THE IMPORTANCE OF MATCHING THE
SPATIAL SCALES OF PROBABILISTIC
MONITORING DESIGNS WITH
MANAGEMENT QUESTIONS

P. Trowbridge, S. Jones,
H. Walker, and J. Kiddon

EMAP Symposium, Newport RI
May 6, 2004
Acknowledgments

Co-Authors

• Dr. Stephen Jones, UNH
• Dr. Hal Walker, EPA/ORD
• Dr. John Kiddon, EPA/ORD
Objectives

• To investigate the effects of spatial scales on the outcomes of Probability-Based Monitoring (PBM) programs
  – Representativeness
  – Spatial Autocorrelation
  – Confidence Intervals

• To provide practical advice to State managers for implementing PBMs.
Methods

• Use “natural experiment” of NH’s small scale for the National Coastal Assessment
• Conduct fine-scale studies at 4 NH sites
• Make comparisons between results for mercury in sediment at three scales:
  – Gulf of Maine
  – New Hampshire
  – Small study areas
Scale Model of Hexagon Sizes

Maine:
50,000 ha
22 km X 22 km

Massachusetts:
17,500 ha
13 km X 13 km

New Hampshire:
322 ha
2 km X 2 km
Resource Area in each NCA hexagon. The area of estuary inside the hexagon.
Resource Area in NH NCA and Special Study Hexagons

Area (m²)
First Experiment: Representativeness

Outcome: The NCA study design reasonably estimates average values at multiple spatial scales.
Methods

• Compare cumulative distribution function (CDF) from a random subsample of NH NCA data to whole NH NCA dataset
• Compare CDF for 4 intensive study areas to NH NCA samples from the same area
• Use mercury in sediment concentrations as a common parameter
NH NCA Study
Design in Great Bay
Trend stations shown in yellow.
Comparison of Hg CDFs for different scales

- **Hg Concentration (mg/kg)**
- **Percent of Data Below Value**

### Median Values

<table>
<thead>
<tr>
<th></th>
<th>Hg (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trend stns</td>
<td>0.16</td>
</tr>
<tr>
<td>All data</td>
<td>0.14</td>
</tr>
</tbody>
</table>
NH NCA Design with locations of special studies

- Salmon Falls River
- Cochecho River
- Southeast Great Bay
- Piscataqua River
Representativeness Conclusions

- NCA study design captures median values of CDFs but misses extremes.
- Representativeness was demonstrated at two different scales so is likely to be robust.
Second Experiment: Spatial Autocorrelation

Outcome: For stations <20 km apart, spatial correlation is likely.
Spatial Correlation

- As hexagon sizes shrink, the stations converge.
- Adjacent stations provide similar information.
- Autocorrelated stations are not independent.
Methods

• Investigate spatial autocorrelation by charting semivariograms for mercury in sediments at three different scales
  – Gulf of Maine
  – New Hampshire
  – Special study sites
Definition of Semivariogram

Schematic Semi-Variogram

- **Sill** ~ Total variance among distant samples. For completely independent data this simplifies to sample variance $S^2$

- **Nugget Effect** ~ Measure of small scale spatial heterogeneity.

- **Range of Influence** ~ Distance at which samples are uncorrelated.

The semi-varioagram is a measure of dissimilarity:

$$\text{Average } (V_i - V_j)^2 / 2$$

Spatial Correlation = $1 - (\text{Semi-Variogram}) / \sigma^2$

**Source:** John W Kern
Semivariogram for $\ln [\text{Hg}]$

Gulf of Maine

- Range $\sim 20$ km
- Nugget Semivariance $\sim 0.35$
- Sill Semivariance $\sim 1.2$

Range in next slide, 20,000 m
Semivariogram for \( \ln [\text{Hg}] \)
NH NCA (2000-2001)

All NH NCA sites appear to be autocorrelated

Nugget semivar ~0.4

Range from previous slides, 20,000 m

Exponential model: \( C_0 = 0.4340; \ C_0 + C = 1.7170; \ A_c = 39680.00; \ r^2 = 0.646; \ RSS = 0.0671 \)
NH NCA Design with locations of special studies

Salmon Falls River

Cocheco River

Southeast Great Bay

Piscataqua River
Semivariogram for Ln [Hg] 4 Special Study Areas Combined

All Nugget with High Semivariance ~0.8
Semivariogram for Ln [Hg]

SE Great Bay

Decreasing semivariance.
Approaching zero nugget at small scales.

Linear model \( (C_0 = 0.0010, \ C_0 + C = 0.7420, \ A_0 = 4240.00; \ \ r^2 = 0.937, \ \ RSS = 4.508E-03) \)
Semivariogram for $\ln [Hg]$ Cocheco River

All nugget but low semivariance $\sim 0.25$. Uniform variance throughout study area.

Linear model ($C_0 = 0.2569; C_o + C = 0.2569; A_0 = 2665.81; \tau^2 = 0.059$; $RSS = 0.0454$)
Semivariogram for Ln [Hg] Piscataqua R.

All nugget but high semivariance ~1.0. Same order as GOM sill semivariance.

Linear model (C0 = 0.9567; C0 + C = 0.9567; Ao = 2478.68; r2 = 0.677; RSS = 0.471)
Semivariogram for Ln [Hg]
Salmon Falls R

Same as for Piscataqua River

Linear model (Co = 0.9049; Co + C = 0.9049; Ao = 4459.30; r2 = 0.230; RSS = 0.603)
Summary - High Density Hg Data

• SE Great Bay. Median $[\text{Hg}] = 0.22$ Max = 0.30
  – Semivariance suggestive of small nugget at fine scale.

• Cocheco. Median $[\text{Hg}] = 0.52$ Max=0.76
  – All nugget but low semivariance (0.26).
  – Consistently high Hg concentrations.

• Piscataqua. Median $[\text{Hg}] = 0.20$ Max = 0.67

• Salmon Falls. Median $[\text{Hg}] = 0.27$ Max = 0.60
  – All nugget with high semivariance (0.90-0.97)
  – Mixture of high and low Hg concentrations. Heterogeneous.
Spatial Correlation Conclusions

- On a broad scale, stations > 20 km apart are uncorrelated and independent.
- NH NCA stations are ~2 km apart so autocorrelation appears to be present. However, the type and level of autocorrelation is not uniform.
- Drill down into four areas with high Hg in NH NCA survey revealed three different contamination structures.
Third Experiment: Variance Calculations

Outcome: Variance was higher than expected. For small scales, variance algorithm and assumption of independence may be in error.
Variance/Confidence Intervals

- With 82 NCA stations for the 18 miles of NH coastline, we would expect tight confidence intervals on the results.
- However, the Horvitz-Thompson Estimator algorithm produced very large error bars on our estimates.
Mercury in Sediment in NH's Estuaries

Possible range: 14-72%!
<table>
<thead>
<tr>
<th>Scenario</th>
<th>LCL</th>
<th>UCL</th>
<th>CI</th>
<th>Error Bar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>35.55%</td>
<td>74.34%</td>
<td>38.80%</td>
<td>19.40%</td>
</tr>
<tr>
<td>Orginial with Area X1000</td>
<td>35.55%</td>
<td>74.34%</td>
<td>38.80%</td>
<td>19.40%</td>
</tr>
<tr>
<td>Original with Area X0.001</td>
<td>35.55%</td>
<td>74.34%</td>
<td>38.80%</td>
<td>19.40%</td>
</tr>
<tr>
<td>Smallest 10 Removed</td>
<td>38.90%</td>
<td>71.92%</td>
<td>33.03%</td>
<td>16.51%</td>
</tr>
<tr>
<td>Largest 10 Removed</td>
<td>36.71%</td>
<td>65.38%</td>
<td>28.67%</td>
<td>14.33%</td>
</tr>
<tr>
<td>Smallest and Largest 10 Removed</td>
<td>41.25%</td>
<td>63.14%</td>
<td>21.89%</td>
<td>10.94%</td>
</tr>
<tr>
<td>Medians</td>
<td>43.23%</td>
<td>53.27%</td>
<td>10.04%</td>
<td>5.02%</td>
</tr>
</tbody>
</table>
Example of riverine resource in NCA hexagons
Possible Causes for High Variance

• Is the recommended algorithm for variance calculations correct for the situation in NH? The jury is still out. Something is not quite right here.

Other issues
• Samples in NH are not independent because of scale
  – New analysis methods might be needed to address spatial autocorrelation.
• Are hexagons the appropriate sampling design for the resource in NH?
  – Most NH hexagons are <50% resource
  – Median % resource in a hex is 18%
What does this all mean?

Practical advice for managers
Value of the NCA Approach

- Provides unbiased comparisons of conditions between states.
- Provides accurate representation of median values of a parameter at the state and regional level.
- Cost-effective monitoring strategy for assessing 100% of surface waters for §305(b) reporting.
The NCA program provides valuable insight into the bigger picture of Hg in the Northeast.
Potential Problems with the NCA Approach at Small Scales

- For stations <20 km apart, autocorrelation is likely. May violate assumption of independence.
- Inflated variance/confidence intervals can develop at small spatial scales. Possible errors in variance estimator algorithm at this scale.
- State managers need to consider these factors when using NCA data or data from small scale PBMs for CWA reporting requirements.
Questions/Comments

Phil Trowbridge
NH Estuaries Project
Department of Environmental Services
29 Hazen Drive
Concord, NH 03302-0095
Tel: 603/271-8872
ptrowbridge@des.state.nh.us
Extra slides in case questions come up...
Resolution of Spatial Features

What hex size is needed?
Methods

• “Sampled” bathymetry coverage to determine estimated CDFs for different hex sizes.
• Compared estimated CDFs to “true” CDF.
• Identified unresolved features and relationship to hex size
Red hex=20 hex design
Blue hex=160 hex design
Yellow=>20 ft channel
Bathymetry CDFs for Different Designs

CDFs diverge at >20ft depth

160 hex design matches true CDF well.

Therefore, deep channel is the unresolved feature.
## Percent of hex covered by unresolved feature

<table>
<thead>
<tr>
<th>Study Design</th>
<th>% of Resource Area with &gt;20 ft depth</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 hexs</td>
<td>17%</td>
</tr>
<tr>
<td>40 hexs</td>
<td>23%</td>
</tr>
<tr>
<td>80 hexs (NH NCA)</td>
<td>25%</td>
</tr>
<tr>
<td>160 hexs</td>
<td>32% (able to resolve feature)</td>
</tr>
</tbody>
</table>
Spatial Resolution

Conclusions

• In this case, to resolve a given bathymetric feature, need hexagons that are ~1/3 filled by the feature.
• This result may not be translatable to other cases/situations.
• Could generate a rule of thumb if this result were repeated with different parameters in different locations.
Semivariogram for Ln [Hg] for the Northeast (DE through ME)

Range ~20,000 m
Nugget Semivar ~0.3

Distance of 20 km corresponds to ~400 km² grid size (assuming square). Median hex size for Northeast states (except NH) is ~300 km². For NH, median size is <100 km².