Analysing the spatio-temporal distribution of crime in Lancashire

Irene Kaimi, Peter Diggle and Alexandre Rodrigues

Overview

- The MADE project
- Data
- Statistical Formulation
- Results
- Work in progress

The MADE project

Multi Agency Data Exchange

A data warehouse tool for all the datasets which are relevant to crime and disorder and are available throughout Lancashire.

Goal

To help people within Lancashire to make a more informed decision about community safety issues in their neighbourhood.

Objectives

- Develop a statistical model for the spatio-temporal distribution of recorded crimes
- Implement predictive inference as R code
- Develop web-based real- probabilistic mapping of local (in space and time) variations in crime-rate

The MADE Data

Information, for each reported crime:

- location (lower super-output area)
  LSOA: Minimum population 1000,
  mean population 1500;
  built from Output Areas
- time (day, hour, minute)
- type of crime:
  - other wounding (19%)
  - criminal damage (51%)
  - serious acquisitive crime (30%)

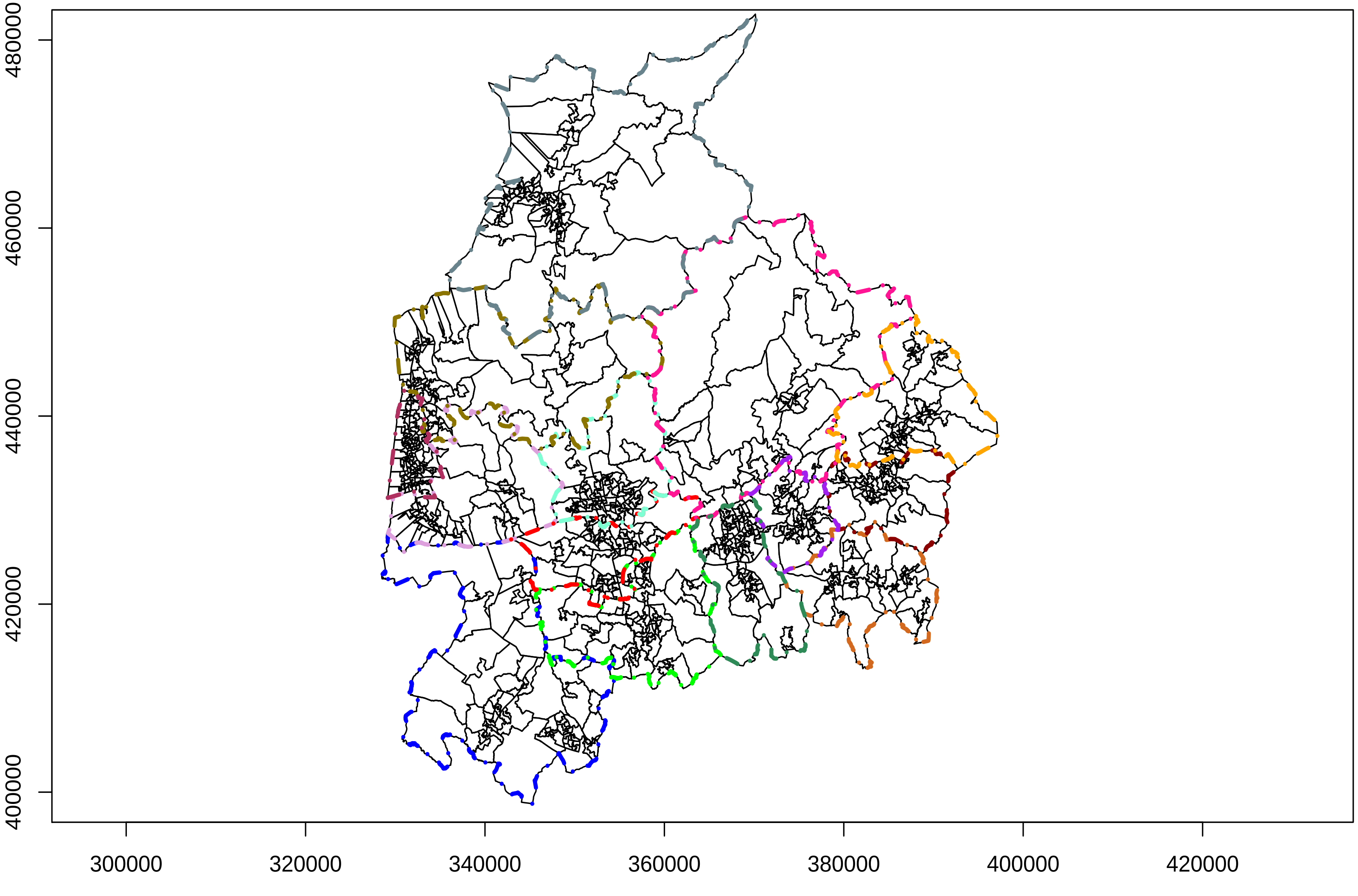
+ LSOA population
+ Spatial covariates at LSOA level

The MADE Data

- Data cover whole of Lancashire, divided into 940 LSOA’s
- Time-period: 1 April 2003 to 31 March 2009 (412,589 records)

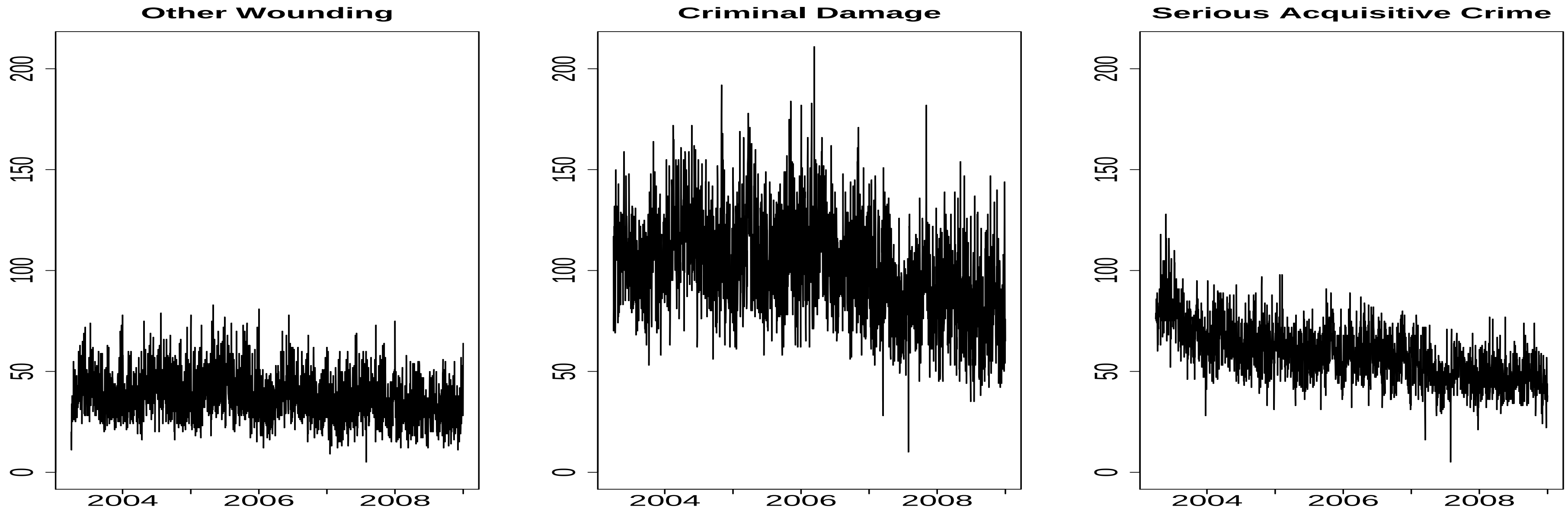
Exploratory Analysis

LSOA's in Lancashire



Exploratory Analysis

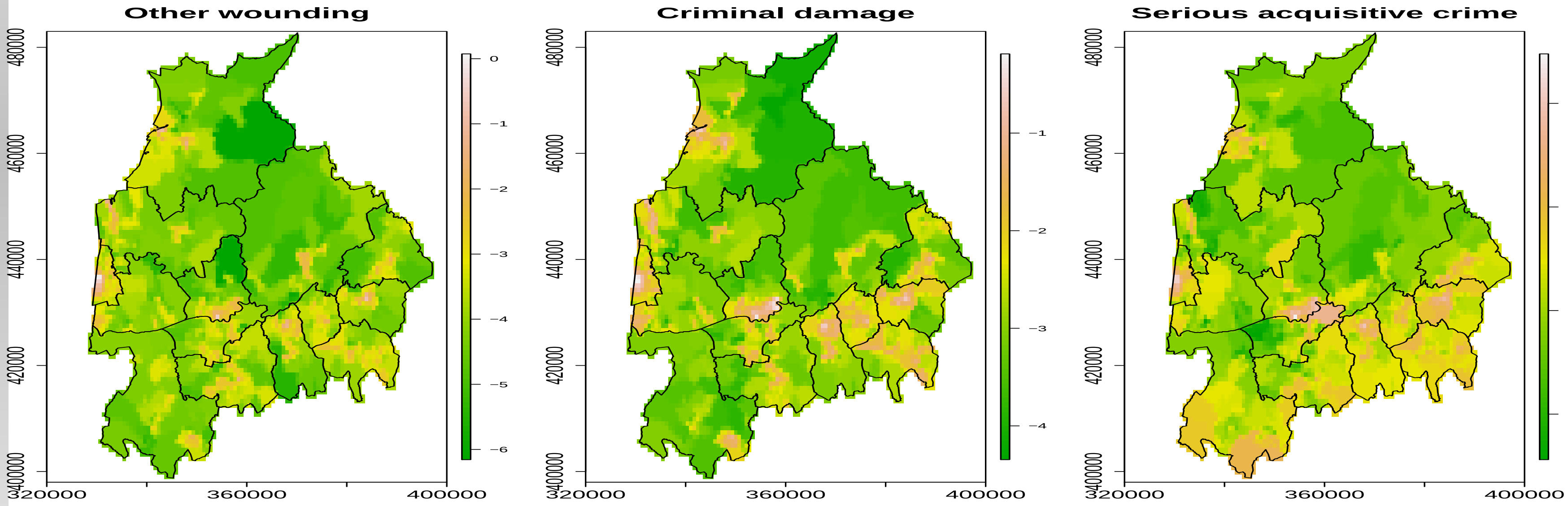
Time series of daily crime counts by category



The three categories show qualitatively different behaviour $\Rightarrow$ analyse separately.

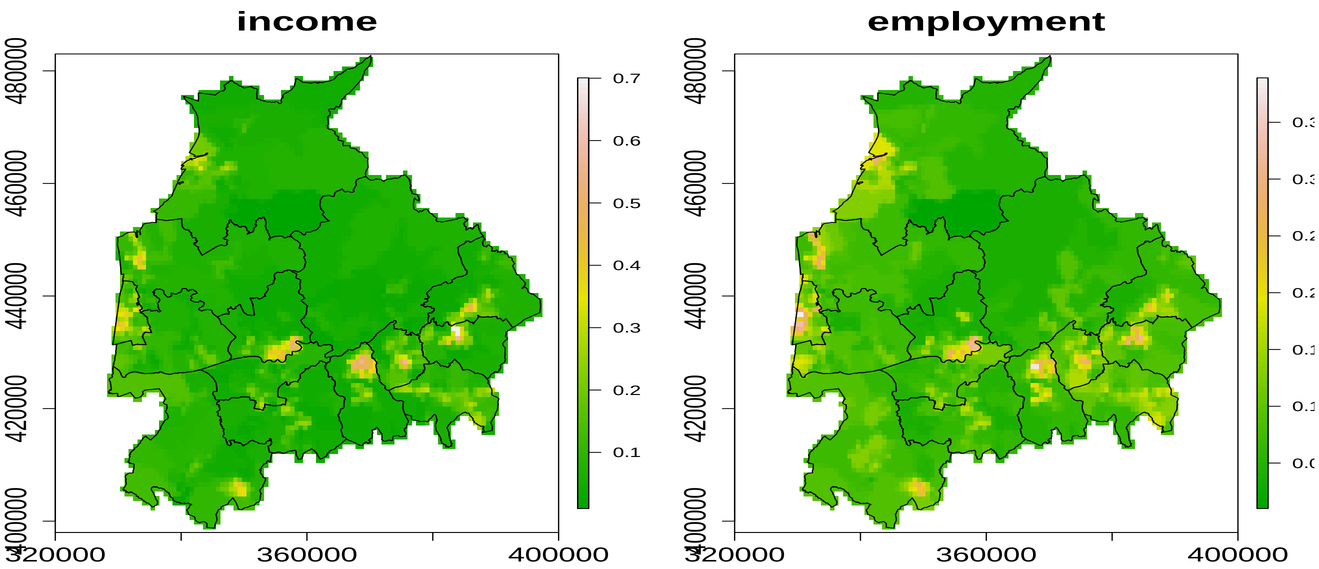
Exploratory Analysis

Rates of crimes



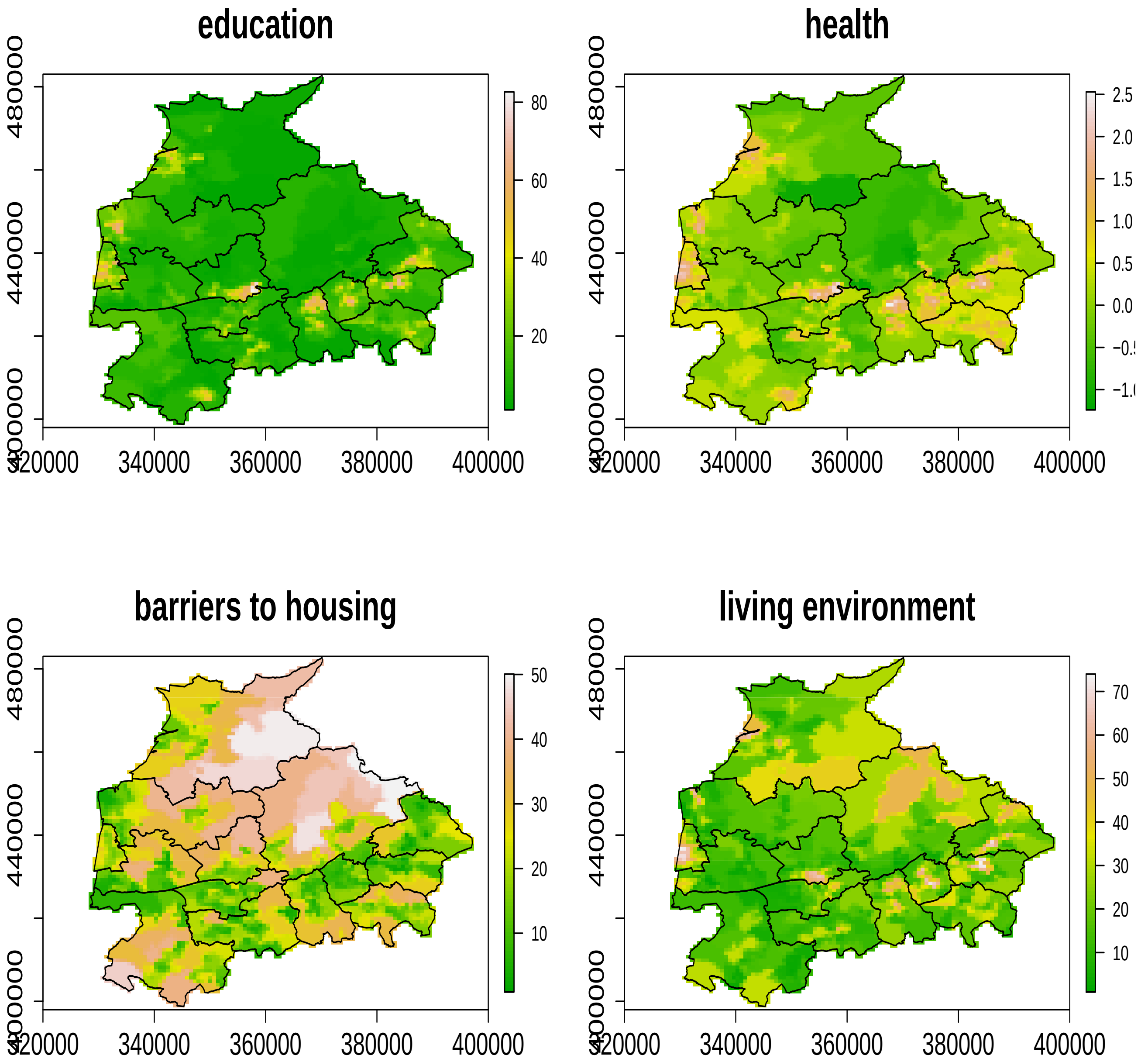
Exploratory Analysis

Spatial covariates: Deprivation rates



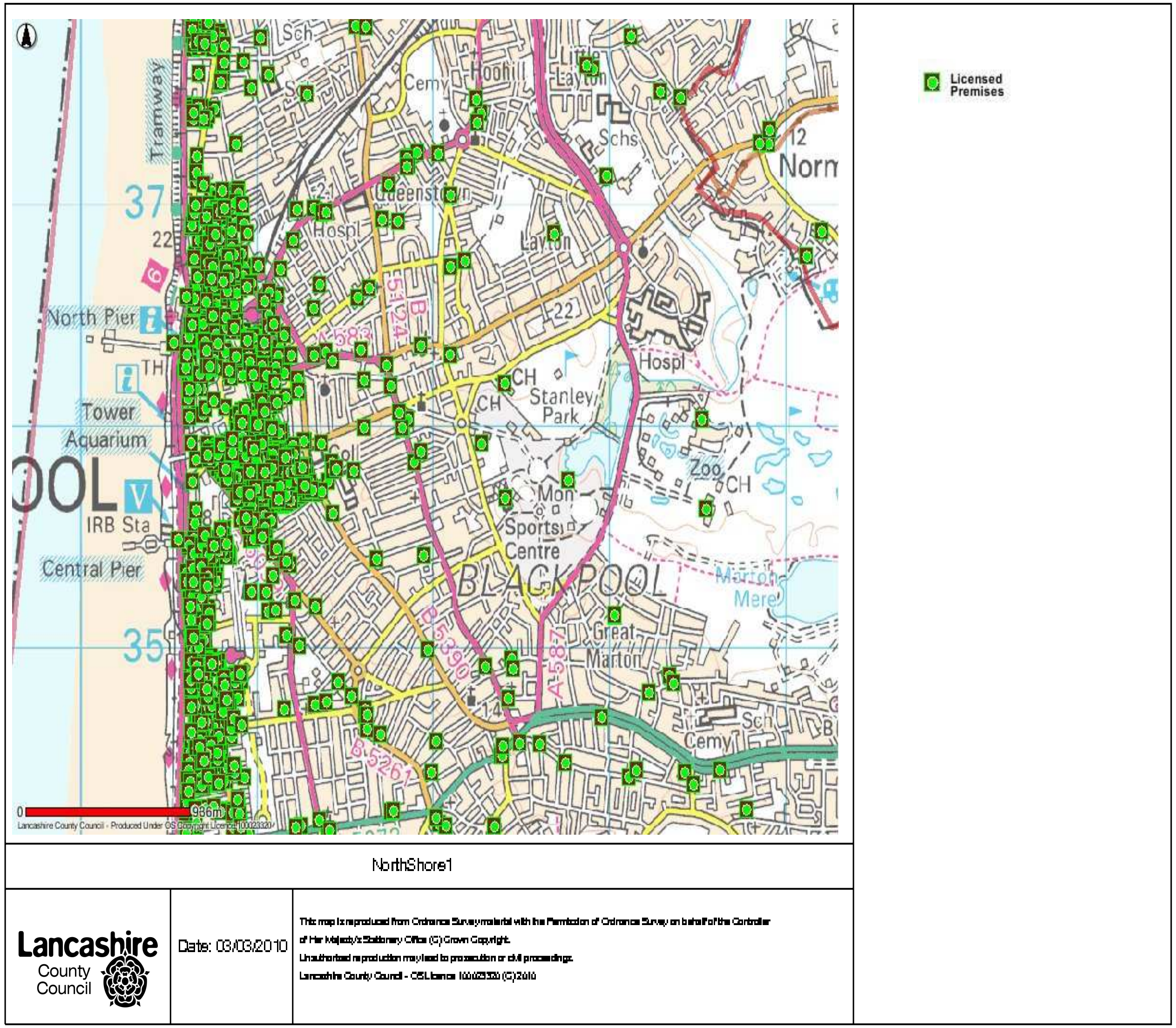
Exploratory Analysis

Spatial covariates - Deprivation indices



Exploratory Analysis

Blackpool North shore overview - licensed premises



Statistical Formulation

The underlying spatio-temporal point process that generates the number of crimes Y_{it} within LSOA $i; i = 1, \ldots, N$ at the time point $t; t = 1, \ldots, T$ has intensity

$$\lambda(\mathbf{x}, t) = \mu(\mathbf{x}, t) R(\mathbf{x}, t), \mathbf{x} \in \mathcal{R}^2, t \in \mathcal{R}$$

- $\mu(\mathbf{x}, t)$: deterministic spatio-temporal variation in the mean number of incident crimes per unit time
- $R(\mathbf{x}, t)$: a spatio-temporal stochastic process
  - * models the residual spatio-temporal variation
  - * its covariance function determines the form of dependence between space and time

Statistical Formulation

Assume multiplicative spatial and temporal deterministic variation,
i.e. $\mu(\mathbf{x}, t) = \lambda(\mathbf{x})\mu(t)$ where

- $\mu(t)$ temporal variation in the spatially averaged incidence rate
- $\lambda(\mathbf{x})$ overall purely spatial variation in the intensity of reported crimes
  Local variations within LSOA's cannot be identified,
  $\Rightarrow \lambda(\mathbf{x}) = \lambda_i$ (constant) for all $\mathbf{x}$ in $LSOA_i$

Statistical Formulation

The process that generates the crimes is assumed to be a spatio-temporal log-Gaussian Cox Process.

Hence,

$$R(\mathbf{x}, t) = \exp\{S(\mathbf{x}, t)\},$$

- $S(\mathbf{x}, t)$ is a stationary spatio-temporal Gaussian process such that $E(\exp\{S(\mathbf{x}, t)\}) = 1$.
- $S(\mathbf{x}, t)$ has covariance function $\gamma(u, v) = \sigma^2 \rho(u, v)$ where $\rho(\cdot, \cdot)$ is a spatio-temporal correlation function, and u and v denote spatial and temporal lags, respectively.

Statistical Formulation

Take $t = 1, \ldots, M$ days.
Scale $\lambda(\mathbf{x})$ such that $\int_A \lambda(\mathbf{x}) = 1$
$\rightarrow \mu(t) =$temporal variation in the mean number of incident crimes per day
$\Rightarrow$ Data: Y_{it} : number of crimes on day t;
$t = 1, \ldots, M$, in $LSOA_i; i = 1, \ldots, N$.
Conditional on the unobserved process $R(\cdot)$,

$$Y_{it}|R(\cdot) \sim Poisson\left(\lambda_i \mu(t) \int_{LSOA_i} R(\mathbf{x}, t) d\mathbf{x}\right)$$

- Poisson number of counts
- Straightforward calculation of the covariance structure

Statistical Formulation

For our log- Gaussian Cox process the second-order intensity function

$$\lambda_2(u,v) = \exp\{\gamma(||x-y||, v)\},$$

where $\gamma(||x-y||, v) = \sigma^2 \rho(u,v)$. Then,

$$\text{Cov}\{Y(i,t), Y(j,t-v)\} = \mu(t)\lambda_i \mu(t-v)\lambda_j \left[\int_{x,y \in A_i \times A_j} \exp\{\gamma(||x-y||, v)\} dx dy - |A_i||A_j| \right], \tag{1}$$

where A_i represents the i^{th} LSOA and $|A_i|$ is the area of the region A_i. The variance is given by

$$\text{Var}\{Y(i,t)\} = \{\mu(t)\lambda_i\}^2 \left[\int_{x,y \in A_i} \frac{\exp\{\gamma(||x-y||, 0)\} dx dy}{|A_i|^2} - 1 \right] + \mu(t) p_i, \tag{2}$$

where $p_i = \lambda_i A_i$.

Estimation of $\mu(t)$

We first fit a semi-parametric model for $\mu(t)$ of the form

$$\log(\mu_t) = Z_t'\beta + f(t) \tag{3}$$

where Z_t is a vector of covariates at time t and f is a smooth, but otherwise unspecified, function of time. Explanatory variables:

- day-of-week effect, $\delta_{d(t)},\ d(t) = 0, 1, ..., 6$ as a seven-level factor,
- sine-cosine terms with periods of twelve and six months to capture seasonal effects and
- low-order polynomial time-trends.

Estimation of $\lambda(\mathbf{x})$

- $y_i; i = 1, \ldots, N$ the number of crimes in $LSOA_i$
- $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_N)$ the matrix of q spatial covariates.

$Y_i \sim$ Poisson with mean $N_i \lambda_i$, and

$$\lambda_i = \exp(\boldsymbol{\beta}_i \mathbf{w}_i), \quad (4)$$

- the $\boldsymbol{\beta}_i$'s are parameters to be estimated and
- N_i is the population of the i^{th} LSOA, $\Rightarrow \lambda_i$ the crime-rate in the i^{th} LSOA.

Covariates:

- density of licensed premises
- deprivation rates/scores for six domains

Estimation of $S(\mathbf{x}, t)$

- $\rho(u, v)$ is separable, i.e. $\rho(u, v) = \rho_S(u)\rho_T(v)$,

$C_{i,j}(t, t - v) = \text{Cov}\{Y(i, t), Y(j, t - v)\}$ the moment-based estimates of σ^2 and θ_S minimise the criterion

$$\sum_t \sum_i \sum_j \left\{ \widehat{C_{i,j}(t, t)} - C_{i,j}(t, t) \right\}^2, \qquad (5)$$

$$\widehat{C_{i,j}(t, t)} = Y(i, t)Y(j, t) - \mu(t)p_i\mu(t)p_j.$$

- non-separable covariance function $\rho(u, v)$

  Minimise with respect to model parameters the expression

$$\sum_{v=1}^{v_0} \sum_{t=v+1}^{T} \sum_i \sum_j \left\{ \widehat{C_{i,j}(t, t - v)} - C_{i,j}(t, t - v) \right\}^2. \qquad (6)$$

Estimation of $S(\mathbf{x}, t)$

Making things simpler

- $\int_{x,y \in A_i \times A_j} \exp\{\gamma(||x-y||, v)\} dxdy = \exp\{\gamma(||c_i - c_j, v||)\} A_i A_j$,
  where c_i is the centroid of area A_i
- $\text{Cov}\{Y(i,t), Y(j,t-v)\} = \mu(t) p_i \mu(t-v) p_j [\exp\{\gamma(||c_i - c_j||, v)\} - 1]$
- Denote $Z(i,j,t,v) = \frac{Y(i,t)Y(j,t-v)}{\mu(t)p_i\mu(t)p_j}$
- $E[Z(i,j,t,v)] = \exp\{\gamma(||c_i - c_j||, v)\}$
- Hence,

$$\frac{1}{T-v} \sum_{t=v+1}^{T} Z(i,j,t,v) \rightarrow \exp\{\gamma(||c_i - c_j||, v)\}$$

Results

Overall temporal variation $\mu(t)$

Models:

- Semi-parametric:
  - $\log(\mu_t) = \delta_{d(t)} + f(t)$
  - $\log(\mu_t) = \delta_{d(t)} + \alpha_1 \cos(\omega t) + \beta_1 \sin(\omega t) + \alpha_2 \cos(2\omega t) + \beta_2 \sin(2\omega t) + f(t)$
- Parametric:
  - $\log(\mu_t) = \delta_{d(t)} + \alpha_1 \cos(\omega t) + \beta_1 \sin(\omega t) + \alpha_2 \cos(2\omega t) + \beta_2 \sin(2\omega t) + \epsilon_1 t + \epsilon_2 t^2.$

Results

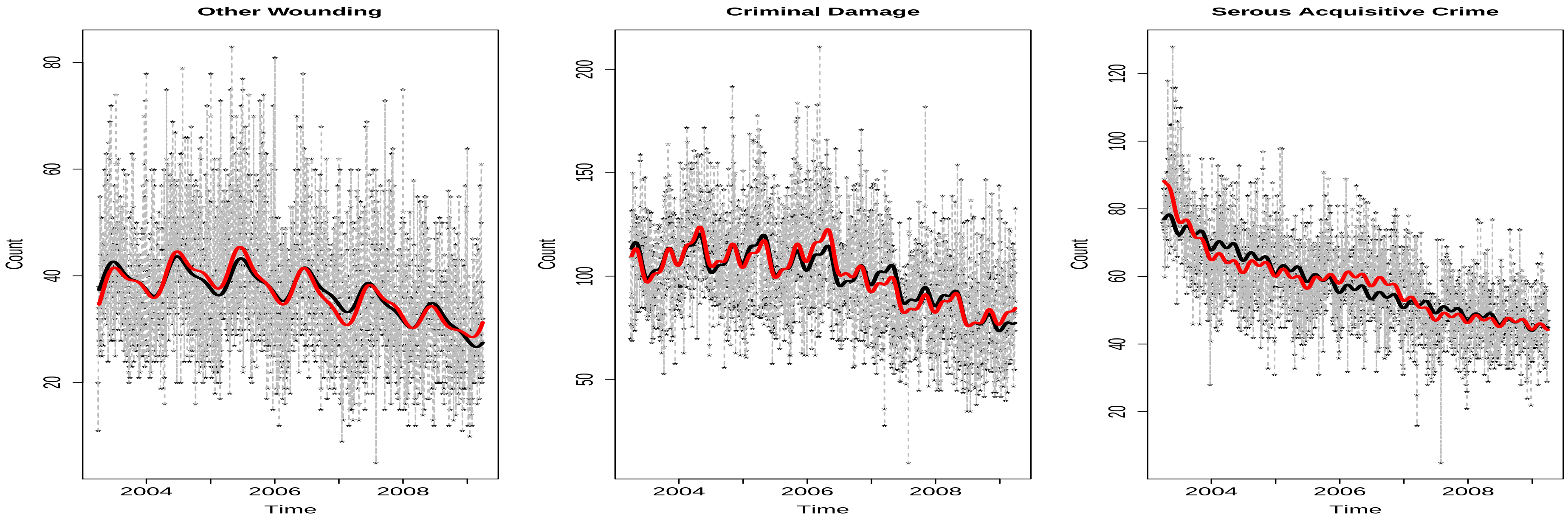
Overall temporal variation $\mu(t)$

- Strong and significant day of week effects, Thursday (lowest) - Sunday (highest)
- Log-linear time trend significant;
  log-quadratic time trends gives unequivocal significant improvement in model fit for all three crime categories
- sine and cosine terms significant;
  different seasonal pattern for each crime category

Results

Overall temporal variation $\mu(t)$

Average weekly fit of GLM (black line) and GAM (red line) compared with the actual number of cases



Results

Overall spatial variation $\lambda(\mathbf{x})$

- The effect of density of licensed premises is statistically significant for all three types of crime (p - value $<< 0.0001$).
- Deprivation indices/rates effects vary in size and significance for the three categories of crime

Spatial regression - Results

Other wounding

- Not significant: Income and housing barriers effects
- Significant: Employment, health, living environment, education
- Employment deprivation rate effect high (2.8). Rate of other wounding crime in a LSOA in Blackburn (employment deprivation $= 50\%$) is 4.1 times the rate in a LSOA in Lancaster (employment deprivation $= 1\%$)

Spatial regression - Results

Criminal damage

- Not significant: Employment
- Significant: Income, health, barriers to housing and benefits, education, living environment,

Spatial regression - Results

Serious acquisitive crime

- Not significant: Employment, barriers to housing, income
- Significant: Health, living environment, education
- Size of health and disability deprivation index effect: 0.64
- e.g. index of health deprivation in a LSOA in Ribble Valley is -1.24, whereas index of deprivation in a LSOA in Blackburn is 3.23 $\Rightarrow$ rate of serious acquisitive crime in the LSOA in Blackburn is $\exp(0.64 \times 4.47) = 17.5$ times greater than the rate in the LSOA in Ribble Valley.

Individual districts

- 14 local authority districts
- Both urban and rural districts
- Wide range of socio-economic conditions
- The pattern of crime varies considerably over the 14 districts
- The geographical region covered by each district is different
- Different geographical shape of each district, number of LSOA's forming the district, and sizes affect the form of the spatial dependence between LSOA's within the same district.

Results

Lancaster - Preston - Blackpool

- Different seasonal pattern
- The intercept term of the model is different in each case
- Different form for the quadratic time function in each case.
- The weekday effects only marginally distinct
- The significance of the density of licensed premises is consistently high for the three districts
- The rates and scores of the six domains of deprivation have variable statistical significance and size of effects.

$\therefore$ Effects of temporal and spatial covariates and spatio-temporal correlation are not the same throughout the county of Lancashire

Results

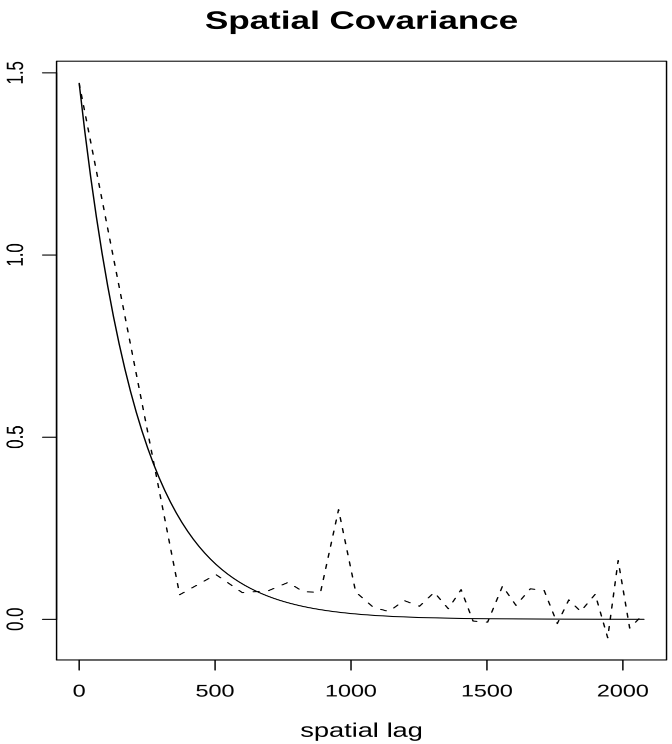
Spatio-temporal interaction

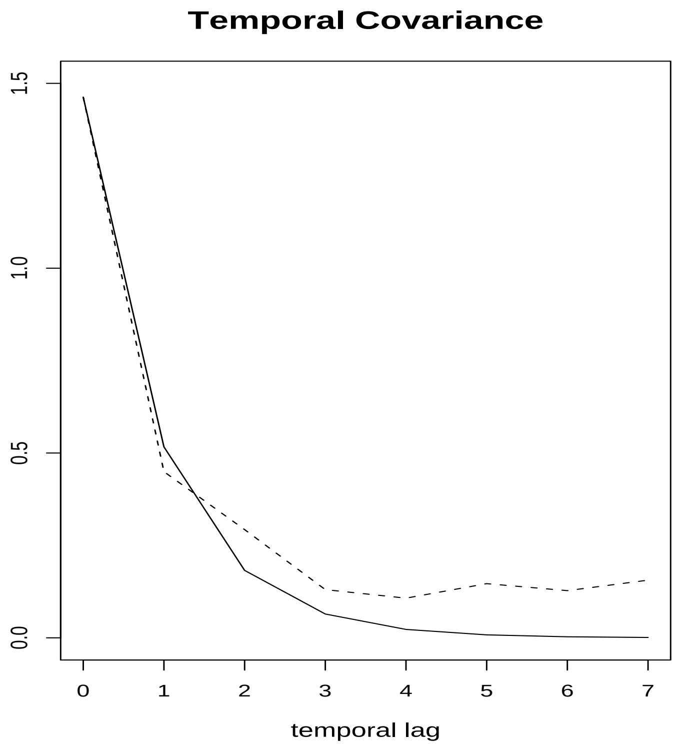
Match theoretical and empirical descriptors of the spatial covariance structure of the point process model to find its form

Results

Spatio-temporal interaction

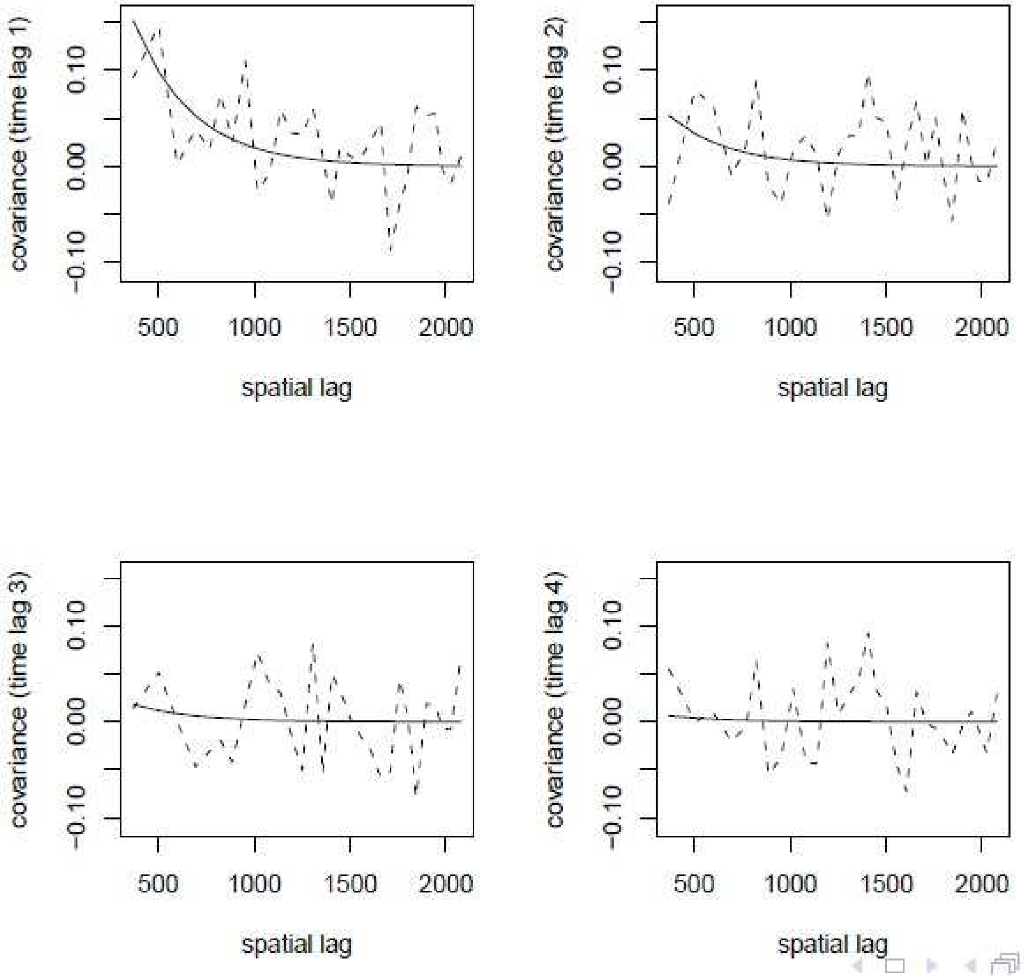
$$\gamma(0, v) \propto \exp(-v/\phi_T) \qquad \gamma(u, 0) \propto \exp(-u/\phi_S)$$





Results

$$\gamma(u, v) = \sigma^2 \exp(-u/\phi_S) \exp(-v/\phi_T)$$

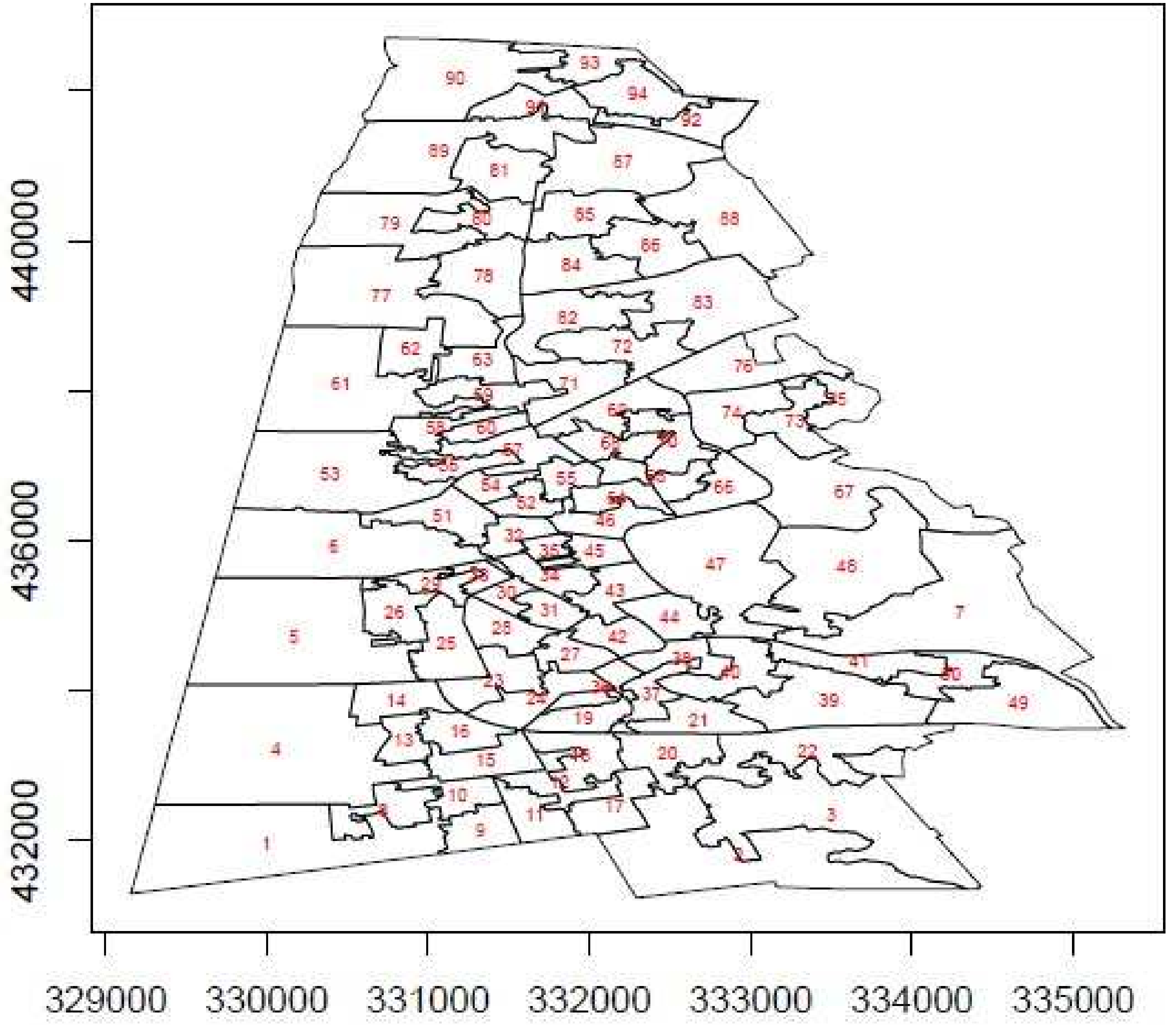


Results

Separable model

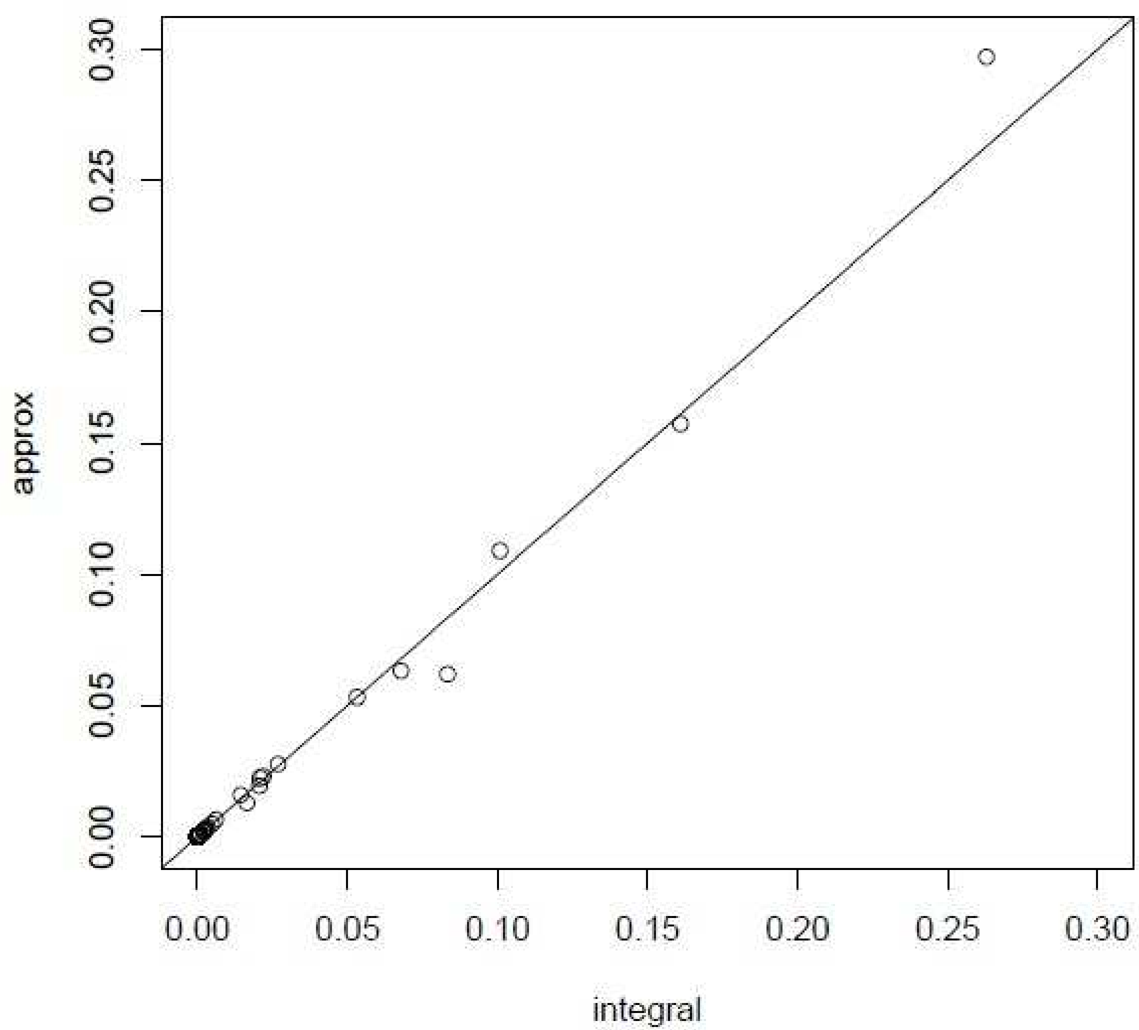
- $\gamma(u, v) = \sigma^2 \exp(-u/\phi_S) \exp(-v/\phi_T)$
- Minimise $\sum_t \sum_i \sum_j \left\{ \widehat{C_{i,j}(t,t)} - C_{i,j}(t,t) \right\}^2$,
- Consider pair (i, j) such as $||c_i - c_j|| < 3000$ meters

Results


432000
436000
440000
329000
330000
331000
332000
333000
334000
335000

Results

Highest correlation 33, 26, 30, 6



Work in progress

Prediction

- Use a Markov Chain Monte Carlo algorithm to generate a sample from the predictive distribution of the spatio-temporal surface $S(\mathbf{x}, t)$ conditional on the observed spatio-temporal pattern of crimes up to and including time t.
- Find space-time clusters of crimes, by evaluating the predictive probability $\mathrm{Pr}(R(\mathbf{x}, t) > c|\mathrm{data})$, where c is a threshold value above which an alarm is triggered.
- Plot the exceedance probabilities as a colour-coded map to highlight LSOA's in which these probabilities are high.