



# Stat 534 Project: Extrapolation from Poisson Process Intensity Surface Models

Kenny Flagg

- Goals
  - Estimate inhomogeneous intensity surface from events in a subregion
  - Infer intensity across entire region
- Applications
  - Mapping where endangered species are located
  - Mapping geomagnetic anomalies prior to an unexploded ordnance (UXO) remediation

# Maximum Likelihood Intensity Surface Fitting

- General point processes
  - The theory is not too complicated but the computation is very difficult
- Poisson processes
  - Doable with numerical methods
  - Log-likelihood of Poisson with intensity  $\lambda(\mathbf{s})$  on region  $D$  (note typos in Diggle (2013))

$$\begin{aligned}\ell(\lambda) &= \{-\mu + n \log(\mu) - \log(n!)\} + \sum_{i=1}^n \{\log(\lambda(\mathbf{s}_i)) - \log(\mu)\} \\ &= \sum_{i=1}^n \log(\lambda(\mathbf{s}_i)) - \int_D \lambda(\mathbf{s}) d\mathbf{s} - \log(n!).\end{aligned}$$

where  $\mu = \int_D \lambda(\mathbf{s}) d\mathbf{s}$

- Assuming events are independent (conditional on the intensity function),

$$\log(\lambda(\mathbf{s})) = \mathbf{x}(\mathbf{s})^T \boldsymbol{\beta}$$

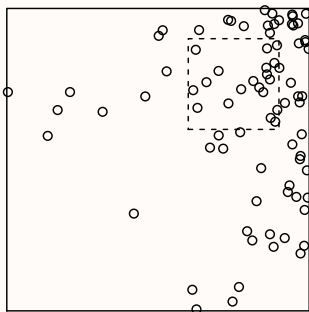
where  $\mathbf{x}(\mathbf{s})^T$  is a row of predictors at location  $\mathbf{s}$

- Predictors can include covariates, but they must be known across the whole region
- Berman and Turner (1992) use dummy points and quadrature to set up an approximation as a weighted Poisson regression
- Their method is implemented in `spatstat`'s `ppm`, with `glm` from base R or `gam` from `mgcv` as the back-end

# Simple Example

- True model
  - Poisson process on the unit square
  - $\log(\lambda(x, y)) = 5x + 2y$
- But we don't observe  $0.6 < x < 0.9, 0.6 < y < 0.9$

**Event Locations**

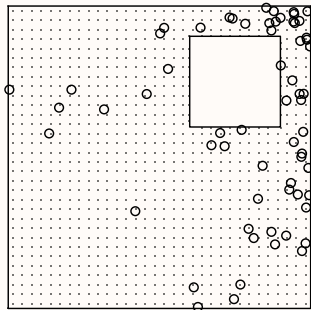


# Estimated Model

- Fit the model  $\log(\lambda(x, y)) = \beta_0 + \beta_1 x + \beta_2 y$

	Estimate	S.E.
$\hat{\beta}_0$	0.20	0.56
$\hat{\beta}_1$	4.54	0.59
$\hat{\beta}_2$	2.00	0.44

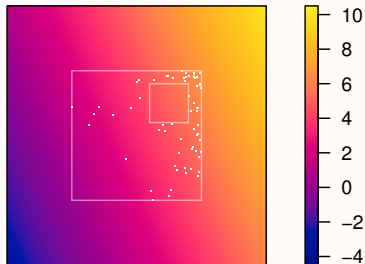
**Data and Dummy Points**



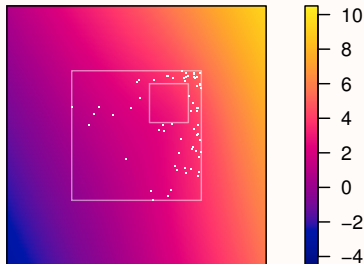
# Extrapolate the Surface

- Use the predict method
- Specify a new window  $-0.5 < x < 1.5, -0.5 < y < 1.5$

$$\log(\hat{\lambda}(x, y))$$

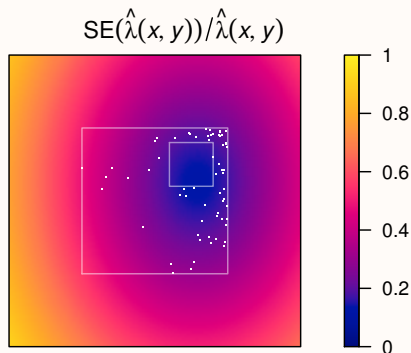


$$\log(\text{SE}(\hat{\lambda}(x, y)))$$



# Where is the Uncertainty?

- Relative standard error is lowest where the highest intensity was observed

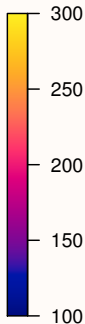
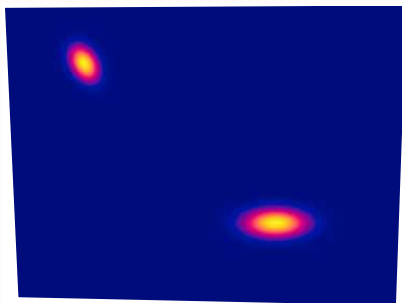




# Simple UXO Site

- 952.38 acre region (roughly 7,625 ft by 5,709 ft)
- High density of geomagnetic anomalies around targets
- Low density of background anomalies

**True Intensity Surface**

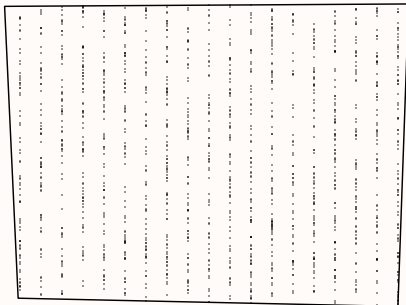


Anomalies per Acre

# Observed Events

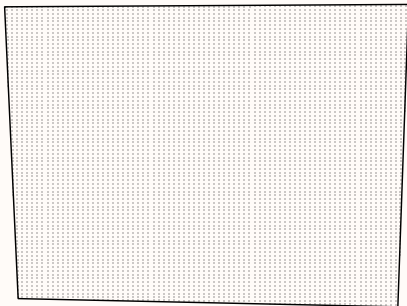
- Metal detectors record anomalies in six foot wide strips along parallel transects with 396 feet between centerlines
- Observed 14.7 acres, 1.5% of the site

## Observed Geomagnetic Anomalies

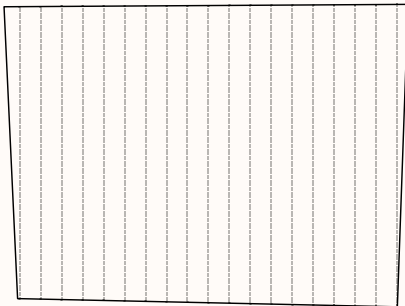


- Polynomial models  $\log(\lambda(x, y)) = \sum_{i=0}^p \sum_{j=0}^{p-i} \beta_{ij} x^i y^j$  for  $p = 2, 3, \dots, 12$

**Default Dummy Points**



**Manual Dummy Points**



# Problems with Implementation

- Needed to rescale coordinates
- For  $p \geq 10$ , `pmm` cannot compute SEs because the “Fisher information matrix is singular”
- The `plot` method gives an error about infinite values
- When the window isn't specified, the `predict` method's default grid misses all but two transects
- The `predict` method does not work with spline smoothers
- The magnitudes of the predictions are much too large

- Berman, Mark and Rolf Turner (1992). “Approximating point process likelihoods with GLIM”. In: *Applied Statistics*, pp. 31–38.
- Diggle, Peter J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. 3rd ed. CRC Press.