Imperial College London

Areal Relevance with Kernels

Yidan Xu

Supervised by Dr Seth Flaxman,
Department of Mathematics

Case Study: Primary Biliary Cirrhosis in Newcastle-upon-tyne

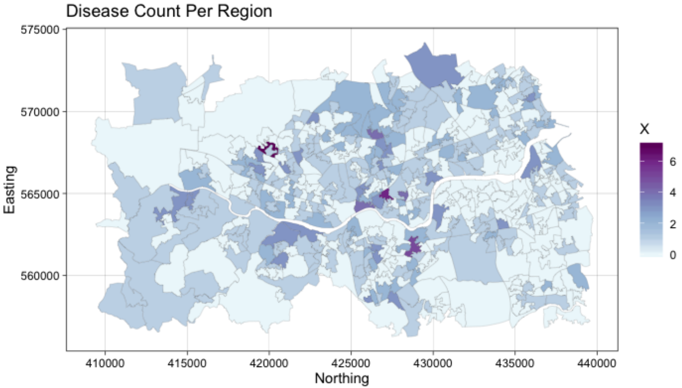


Figure 1: Number of diagnosed disease cases in each region.

Implicit Conditional Autoregressive (ICAR)

Spatial interactions between a pair of areal units i and j can be modelled conditionally as a spatial random variable $\phi = (\phi_1, \phi_2, \ldots, \phi_n)^t$.

$$\mathcal{P}(\phi_i | \phi_j, j \neq i, \tau_i^{-1}) = N\left(\frac{\sum_{i \sim j} \phi_i}{d_{i,i}}, \frac{1}{d_{i,i}\tau_i}\right) \tag{1}$$

$$\begin{gathered} Y_i | \varphi_i \sim Poisson(E_i e^{\varphi_i}) \\ \varphi = x\beta + \theta + \phi \end{gathered} \tag{2}$$

τ_i is the mean and precision parameter for i;
$d_{i,i}$ is the number of neighbours for region i.

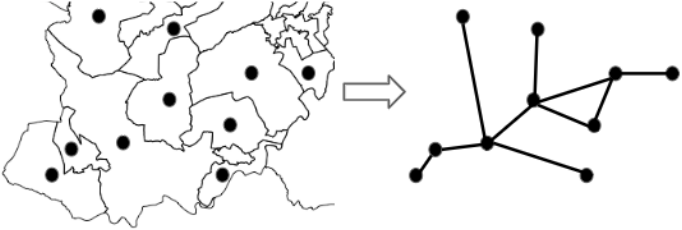


Figure 2: ICAR model gives simplification of the spatial relations as a graphical model.

Log-Gaussian Cox Process

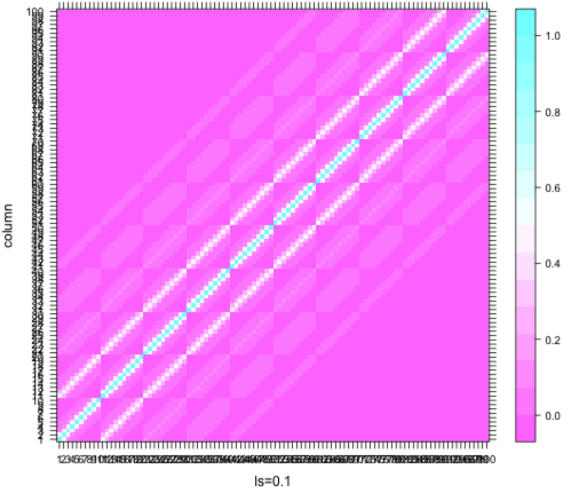
The LGCP is a doubly stochastic point process model.

Definition

For any space region $S \subset W$, $N(S)$ a Poisson distributed random variable, counting the number of points in S.

$$\begin{aligned} \log \Lambda &\sim GP(\mu, k_\theta(\cdot, \cdot)) \\ N(S)|\Lambda &\sim Poisson(\int_S \Lambda(s)\,\mathrm{d}s) \end{aligned} \tag{3}$$

$k_\theta(\cdot, \cdot)$ the covariance function with respect to lengthscale θ.



Matérn 5-2 Kernel with varying lengthscale

Areal Relavance with Kernels

Consider disjoint regions R_1 and R_2, with population density distribution W_1 and W_2 respectively.

$$\begin{aligned} \kappa_\theta(R_1, R_2) &= \int_{R_1}\int_{R_2} k_\theta(u, v)\ \mathrm{d}W_1(u)\mathrm{d}W_2(v) \\ &\approx \sum_{n_1}\sum_{n_2} k_\theta(u_i, v_j) w_1(u_i) w_2(v_j) h^2 \end{aligned} \tag{4}$$

u_i and v_j is the centroids of the grid cells.
h is the cell width.

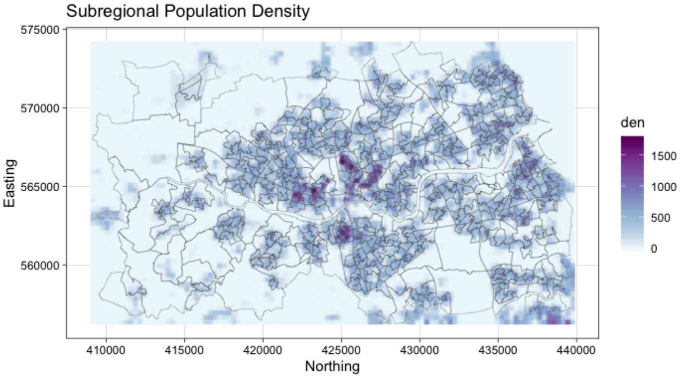
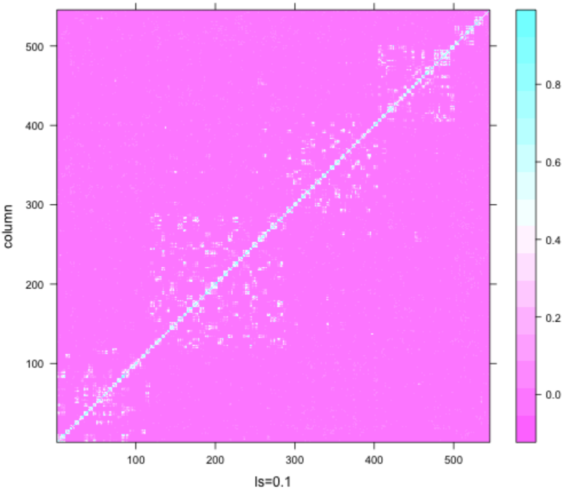


Figure 3: Sub-regional population density over a grid with 300m by 300m resolution.



Kernel $\kappa_\theta(\cdot, \cdot)$ with varying lengthscale.

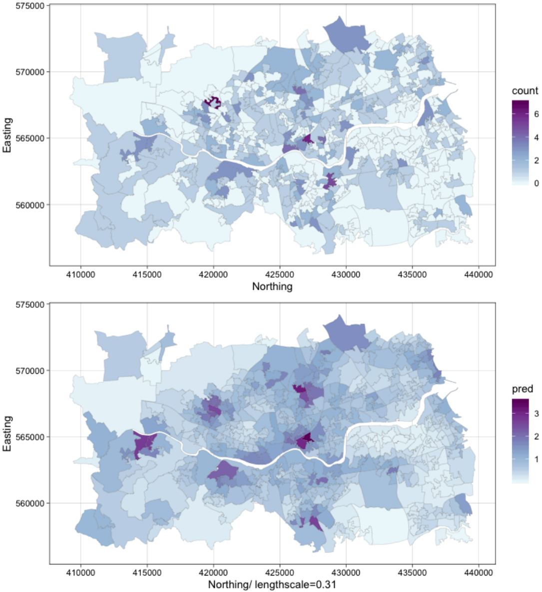


Figure 4: Original count data and the mean of posterior LGCP model fitted with cross-validated lengthscale.