Generalized Random Tessellation Stratified (GRTS)
Spatially-Balanced Survey Designs for Aquatic Resources

Anthony (Tony) R. Olsen
USEPA NHEERL
Western Ecology Division
Corvallis, Oregon
Voice: (541) 754-4790
Email: olsen.tony@epa.gov
Co-Developers

- Tony Olsen, EPA WED
- Don Stevens, Oregon State U
- Denis White, EPA WED
- Tom Kincaid, Dynamac Corp
- Christian Platt, Dynamac Corp
- Richard Remington, EPA WED
- Barbara Rosenbaum, INDUS Corporation
- David Cassell, CSC

- EMAP Surface Waters Research Group
- Users: States, EPA Regions, & others
Overview

• Aquatic resource characteristics
• Sample frame
  ▪ GIS coverages
  ▪ Imperfect representation of target population
• GRTS theory
• GRTS implementation
### Features in Space as GIS objects

<table>
<thead>
<tr>
<th>Feature</th>
<th>Points</th>
<th>Lines</th>
<th>Polygons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lakes</td>
<td>Individual lakes</td>
<td></td>
<td>Lake area</td>
</tr>
<tr>
<td>Streams</td>
<td>Segments</td>
<td>Linear network</td>
<td></td>
</tr>
<tr>
<td>Estuaries</td>
<td></td>
<td></td>
<td>Estuarine area</td>
</tr>
<tr>
<td>Wetlands</td>
<td>Depressional wetlands</td>
<td></td>
<td>Wetland area</td>
</tr>
<tr>
<td>Roads/trails</td>
<td></td>
<td>Linear network</td>
<td></td>
</tr>
<tr>
<td>Hydrologic Units</td>
<td>As points</td>
<td></td>
<td>As areas</td>
</tr>
<tr>
<td>Terrestrial</td>
<td></td>
<td></td>
<td>As areas</td>
</tr>
<tr>
<td>Vegetation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions*
Natural Resource Characteristics

• Types of natural resources
  ▪ Area polygons: large lakes and reservoirs, estuaries, coastal waters, everglades, forests, landscapes
  ▪ Linear networks: streams, rivers, roads, trails
  ▪ Discrete points: small lakes, stream reaches, prairie pothole wetlands, hydrologic units (“watersheds”), buildings

• Target population
  ▪ Finite in a bounded geographic region: collection of points
  ▪ Continuous in a bounded geographic region
    • As linear network
    • As collection of polygonal areas

• Generalizations
  ▪ Geographic region may be 1-dimensional (p-dimensional)
  ▪ “Space” may be defined by other auxiliary variables
**Typical Aquatic Sample Frames**

- GIS coverages do exist for aquatic resources
- National Hydrography Dataset (NHD)
  - Based on 1:100,000 USGS maps
  - Combination of USGS Digital Line Graph (DLG) data set and USEPA River Reach File Version 3 (RF3)
  - Includes lakes, ponds, streams, rivers
- Sample frames derived from NHD
  - Use GIS to extract frame to match target population
  - Enhance NHD with other attributes used in survey design
- Issues with NHD
  - Known to include features not of interest (over-coverage)
  - Known to exclude some aquatic resources (under-coverage)
Other Natural Resource Frames

- Omernik ecoregions digital maps
- National Wetland Inventory (NWI) digital maps
- Landcover/landuse digital maps
Generalized Random Tessellation Stratified (GRTS) Survey Designs

• Probability sample producing design-based estimators and variance estimators
• Gives another option to simple random sample and systematic sample designs
  ▪ Simple random samples tend to “clump”
  ▪ Systematic samples difficult to implement for aquatic resources and do not have design-based variance estimator
• Emphasize spatial-balance
  ▪ Every replication of the sample exhibits a spatial density pattern that closely mimics the spatial density pattern of the resource
GRTS Implementation Steps

- Concept of selecting a probability sample from a sampling line for the resource
- Create a hierarchical grid with hierarchical addressing
- Randomize hierarchical addresses
- Construct sampling line using randomized hierarchical addresses
- Select a systematic sample with a random start from sampling line
- Place sample in reverse hierarchical address order
Ohio River GRTS
Create Straight Line
All Reaches

Create Line for Ohio River (length)
Divide into 4 segments
Create Random Sequence (3, 0, 2, 1)
Assign address & color
Sort address
Repeat for each segment
Repeating Process

Divide each segment into 4 segments
Create New Random Sequences
(2,0,3,1) (1,2,0,3) (0,3,1,2) (2,3,1,0)
Assign address & colors
Sort
Selecting 16 Sample Points

Subdivide, Create Random Sequence, Assign Address & Colors
Sort Addresses
Random Starting point, Uniformly Sample Line
Assign Sequence Number to Each Point
## Reverse Hierarchical Order

<table>
<thead>
<tr>
<th>Original Order</th>
<th>Base4</th>
<th>Reverse Base4</th>
<th>Sort</th>
<th>RHO Site Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
<td>00 01 02 03 10 11 12 13 20 21 22 23 30 31 32 33</td>
<td>00 10 20 30 01 11 21 31 02 12 22 32 03 13 23 33</td>
<td>00 01 02 03 10 11 12 13 20 21 22 23 30 31 32 33</td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16</td>
</tr>
</tbody>
</table>

### Instructions
- Create Base4 Addresses
- Reverse Address Digits
- Sort
- Assign RHO Site Nos.
Map Sites

RHO

Original Line (unsorted)
Selecting a Probability Sample from a Sampling Line: Linear Network Case

- Place all stream segments in frame on a linear line
  - Preserve segment length
  - Identify segments by ID
- In what order do place segments on line?
  - Randomly
  - Systematically (minimal spanning tree)
  - Randomized hierarchical grid
- Systematic sample with random start
  - \( k = \frac{L}{n} \), \( L \) = length of line, \( n \) = sample size
  - Random start \( d \) between \([0,k)\)
  - Sample: \( d + (i-1)*k \) for \( i=1,…,n \)
Selecting a Probability Sample from a Sampling Line: Point and Area Cases

- **Point Case:**
  - Identify all points in frame
  - Assign each point unit length
  - Place on sample line

- **Area Case:**
  - Create grid covering region of interest
  - Generate random points within each grid cell
  - Keep random points within resource (A)
  - Assign each point unit length
  - Place on sample line
Randomized Hierarchical Grid

• Step 1: Frame: Large lakes: blue; Small lakes: pink; Randomly place grid over the region
• Step 2: Sub-divide region and randomly assign numbers to sub-regions
• Step 3: Sub-divide sub-regions; randomly assign numbers independently to each new sub-region; create hierarchical address. Continue sub-dividing until only one lake per cell.
• Step 4: Identify each lake with cell address; assign each lake length 1; place lakes on line in numerical cell address order.
Hierarchical Grid Addressing

213: hierarchical address
Population of 120 points
Reverse Hierarchical Order

- Construct reverse hierarchical order
  - Order the sites from 1 to n
  - Create base 4 address for numbers
  - Reverse base 4 address
  - Sort by reverse base 4 address
  - Renumber sites in RHO
- Why use reverse hierarchical order?
  - Results in any contiguous set of sample sites being spatially-balanced
  - Consequence: can begin at the beginning of list and continue using sites until have required number of sites sampled in field

<table>
<thead>
<tr>
<th>RHO</th>
<th>Reverse Base4</th>
<th>Base4</th>
<th>Original Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>00</td>
<td>00</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>01</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>02</td>
<td>20</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>03</td>
<td>30</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>01</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>21</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>31</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>20</td>
<td>02</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>21</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>23</td>
<td>32</td>
<td>15</td>
</tr>
<tr>
<td>13</td>
<td>30</td>
<td>03</td>
<td>4</td>
</tr>
<tr>
<td>14</td>
<td>31</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>32</td>
<td>23</td>
<td>12</td>
</tr>
<tr>
<td>16</td>
<td>33</td>
<td>33</td>
<td>16</td>
</tr>
</tbody>
</table>
Unequal Probability of Selection

- Assume want large lakes to be twice as likely to be selected as small lakes
- Instead of giving all lakes same unit length, give large lakes twice unit length of small lakes
- To select 5 sites divide line length by 5 (11/5 units); randomly select a starting point within first interval; select 4 additional sites at intervals of 11/5 units
- Same process is used for points and areas (using random points in area)
Complex Survey Designs based on GRTS

- Stratified GRTS: apply GRTS to each stratum
- Unequal probability GRTS: adjust unit length based on auxiliary information (e.g., lake area, strahler order, basin, ecoregion)
- Oversample GRTS:
  - Design calls for n sites; some expected non-target, landowner denial, etc; select additional sites to guarantee n field sampled
  - Apply GRTS for sample size 2n; place sites in RHO; use sites in RHO
- Panels for surveys over time
- Nested subsampling
- Two-stage sampling using GRTS at each stage
  - Example: Select USGS 4th field Hucs; then stream sites within Hucs
Two GRTS samples: Size 30
Spatial Balance: 256 points
Spatial Balance: With oversample

Voronoi Polygons

Uniform Sample

GRTS Sample

Small ← Polygon Area → Large
Ratio of GRTS to SRS Voronoi polygon size variance

- Continuous domain with no voids
- Constant polygon size, total perimeter = 88.4
- Linearly increasing polygon size, total perimeter = 84.9
- Exponentially increasing polygon size, total perimeter = 43.1
Impact on Variance Estimators of Totals
RF3 Stream Length: EMAP West

- Total
- 1st
- 2nd
- 3rd
- 4th +

Strahle Length (1,000 km)

- Non Perennial
- Perennial

Length (1,000 km)

RESEARCH & DEVELOPMENT
Building a scientific foundation for sound environmental decisions
Perennial Streams GRTS sample
Variable Density Sample for the Central Region of the Southern California Bight

Levels of Response Evaluation
- Potential sample point
  - Level 1: Water quality and sediment-based indicators
  - Level 2: Fish assemblages + Level 1 indicators
  - Level 3: Sediment toxicity and fish tissue + Level 2 indicators

Publicly Owned Treatment Works
- Hyperion
- Los Angeles County

River and Storm Outflow
- Malibu Creek
- Ballona Creek

Bathymetry (m)
- 10
- 100
- 200

Western Ecology Division
Corvallis, Oregon
April 18, 2000

Building a scientific foundation for sound environmental decisions
### RF3 Sample Frame: Lakes

<table>
<thead>
<tr>
<th>Lake Area (ha)</th>
<th>Number of Lakes</th>
<th>Percent</th>
<th>Cumulative Number of Lakes</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–5</td>
<td>172,747</td>
<td>63.8</td>
<td>172,747</td>
<td>63.8</td>
</tr>
<tr>
<td>5–10</td>
<td>44,996</td>
<td>16.6</td>
<td>217,743</td>
<td>80.4</td>
</tr>
<tr>
<td>10–50</td>
<td>40,016</td>
<td>14.8</td>
<td>257,759</td>
<td>95.2</td>
</tr>
<tr>
<td>50–500</td>
<td>11,228</td>
<td>4.1</td>
<td>268,987</td>
<td>99.3</td>
</tr>
<tr>
<td>500–5000</td>
<td>1,500</td>
<td>0.6</td>
<td>270,387</td>
<td>99.9</td>
</tr>
<tr>
<td>&gt;5000</td>
<td>274</td>
<td>0.1</td>
<td>270,761</td>
<td>100.0</td>
</tr>
</tbody>
</table>
National Fish Tissue Contaminant Lake Survey
## Sample Selected: Lakes

<table>
<thead>
<tr>
<th>Lake Area (ha)</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>All Years</th>
<th>Expected Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>39</td>
<td>41</td>
<td>47</td>
<td>47</td>
<td>174</td>
<td>938.84</td>
</tr>
<tr>
<td>5-10</td>
<td>44</td>
<td>40</td>
<td>47</td>
<td>46</td>
<td>177</td>
<td>261.61</td>
</tr>
<tr>
<td>10-50</td>
<td>32</td>
<td>47</td>
<td>46</td>
<td>25</td>
<td>150</td>
<td>256.51</td>
</tr>
<tr>
<td>50-500</td>
<td>34</td>
<td>37</td>
<td>29</td>
<td>34</td>
<td>134</td>
<td>85.06</td>
</tr>
<tr>
<td>500-5000</td>
<td>36</td>
<td>30</td>
<td>31</td>
<td>41</td>
<td>138</td>
<td>11.36</td>
</tr>
<tr>
<td>&gt;5000</td>
<td>40</td>
<td>30</td>
<td>25</td>
<td>32</td>
<td>127</td>
<td>2.21</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>225</td>
<td>225</td>
<td>225</td>
<td>225</td>
<td>900</td>
<td></td>
</tr>
</tbody>
</table>
GRTS Sample of Streams

All Streams and All Sites

Category 1

Category 2

Category 3
GRTS Site Selection using psurvey.design

• Create sample frame as shape file
  ▪ ArcGIS, Arcview
  ▪ Shape file must contain points, lines, or polygons
  ▪ Make sure a “.prj” file is produced so have projection information

• Select sample: R program
  ▪ Start R program
  ▪ Load psurvey.design library (install first if not done yet)
  ▪ Input:
    ▪ Read dbf file in R
    ▪ Specify the survey design
    ▪ Select sites using “grts” command
  ▪ Output:
    ▪ Point shapefile of selected sites
# Load psurvey.design library
# Read dbf file
att <- read.dbffile('eco_l3_ut')
head(att)
# specify design
Equaldsgn <- list(None=list(panel=c(PanelpOne=115),
                        seltype='Equal'))

Equalsites <- grts(design=Equaldsgn,
                    src.frame='shapefile',
                    in.shape='eco_l3_ut',
                    att.frame=att,
                    type.frame='area',
                    DesignID='UTEco3EQ',
                    shapefile=TRUE,
                    prj='eco_l3_ut',
                    out.shape='Eco3.EqualSites')

Example
R script to sample Utah landscape
Comments

• GRTS using R can be applied in one dimension
• GRTS conceptually extends to sampling 3-d or greater dimensions
• X,Y coordinates can be any continuous variables
GRTS Implementation Process

- Randomly place a square over the sample frame geographic region
- Construct hierarchical grid (e.g. ‘quadtree’) with hierarchical addressing in the square
- Construct Peano mapping of two-space to one-space using hierarchical addressing
- Complete hierarchical randomization of Peano map
- Place sample frame elements on line in one-space using hierarchical randomization order, assigning length to element based on frame and inclusion density (unequal probability)
- Select systematic sample with random start from line
- Place sample in reverse hierarchical order