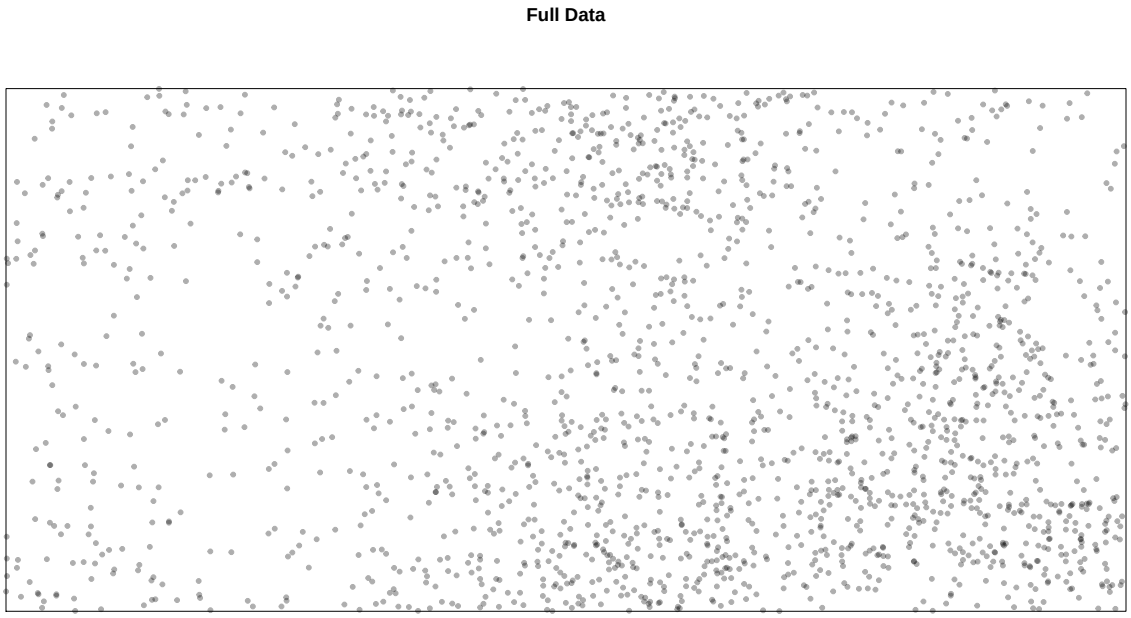
Example Spatial Sampling Paths

Kenny Flagg

May 15, 2020

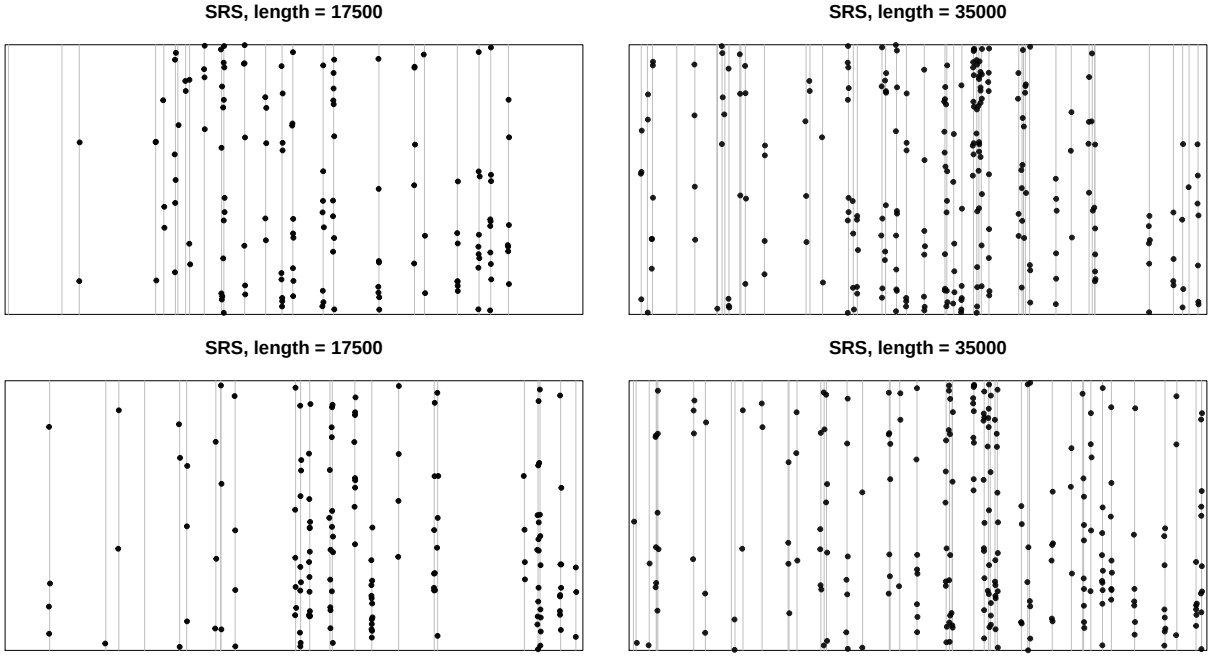
Full Dataset



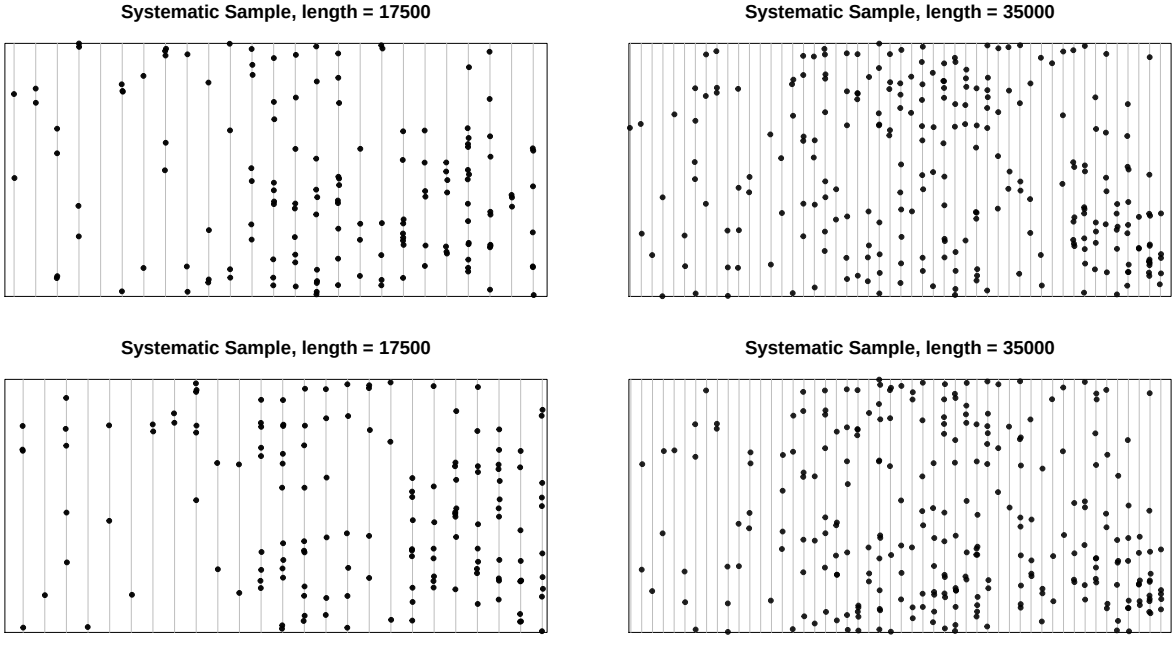
The dataset is in a 1500×700 rectangular region. It will be surveyed such that only events within 2 units of a path are observed.

- Parallel transect designs
  - Simple random sample
  - Systematic sample, evenly spaced, with random starting point
  - Inhibitory plus close pairs
- Hilbert curve, a deterministic space-filling curve, with random starting point
- Shortest path (traveling salesperson) through a Latin hypercube sampling design
- Random particle movement, random angle and direction for each segment, with a preference for new locations

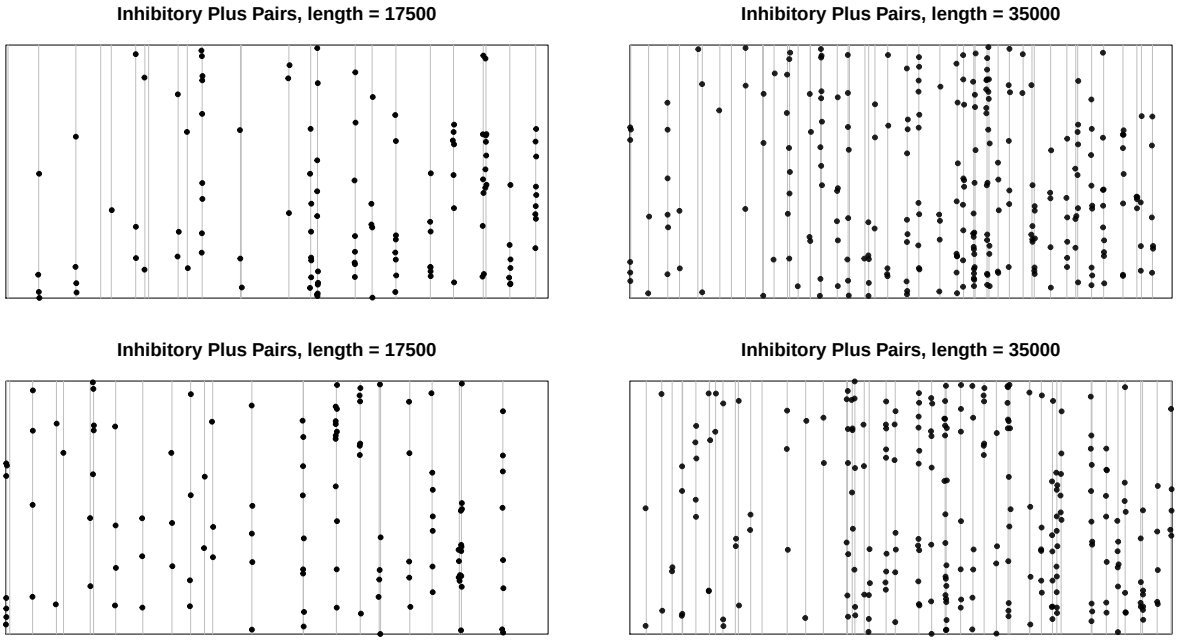
Simple Random Sample of Parallel Transects



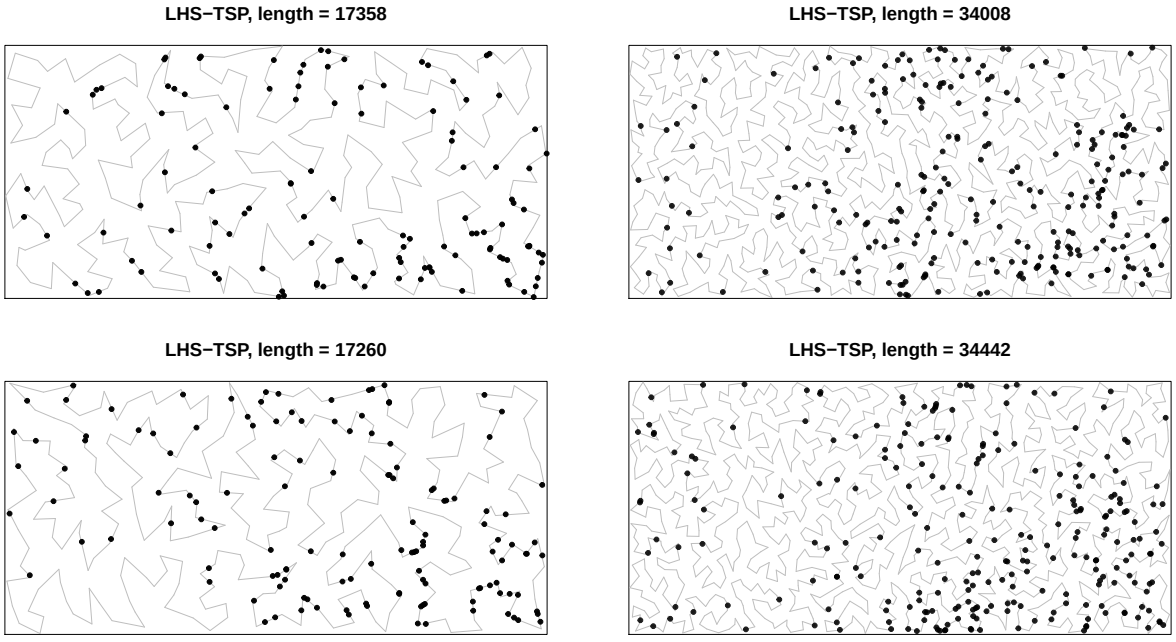
Systematic Sample of Parallel Transects



Inhibitory Plus Close Pairs of Parallel Transects

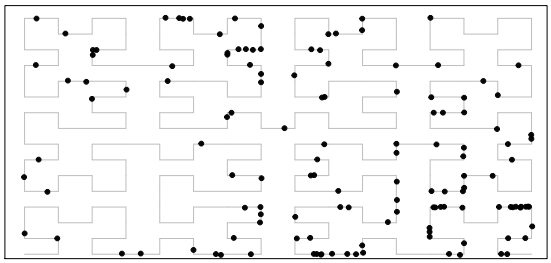


Latin Hypercube Sampling-Traveling Salesperson

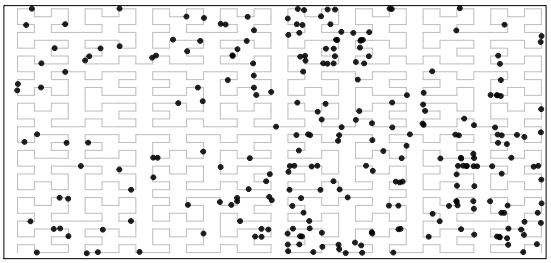


Hilbert Curve

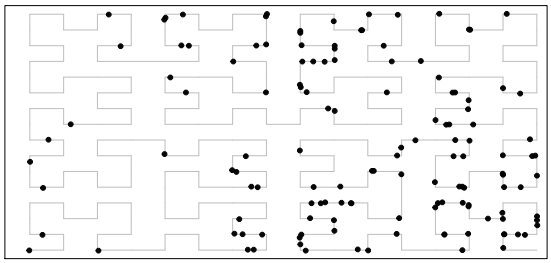
Hilbert Curve, length = 17442



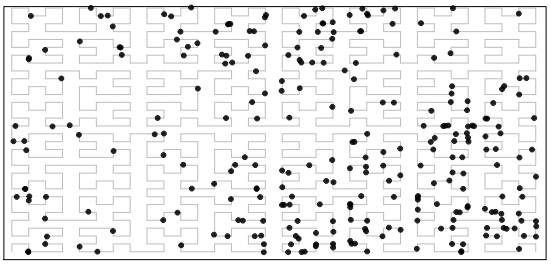
Hilbert Curve, length = 35025



Hilbert Curve, length = 17442

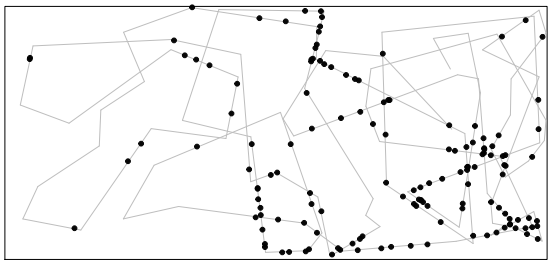


Hilbert Curve, length = 35025

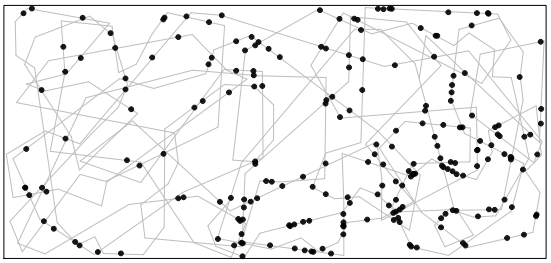


Particle Movement

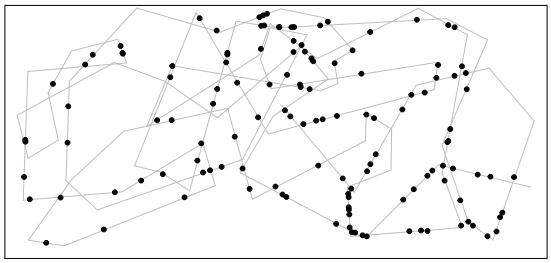
Random Particle Movement, length = 17196



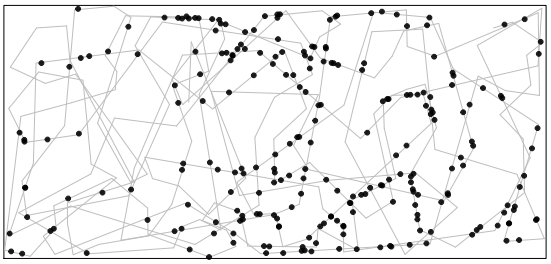
Random Particle Movement, length = 34513



Random Particle Movement, length = 17148



Random Particle Movement, length = 34538



Simulation Study Design

Generative Models

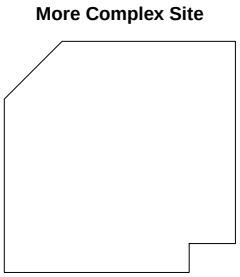
- Log-Gaussian Cox process (LGCP)
- LGCP with cluster process overlaid (i.e. LGCP plus hotspots)

Model Fitting

- Bayesian LGCP model with no predictors
- Fit in R with the `inla` package using integrated nested Laplace transformation (INLA) and Gaussian Markov random field (GMRF) approximation of GP
- Need to decide priors
  - Uninformative/weakly-informative priors not easily implemented in `inla` for the GMRF approximation
  - Want to emmphasize spatial prediction without ignoring parameter estimation

Study Regions

- Rectangle
- Square with two corners cut out



Criteria Suite

- Nearest neighbor distance for poitns on the path
  - Maximize the minimum and average
- Path length
  - Minimize the distance traveled
- Prediction variance of the GP
  - Minimize the maximum and average
- Posterior variance of parameters
  - Minimize variance of intercept, variance, and range

For the LGCP with hotspots, include additional criteria based on thresholding the predicted intensity surface at a set level to delineate hotspots

- AUC, F1, sensitivity, specificity, precision, etc
  - Maximize these
- Threshold-penalized average predictive variance (don't know if it has a name already so this is what I call it)

$$\text{TPAPV} = \frac{1}{|\mathcal{R}|}\int_{\mathcal{R}} \text{Var}\left[\lambda(u)\right] p^{|\lambda(u)-A|}\text{d}u$$

  - A is the action level/decision threshold
  - $0 < p < 1$ penalizes uncertainty about the boundary used for thresholding
  - Minimize this

Sampling Schemes

- SRS of parallel transects
  - Vary number of transects
- Systematic sample of parallel transects
  - Vary number of transects
- Inhibitory plus close pairs of parallel transects
  - Vary number of transects
  - Vary proportion of transects that are paired
- Latin hybercube sampling
  - Vary number of bins
- Hilbert curve
  - Vary number of iterations in curve construction
- Random particle movement
  - Vary maximum path length
  - Vary correlation between adjacent segment length (uncorrelated or strong negative correlation)
  - Vary mean/SD of turn angle