Small area estimation of indicators of stream condition for MAIA using hierarchical Bayes prediction models

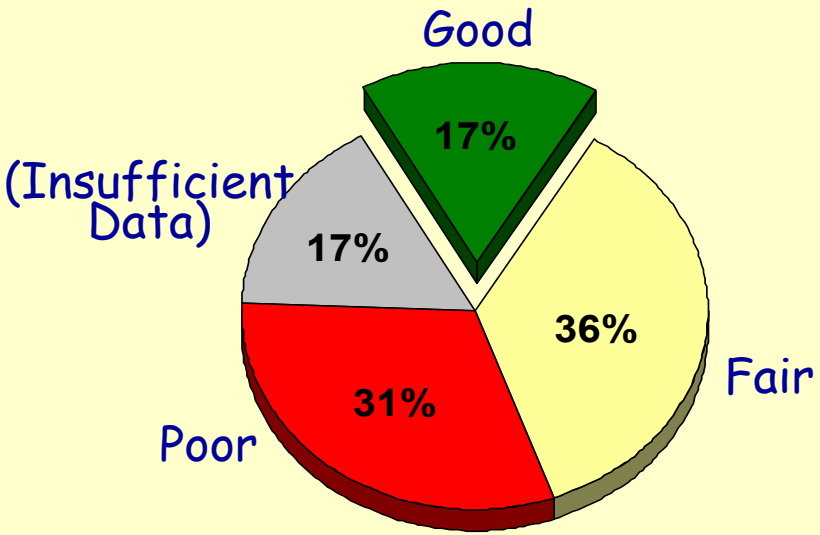
Anthony (Tony) R. Olsen
USEPA NHEERL
Western Ecology Division
Corvallis, Oregon
Email: olsen.tony@epa.gov

Collaborators

- **Mark Handcock, University of Washington**
- **Don L. Stevens, Jr, Oregon State University**
- **Steve Carroll, Oregon State University**
- **Jeanne Sifneos, Oregon State University**
- **Alan Herlihy, Oregon State University**
- **Alix Gitelman, Oregon State University**
- **Deb Chaloud, USEPA NERL Landscape Ecology**

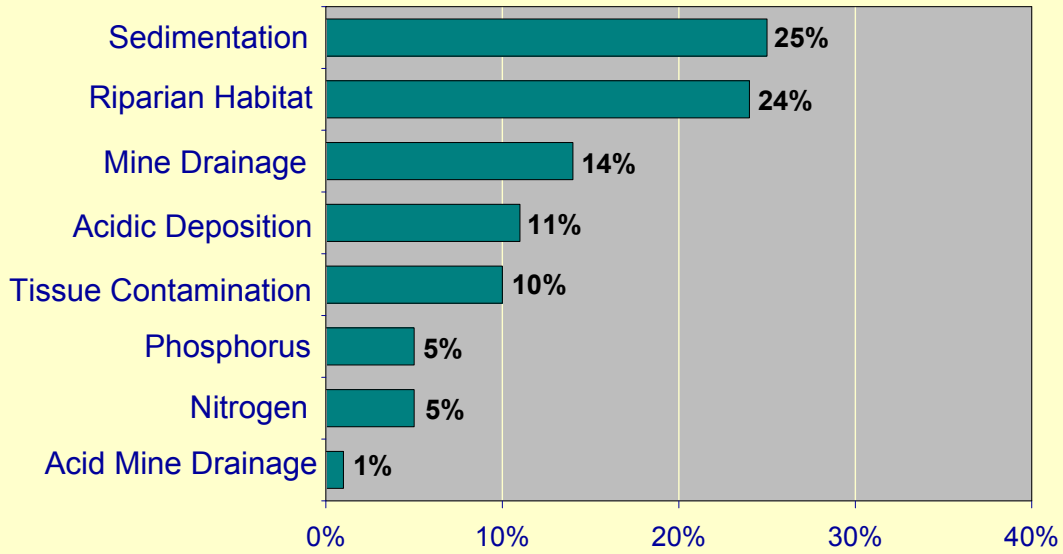
EMAP Probability Survey Example Results (complex)

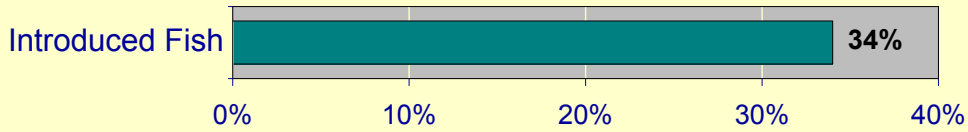
Fish Index of Biotic Integrity



Proportion of Stream Length

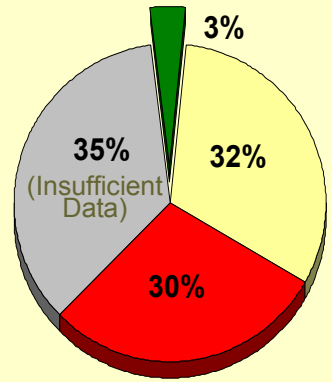
Relative Ranking of Stressors



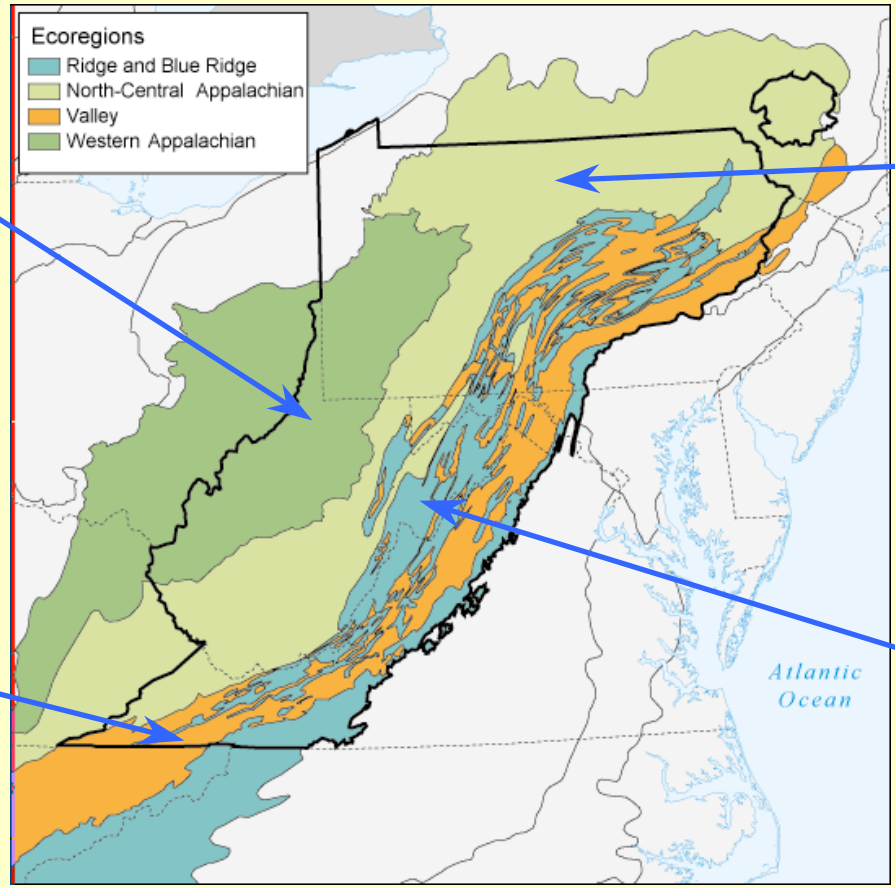


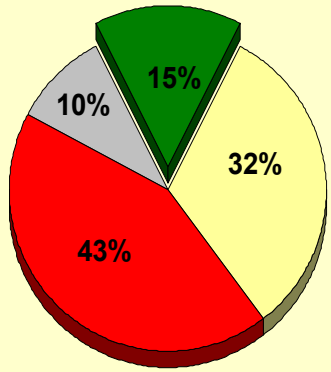
Geographic Targeting

Where does Fish IBI suggest problems?

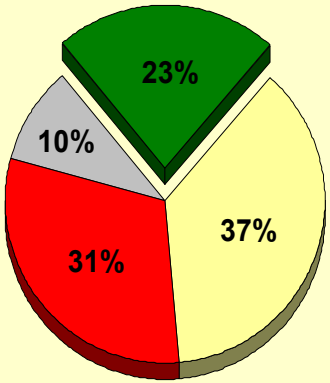


Western Appalachians

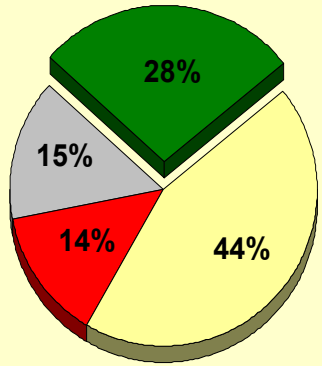




North-Central Appalachians



Valleys



Ridge and Blue Ridge

Additional Information Needs

- **Improve estimates using auxiliary information**
  - **Watershed variables known on entire sample frame**
  - **Model-assisted estimators**
- **Estimates for smaller geographic areas**
  - **Limitation is small sample size**
  - **Costs prohibitive to increase sample size**
- **Spatially-explicit prediction**
  - **Want a "map" of predicted response variables**
  - **Model-based prediction**
  - **Potentially large uncertainty in predictions**

Major Hydrography

Mid-Atlantic Study Area

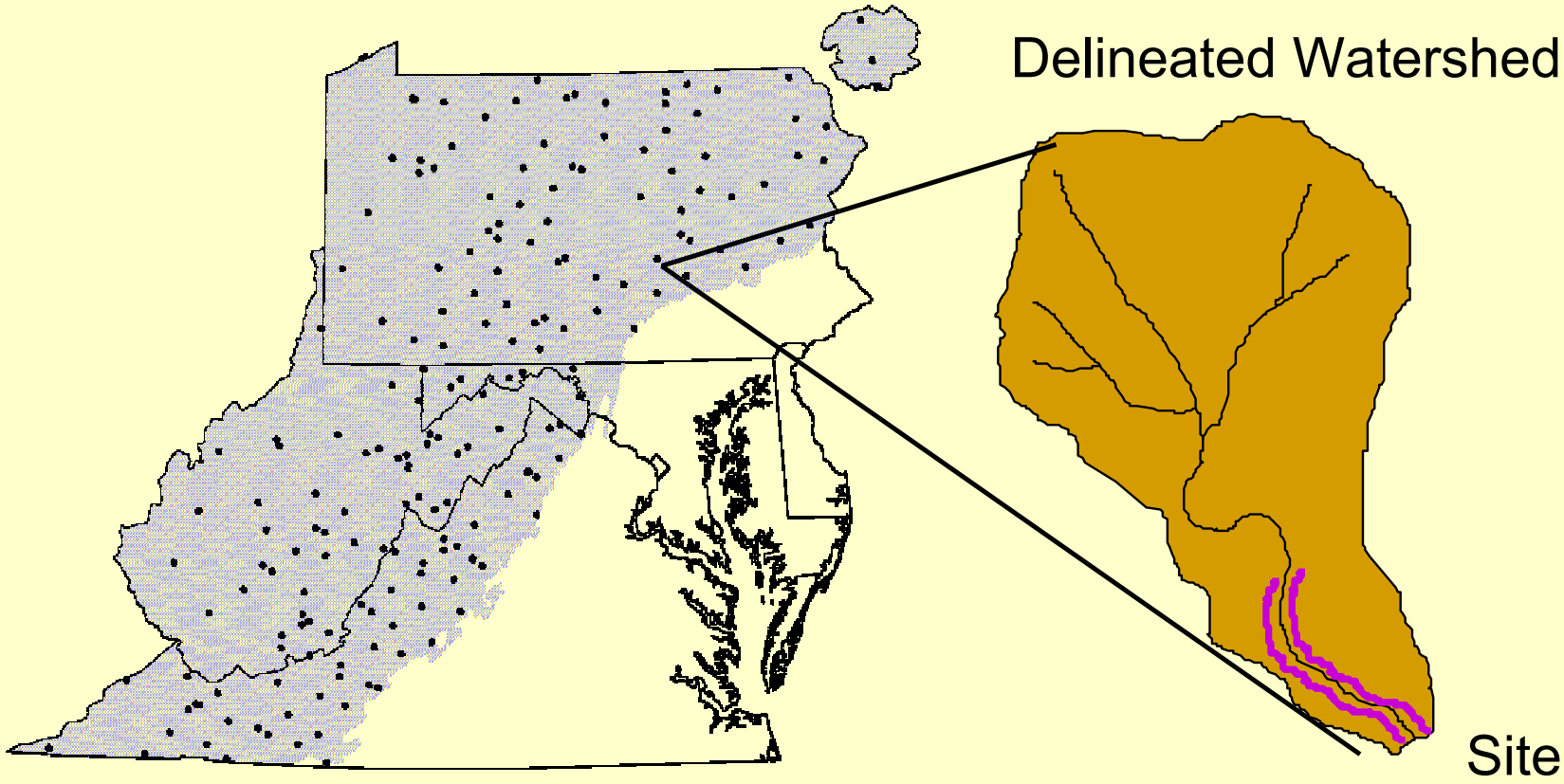
Small Area Estimation Approach: Model Building Step

- **Build Bayesian hierarchical spatial model relating response variable to variables available on entire stream network**
  - **300-800 stream sites from EMAP MAIA probability sample**
  - **Delineated watersheds for each site**
  - **Calculate predictor variables for site, reach, watershed**
  - **Calculate spatial association structure**
  - **Fit model**
- **Determine when model applies: geographically, stream order, and watershed size**

Indicator Variables

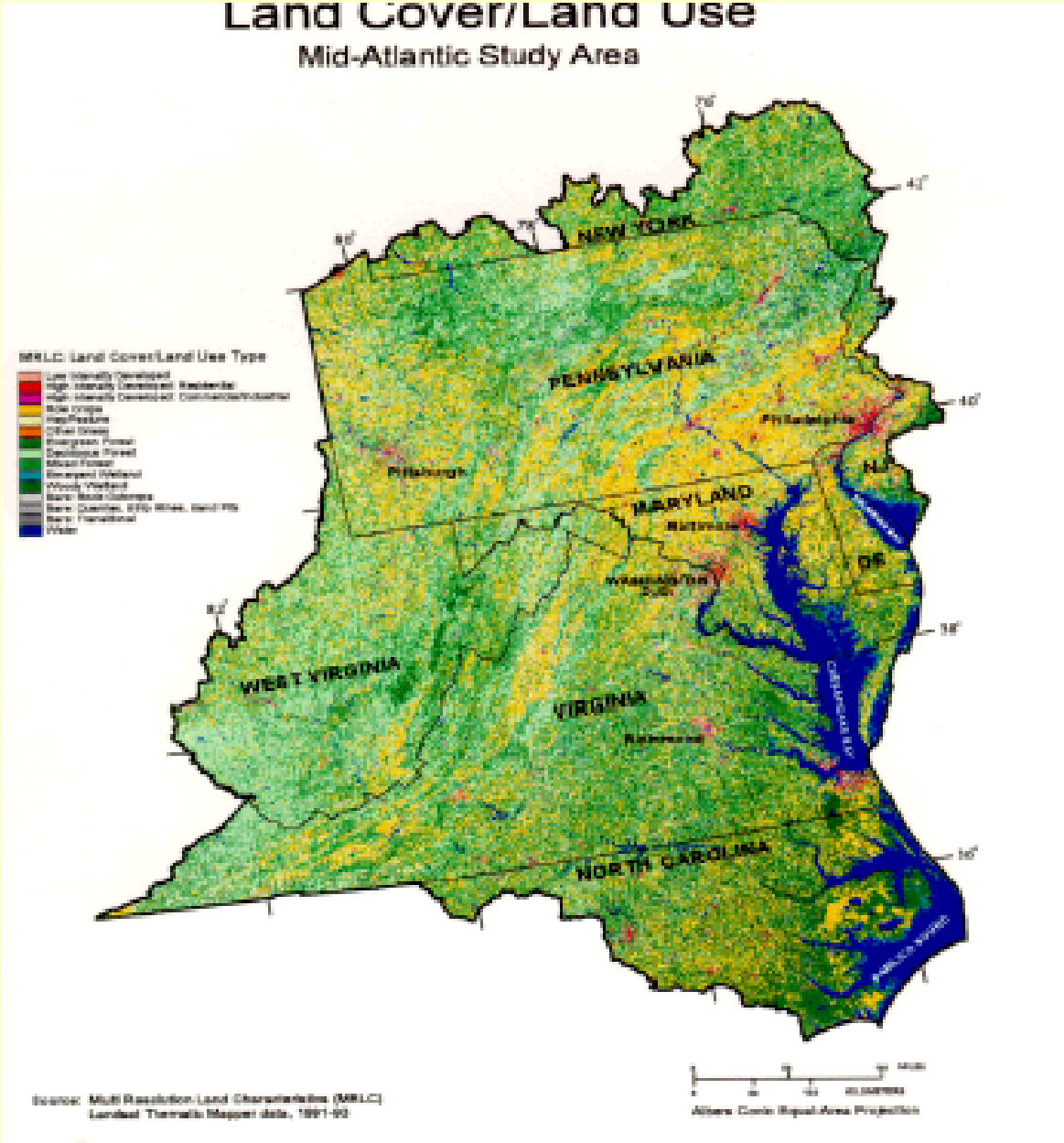
- **Benthic macro invertebrates**
  - **EMAP IBI**
  - **Metrics used in IBI**
- **Fish assemblage IBI**
  - **EMAP IBI**
  - **Metrics used in IBI**
- **Water chemistry**
  - **Total nitrogen**
  - **Total phosphorus**

Watershed Variable Design



Predictor Variables: Tentative

- **Site variables**
  - **Location (x, y)**
  - **Elevation**
  - **Omernik Ecoregion Level III**
- **Watershed variables**
  - **%Agriculture from MRLC**
  - **%Urban from MRLC**
  - **%Forest from MRLC**
  - **Erodibility Index derived from STATSGO**
  - **Human population density**
  - **Road density**
  - **Majority membership Omernik Ecoregion Level III**
  - **Area**


Land Cover/Land Use
Mid-Atlantic Study Area
MRLC Land Cover/Land Use Type
NEW YORK
PENNSYLVANIA
MARYLAND
DE
NJ
WEST VIRGINIA
VIRGINIA
NORTH CAROLINA
Pittsburgh
Philadelphia
Richmond
Albers Conic Equal-Area Projection

Spatial Association Structures

Hydrologic Associations

Watershed Area Based

$$\eta(w_i, w_j) = 1 - \frac{OverlapArea(w_i, w_j)}{MaxArea(w_i, w_j)}$$

Stream length within watershed based

$$\eta(w_i, w_j) = 1 - \frac{OverlapLength(w_i, w_j)}{MaxLength(w_i, w_j)}$$

Distance Association

$$\phi(w_i, w_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Small Area Estimation Approach: Site Prediction Step

- **Select set of prediction sites using probability survey design over MAIA region**
  - **200 sites in each 4th Field HUC**
  - **Total sites approximately 25,000**
- **Delineate watersheds for each site (automatically)**
- **Calculate predictor variables for each site: site, reach, and watershed-based**
- **Calculate spatial structure metrics**
- **Predict response at each site using fitted model**

Hierarchical Bayes Prediction Equation

$$Z(x) = W(x)\beta_1 + S(x)\beta_2 + \eta(x \mid \gamma) + \phi(x \mid \nu) + \varepsilon(x \mid \sigma)$$

x – identifies watershed and outlet location

Z(x) – indicator variable

W(x) – watershed predictor variables (vector)

S(x) – site predictor variables (vector)

$\eta(x \mid \gamma)$ – hydrologic spatial association variable

$\phi(x \mid \upsilon)$ – distance spatial association variable

$\varepsilon(x \mid \sigma)$ – random error

Small Area Estimation Approach: Small Area Prediction Step

- **Predict the indicator variable for 200 sites within each 4th field HUC**
- **For each 4th field HUC estimate the CDF for the indicator variable using the survey design**
- **Estimate % stream km within each 4th field HUC exceeding criterion for indicator**
- **Graphically display results.**

Issues and Discussion

- **Requires automatic delineation of watersheds**
  - **Watersheds for estimating prediction equation (800)**
  - **Watersheds for prediction (25,000)**
- **Requires calculation of watershed overlap**
  - **computer intensive**
  - **can use just area overlap or do need stream overlap?**
- **Site-specific predictions will have large uncertainty**
  - **Focus on HUC level summaries to account for this**
  - **Deconvolution required to adjust for prediction error bias**
- **Indicator variable selection**
  - **Probability of not meeting designated use; change to logistic model**