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**
- **Note: Only surface water and sediment THg models have been completed.**
  - **Both more sensitive to bank stabilization project than before**
- **Fish models pending**

Pre-Model Calculations

- **For each 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**
  - **Max discharge rate within last 6, 12, …, 36 months**

General Methodology

- **Stepwise regression on large number of factors**
- **Data standardized prior to modeling**
  - **main effects standardized to mean 10, std. 1**
- **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**
- **Validity checks**
  - **PCA, residual analysis**
  - **Multiple regression models, all pointing to same general conclusions**

Basic Observations on Models

- **The river is a dynamic system**
  - **Surface water, sediment, floodplain, discharge, rainfall, porewater, banks, *etc.*, interact**
- **Sediment & bank Hg, discharge rates, other terms, used to predict SW and fish Hg**
  - **Models predict SW and sediment THg much reduced if bank Hg terms omitted**
- **Bank Hg, discharge rates, other factors, used to predict sediment Hg**

Model Development

- **Consistent and coherent models are developed for SW & SED THg & MeHg, and FISH & other biota THg *from all available data***
- **Surface water, sediment, and floodplain mercury are basic elements of all models for biota**
- **River is dynamic system, with surface water, sediment, floodplain, discharge, rainfall, porewater, *etc.* all interacting**
- **Surface water mercury values added to sediment models to account for possible indirect effects of bank mercury**

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

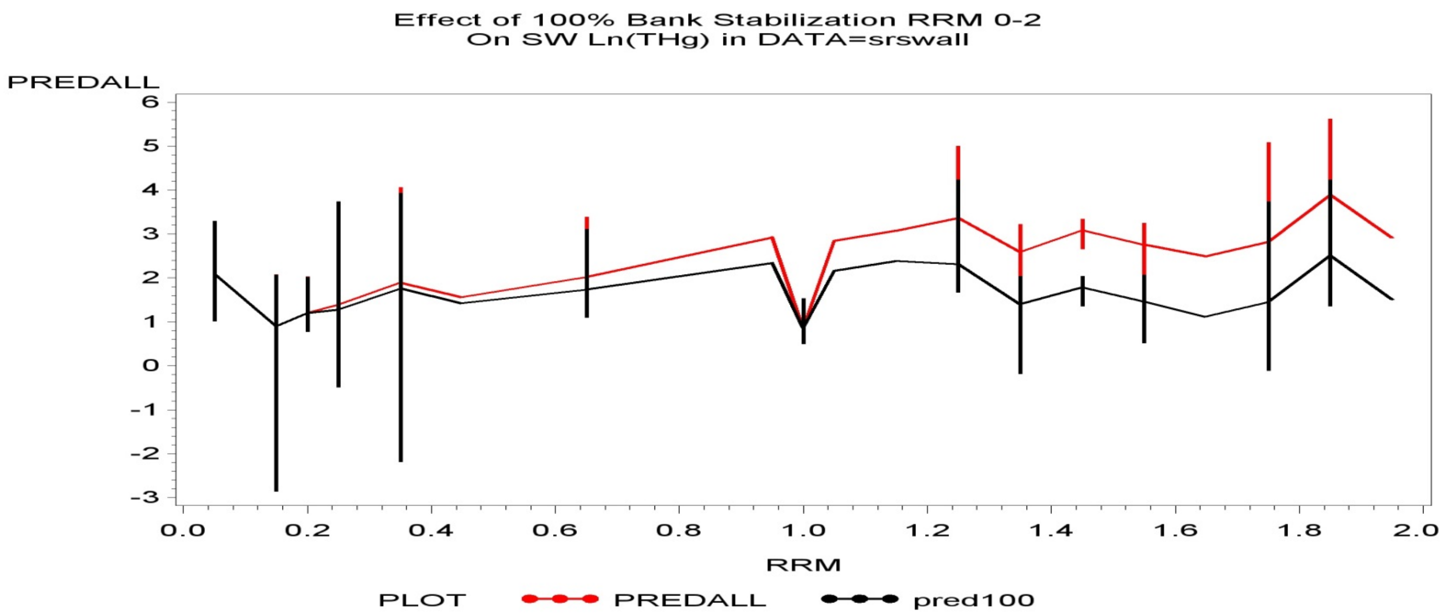
Basic Conclusions from Models

- **Riverbank mercury loading extremely important predictor of SW (THg & MeHg), sediment (THg & MeHg), and (presumably) fish tissue (THg) mercury 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 landuse 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**
  - **This remains to be verified under new model**

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 (Unfiltered)



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).

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=1.2** * **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, the predicted values are Ln(31)=3.16, Ln(14)=2.09**
  - **Diff=1.07, so N=10**
  - **This is the number of samples required to be 95% confident that the sample means will be *significantly* different.**

*** From regression, but this value of *S* takes in many sources of variation. *S*=1.12 for sediment.**

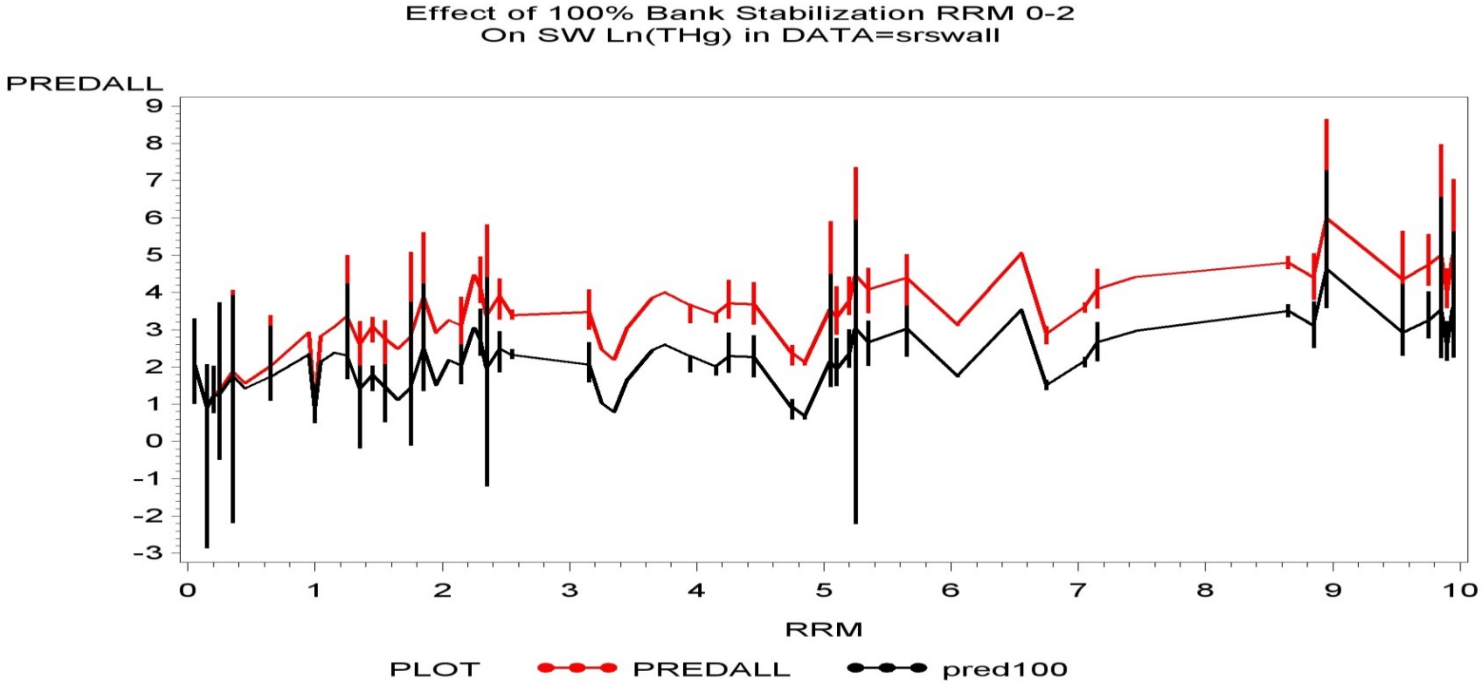
Reducing Sample Sizes

- **Sample size = 10 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 10 SW samples around RRM 2 during a one month period to establish a baseline.**
- **Compute standard deviation, *S*, of Ln(THg) values**
- **Compare this to value used in calculation (0.46).**
  - **If S >> 0.46 or S <<0.46, revise sampling size for subsequent years**
  - **In any case, compare sample means at same month under similar conditions (discharge rate, etc.) across years**

Predicted Impact Continues Down River



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 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).

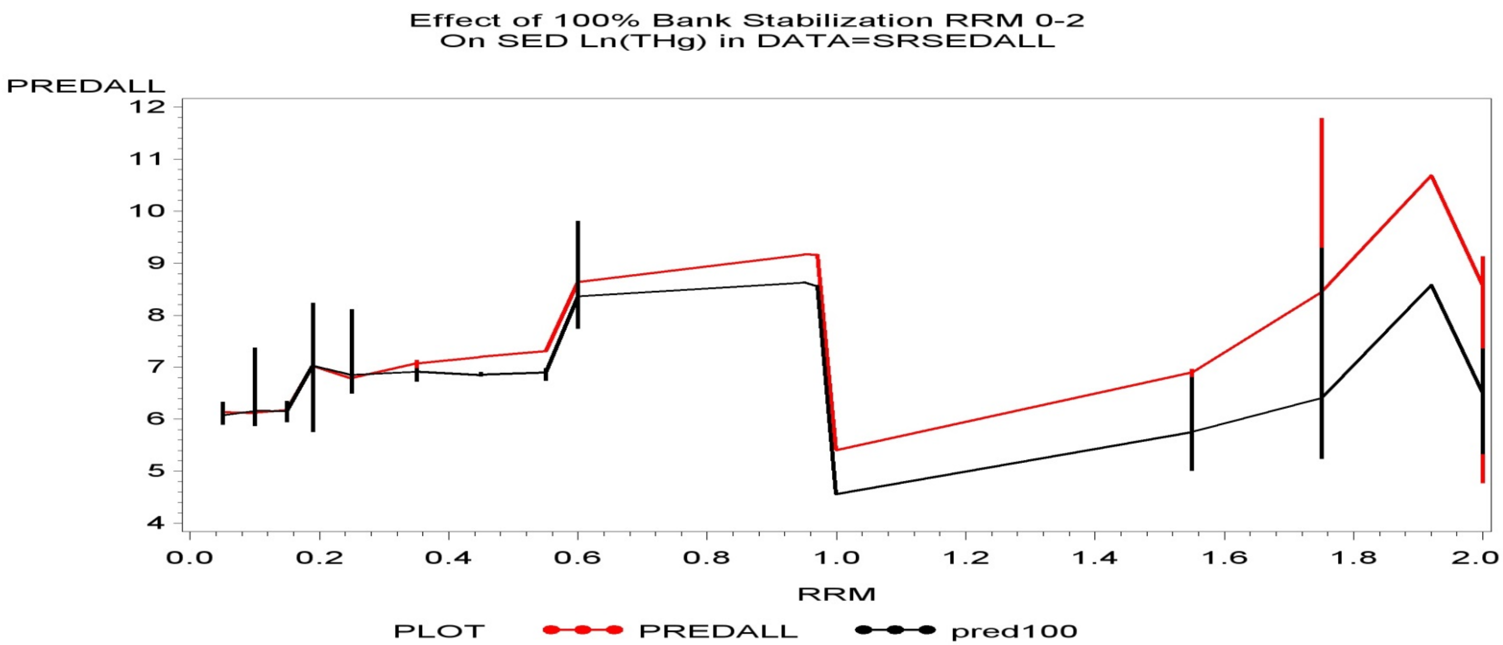
Prediction assumes total removal (Pred100) of bank Hg loading to the river in 0-2 RRM stretch compared to no stabilization (PredAll).

SW THg Reduction with Partial Stabilization in RRM 0-2

% Reduction	RRM	Logn Predall	Logn PredRed	**Predall**	**PredRed**	**Diff**	LCB	UCB
25	2							
25	10							
50	2	3.16	2.57	**23.5**	**13.1**	**10.4**	4.9	16.7
50	10	4.55	3.85	**94.9**	**46.9**	**48**	12.3	76.9
75	2							
75	10							
100	2	3.16	2.09	**23.5**	**8.1**	**15.4**	9.8	21
100	10	4.55	3.14	**94.9**	**23.1**	**71.8**	42.9	100.6

Predall is predicted SW THg value (ng/L) with no bank stabilization.
PredRed is predicted SW THg with bank stabilization.
%Reduction is the percent reduction of bank Hg entering river from RRM 0-2.

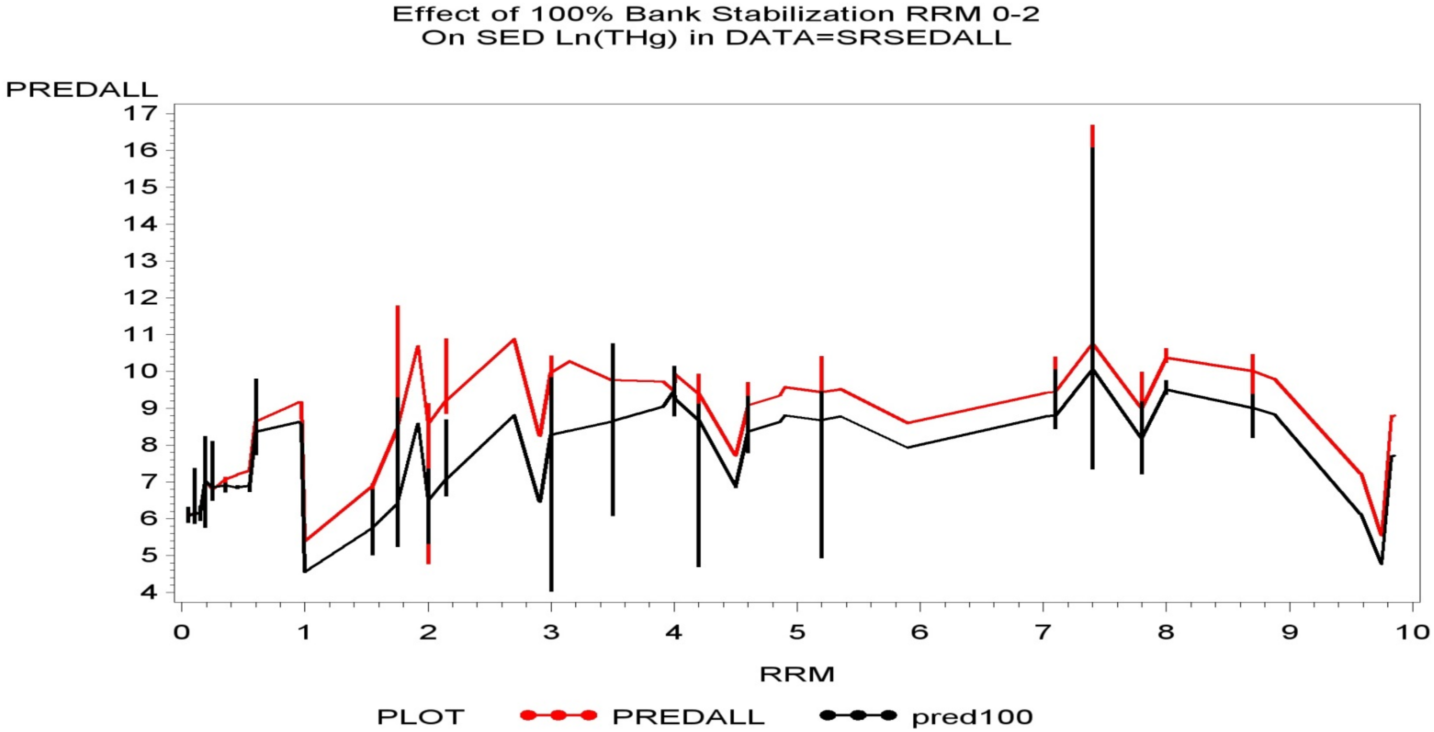
Sediment THg Reduction



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 decrease at RRM 2 is predicted to be 5850 ug/kg, from 6602 to 752 (89% REDUCTION), with approximate 95% Conf. Bounds=(2579,9099).

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 RRM 0-10



The effect on sediment THg persists down stream.

SED THg Reduction with Partial Stabilization in RRM 0-2

%Reduction	RRM	Logn Predall	Logn PredRed	**Predall**	**PredRed**	**Diff**	LCB	UCB
25	2							
	10							
50	2	8.8	7.02	**6602**	**1117**	**5485**	2230	8762
	10	8.8	7.71	**6675**	**2241**	**4434**	1787	8080
75	2							
	10							
100	2	8.8	6.62	**6602**	**752**	**5850**	2579	9099
	10	8.8	7.71	**6675**	**2241**	**4434**	1787	8080

Predall is predicted SED THg value (ug/kg) with no bank stabilization.
PredRed is predi10cted SED THg with bank stabilization.
%Reduction is the percent reduction of bank Hg entering river from RRM 0-2.
LCB and UCB are 95% Conf.Bnds on Diff.
Sample size N=7 adequate to find significant decrease at RRM 10.

Conclusion

- **The models for surface water and sediment THg 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.**