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

- Don Stevens, Oregon State U
- Denis White, EPA WED
- Richard Remington, EPA WED
- Barbara Rosenbaum, INDUS Corporation
- David Cassell, CSC

- EMAP Surface Waters Research Group
- State monitoring staff

Overview

- Aquatic resource characteristics
- Sample frame
  - GIS coverages
  - Imperfect representation of target population
- GRTS theory
- GRTS implementation
  - Old: ArcInfo, SAS, C-program
  - New: R program with GIS coverage preparation

Aquatic Resource Characteristics

- Types of aquatic resources
  - Area polygons: large lakes and reservoirs, estuaries, coastal waters, everglades
  - Linear networks: streams and rivers
  - Discrete points: small lakes, stream reaches, prairie pothole wetlands, hydrologic units (“watersheds”)
- 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)

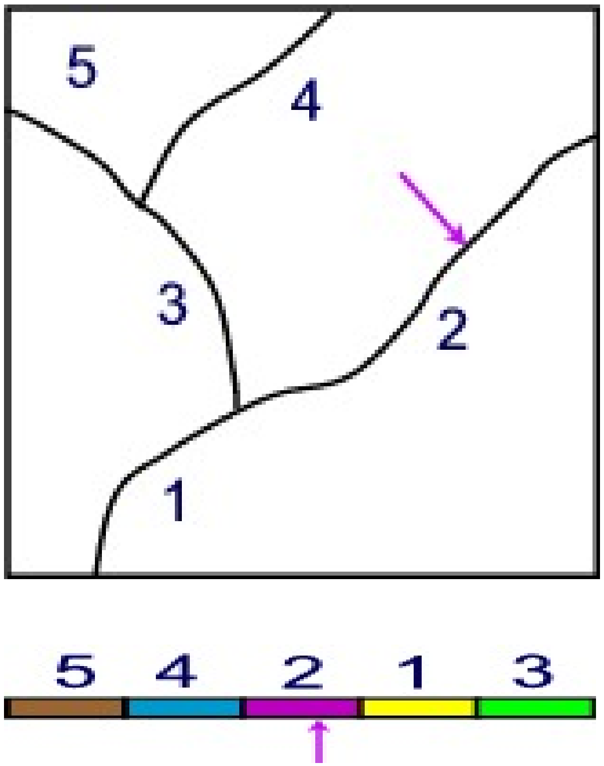
Generalized Random Tessellation Stratified (GRTS) Survey Designs

- Probability sample producing design-based estimators and variance estimators
- Give 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

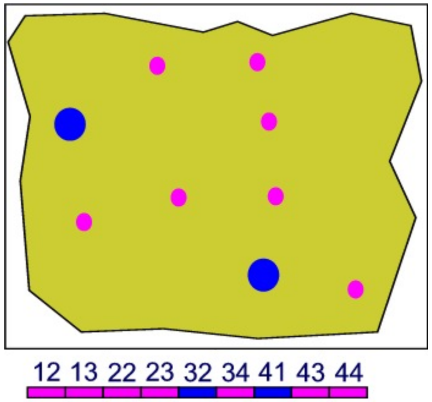
- Concept of selecting a probability sample from a sampling line for the resource
- Create a hierarchical grid with hierarchical addressing
- Randomize hierarchical addresses
- Construct sampling line using randomized hierarchical addresses
- Select a systematic sample with a random start from sampling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

Selecting a Probability Sample from a Sampling Line: Linear Network Case

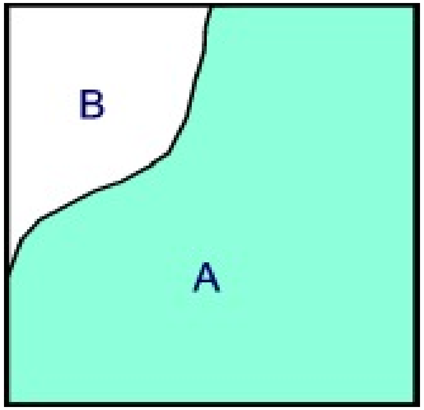


- Place all stream segments in frame on a linear line
  - Preserve segment length
  - Identify segments by ID
- In what order do place segments on line?
  - Randomly
  - Systematically (minimal spanning tree)
  - Randomized hierarchical grid
- Systematic sample with random start
  - k=L/n, L=length of 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

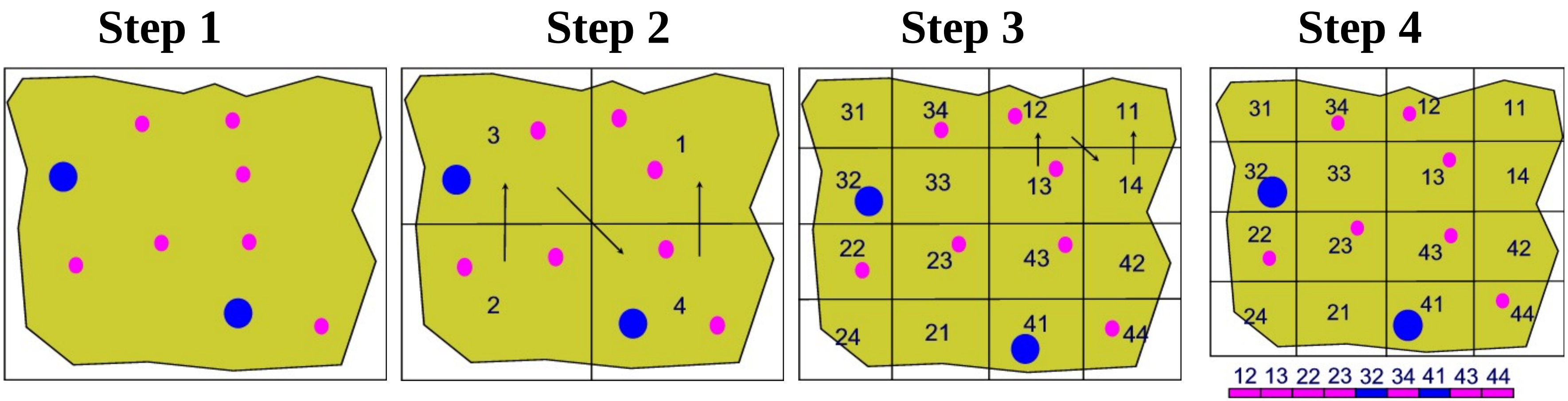


- Point Case:
  - Identify all points in frame
  - Assign each point unit length
  - Place on sample line



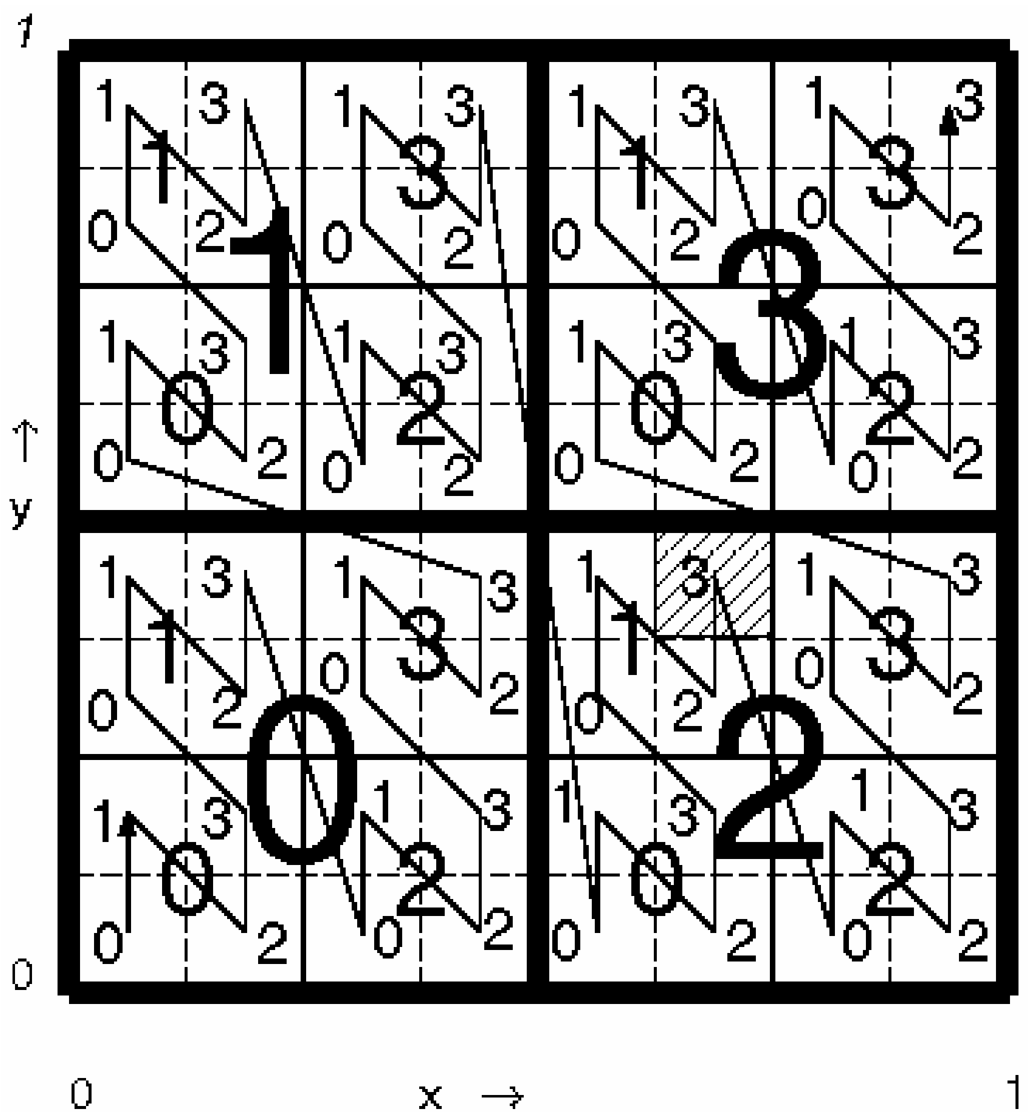
- Area Case:
  - Create grid covering region of interest
  - Generate random points within each grid cell
  - Keep random points within resource (A)
  - Assign each point unit length
  - Place on sample line

Randomized Hierarchical Grid



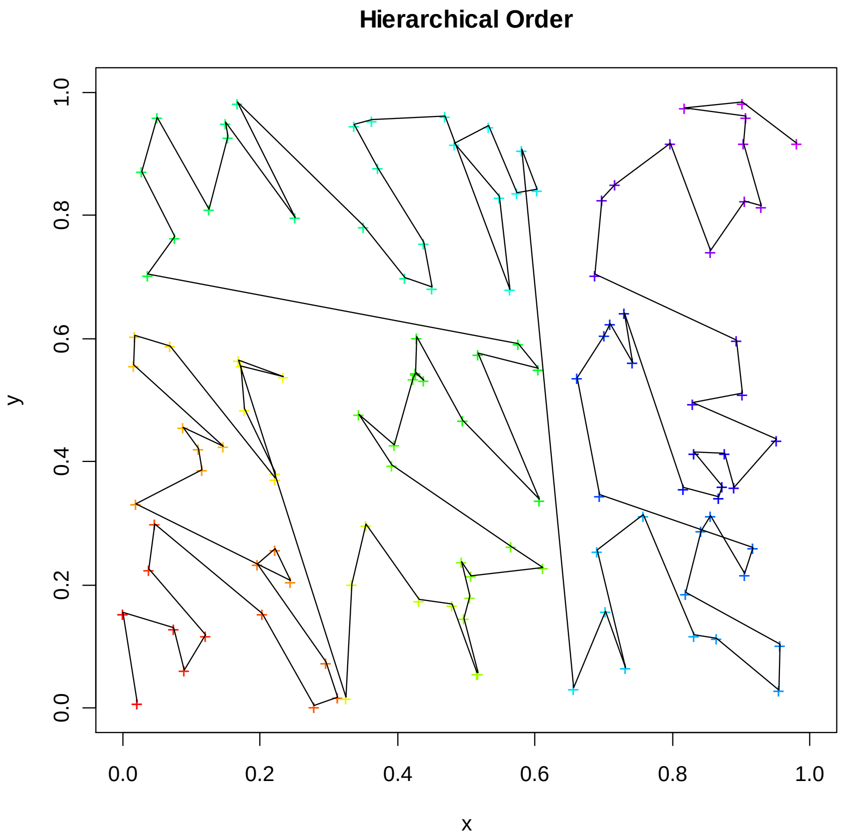
- Step 1: Frame: Large lakes: blue; Small lakes: pink; Randomly place grid over the region
- Step 2: Sub-divide region and randomly assign numbers to sub-regions
- Step 3: Sub-divide sub-regions; randomly assign numbers independently to each new sub-region; create hierarchical address. Continue sub-dividing 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.

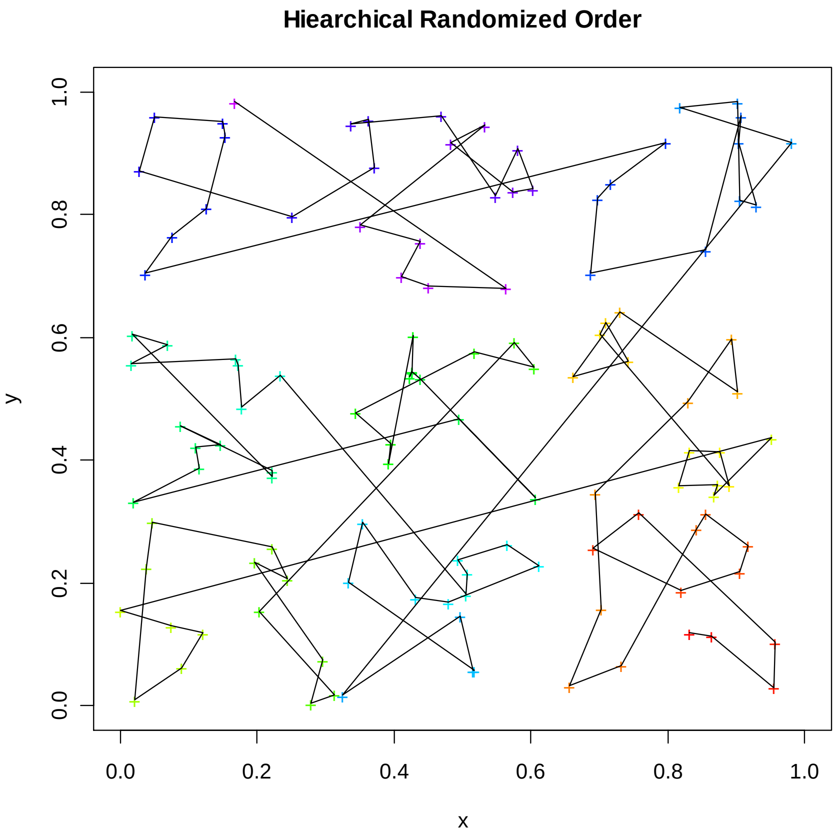
Hierarchical Grid Addressing



213: hierarchical address

Population of 120 points



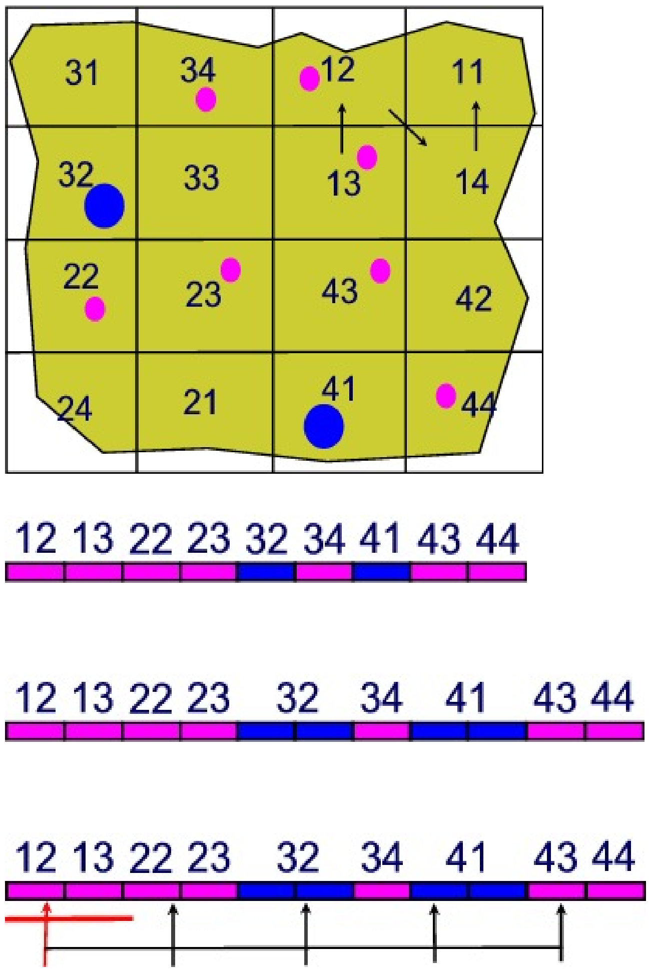


RHO	Reverse Base4	Base4	Original Order
1	00	00	1
2	01	10	5
3	02	20	9
4	03	30	13
5	10	01	2
6	11	11	6
7	12	21	10
8	13	31	14
9	20	02	3
10	21	12	7
11	22	22	11
12	23	32	15
13	30	03	4
14	31	13	8
15	32	23	12
16	33	33	16

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

Unequal Probability of Selection

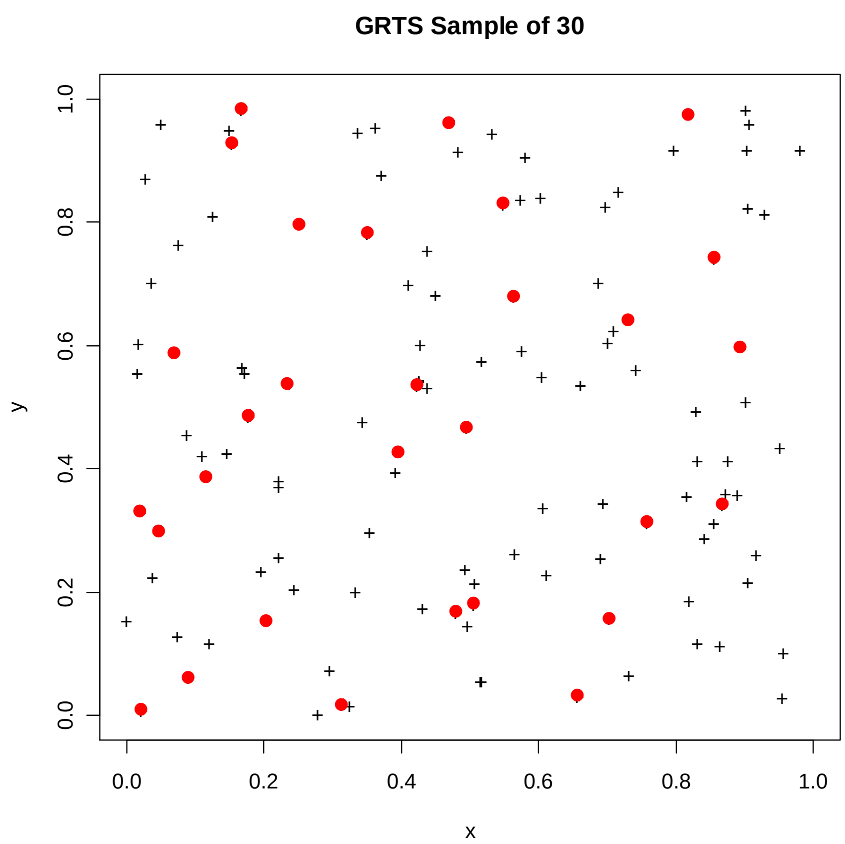


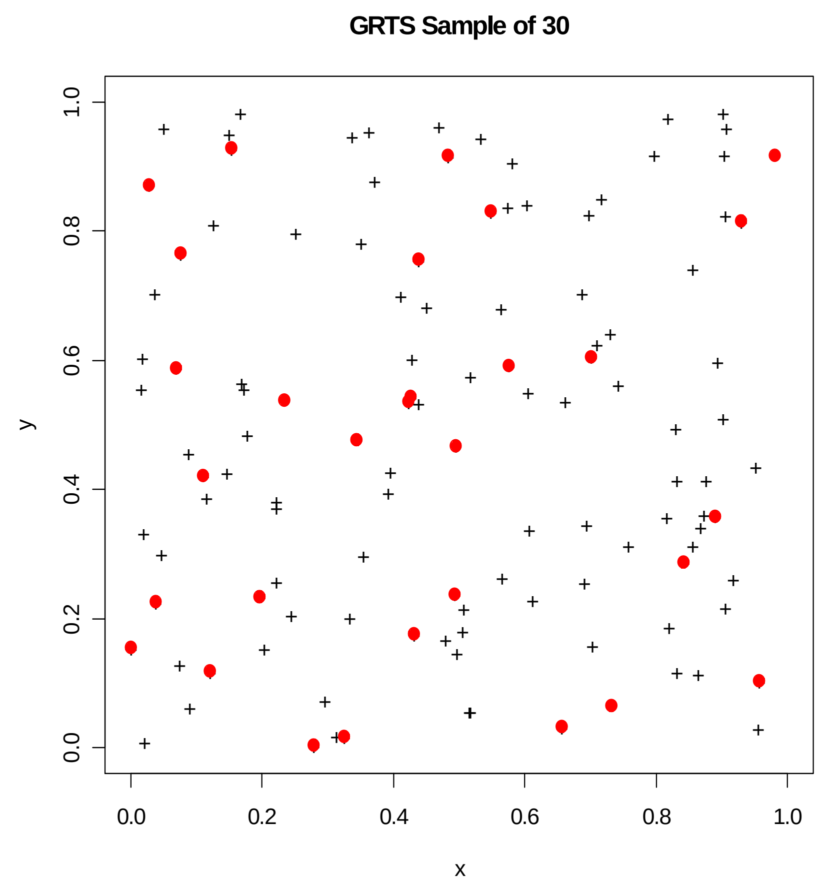
- Assume want large lakes to be twice as likely to be selected as small lakes
- Instead of giving all lakes same unit length, give large lakes twice unit length of small lakes
- To select 5 sites divide line length by 5 (11/5 units); randomly select a starting point within first interval; select 4 additional 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

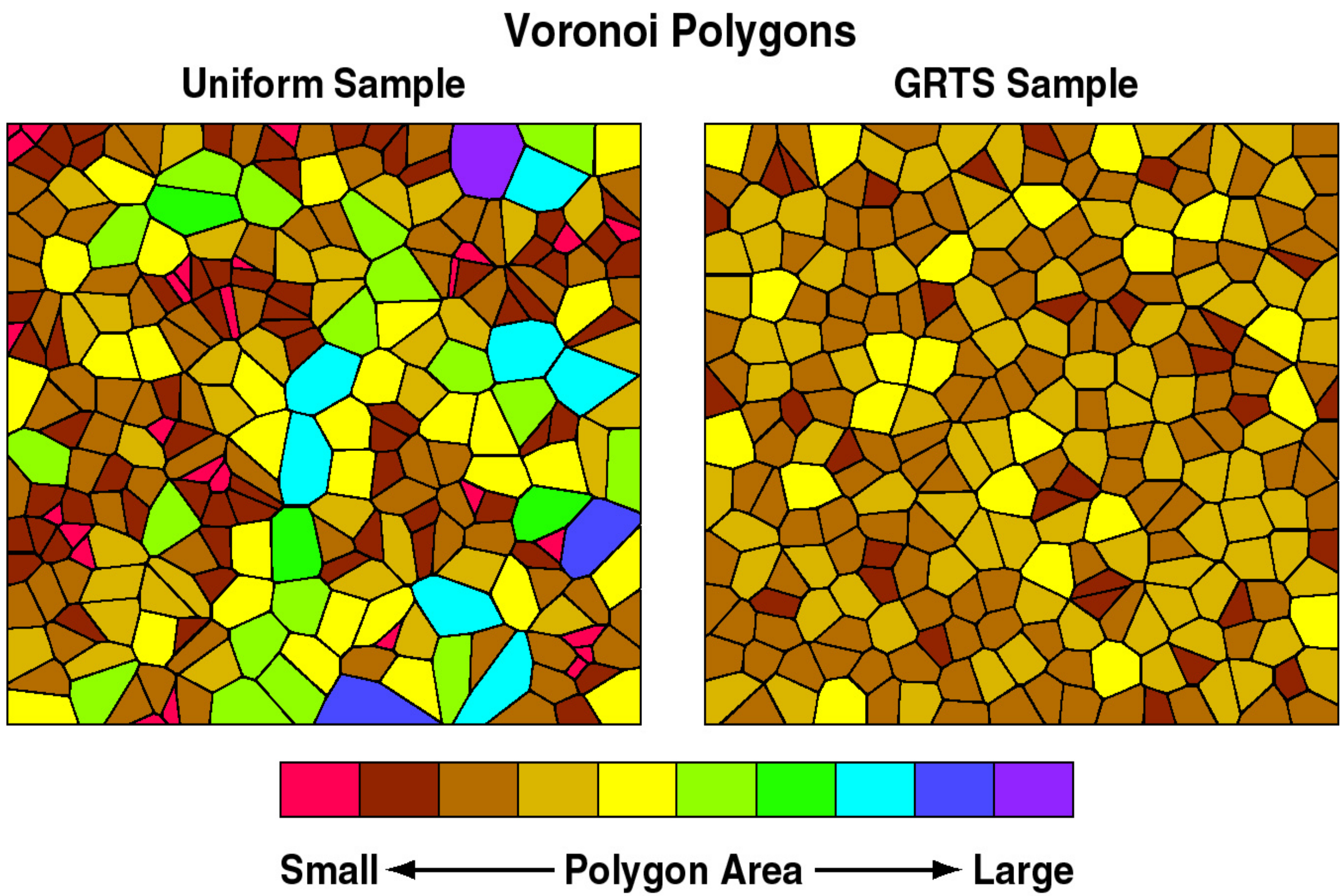
- Stratified GRTS: apply GRTS to each stratum
- Unequal probability GRTS: adjust unit length based on auxiliary information (eg lake area, strahler order, basin, ecoregion)
- Oversample GRTS:
  - Design calls for n sites; some expected non-target, landowner denial, etc; select additional sites to guarantee 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 4^{th} field Hucs; then stream sites within Hucs

Two GRTS samples: Size 30

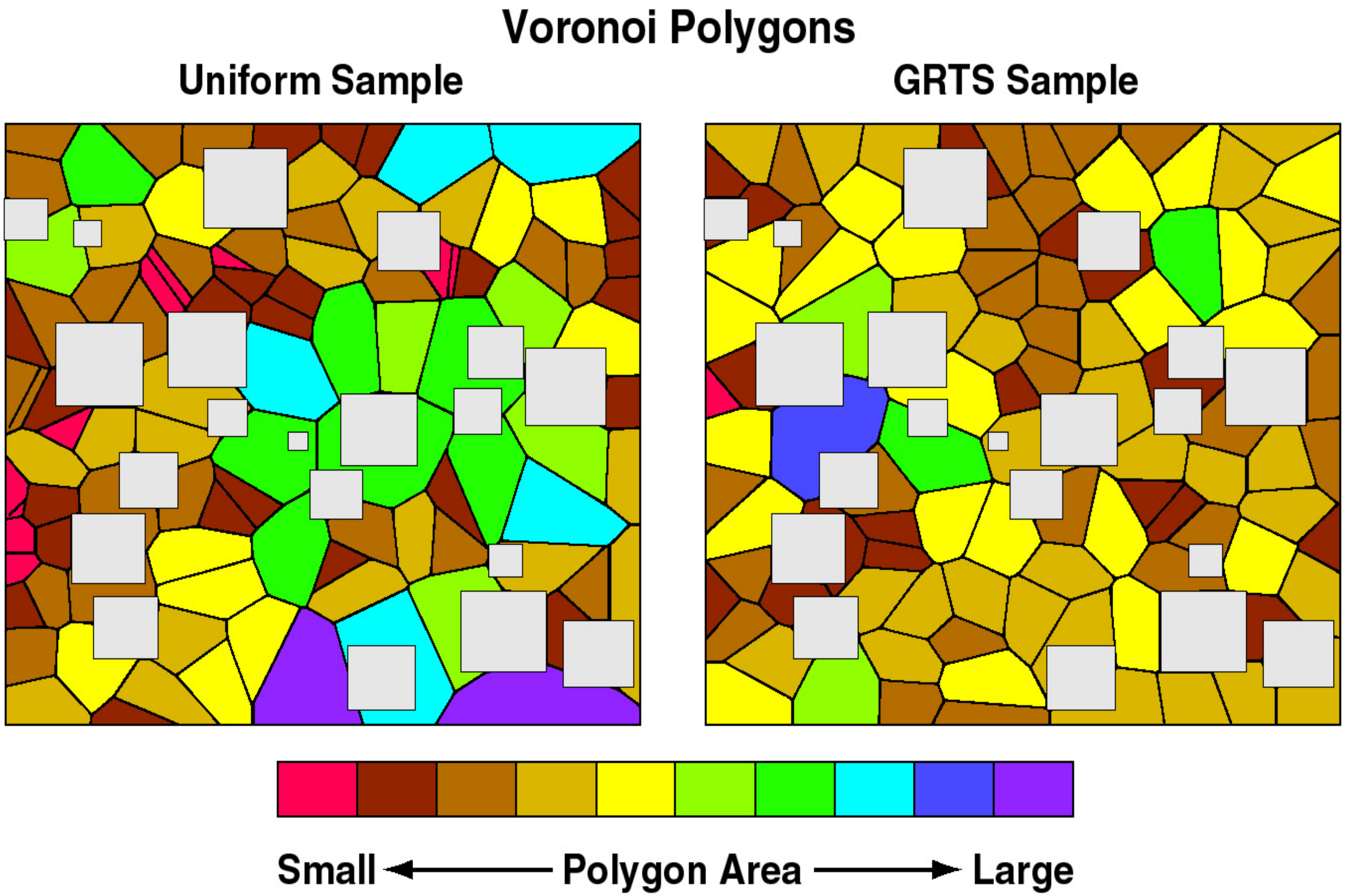




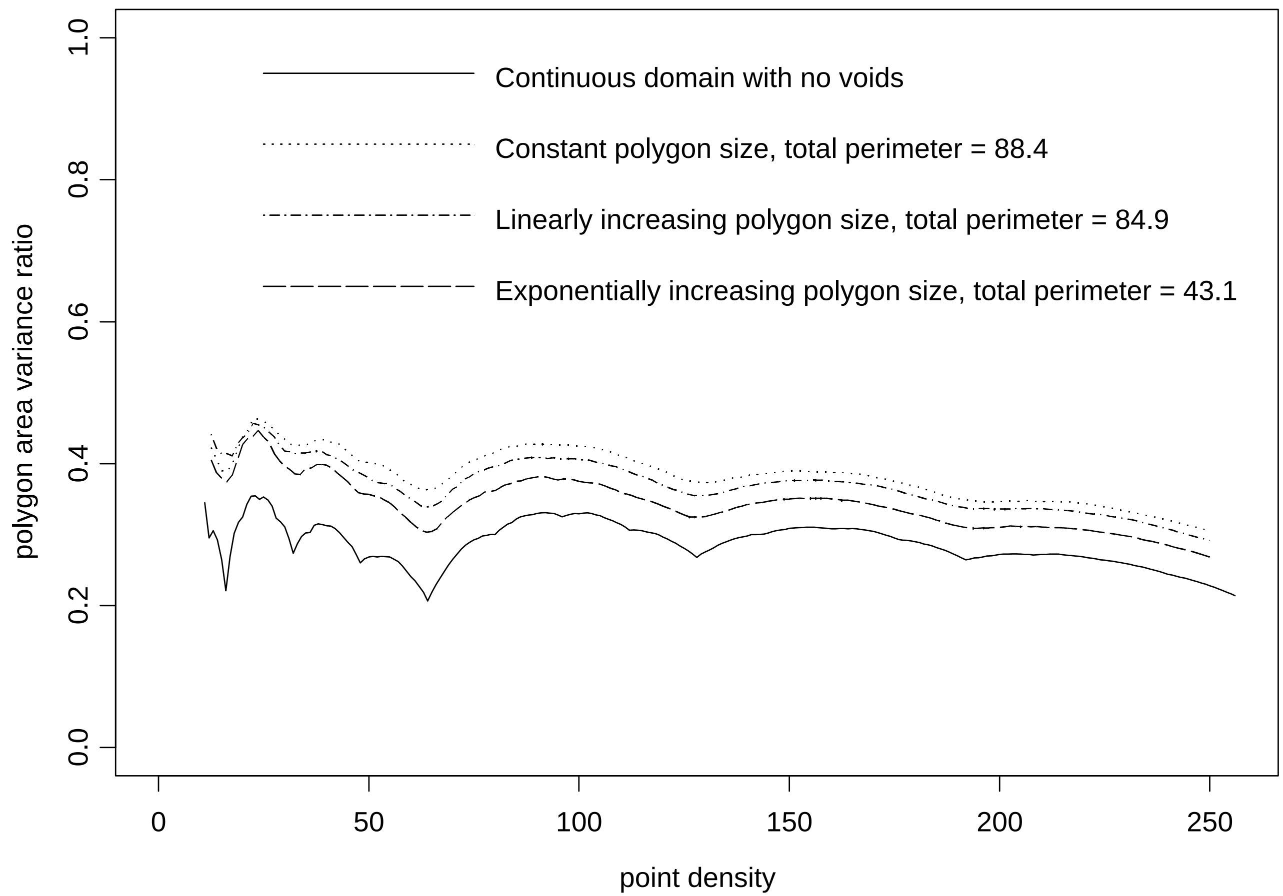
Spatial Balance: 256 points



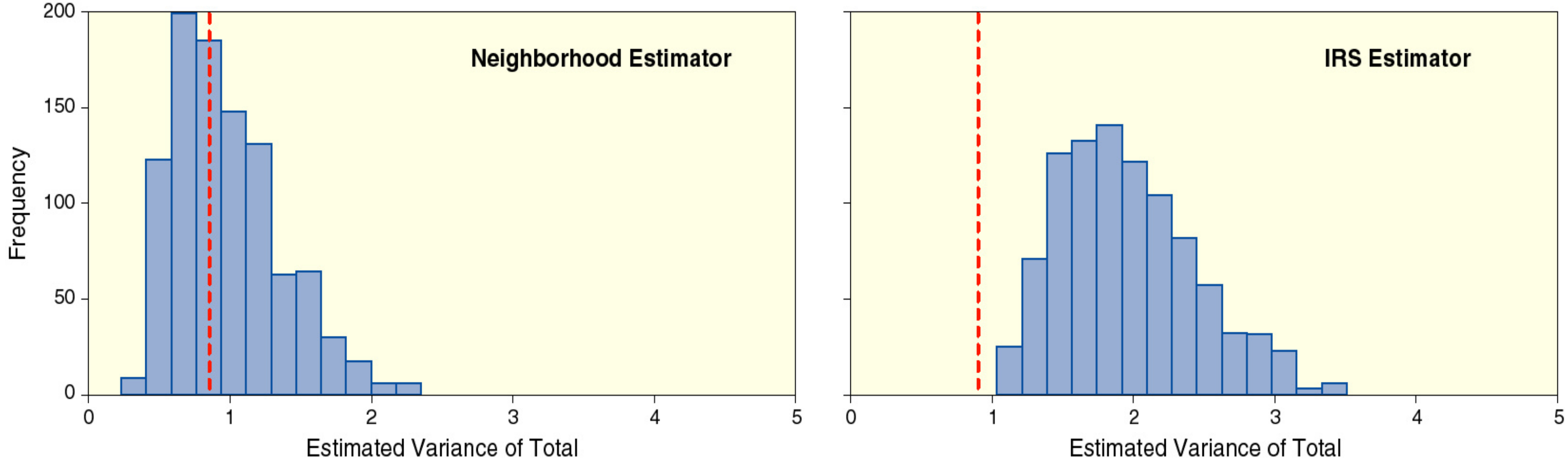
Spatial Balance: With oversample



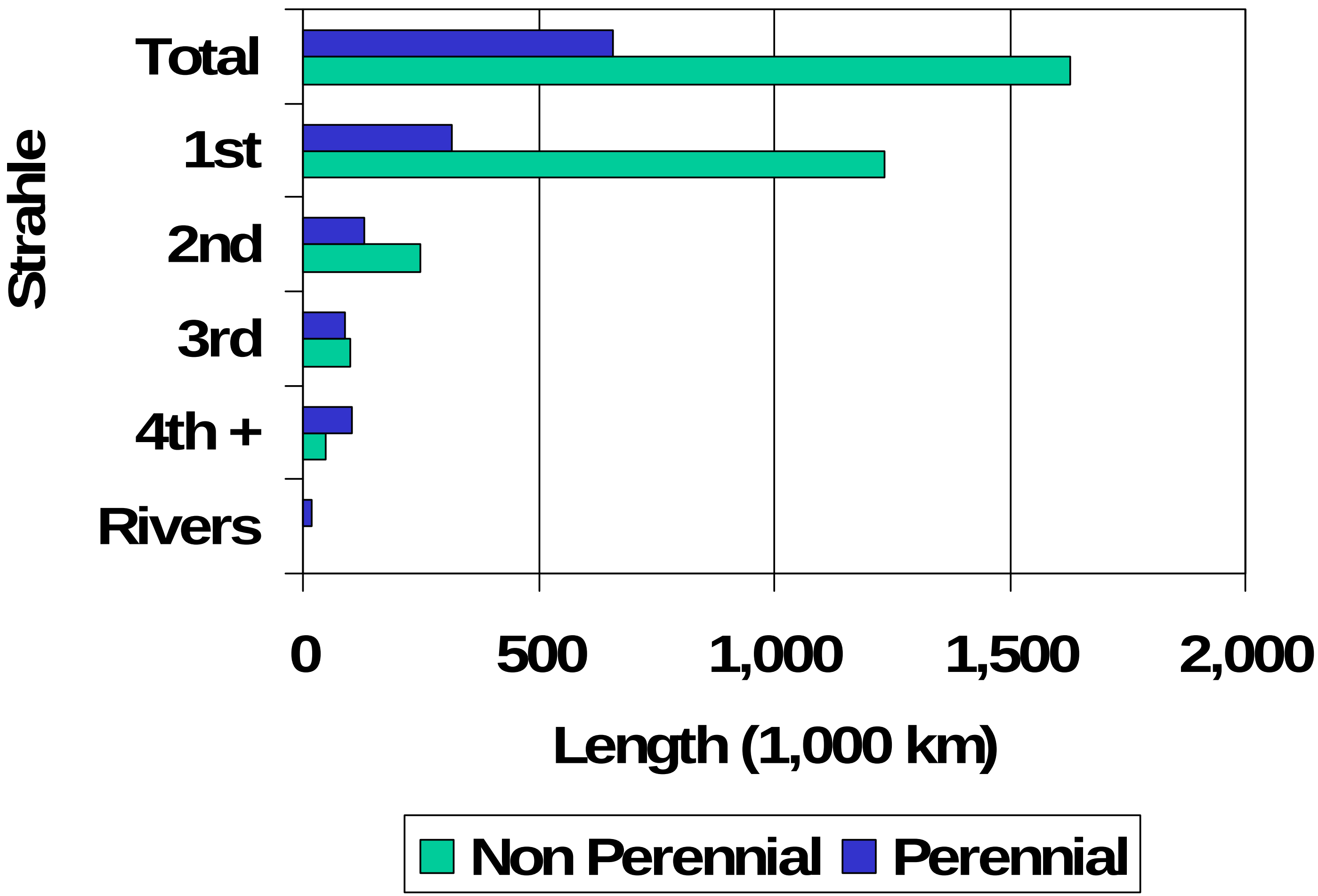
Ratio of GRTS to SRS Voronoi polygon size variance



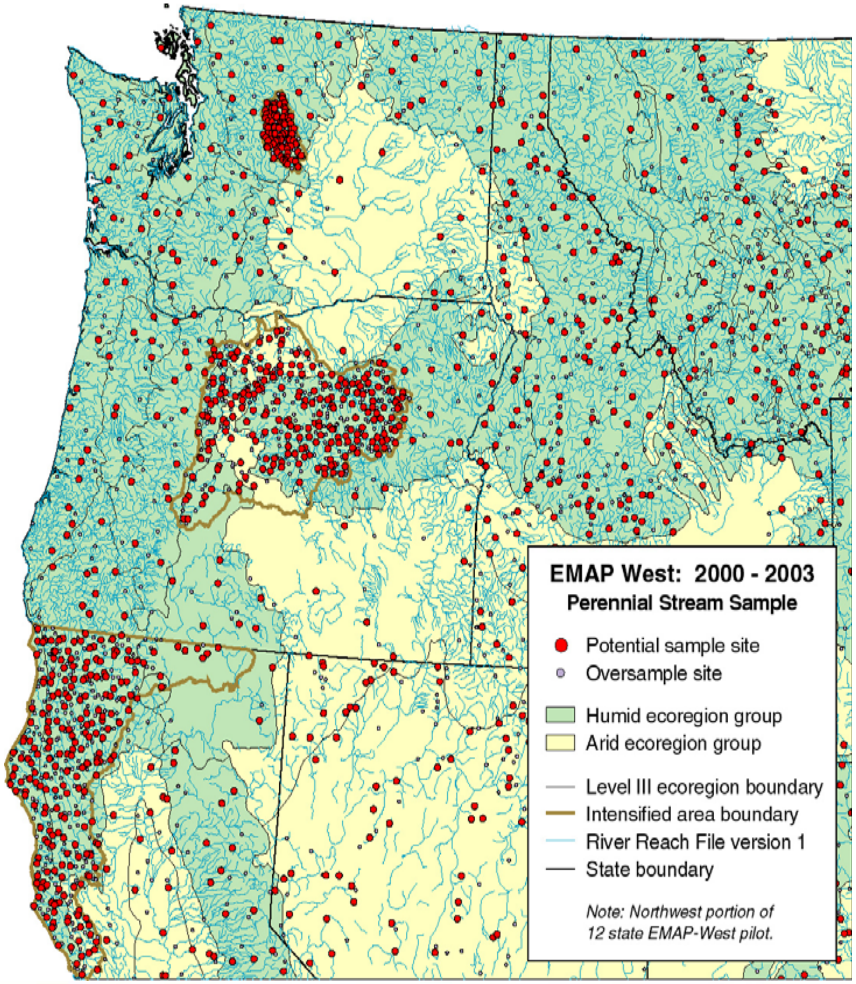
Impact on Variance Estimators of Totals

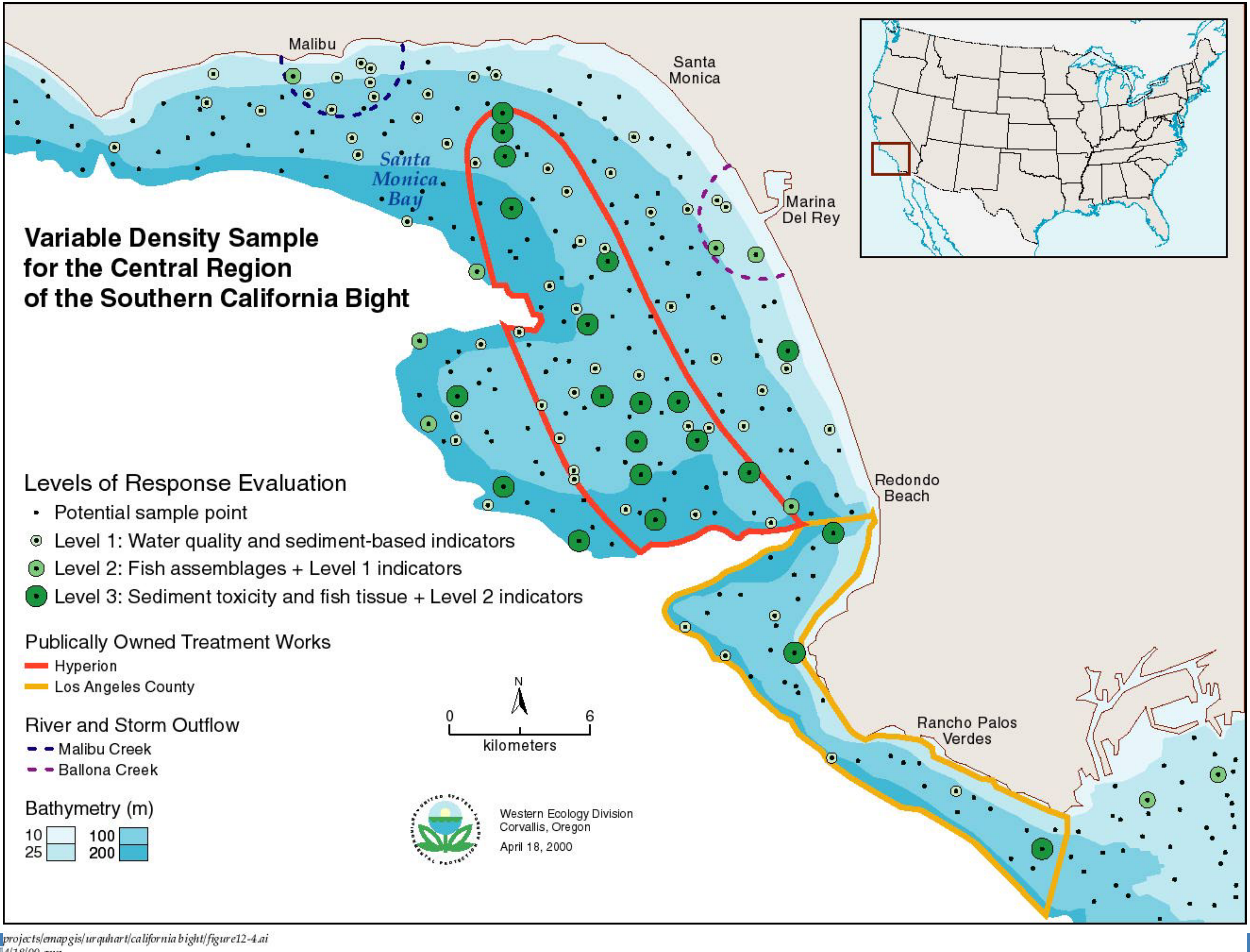


RF3 Stream Length: EMAP West



Perennial Streams GRTS sample

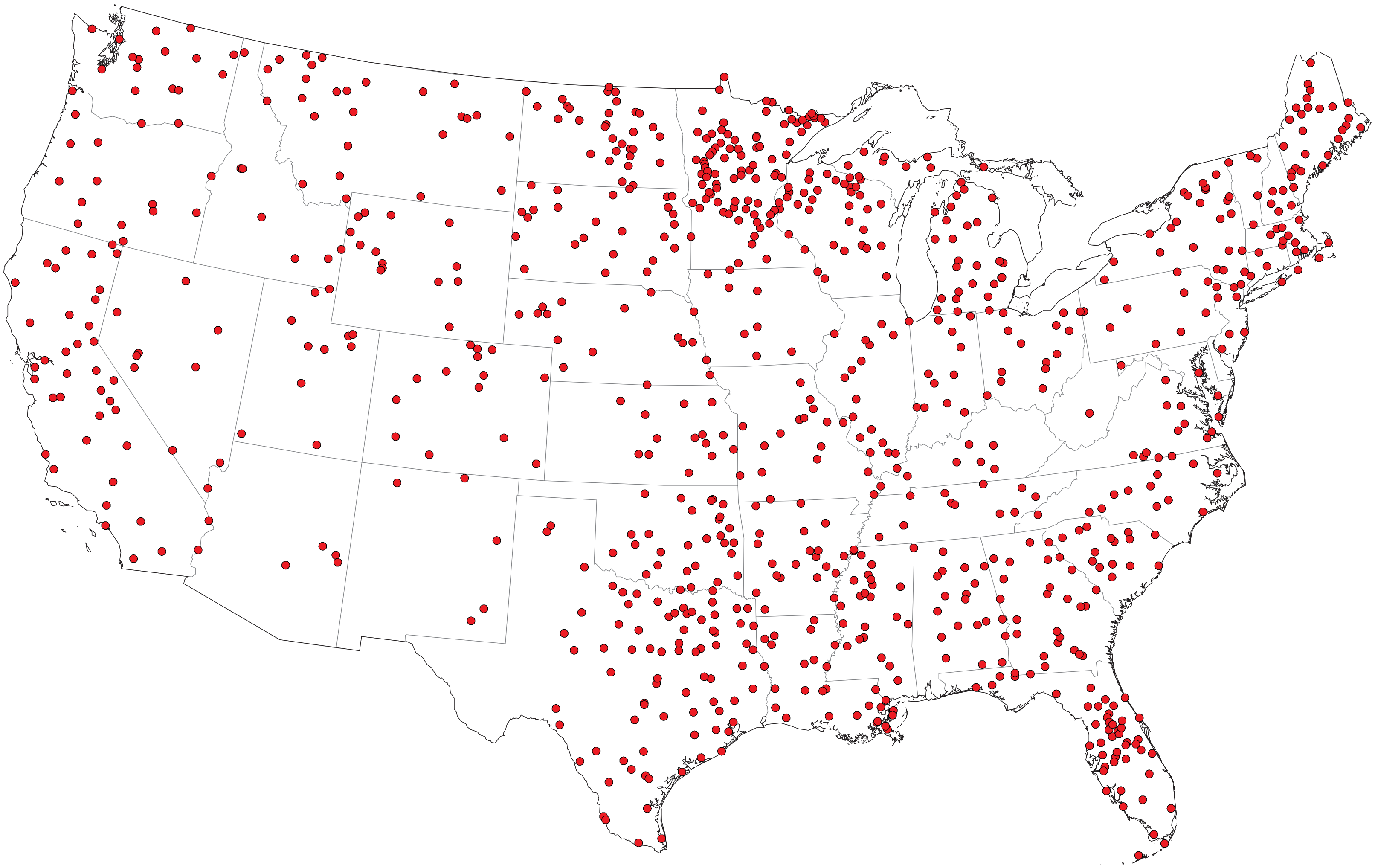



Malibu
Santa Monica
Santa Monica Bay
Marina Del Rey
Redondo Beach
Rancho Palos Verdes
Variable Density Sample for the Central Region of the Southern California Bight
Levels of Response Evaluation
Potential sample point
Level 1: Water quality and sediment-based indicators
Level 2: Fish assemblages + Level 1 indicators
Level 3: Sediment toxicity and fish tissue + Level 2 indicators
Publically Owned Treatment Works
Hyperion
Los Angeles County
River and Storm Outflow
Malibu Creek
Ballona Creek
Bathymetry (m)
10
25
100
200
N
0
6
kilometers
Western Ecology Division
Corvallis, Oregon
April 18, 2000
projects/emapgis/urquhart/californiabight/figure12-4.ai
4/18/00 smp

RF3 Sample Frame: Lakes

Lake Area (ha)	Number of Lakes	Percent	Cumulative Number of Lakes	Cumulative Percent
1–5	172,747	63.8	172,747	63.8
5–10	44,996	16.6	217,743	80.4
10–50	40,016	14.8	257,759	95.2
50–500	11,228	4.1	268,987	99.3
500–5000	1,500	0.6	270,387	99.9
>5000	274	0.1	270,761	100.0

National Fish Tissue Contaminant Lake Survey

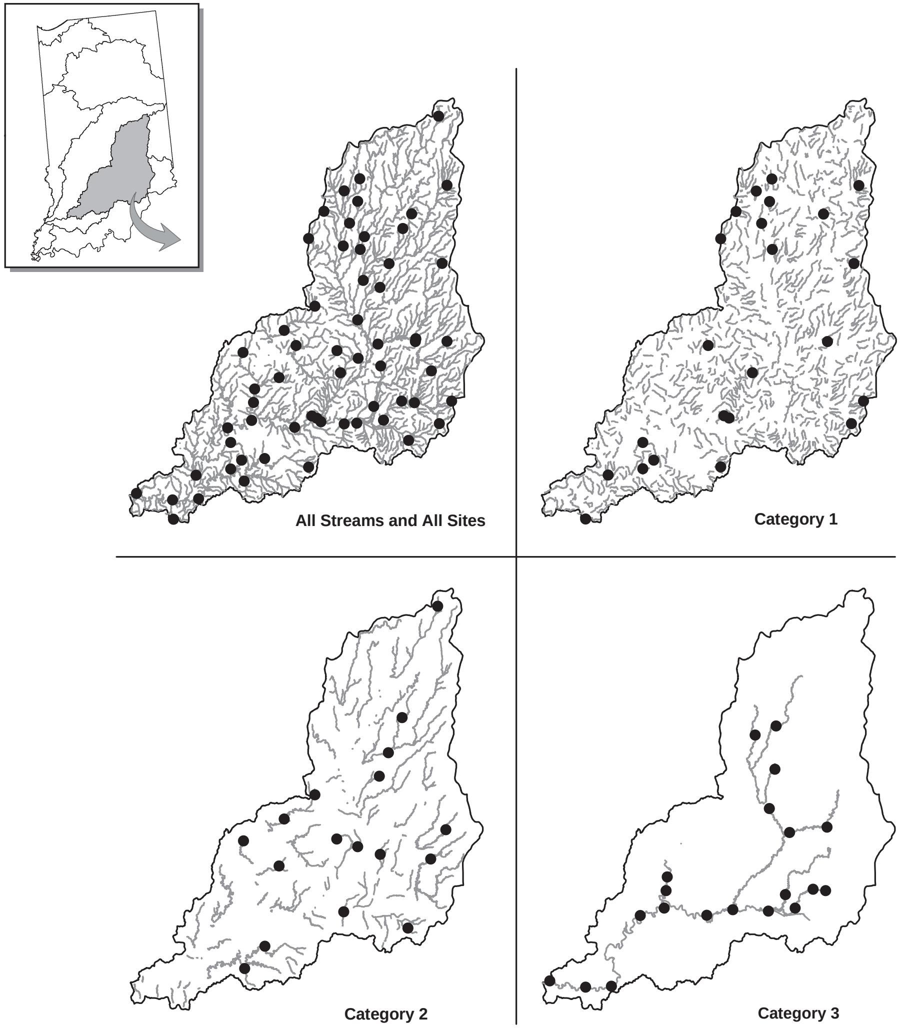


US EPA NHEERL-WED EMAP Stat&Design j199.ow.lakes/plots/owsampall.ai 5/5/1999

Sample Selected: Lakes

Lake Area (ha)	1999	2000	2001	2002	All Years	Expected Weight
1-5	39	41	47	47	174	938.84
5-10	44	40	47	46	177	261.61
10-50	32	47	46	25	150	256.51
50-500	34	37	29	34	134	85.06
500-5000	36	30	31	41	138	11.36
>5000	40	30	25	32	127	2.21
Total	225	225	225	225	900	

GRTS Sample of Streams



Initial Software Implementation

- Hierarchical grid creation: C-program
- Extract sample frame from RF3/NHD: Arc/Info
- Intersect frame with hierarchical grid: Arc/Info
- Export data and summarize frame to determine inclusion densities for unequal probability sampling: SAS
- Complete hierarchical randomization, systematic sample of line, reverse hierarchical ordering: SAS
- Import sample to create design file with geographic coordinates: Arc/Info

New Implementation

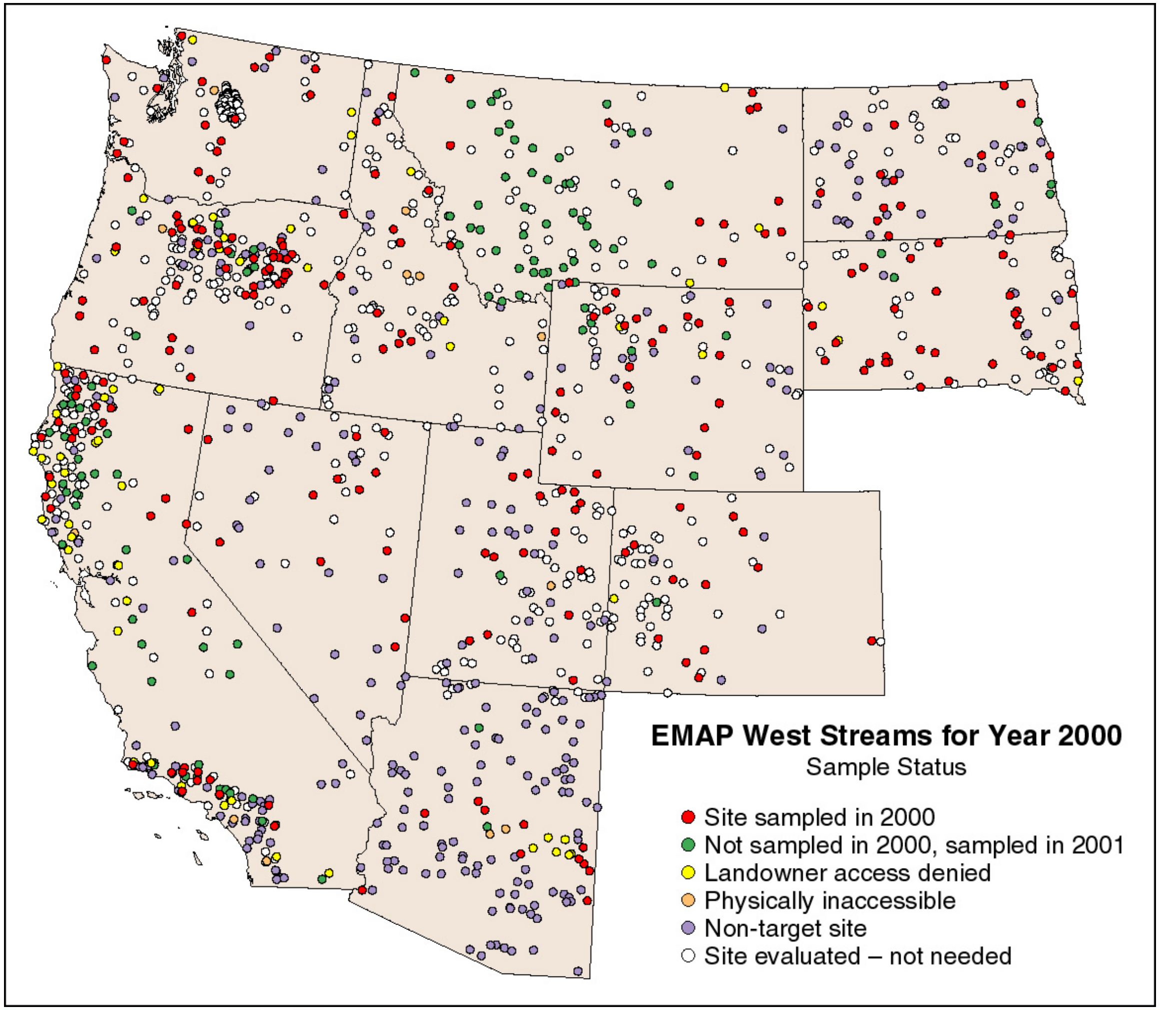
- Extract sample frame from RF3/NHD
  - Arc/Info, Arcview, ….
  - Required data format for points, lines, polygons
- Select sample: R program
  - Input:
    - Survey design specification
    - Sample frame data
  - Output:
    - Data frame with sample

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
- Software implementation
  - GIS preparation followed by R program
    - Incorporate simple GIS operations in R operation
  - Add extension to Arcview

VISIT PATTERN FOR ODFW PANELS

Over-sample: use in implementation



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