Application of Risk Assessment Outcomes to Remedial Decisions: Importance of Natural Recovery Rate Estimates in Sediment

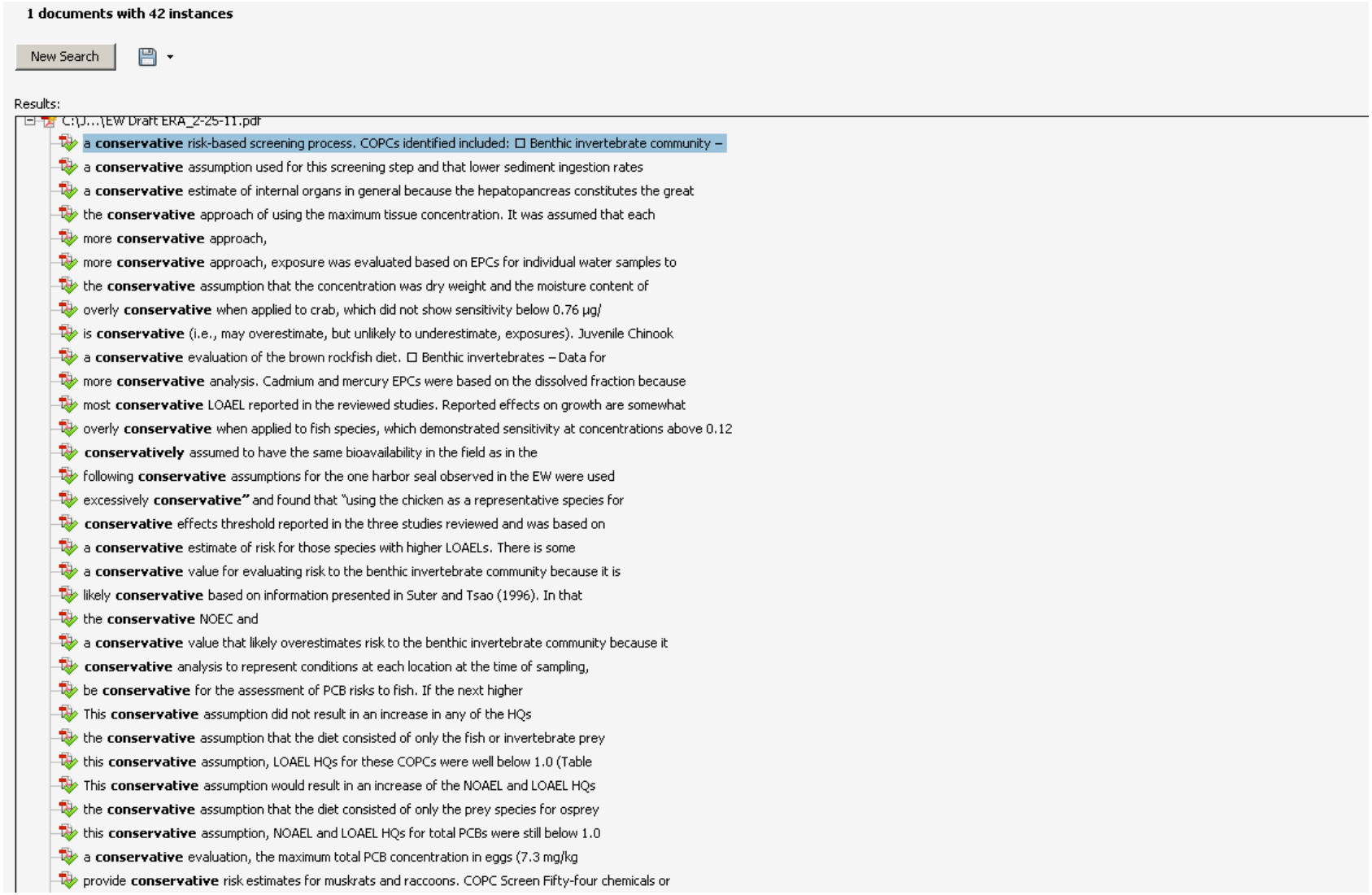
Jay Field
NOAA Office of Response and Restoration, Seattle, Washington
John Kern
Kern Statistical Services, Inc., Sauk Rapids, Minnesota
Lisa Rosman
NOAA Office of Response and Restoration, New York City, New York

SETAC 2011
Boston, MA
November 16, 2011

Evaluating Risk Assessments and Remediation Decisions

- Remedial decision-making requires an understanding of the risk associated with current conditions (baseline) and the projected changes in risk for an array of remedial alternatives
- For aquatic sites, comparative evaluation of risk/benefits of different remedial alternatives often involves natural recovery models for sediment
- So, how do you know when the risk assessment or comparison of alternatives is conservative?

How Do You Know If Risk Assessment is Conservative?

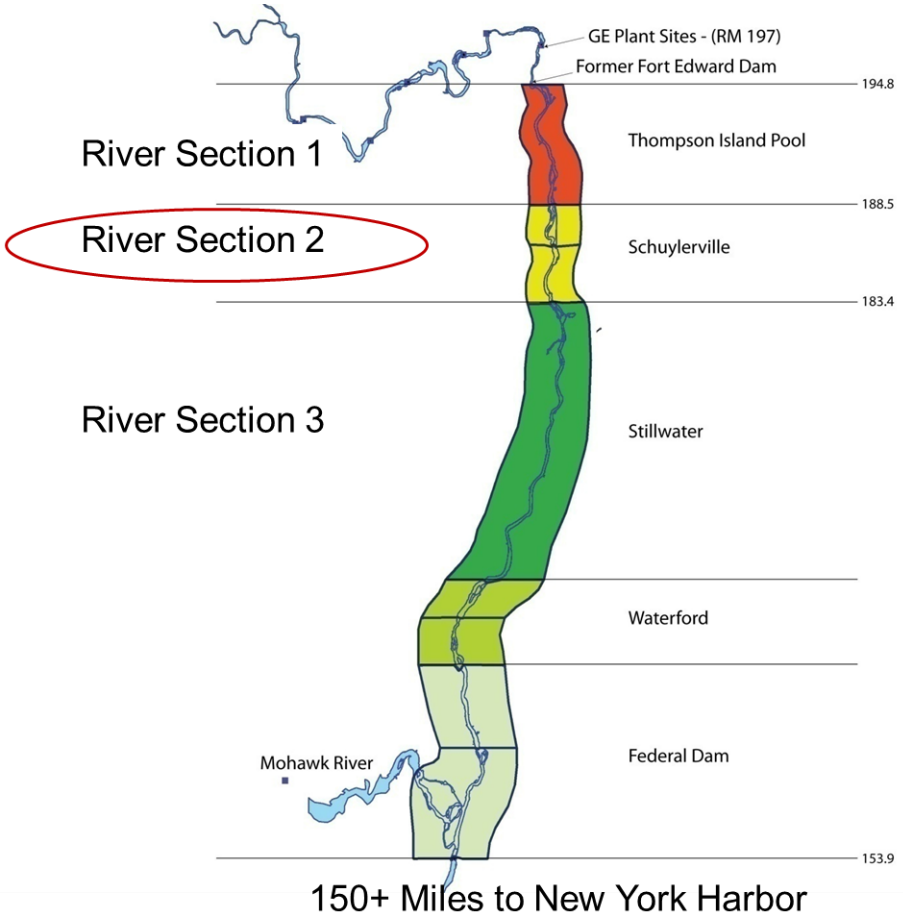


What I'm going to talk about

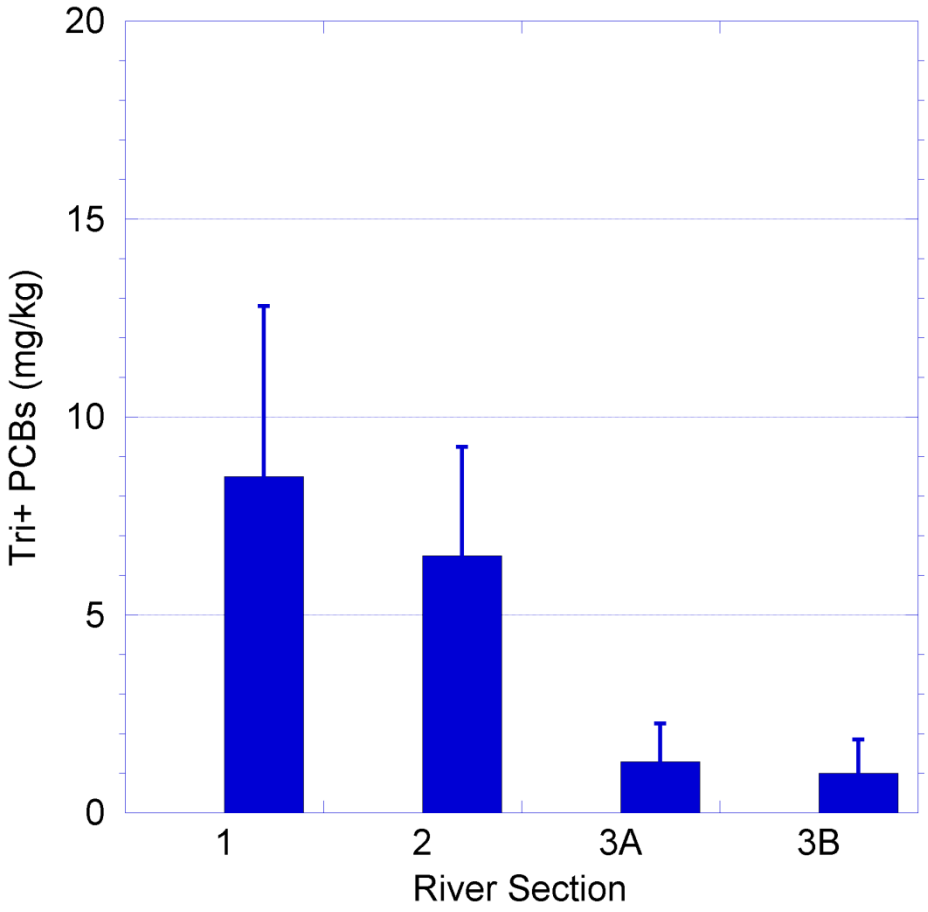
- Hudson River example: Natural recovery models played an important role in the comparative evaluation of the remedial alternatives
  - The implications of overestimating rate of natural recovery for decision-making
- Data trumping/10-foot rule approach commonly used to replace older data from a baseline sediment database or provide information on natural recovery
  - The implications of biased re-sampling to support assumptions of natural recovery

Model Predictions: Surface Sediment Concentrations

- Remedial Alternatives Comparison
  - Modeled surface Tri+ PCBs
    - Pre-Dredging (2003)
    - Post-Dredging (2008-2010)
- Remedial Design sediment cores collected from the Upper Hudson River (UHR) 2002 to 2007
  - Systematic sampling grid used to characterize cohesive sediment throughout the UHR
  - Arithmetic average surface Tri+ PCB in top 5 cm (n=8167).

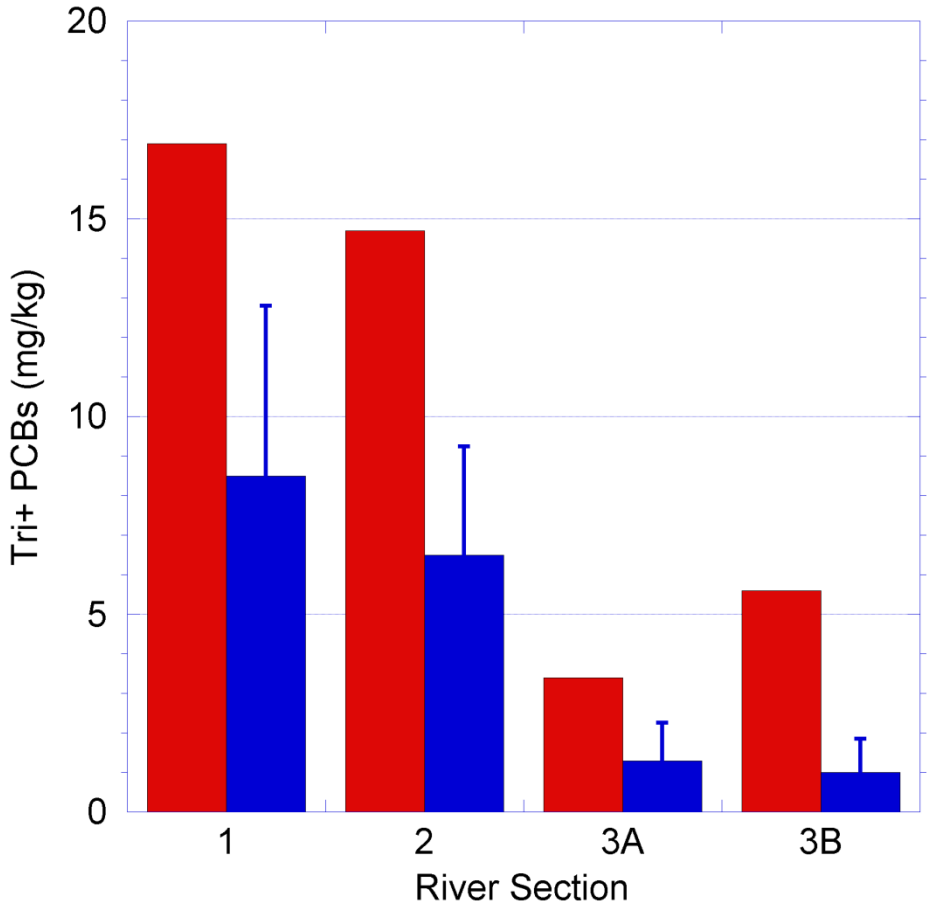


Model Predictions: Pre-Dredging Surface Sediment Concentrations



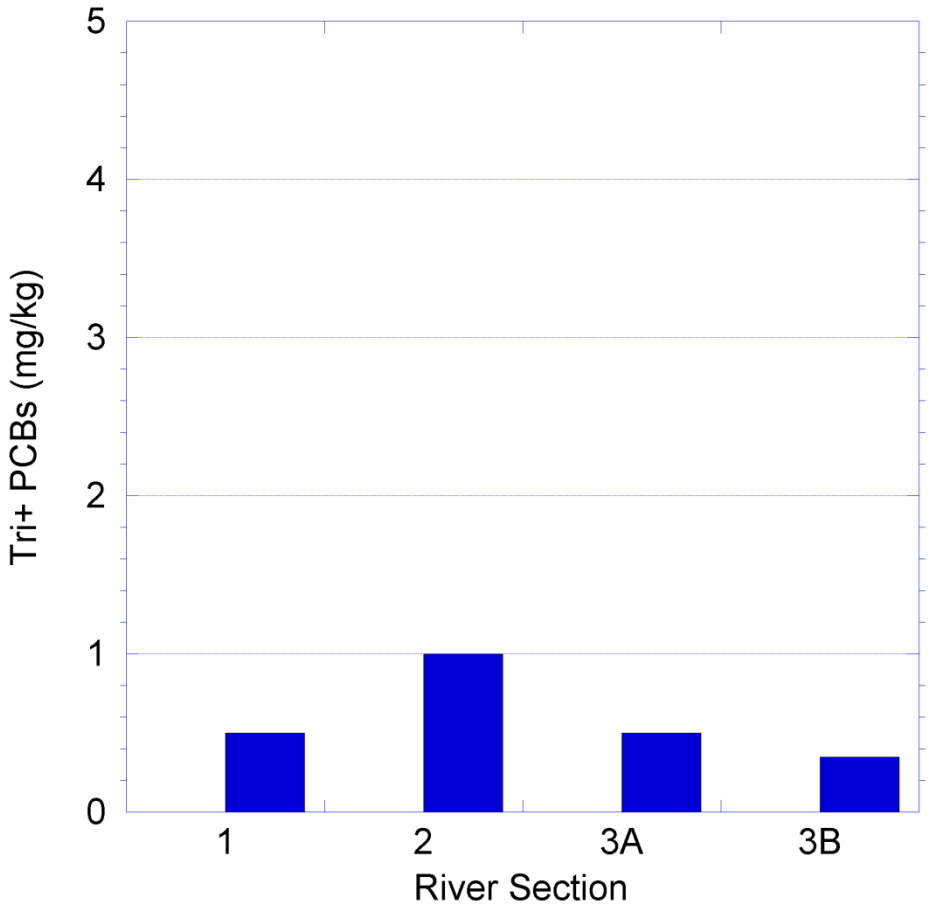
- MNA model predictions of Tri+ PCB concentrations in cohesive (fine-grained) sediment for the surface (top 5 cm) by River Section before the start of dredging.
- Samples collected to define dredge areas in River Sections 2 and 3 targeted cohesive sediment.
- Error bars represent the upper bound for No Action alternative

Model Predictions Compared to Estimated Pre-Dredging Surface Concentrations



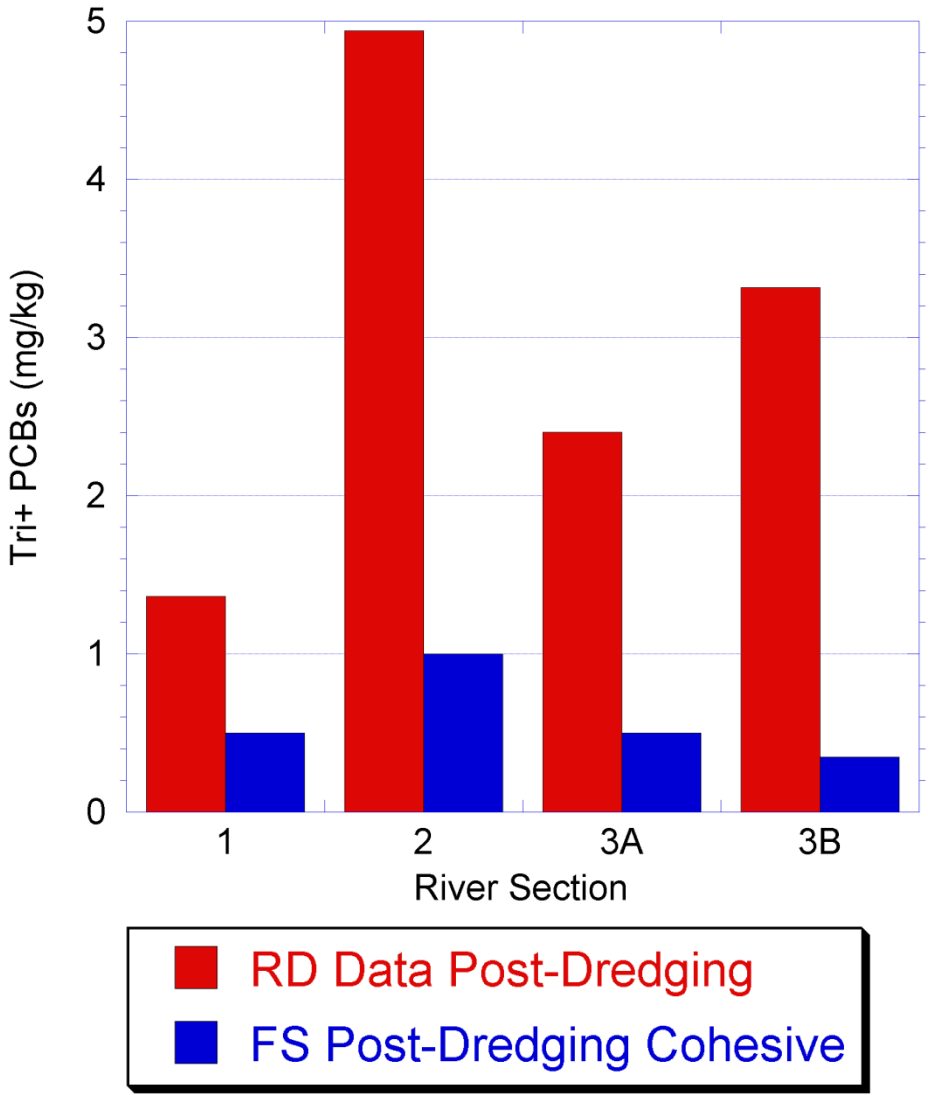
- Surface (top 5 cm) sediment concentrations (red bars) exceeded the No Action upper bound of model predictions (blue error bars) and were more than 2X the mean concentration predicted for cohesive sediments in all 3 river sections (blue bars).
- Models overestimated the rate of natural recovery.
- Widespread burial of PCBs in the surface sediment was not observed.

Model Predictions: Post-Dredging Surface Concentrations



- The Record of Decision expected that the selected dredging alternative would result in <1 ppm mean Tri+ PCB concentrations in cohesive sediments throughout the Upper Hudson.

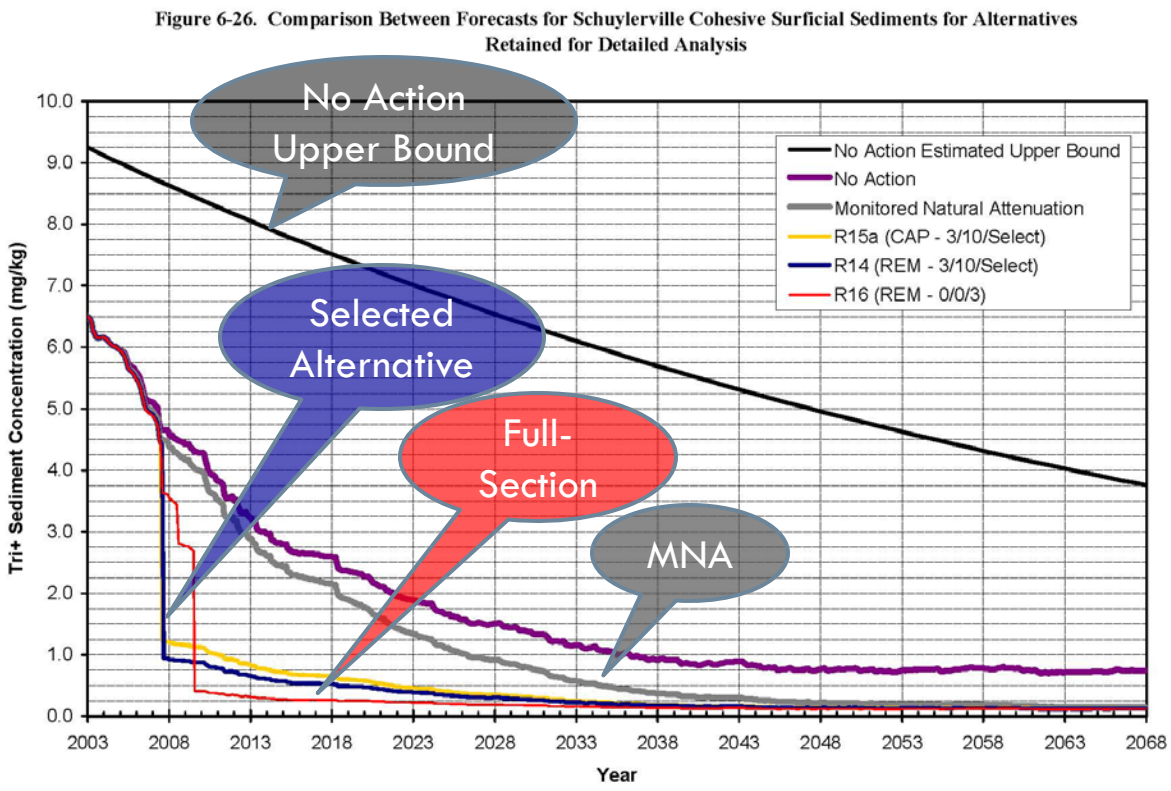
Model Predictions Compared to Estimated Post-Dredging Surface Concentrations



Estimated post-dredging concentrations (red bars)

- River Section 1:
  ~ 2.5X higher than model predictions
- River Sections 2 and 3:
  ~ 5X higher than model predictions

Comparison of Remedial Alternatives Model Forecasts: Sediment

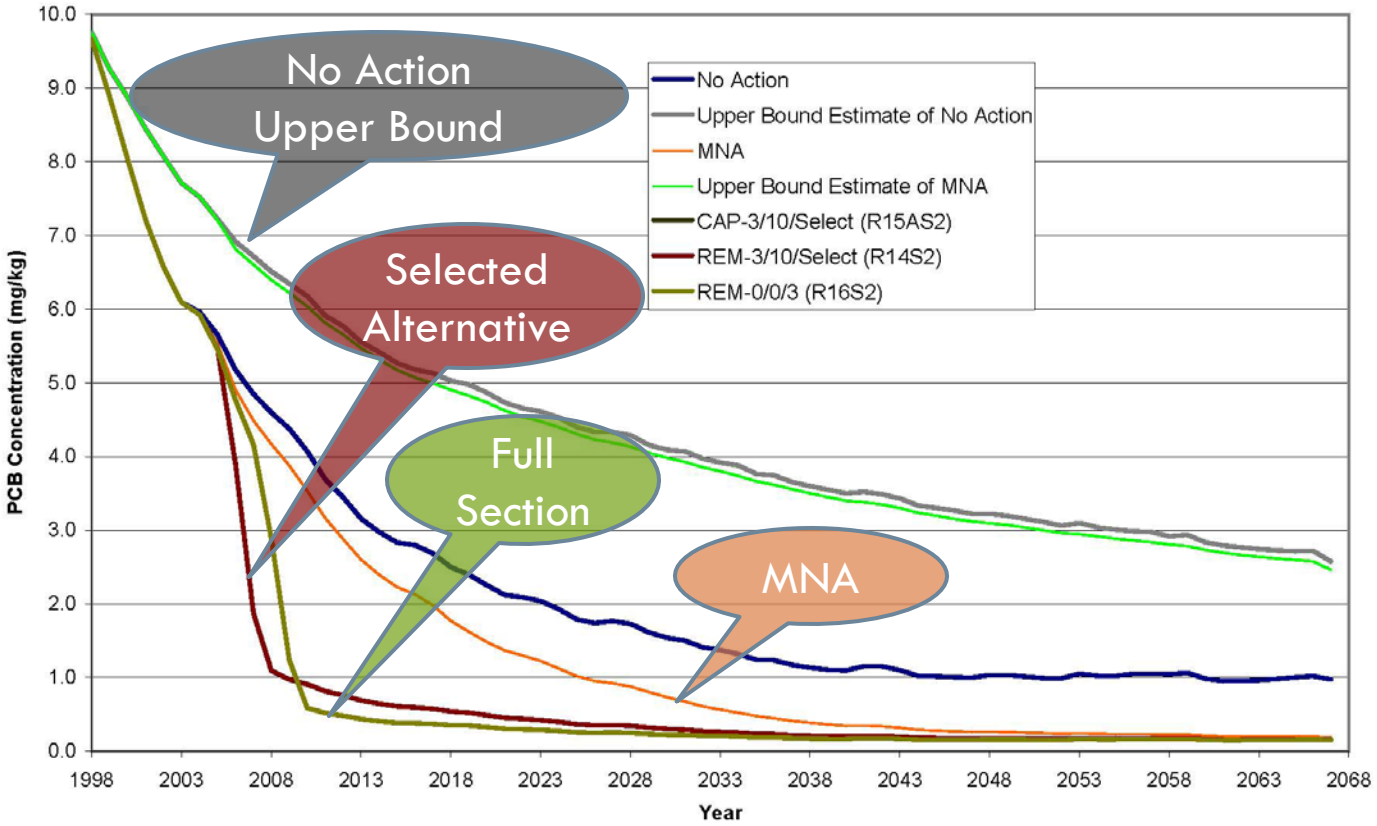


River Section 2 Cohesive Sediment

- Alternatives
  - No Action Upper Bound
  - MNA
  - Selected Alternative
  - Full Section Remedy

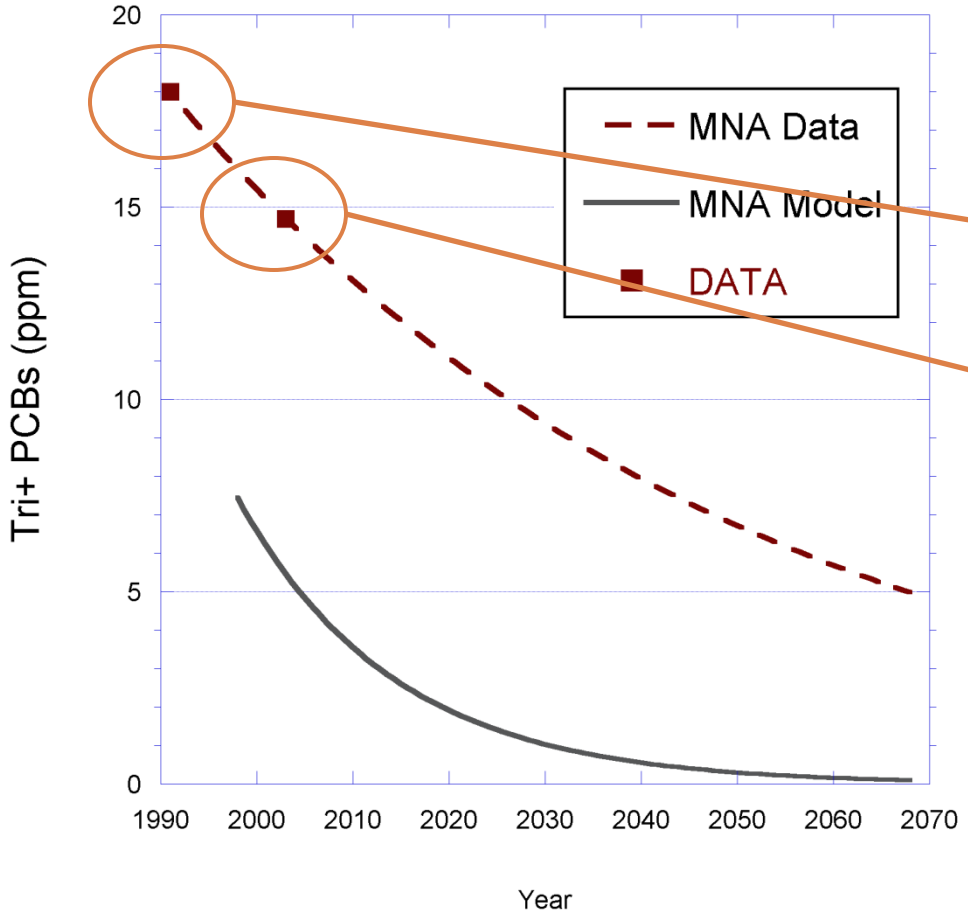
Comparison of Remedial Alternatives Model Forecasts: Fish

Figure 6-39. Comparison between Species-Weighted Fish Fillet Average PCB Concentration in River Section 2 for Alternatives Retained for Detailed Analysis



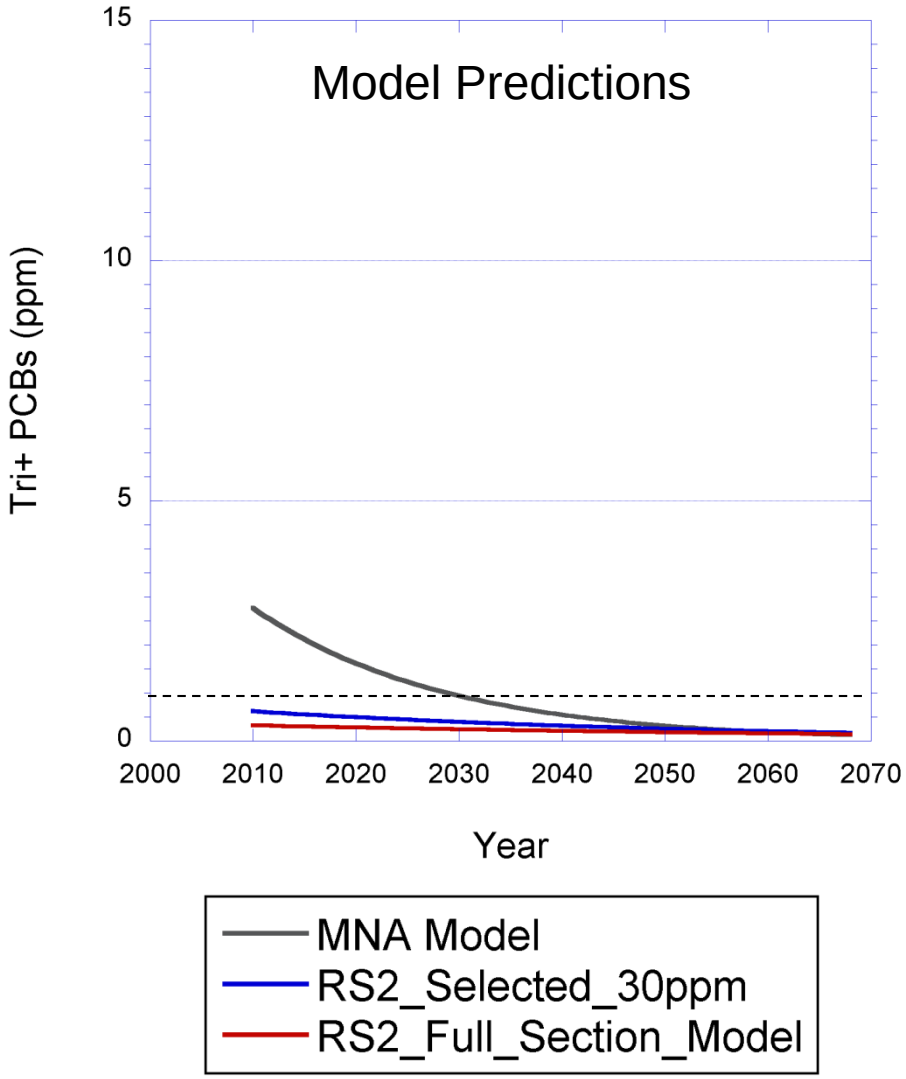
River Section 2

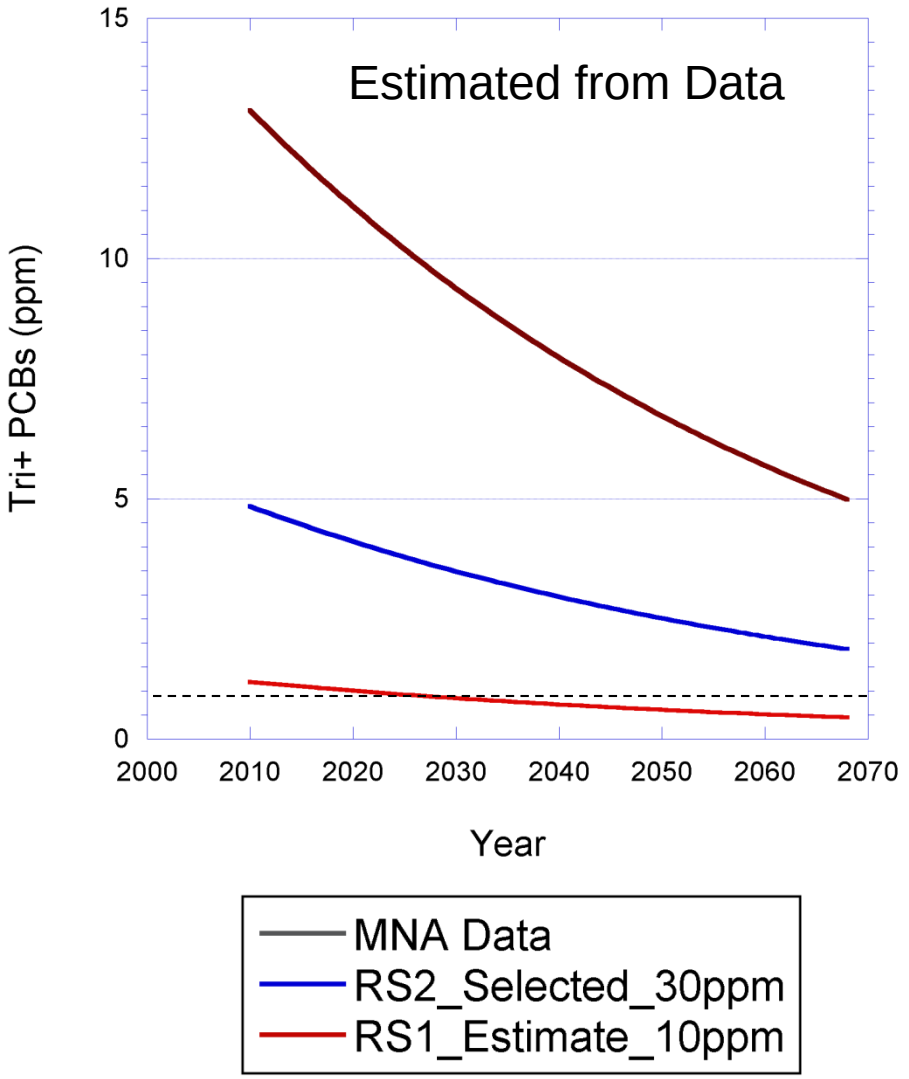
Estimated Monitored Natural Attenuation



- River Section 2 Cohesive Sediment
  - GE large-scale sediment transect survey (1991)
  - Remedial Design data (~2003; N=1541)
- Decay rate from MNA model about 5X the rate estimated from data

Post-Dredging Comparison of Alternatives Based on MNA Models





Summary: Natural Recovery Rate

- Models overestimated rate of natural recovery, and, as a result
  - Underestimated surface concentrations prior to the initiation of dredging and following removal
  - Reduced ability to discriminate among remedial alternatives, including the time required to achieve risk thresholds

Data Trumping and 10-foot Rule

- Re-sample previously sampled locations with elevated surface sediment chemical concentrations to determine “whether these elevated surface sediment chemical concentrations that existed in the past still exist”
- Exclude older samples from surface sediment baseline if newer samples collected within 10 feet of original location

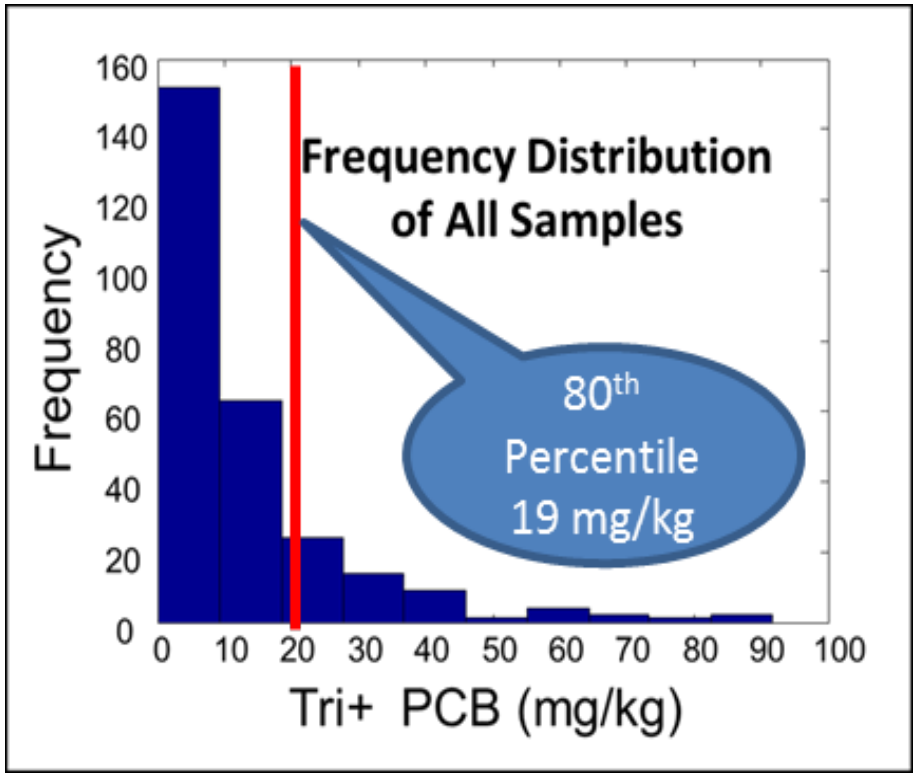
Assumptions:

- "more recent results more accurately represent current conditions”
- Any differences in concentration reflect temporal change, not spatial heterogeneity

Testing the 10-foot Rule
What's Bias Got to Do With It?

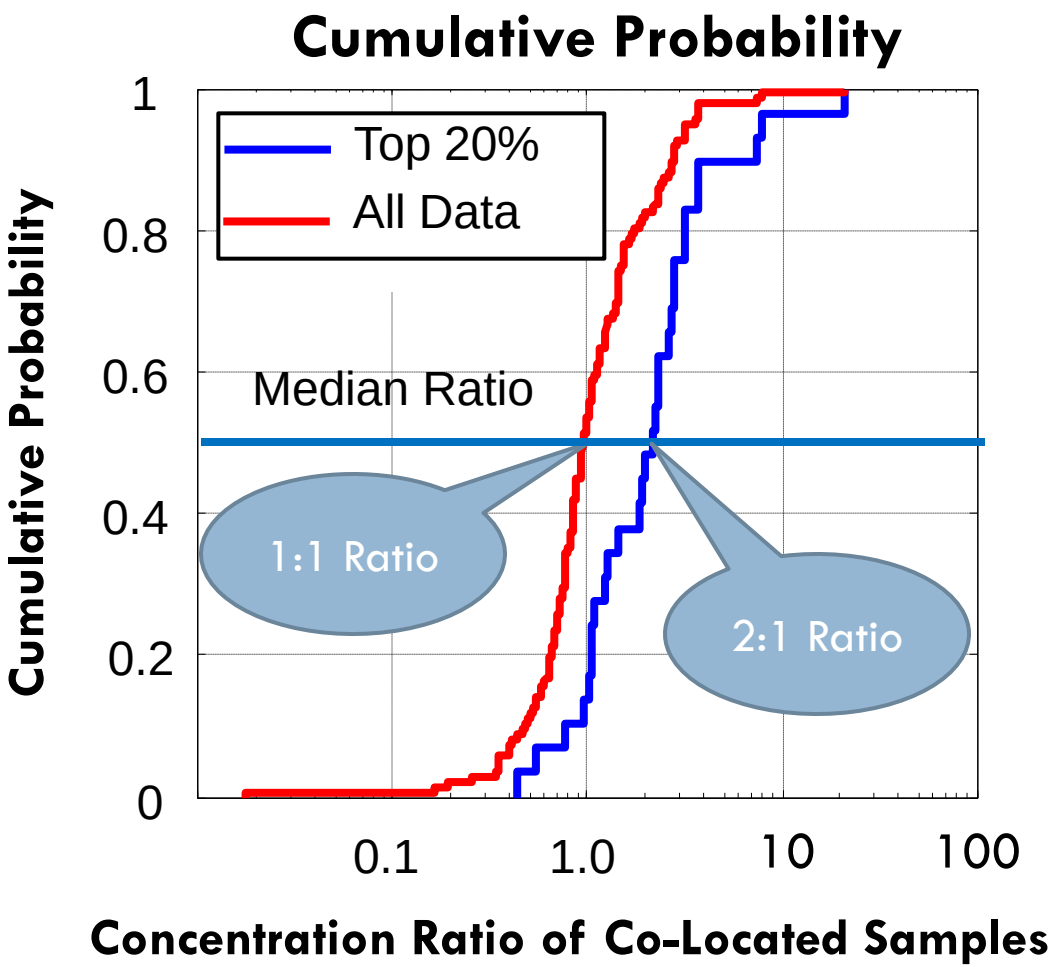
- Evaluate the accuracy of this approach using paired co-located surface sediment samples from 136 locations throughout the Upper Hudson
- First sampled in 2002-2003 and re-sampled in 2004-2005
- Second sample was within 10 feet of initial sample
- Samples from the upper 20th percentile of the concentration distribution were compared to their co-located sample to represent the re-sampling of locations with elevated concentrations

Comparison of Paired Samples



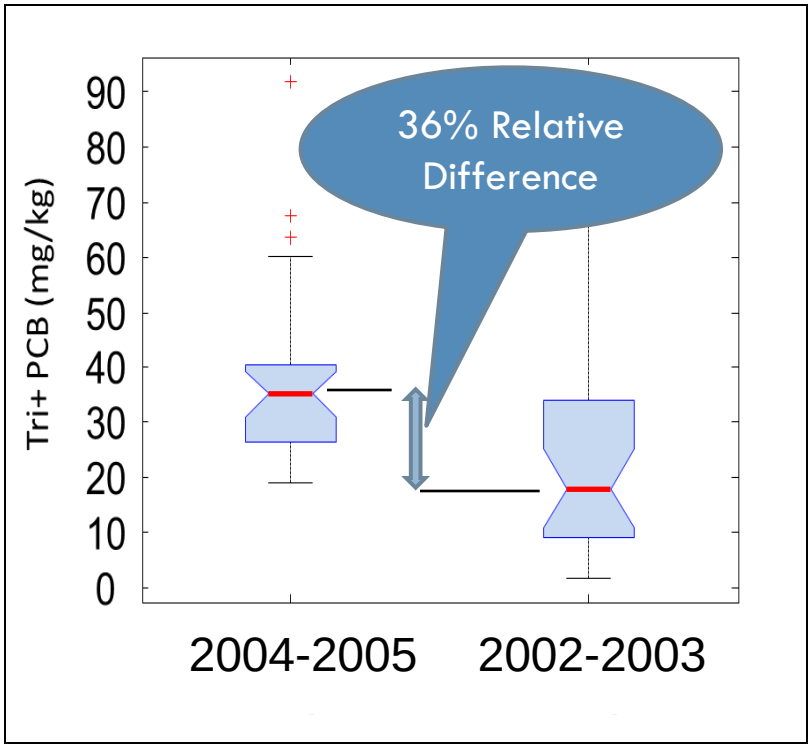
- Select 2nd samples with concentration > 80th percentile of the distribution
- Compare to samples collected 1-3 years earlier within 10 ft

Distribution of Sample-Pair Ratios



- Median Ratio for all data is 1:1
- Median Ratio for preferentially selected top 20% is ~2:1
- Comparison of secondary sample at locations of top 20% of first sample virtually guarantees apparent decreasing temporal trends

Results of Paired-Sample Comparison



- Median concentration for the upper 20th percentile of 2004-2005 sample distribution is 36% higher than the median for paired samples collected 1-3 years earlier.
- Result is an artifact of the biased sampling used to obtain the test set.

Temporal Change or Spatial Heterogeneity?

An analysis of spatial and temporal changes in surface sediment concentrations for the Portland Harbor Superfund site concluded:

- "concentration differences between paired samples do not vary in magnitude with time"
- "concentration differences measured in same-day pairs approximately equals or exceeds the range measured in samples collected up to 3000 days apart."
- "...much of the heterogeneity in collocated measurements appears to reflect actual small-scale heterogeneity in the system, the heterogeneity associated with the sediment sampling and analysis process, or both."

Source: Integral et al 2007

Summary:
Data Trumping and10-ft Rule

- Re-sampling locations with elevated concentrations virtually guarantees apparent decreasing temporal trend
- Replacing older sediment data by re-sampling locations with elevated concentrations based on an assumption of natural recovery is not justified
- Biased re-sampling approach is not useful for determining rate of recovery.
- The assumption that concentration differences for samples collected within 10 ft reflect temporal change, not spatial heterogeneity should be tested.

“All models are wrong, some are useful”

- For a decision-maker, useful models provide the ability to discriminate differences in outcome for an array of alternatives
- How do you know?
- Need good data, including data for baseline conditions and temporal rate of change in surface sediment concentrations that are representative of the area of concern

Design Recommendations for Sediment Temporal Trend Monitoring Plan

- Incorporate trend monitoring early in site assessment
- Use unbiased sampling procedures
  - Identify important strata boundaries at the outset of the monitoring program
  - Determine sample size using variability of existing data to quantify temporal decay rates with adequate precision for comparisons of remedial alternatives
- Monitor same locations at ~ 5 year intervals
  - Use paired and repeated measures statistical analyses within strata to evaluate local trends
  - Combine results across strata to develop global statements about trend in overall average (SWAC).
  - Interpolation is unnecessary because sampling is unbiased

Motto

- “In God we trust---All others must bring data”

Source: attributed to W. Edwards Deming