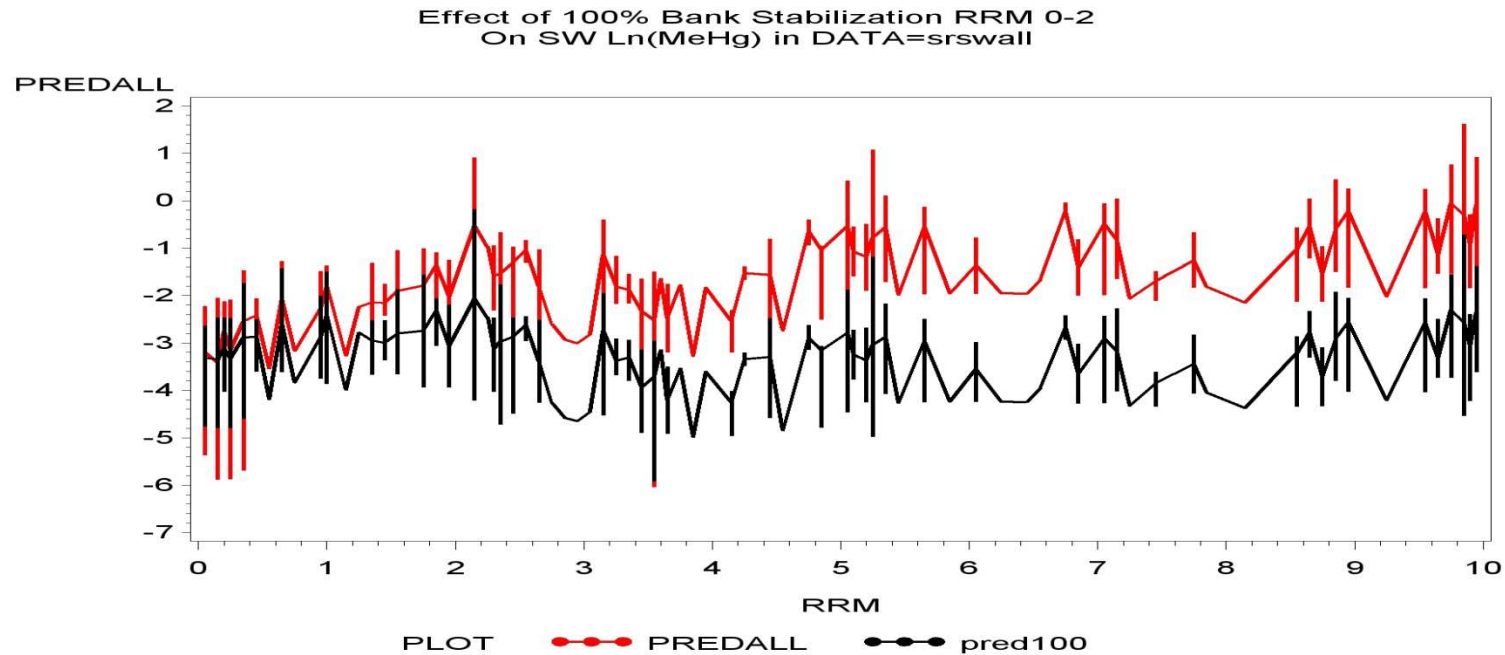
Predall is predicted SW MeHg value (ng/L) with no bank stabilization.

PredPct is predicted SW MeHg with bank stabilization.

LCBADJ and **UCBADJ** =95% confidence bounds on **PredPct**

%Reduction is the percent reduction of bank Hg entering river from RRM 0-2.

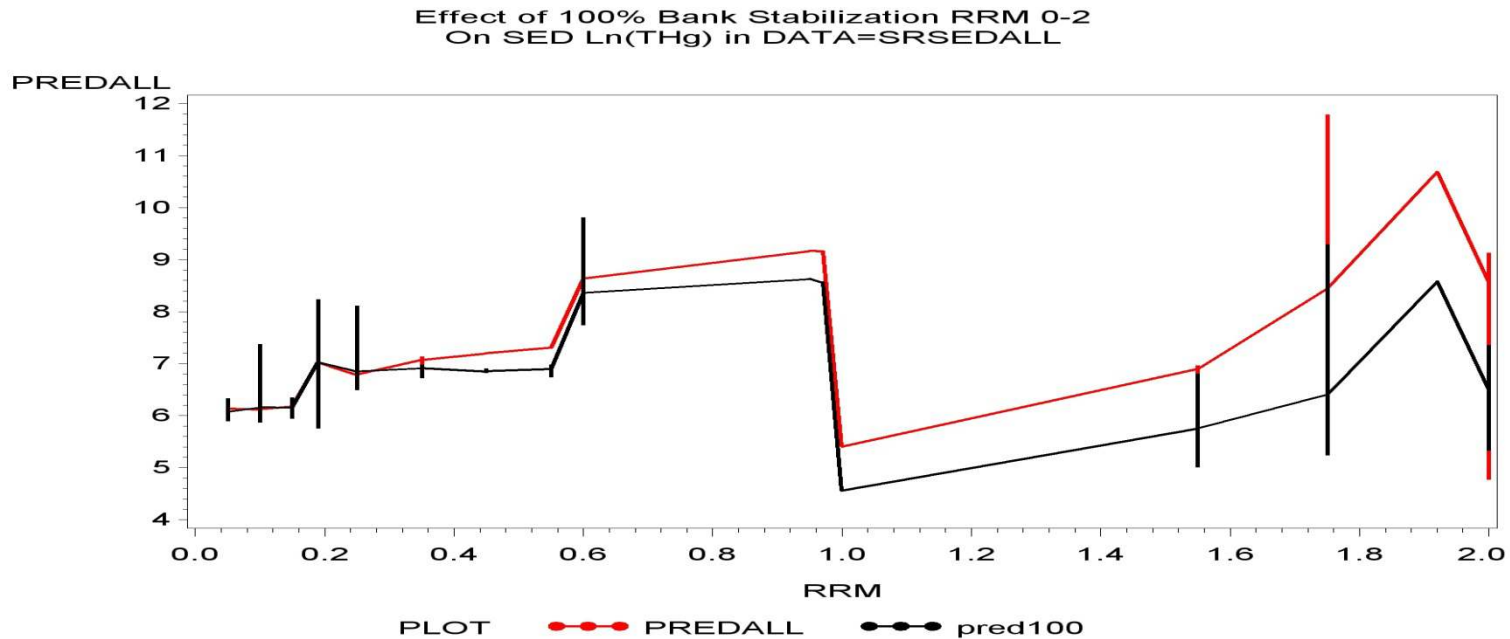
Predicted Impact on SW MeHg Continues Down River



Generally, there is a reduction in SW MeHg in the first 25 RRM's below the plant site. The plot is on a natural log scale. The actual predicted decrease at RRM 10 is approximately **0.39 ng/L**, from 0.43 to 0.04 (**90% REDUCTION**).

Prediction assumes total removal (Pred100) of bank Hg loading to the river in 0-2 RRM stretch compared to no stabilization (PredAll).

Sediment THg Reduction



Generally, there is a reduction in SED THg in the region of the bank stabilization. The plot is on a natural log scale. The actual predicted decrease at RRM 2 is approximately **87%**, from **5160** to **666** ug/kg.

Prediction assumes total removal (Pred100) of bank Hg loading to the river in 0-2 RRM stretch compared to no stabilization (PredAll).

SED THg Reduction with Partial Bank Stabilization in RRM 0-2

%Reduction	RRM	Predall	Predpct	Diff	Diff%	LCBADJ	UCBADJ
25	2	5160	2017	3143	61	1213	3356
	8.7	17612	15367	2245	13	11233	21021
50	2	5160	1025	4135	80	616	1705
	8.7	17612	9338	8273	47	6827	12774
75	2	17612	10227	7385	42	7476	13990
	8.7	1531	575	957	62	318	1037
100	2	5160	666	4493	87	401	1109
	8.7	17612	8207	9405	53	5999	11226

Predall is predicted SED THg value (ug/kg) with no bank stabilization.

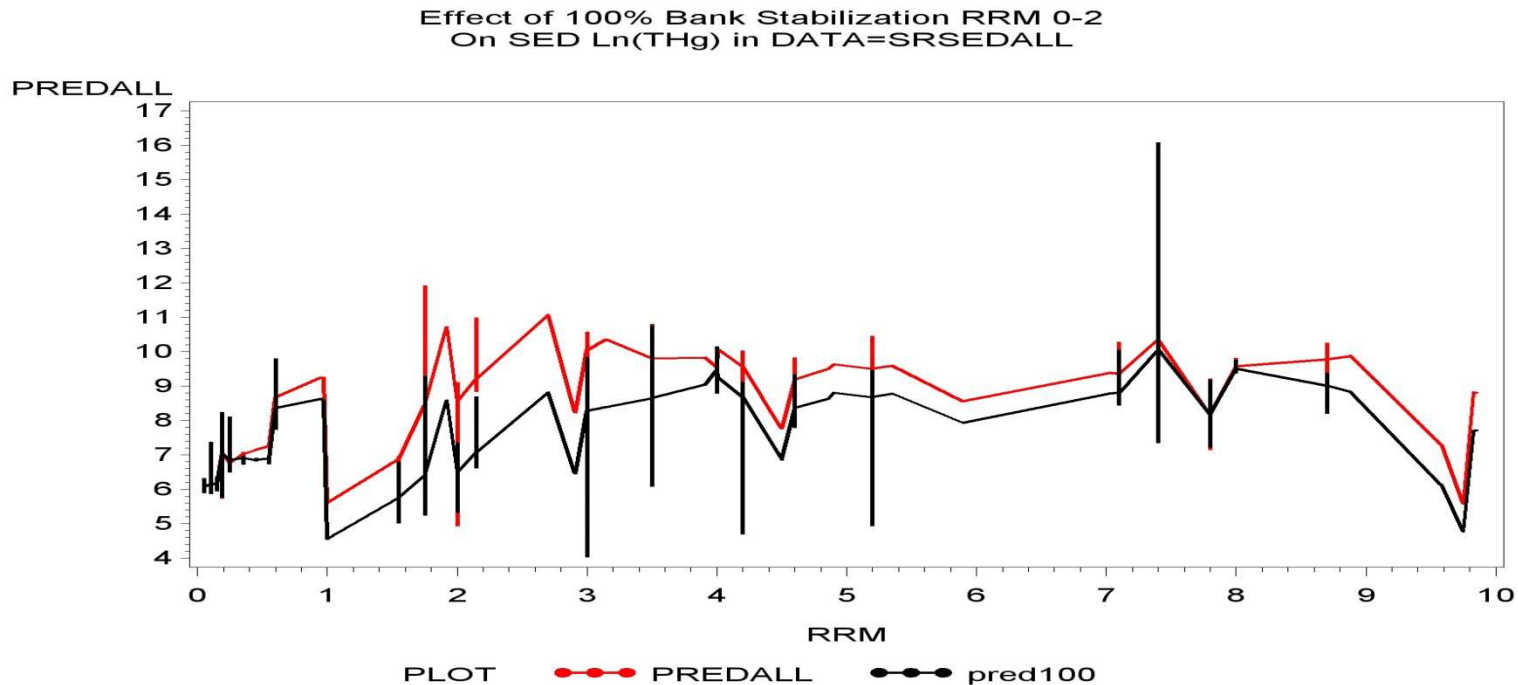
PredPct is predicted SED THg with bank stabilization.

%Reduction is the percent reduction of bank Hg entering river from RRM 0-2.

LCBADJ and UCBADJ are 95% Conf.Bnds on PredPct.

Sample size N=21 adequate to find significant decrease at RRM 8.7 @ 100% reduction

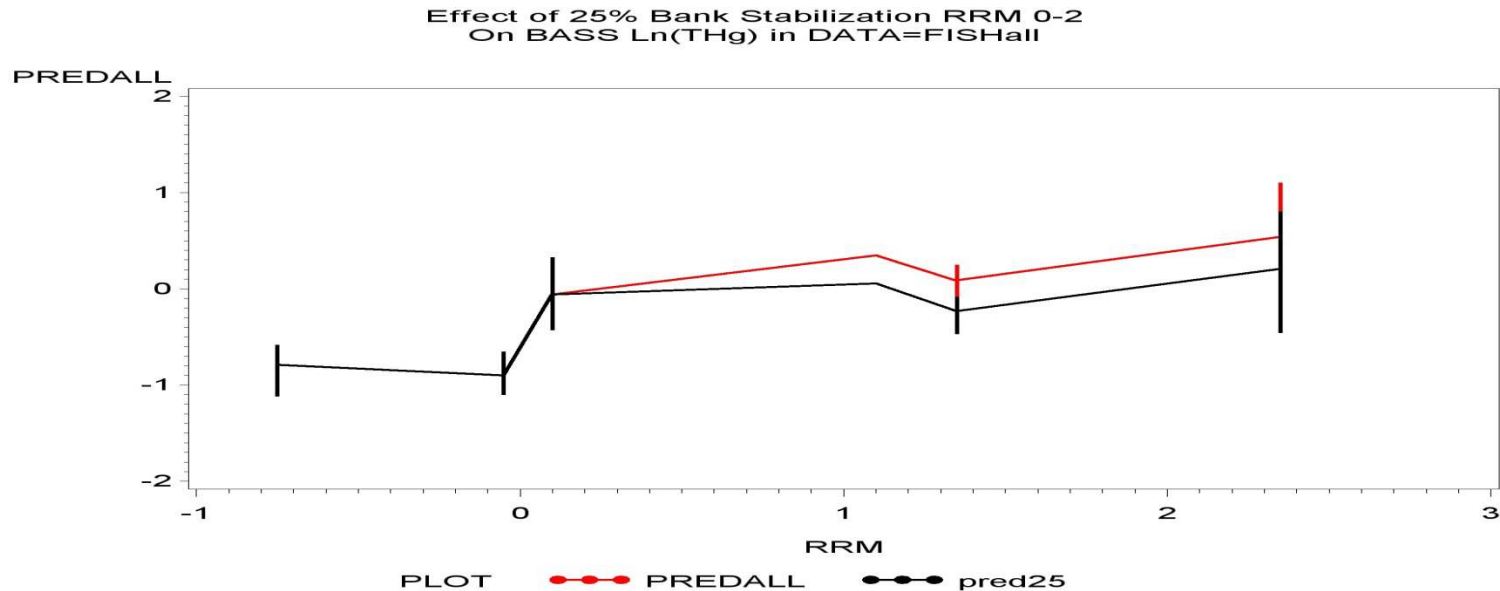
Predicted Impact on Sed THg Continues Down River



At RRM, the model predicts an approximate 58% reduction in sediment THg, from 17,000 to 7000 ug/kg.

As before, this figure assumes 100% reduction in bank loading in RRM 0-2. The effects are smaller if a smaller reduction in bank loading occurs.

Fish THg Reduction



Plot is not as helpful as it might be. There were only three locations for fish samples between RRM 0 and 2.35. The figure following shows the predicted large decrease at 2.35 does not persist.

Fish THg Reduction with Partial Bank Stabilization in RRM 0-2

%Reduction	RRM	Predall	Predpct	Diff	Diff%	LCBADJ	UCBADJ
25	2.35	1.71	1.23	0.48	28	0.94	1.61
	11.8	3.32	2.96	0.37	11	2.49	3.51
50	2.35	1.71	1.05	0.67	39	0.80	1.37
	11.8	3.32	2.63	0.69	21	2.22	3.12
75	2.35	1.71	0.92	0.80	47	0.70	1.20
	11.8	3.32	2.34	0.98	30	1.98	2.78
100	2.35	1.71	0.80	0.91	53	0.61	1.05
	11.8	3.32	2.08	1.24	37	1.76	2.47

Predall is predicted bass THg value (mg/kg) with no bank stabilization.

PredPct is predicted bass THg with bank stabilization.

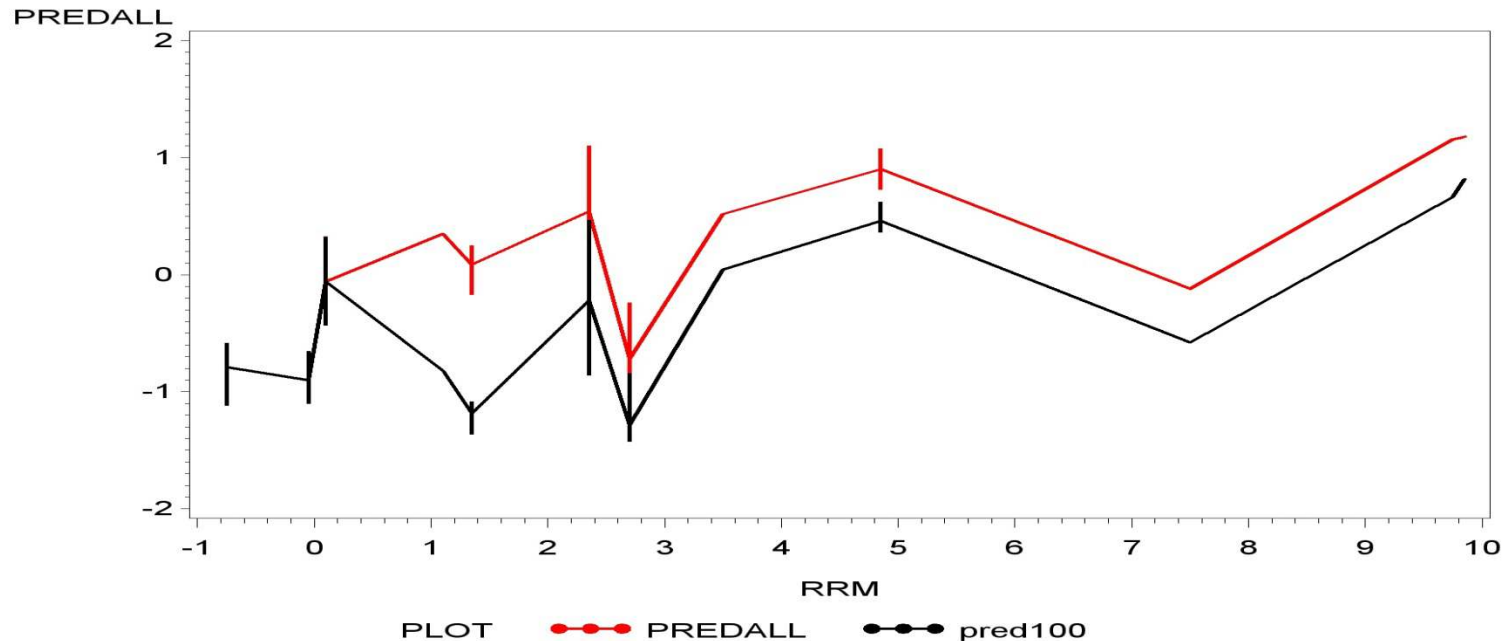
%Reduction is the percent reduction of bank Hg entering river from RRM 0-2.

LCBADJ and UCBADJ are 95% Conf.Bnds on PredPct.

Sample size N=15 to find significant decrease at RRM 11.8 @ 100% reduction

Fish THg Reduction Continues Downstream

Effect of 100% Bank Stabilization RRM 0-2
On BASS Ln(THg) in DATA=FISHall



At RRM 11.8, the model predicts a 37% decrease in fish tissue THg after bank stabilization effects are fully realized. The same % decrease persists to RRM 23.5.

Note: Plot shows THg on a log scale, which is how the data were modeled.

Percent Fish w/ THg<0.3 mg/kg

			Model Predicted PCT Fish THg<0.3				
Reach	#FISH	OBS%	BankRed% 0	BankRed% 25	BankRed% 50	BankRed% 75	BankRed% 100
0	85	45	30	30	30	30	30
1	210	6	5	6	7	9	12
2	96	22	12	16	20	25	29
3	54	0	0	0	1	1	1
4	179	0	0	0	0	0	0
5	178	10	2	3	4	5	6
6	101	2	0	0	0	1	1

Percent fish currently w/ THg level<0.3 mg/kg (OBS%) underestimated several reaches.

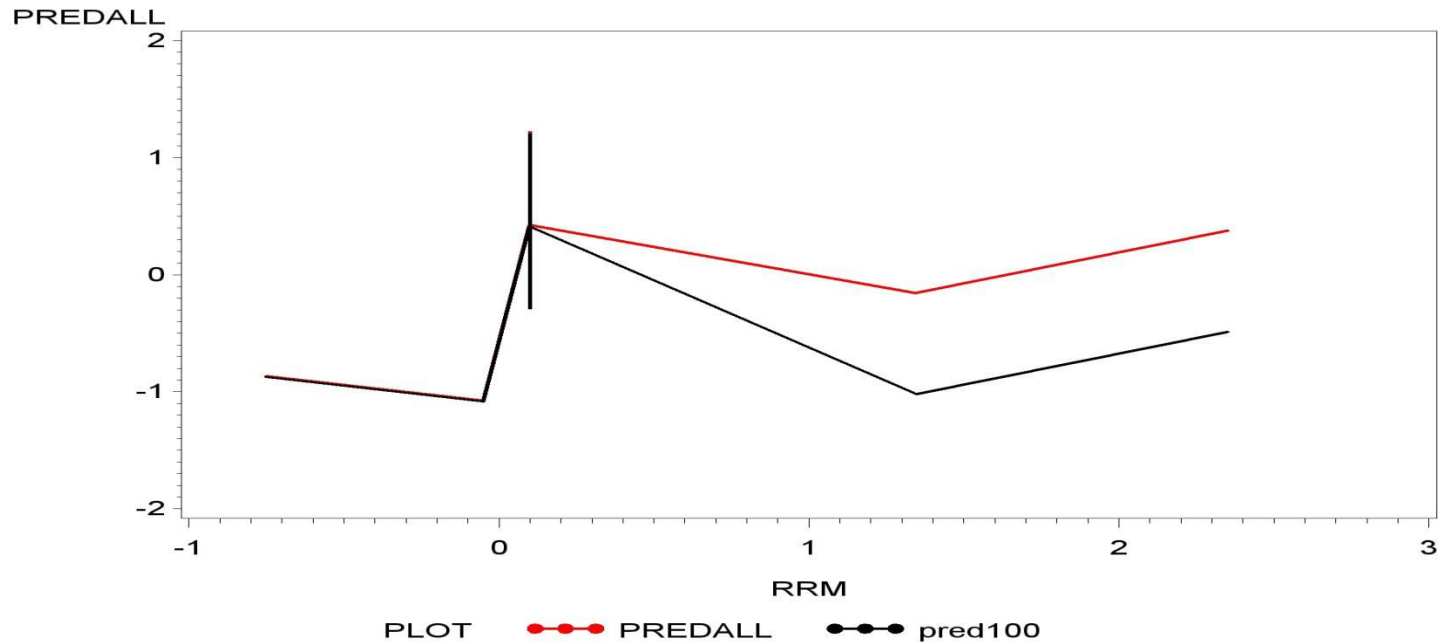
Percent in reaches 1, 2 and 5 predicted to fall below 0.3 mg/kg to more than double under 100 % reduction in river bank loading in RRM0-2.

Other reaches predicted to stay approximately the same.

Model do not predict change in THg level above plant site.

Fish MeHg Reduction

Effect of 100% Bank Stabilization RRM 0-2
On BASS Ln(MeHg) in DATA=FISHall



Plot is not as helpful as it might be. There were only two locations for fish samples between RRM 0 and 2.35. The figure following shows the predicted huge decrease at 2.35 does not persist.

NOTE: There is greater uncertainty about the model for fish MeHg than for any other response.

Fish MeHg Reduction with Partial Bank Stabilization in RRM 0-2

%Reduction	RRM	Predall	Predpct	Diff	Diff%	LCBADJ	UCBADJ
25	2.35	1.46	1.17	0.29	20	0.45	3.01
	11.8	3.71	3.25	0.47	13	2.66	3.97
50	2.35	1.46	0.94	0.51	35	0.37	2.43
	11.8	3.71	2.87	0.85	23	2.35	3.51
75	2.35	1.46	0.76	0.70	48	0.29	1.96
	11.8	3.71	2.53	1.18	32	2.07	3.10
100	2.35	1.46	0.61	0.84	58	0.24	1.58
	11.8	3.71	2.24	1.48	40	1.83	2.74

Predall is predicted bass MeHg value (ng/g) with no bank stabilization.

PredPct is predicted bass THg with bank stabilization.

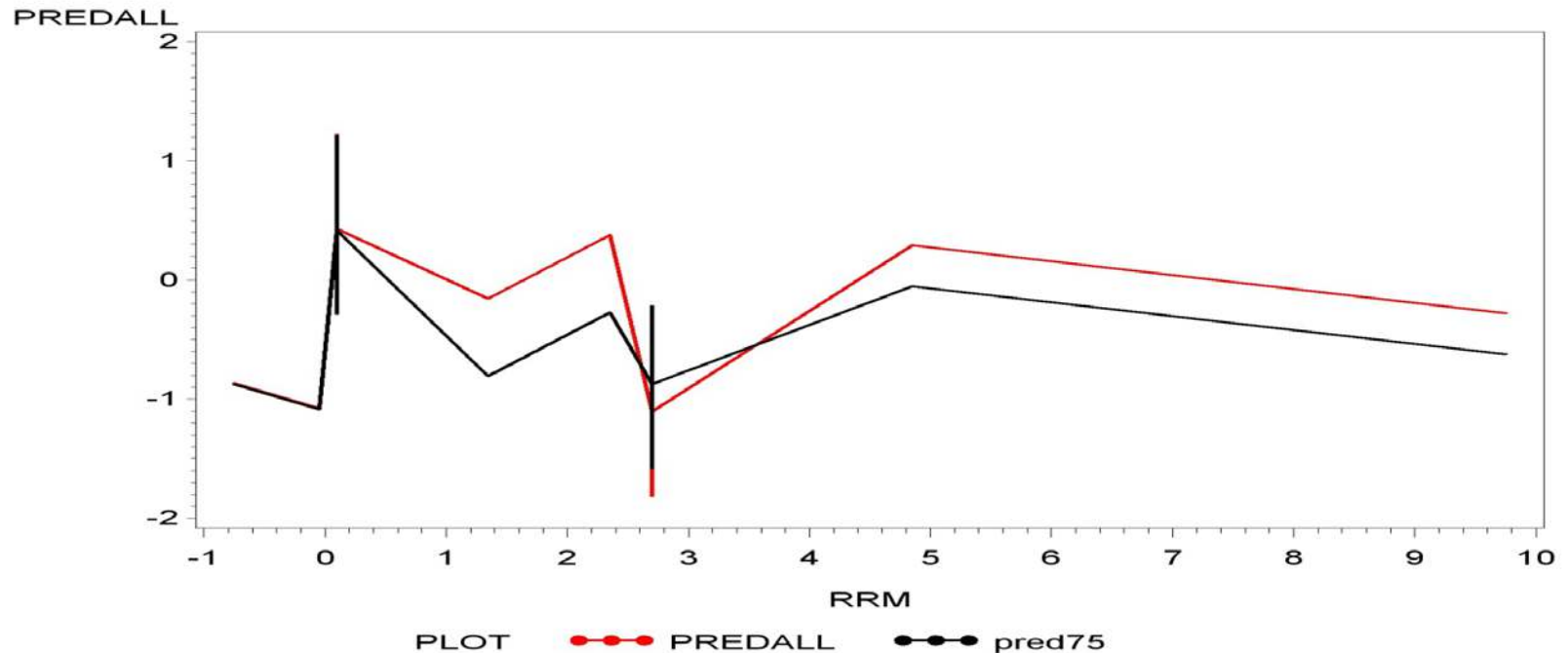
%Reduction is the percent reduction of bank Hg entering river from RRM 0-2.

LCBADJ and UCBADJ are 95% Conf.Bnds on PredPct.

Sample size N=14 to find significant decrease at RRM 11.8 @ 100% reduction

Reduction in Fish MeHg Reduction Continues Downstream

Effect of 75% Bank Stabilization RRM 0-2
On BASS Ln(MeHg) in DATA=FISHall



At RRM 11.8, predicted bass MeHg level is reduced 40% after bank stabilization effect is fully realized. Plot shows MeHg on a natural log scale.

MODEL SUMMARIES

SW THg Model

- **$R^2=0.54$, 1688 observations**
- **Bank loading enters as total load up-river and right-bank load within 1 RRM upriver**
- **Other important factors are rainfall & floodplain Hg including their interactions, seasonal variation, sediment and tributary Hg & sediment chemistry & composition**

SW MeHg Model

- $R^2=0.47$, 2184 observations
- Bank loading enters as total right bank THg upriver & interaction of rainfall within 5 days and right bank THg within 3 RRM upstream
- Other factors include sediment and trib MeHg & THg & other chemistry plus their interactions with discharge rate, floodplain & seasonality

Sediment THg Model

- $R^2=0.64$, 375 observations
- Bank loading enters as right bank load within 2 RRM upstream & interaction of total right bank and discharge rate within 5 days prior
- Other factors include river gradient & floodplain area, Trib THg, floodplain use, seasonality

Bass THg Model

- $R^2=0.56$, 903 observations
- Bank loading enters as total loading rate within 1 RRM and total upstream & interaction with maximum discharge rate within 1-3 years
- Other factors include sediment, trib & porewater Hg, aquatic vegetation, food items, rainfall, discharge rate, floodplain use, seasonality

Bass MeHg Model

- $R^2=0.62$, 286 observations
- Bank loading enters as right bank load within 1 RRM and total bank load within 2 RRM upstream & interactions with maximum discharge rate within 1 year prior
- Other factors include fish length, maximum discharge rate within 18 months prior

Conclusions

- The models for surface water, sediment, and fish THg and MeHg predict significant decreases will be observed with feasible sample sizes.
- These models do NOT indicate how much time will be required before these decreases will be observed.

Conclusions 2

- These predicted effects assume the models are correct.
- Confidence bounds do not capture all the uncertainty in the models. They do not capture model misspecification or other aspects of model uncertainty. They reflect only the uncertainty in the effects *if* the models are correct.