Design-based/Model-assisted Survey Methodology for Aquatic Resources
An EPA STAR Cooperative Research Program

Don L. Stevens, Jr.
Department of Statistics
Oregon State University
Presented at Statistical Survey Design And Analysis For Aquatic Resources
Fort Collins, CO
September 20-21, 2002
This presentation was developed under STAR Research Assistance Agreement No. CR-829096 awarded by the U.S. Environmental Protection Agency to Oregon State University. It has not been formally reviewed by EPA. The views expressed in this document are solely those of the author and EPA does not endorse any products or commercial services mentioned in this presentation.
Discussion Outline

• Program description
  – What we are trying to do

• Program context
  – Where we are & how we got here

• Integration & Outreach
  – Who we are working with

• Accomplishments
Program Objectives

• Advance the state of aquatic survey design and analysis
  – Develop methods to incorporate prior knowledge at both design and analysis stages via design-based / model-assisted techniques
  – Develop design and analysis methods for surveys
    • Over time
    • Through space
    • At multiple scales
Program Objectives

• Develop and extend the expertise on aquatic survey design and analysis to States and Tribes
  – Research driven by applications
  – Demonstration projects
  – Survey design and analysis software tools
  – Workshops, seminars, handbooks, web pages
Program Objectives

• Train a cadre of environmental statisticians
  – Post docs
  – Graduate research assistants
  – Faculty
  – State & Tribal personnel
Faculty Investigators

Oregon State University
Don L. Stevens, Jr    Virginia Lesser
Alix Gitelman        Paul Murtaugh
Alan Herlihy         Robert Hughes
Jeanne Sifneos       Steve Carroll

Colorado State University
N. Scott Urquhart    F. Jay Breidt

University of Washington
Loveday Conquest

Iowa State University
Jean Opsomer
Post-docs & Grad Students

**Oregon State University**
- Breda Munoz-Hernández
- Beverly Benham
- Steve Jensen
- Rubén Smith
- Cynthia Cooper

**Colorado State University**
- 0.5 FTE Post Doc

**University of Washington**
- Jean-Yves Courbois
- Graduate Research Assistant
Cooperators

San Francisco Estuary Institute
California State Water Resources Control Board
California Department of Fish & Game
University of California – Davis
University of California – Santa Cruz
Oregon Department of Fish & Wildlife
Florida International University
Miccosukee Tribe of Indians
US EPA National Health & Environmental Effects Research Lab
  Atlantic Ecology Division
  Mid-Continent Ecology Division
  Western Ecology Division
US EPA Region 4
US EPA Region 9
Great Lakes Environmental Indicators Project
Cooperators

San Francisco Estuary Institute
California State Water Resources Control Board
California Department of Fish & Game
University of California – Davis
University of California – Santa Cruz
Oregon Department of Fish & Wildlife
Florida International University
Miccosukee Tribe of Indians
US EPA National Health & Environmental Effects Research Lab
  Atlantic Ecology Division
  Mid-Continent Ecology Division
  Western Ecology Division
US EPA Region 4
US EPA Region 9
Great Lakes Environmental Indicators Project
Program Context or Why are we interested in aquatic survey design and analysis?
Clean Water Act (CWA)

• Basic structure for regulating discharges of pollutants into the waters of the United States.
• Passed in 1956 as Federal Water Pollution Control Act
CWA Requires Monitoring, Reporting, and Remediation by the States

• Section 305: Water Quality Inventory
  – (a) (1) describe the specific quality of all navigable waters
  – (a)(3) identify specifically those navigable waters, the quality of which is adequate to provide for the protection and propagation of a balanced population of shellfish, fish, and wildlife and allow recreational activities in and on the water;
CWA Requirements

• (b) (1) Each State shall prepare and submit a (biennial) report which shall include --
  – (A) a description of the water quality (of the State’s waters)
  – (B) an analysis of the extent to which all navigable waters (meet designated uses)
  – (C) an analysis of the extent to which (pollutants have been eliminated & water quality attained)
Sec. 303(d): Water Quality Standards and Implementation Plans.

- Each State shall identify those waters for which the effluent limitations are not stringent enough to implement any water quality standard applicable to such waters.

- Each State shall establish for the waters identified (above) the total maximum daily load for those pollutants (identified by the EPA Administrator).
Historical Context

• CWA requires states to monitor & report on water quality, but leaves methodology up to the states
Environmental Monitoring and Assessment Program (EMAP)

- EMAP began in 1989 with the concept of using probability surveys to assess condition and determine trends in condition of environmental resources
- High priority research areas (1997 Research Plan)
  - ecological indicators
  - monitoring design
  - integration and synthesis of environmental data
Present Status

• Monitoring Design
  – Now have a general purpose design framework, the Generalized Random Tessellation Stratified Design (GRTS) with Reverse Hierarchical Ordering (RHO)
  – Evolved from EMAP’s original hexagon-based grid design
  – GRTS + RHO has been basis for EMAP aquatic sampling for several years
Generalized Random Tessellation
Stratified Design

• Uses a spatial address based on recursive partitioning to order points in 2-space
  – Split a polygon into similar sub-polygons
  – Split each sub-polygon into sub-sub-polygons
  – And so on, *ad infinitum*
Generalized Random Tessellation Stratified Design

• Address follows splitting order
  – Number the sub-polygons
  – First digit identifies sub-polygon
    • Second digit identifies sub-sub-polygon
    • And so on
HIERARCHICAL RANDOMIZATION

- At every step in the recursive partition, randomly permute the order in which sub-polygons are assigned
- Permutations chosen at random and independently from the set of all possible permutations
Let
\[ J = (0, 1] \quad J^2 = (0, 1] \times (0, 1] \]

- \( \varphi(s) \) be a measure on \( J^2 \)
- \( \pi(s) \) be an inclusion intensity function on \( J^2 \)
- \( f : J^2 \to J \) be a hierarchically randomized (quadrant) recursive function
GRTS DESIGN

• Map population domain $R$ into $(0, \frac{1}{2}] \times (0, \frac{1}{2}]$, add random offset to get image $R^* \subset \mathcal{J}^2$

• Set $F(x) = \int_{f^{-1}((0, x))} \pi(s) \phi(s) \, ds$

• $F(x)$ is a random distribution function with range $(0, M)$
GRTS DESIGN

• Pick $u_1 < u_2 < \cdots < u_M$ as a systematic sample with random start from $(0, M]$

• Set $v_i = F^{-1}(u_i)$ and $s_i = (s_{x,i}, s_{y,i}) = f^{-1}(v_i)$

• Then $S = \{s_1, s_2, \ldots, s_M\}$ is a spatially well-balanced sample from the target population with the desired inclusion function.
quadrant-recursive, hierarchical random map

\[ x = f(s) \]

\[ B(x) = \{ s \mid 0 \leq f(s) \leq x \} \]

\[ F(x) = \int_{B(x)} \pi(s) d\varphi(s) \]

systematic sample

\[ s = f^{-1}(x) \]

\[ y_1, y_2, \ldots, y_M \]

\[ x_1, x_2, \ldots, x_M \]

\[ x = F^{-1}(y) \]
GRTS DESIGN

• Because the $u_i$ are ordered, the $s_i$ are in “Quadrant-Recursive Order”, that is, successive quarters of the points will all be in the same quadrant (equi-probable case)

• Re-order using “Reverse Hierarchical Order”:
  – Pick the first point in each quarter of the sample
  – Pick the next 4 points, one from each quarter of the sample
  – and so on
Present Status

- **GRTS features**
  - Spatially balanced sample points
  - Can accommodate variable probability
  - Have an (approximately) unbiased variance estimator based on local neighborhoods
  - RHO points can be added to sample one at a time while maintaining all of the above
Present Status

• GRTS + RHO has the flexibility of SRS
  – Can do explicit spatial stratification
  – Can incorporate panel structures
  – Can dynamically adjust sample for lack of access (physical or legal)
  – Dynamically adjust for non-target sample elements
  – Can adjust sample size after sample selection while maintaining spatial balance
GRTS DESIGN

• Details:


Present Status

• Integration and Synthesis
  – Basic design-based tools in place
  – Need to utilize more sophisticated statistical tools, e.g., model-assisted and model-based approaches
Assessing the TMDL Approach to Water Quality Management

– National Research Council, June, 2001

• EPA needs to develop a uniform, consistent approach to ambient monitoring and data collection across the states

• EPA should endorse statistical approaches to defining all waters, proper monitoring design, data analysis, and impairment assessment

• Monitoring and data collection programs need to be coordinated with anticipated water quality and TMDL modeling requirements
Today, the majority of the nation's waters remain unmonitored and unassessed. Yet Section 305(b) of the CWA requires that all waters be assessed every two years... EPA believes that a probabilistic monitoring design applied over large areas, such as a state or territory, is an excellent approach to producing ...(a)… statistical representation of the extent of waters that may or may not be impaired… States and territories are encouraged to use probabilistic designs for water quality assessments
Present Status

• Issue in reaching goals of the CWA is not the existence of statistical tools
  – getting the tools into the hands of the state & tribal agencies
  – convincing the agencies to use the tools
Program Objectives

• Develop and extend the expertise on aquatic survey design and analysis to States and Tribes
  – Research driven by applications
  – Demonstration projects
  – Survey design and analysis software tools
  – Workshops, seminars, handbooks, web pages
Integration & Outreach
(OSU Project 2)

• Research projects have a specific focus, but share a common theme – design basis
• All development work is on topics provided by clients
  – Projects will work with data sets generated by clients
  – Same data sets used by multiple projects
• Monthly program meetings for OSU projects
• Frequent communication with CSU Program Director
Integration & Outreach

• Success of Program depends on having clients use techniques
  – Demonstration projects
  – Tool development
    • WED
  – Foster client use of tools
    • Design
    • Analysis
• Have contacted state agencies, EPA Laboratories, EPA Regions, STAR Grantees, NMFS
  – Actively working with some
  – Will initiate others now that we have staff
Demonstration Projects

• Sierra Nevada Amphibians – California Department of Fish & Game

• West Coast Tidal Wetland Monitoring and Assessment Venture
  – SF Bay, S. CA Coast, Coos Bay
  – WED, Reg 9, SFEI, SCCWRP, ODEQ
  • + host of others
Demonstration Projects

• Ohio EPA/Kenyon College – Cuyhoga River Wetlands
  – Ohio Rapid Assessment Method (ORAM)
• Montana Department of Environmental Quality – Wetlands Sample
• UC - Davis – Salinas River
Case Studies

• Sampling Coho salmon in Oregon coastal streams
• San Francisco Estuary Regional Monitoring Program for Trace Substances
• Surface Waters Ambient Monitoring Program
ODFW Coho Sampling Design

• There are two primary objectives of the sampling
  – describe current status
  – describe population trend

• Relevant Information:
  – There is spatial pattern in the population
  – Coho have a 3-year life cycle
  – Natural regional stratification into Monitoring Areas (MA’s) which partition ESU’s
Sampling Design

- A rotating panel design was used to balance between extensive population coverage (for status) and repeat visits to same site (for trend)

- Panel definition:
  - One panel of sites visited every year
  - Three panels of sites visited every 3 years
  - Nine panels of sites visited every 9 years
  - Twenty-seven panels of sites visited every 27 years

- Four panels visited each year: the annual panel, a 3-year panel, a 9-year panel, and a 27-year panel
Sampling Design

• Sites on streams were selected using a GRTS design
• Sites were apportioned to panels so that
  – Each panel is in-itself spatially well balanced
  – Each annual sample is spatially well balanced
  – Equi-probable sample within MA
Sampling Design

Design can also be viewed as a nested, multi-stage design:

Stage I: Sites visited at least once in 27 years
\[ S_1 = \text{all panels} \ (27 + 9 + 3 + 1 = 40 \text{ panels}) \]

Stage II: Sites visited at least 3 times in 27 years
\[ S_2 = 1, 3, \text{ and } 9 \text{ year panels} \ (13 \text{ panels}) \]

Stage III: Sites visited at least 9 times in 27 years
\[ S_3 = 1 \text{ and } 3 \text{ year panels} \ (4 \text{ panels}) \]

Stage IV: Sites visited 27 times in 27 years
\[ S_4 = 1 \text{ year panel} \]

\[ S_4 \subset S_3 \subset S_2 \subset S_1 \]
Integrated Biological & Physical Monitoring

• Original design sampled spawning coho salmon
• Modified to sample juvenile coho, coho habitat, & water quality
ODFW Collaboration

• Details described in ODFW report
• Report includes
  – Design background
  – Heavily annotated example design-based analysis
  – S-Plus functions for doing analysis
The research described in this article has been funded by the U.S. Environmental Protection Agency. This document has been prepared at the EPA National Health and Environmental Effects Research Laboratory, Western Ecology Division, in Corvallis, Oregon, through Contract 68-C6-0005 to Dynamac International, Inc and Cooperative Agreement CR82-9096-01 to Oregon State University. It has been subjected to Agency review and approved for publication. The conclusions and opinions are solely those of the authors and are not necessarily the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.
Benefits of Collaboration

• Use Coho data set for
  – Trend analysis
  – Imputation method development
  – Model-assisted analyses

• Report is a prototype for a training module

• ODFW personnel have become envoys
The following series of slides are borrowed from a presentation made by Sarah Lowe at the Annual Meeting of the San Francisco Estuary Regional Monitoring Program for Trace Substances, Oakland, CA, March 22, 2002.
San Francisco Estuary
Regional Monitoring Program
Status & Trends Component
The Design Process

Sarah Lowe, Bruce Thompson,
Rainer Hoenicke, Jon Leatherbarrow,
Robert Smith, Don Stevens,
Cristina Grosso, and the DIWG
Design Process

• Review and evaluate the hydrographic regions
• Determine the number of samples per hydrographic region for water & sediment
• Develop an optimum sampling design to address the new RMP objectives
• Select the sampling locations
Review and Evaluate the Hydrographic Regions

- Evaluating the existing segmentation scheme
- Soliciting the professional opinions
- Performing our own analyses
<table>
<thead>
<tr>
<th>Segment Boundary</th>
<th>Existing</th>
<th>Opinion</th>
<th>Water Cluster</th>
<th>Water Partition</th>
<th>Sediment</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chipps Island</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Benicia Bridge</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Carquinez Bridge</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Carquinez Straight (west end)</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Pt. Pinole</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Pt. San Pablo</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Richmond Bridge</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Angel Island</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Bay Bridge</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>San Bruno Shoal</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>San Mateo Bridge</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Dumbarton Bridge</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Sloughs</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
</tbody>
</table>
The New Segmentation Scheme has Six Main Hydrographic Regions
We determined the final number of samples per region based on:

α Statistical power analyses for key contaminants when compared to specific guidelines

α Regional Board priorities

α Funding
Key contaminants were compared to specific guidelines

α Water: compared dissolved copper to the CA Toxics Rule – WQC

α Sediment: compared copper, mercury and total PAHs to the Effects Range Low guidelines - ERL (Long et al. 1995)
Number of Samples Per Year

Water: 33 total
Sediment: 49 total
Sediment Design: Rotating Panel

- Stratified by segment
- Four panel types based on re-visit schedule: Annual, 5-year, 10-year, 20-year
- One panel of each type visited every year, every segment
- Two sample sites per panel
- Total of 8 samples per segment per year
Sediment Design

- Sample points selected using a GRTS design
- Inclusion intensity varied by segment
- Only fine grained sediment in target population, so sample included replacement sites for coarse sediment sites
SEDIMENT

9
Fixed historical stations
Annual Sample sites