Burglary in London: Insights from Statistical Heterogeneous Spatial Point Processes

Jan Povala

with Seppo Virtanen and Mark Girolami

Imperial College London

February 25, 2020

Outline

Motivation

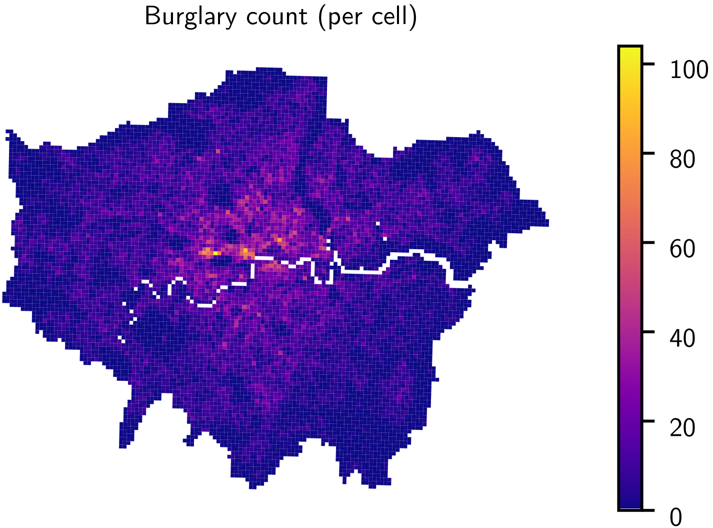
- Model the occurrences of burglary as a spatial point pattern and provide short-term forecasts.
- Provide insights into the intensity of the process.

Two effects in spatial statistics

To avoid biased results and faulty inferences a reasonable spatial model needs to account for:

- **Spatial dependence**: the first law of Geography – "everything is related to everything else, but near things are more related than distant things"(Tobler 1970)
- **Spatial heterogeneity**: phenomena observed on large domains tend to exhibit location-specific dynamics.

The data I



The data II

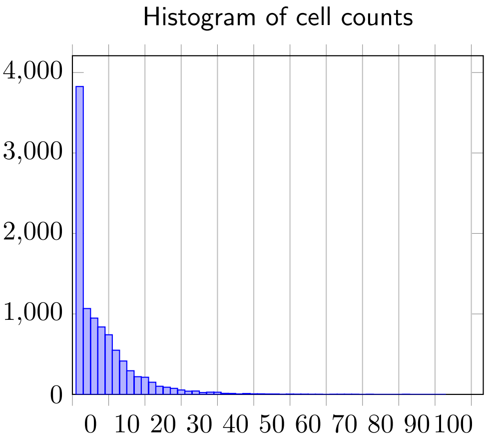


Figure: Data: burglary point pattern of 2015 aggregated at 400mx400m grid cells

Outline

Cox Process

Cox process is a natural choice for an environmentally-driven point process (Cox 1955, Diggle et al. 2013).

Definition

Cox process $\mathsf{Y}(\boldsymbol{x})$ is defined by two postulates:

1. $\Lambda(\boldsymbol{x})$ is a nonnegative-valued stochastic process;
2. conditional on the realisation $\lambda(\boldsymbol{x})$ of the process $\Lambda(\boldsymbol{x})$, the point process $\mathsf{Y}(\boldsymbol{x})$ is an inhomogeneous Poisson process with intensity $\lambda(\boldsymbol{x})$.

Log-Gaussian Cox Process

- Cox process with intensity driven by a fixed component $X(\boldsymbol{x})^\top \boldsymbol{\beta}$ and a latent function $f(\boldsymbol{x})$:

$$\Lambda(\boldsymbol{x}) = \exp\left(X(\boldsymbol{x})^\top \boldsymbol{\beta} + f(\boldsymbol{x})\right),$$

  where $f(\boldsymbol{x}) \sim \mathcal{GP}(0, k_\theta(\cdot,\cdot))$, $X(\boldsymbol{x})$ are socio-economic covariates, and $\boldsymbol{\beta}$ are their coefficients.
- Discretised version of the model:

$$\mathsf{y}_i \sim \text{Poisson}\left(\exp\left[X(\boldsymbol{x}_i)^\top \boldsymbol{\beta} + f(\boldsymbol{x}_i)\right]\right).$$

LGCP limitations

- Fitting this doubly-stochastic model at scale is challening.
- Simplifying assumptions such as stationarity of f may not be appropriate (see Figure 3)

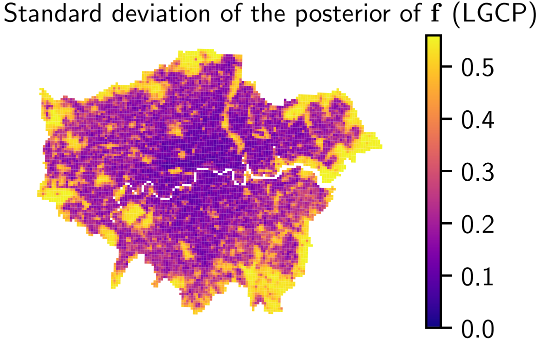


Figure: Standard deviation of the GP

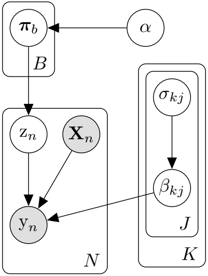
Common approaches to address spatial heterogeneity

- Mixture models with allocation that enforces spatial dependence (Green & Richardson 2002, Fernández & Green 2002, Hildeman et al. 2018).
- Regression coefficients modelled as a Gaussian process (Gelfand et al. 2003, Banerjee et al. 2015).

Both of these approaches have limited scalability.

Our proposed model

$$
\begin{aligned}
\mathsf{y}_n|\mathsf{z}_n = k, \boldsymbol{\beta}, \mathbf{X}_n &\sim \text{Poisson}\left(\exp\left(\mathbf{X}_n^\top \boldsymbol{\beta}_k\right)\right)\\
\mathsf{z}_n|\boldsymbol{\pi} &\sim \text{Categorical}(\boldsymbol{\pi}_{b[n]})\\
\boldsymbol{\pi}_b|\alpha &\sim \text{Dirichlet}(\alpha, \dots, \alpha)\\
\beta_{k,j}|\sigma_{k,j}^2 &\sim \mathcal{N}(0, \sigma_{k,j}^2)\\
\sigma_{k,j}^2 &\sim \text{InvGamma}(1, 0.01)\\
\alpha &= 1/K.
\end{aligned}
$$



Inference

We use Metropolis-within-Gibbs (Geman & Geman 1984, Metropolis et al. 1953) scheme using the following two steps:

1. We sample the regression coefficients $\beta_{k,j}$ jointly for all $k = 1, \ldots, K$ and $j = 1, \ldots, J$. The unnormalised density of the conditional distribution is given as

$$p(\boldsymbol{\beta}|\alpha, \mathbf{X}, \mathbf{y}, \mathbf{z}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{X}, \mathbf{z})p(\boldsymbol{\beta}). \tag{1}$$

   Equation 1 is sampled using Hamiltonian Monte Carlo method (Duane et al. 1987).

2. Mixture allocation can be sampled cell by cell directly

$$p(\mathsf{z}_n = k|\mathbf{z}^{\bar{n}}, \alpha, \mathbf{X}_n\boldsymbol{\beta}, \mathbf{y}) \propto p(\mathsf{y}_n|\mathsf{z}_n = k, \mathbf{X}_n\boldsymbol{\beta}_k)\frac{c^{\bar{n}}_{b[n]k} + \alpha}{K\alpha + \sum_{i=1}^{K} c^{\bar{n}}_{b[n]k}}, \tag{2}$$

   where $c^{\bar{n}}_{b[n]k}$ is the number of cells other than cell n in the encompassing block $b[n]$ assigned to component k, and $\mathbf{z}^{\bar{n}}$ is the allocation vector with the contribution of cell n removed.

Outline

Motivation

Modelling

Experiment

London burglary experiment

- One-year point pattern aggregated to a grid with cell size $400m \times 400m$.
- Covariates $X(\boldsymbol{x})$ chosen based on criminological background.
- Number of mixture components, K, ranges from 1 to 8.
- The blocking structure given by census output areas (MSOA).

Evaluation

We evaluate the performance using these metrics:

- Held-out log likelihood:

$$\text{Held-out log likelihood} = \frac{1}{S}\sum_{s=1}^{S}\frac{1}{N}\sum_{n=1}^{N}\log p(\tilde{\mathrm{y}}_n|\boldsymbol{\theta}^s), \quad (3)$$

- Root mean square error:

$$\text{RMSE} = \frac{1}{S}\sum_{s=1}^{S}\sqrt{\frac{1}{N}\sum_{n=1}^{N}(\mathrm{y}_n^{(s)} - \tilde{\mathrm{y}}_n)^2}. \quad (4)$$

- Predictive accuracy index (PAI): proportion of crimes occurring in marked hotspots divided by the proportion of the study region marked as hotspots (Chainey et al. 2008).
- Predictive efficiency index (PEI): number of crimes predicted by the model for a given area size divided by the maximum number of crimes for the given area size (Hunt 2016).

Results (1 year)

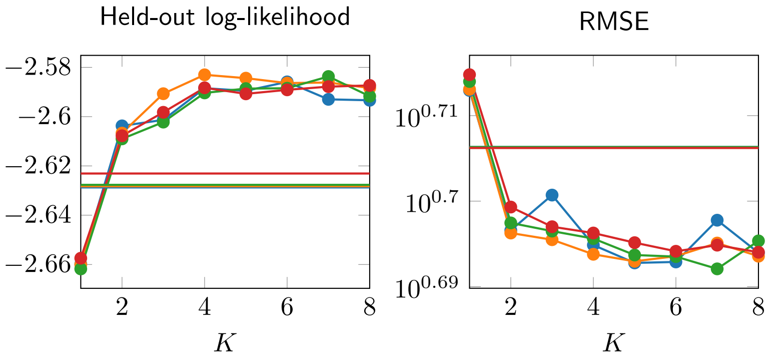


Figure: Evaluation of the performance of SAM-GLM (——), compared to LGCP (- - -)for a one-year dataset. Log-likelihood and root mean square error for the held-out data are shown for different model specifications: specification 1 (——), specification 2 (——), specification 3 (——), specification 4 (——). Blocking: MSOA, training data: burglary 2015, test data: burglary 2016.

Results (3 years)

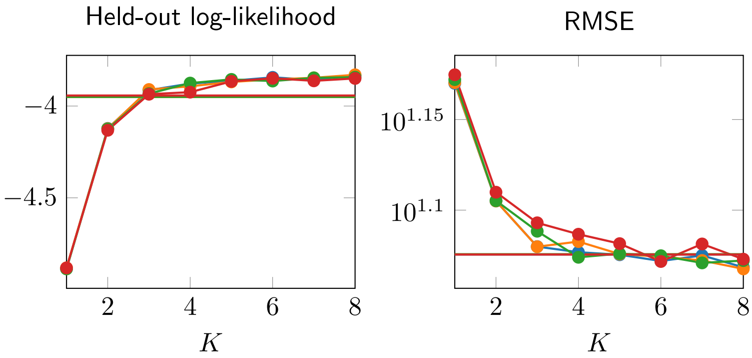


Figure: Evaluation of the performance of SAM-GLM (——), compared to LGCP (- - -) for a three-year dataset. Log-likelihood and root mean square error for the held-out data are shown for different model specifications: specification 1 (——), specification 2 (——), specification 3 (——), specification 4 (——). Blocking: MSOA, training data: burglary 2013-2015, test data: burglary 2016-2018.

Hotspot performance metrics (1 year)

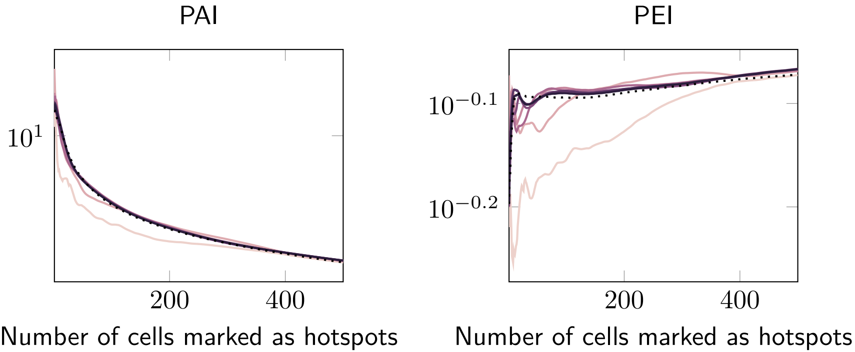


Figure: PAI/PEI performance SAM-GLM (——) and LGCP (·····) models, using specification 4. For the SAM-GLM results, the colour of the line represents the number of components: $K = 1$(——), $K = 2$(——), $K = 3$(——), $K = 4$(——), $K = 5$(——), $K = 6$(——), $K = 7$ (——). Training data: burglary 2015, test data: burglary 2016.

Hotspot performance metrics (3 years)

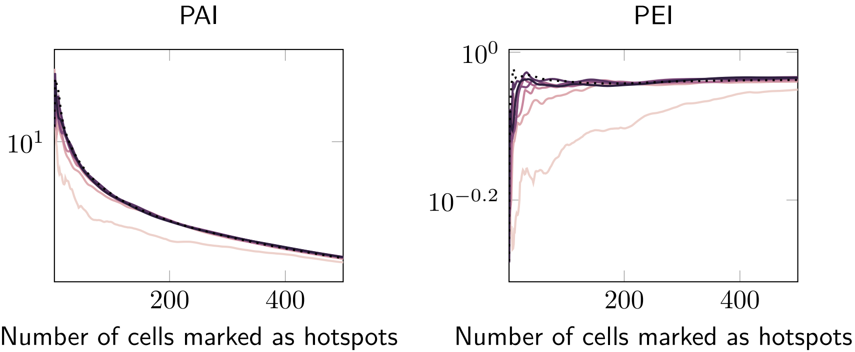


Figure: PAI/PEI performance SAM-GLM (——) and LGCP (······) models, using specification 4. For the SAM-GLM results, the colour of the line represents the number of components: $K = 1$(——), $K = 2$(——), $K = 3$(——), $K = 4$(——), $K = 5$(——), $K = 6$(——), $K = 7$ (——). Training data: burglary 2013-2015, test data: burglary 2016-2018.

Block size sensitivity (1 year)

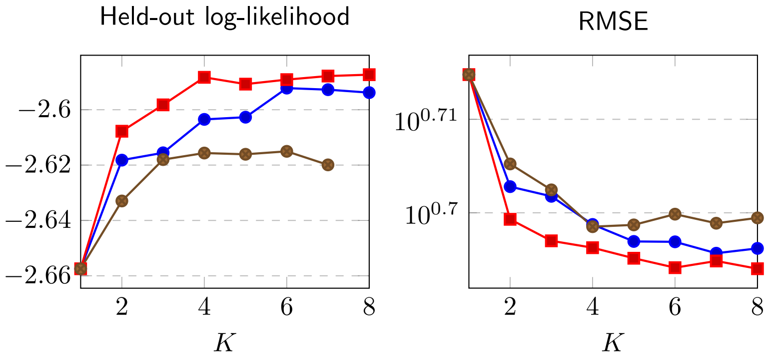


Figure: Log-likelihood and root mean square error for the held-out data for different block sizes: MSOA (—■—), LAD(—●—), single block(—⊗—). Model specification:4, training data: burglary 2015, test data: burglary 2016.

Block size sensitivity (3 years)

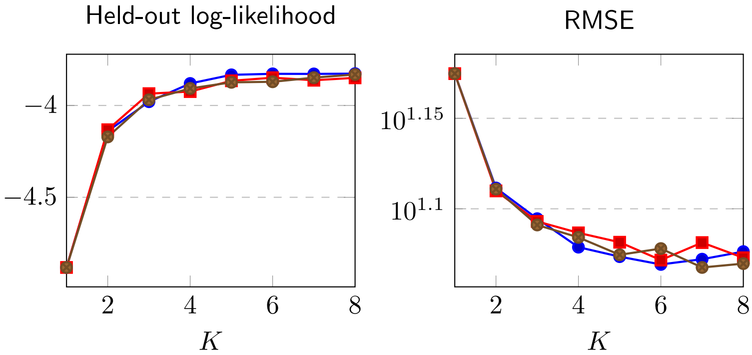


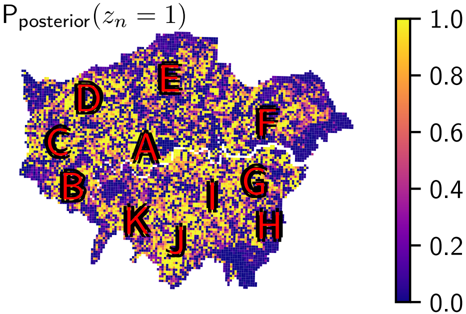
Figure: Log-likelihood and root mean square error for the held-out data for different block sizes: MSOA (red square line), LAD(blue circle line), single block(brown crossed-circle line). Model specification:4, training data: burglary 2013-2015, test data: burglary 2016-2018.

Interpretation of results

To effectively compare the effects of a covariate across different mixture components, we consider a **covariate importance measure**, defined as

$$\mathtt{IMP}_{kj} = 1 - \frac{\sum_n I(\mathsf{z}_n = k)(\mathsf{y}_n - \hat{\mathsf{y}}_{n\tilde{\boldsymbol{\beta}}})^2}{\sum_n I(\mathsf{z}_n = k)(\mathsf{y}_n - \hat{\mathsf{y}}_{n\bar{\boldsymbol{\beta}}^j})^2}, \tag{5}$$

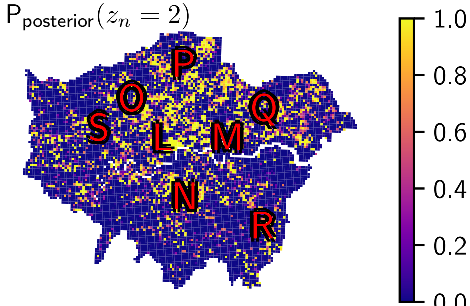
Allocations 1



IMP, component 1		
intercept	0.947 (0.001)	+
log households	0.887 (0.002)	+
log POIs (all)	0.399 (0.022)	+
accessibility	0.225 (0.024)	+
log house price	0.100 (0.017)	+
ethnic heterogeneity	0.083 (0.016)	+
occupation variation	0.025 (0.011)	+
population turnover	0.011 (0.004)	+
(Semi-)detached houses	0.002 (0.002)	+

Kensington, Fulham, and Shepherd's Bush (A); Richmod, and Kingston (2), Southall (C); Edgware (D); Hampstead (E); Barking and Dagenham (F); Bexley (G); Orpington (H); Bromley (I), Croydon, and Purley (J), New Malden, and Morden (K)

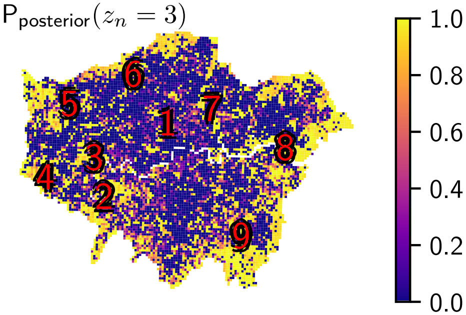
Allocations 2



IMP, component 2		
intercept	0.955 (0.001)	+
log households	0.840 (0.003)	+
accessibility	0.554 (0.012)	+
log POIs (all)	0.510 (0.017)	+
ethnic heterogeneity	0.192 (0.017)	+
occupation variation	0.098 (0.021)	+
population turnover	0.032 (0.006)	-
log house price	0.020 (0.011)	-
(Semi-)detached houses	0.003 (0.002)	+

Soho, Mayfair, Covent Garden, Marylebone, Fitzrovia (L); Shoreditch and Stratford (M); Streatham and Tooting Bec (N); Wembley, and Brent (O); Enfield, Hampstead (P); Romford (Q); Orpington (R); Wembley, Harrow (S)

Allocations 3



IMP, component 3		
log households	0.946 (0.003)	-
intercept	0.906 (0.004)	+
log POIs (all)	0.808 (0.015)	-
occupation variation	0.719 (0.060)	+
log house price	0.680 (0.027)	+
accessibility	0.435 (0.086)	+
ethnic heterogeneity	0.050 (0.024)	+
(Semi-)detached houses	0.002 (0.007)	+
population turnover	0.001 (0.007)	+

Hyde Park, Regent's Park, Hampsted Heath (1), Richmod nad Bushy parks (2), Osterley Park and Kew botanic gardens (3), Heathrow airport (4), RAF Northolt and parks near Harrow (5), Edgware fields (6), Lee Valley (7), industrial zone in Barking and Rainham Marshes (8), parks around Bromley and Biggin Hill airport (9)

Remarks

- The proposed approach allows for fast sampling and achieves performance comparable to LGCP. One posterior sample from the proposed model is of $\mathcal{O}(N \times K)$ time complexity, compared to LGCP's $\mathcal{O}(N^3)$.
- The model gives insights as to which covariate is important for each component.
- The allocation posterior is mostly determined by how well the $\boldsymbol{\beta}$ coefficients explain the log intensity at a given location. The mixture allocation prior does not play a strong role.
- Label-switching, which hampers interpretation, is not present for $K \leq 5$. It is harder to switch modes in higher dimensions.

Conclusions and further work

Conclusions:

- Using stationary GPs is not enough to effectively model point patterns in large urban domains.
- The blocking approach can significantly reduce computation time.
- More details can be found in the submitted arXiv paper: https://arxiv.org/pdf/1910.05212.pdf

Further work:

- Spatial dependence between the blocks.
- Non-blocking models such as Gibbs distribution for mixture allocation.

Bibliography I

Banerjee, S., Carlin, B. P. & Gelfand, A. E. (2015), *Hierarchical modeling and analysis for spatial data*, number 135 *in* 'Monographs on statistics and applied probability', second edition edn, CRC Press, Taylor & Francis Group, Boca Raton.

Chainey, S., Tompson, L. & Uhlig, S. (2008), 'The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime', *Security Journal* **21**(1-2), 4–28.

Cox, D. R. (1955), 'Some Statistical Methods Connected with Series of Events', *Journal of the Royal Statistical Society. Series B (Methodological)* **17**(2), 129–164.

Diggle, P. J., Moraga, P., Rowlingson, B. & Taylor, B. M. (2013), 'Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm', *Statistical Science* **28**(4), 542–563.

Duane, S., Kennedy, A., Pendleton, B. J. & Roweth, D. (1987), 'Hybrid monte carlo', *Physics Letters B* **195**(2), 216 – 222.

Fernández, C. & Green, P. J. (2002), 'Modelling spatially correlated data via mixtures: a Bayesian approach', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**(4), 805–826.

Bibliography II

Gelfand, A. E., Kim, H.-J., Sirmans, C. F. & Banerjee, S. (2003), 'Spatial Modeling With Spatially Varying Coefficient Processes', *Journal of the American Statistical Association* **98**(462), 387–396.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013), *Bayesian Data Analysis*, Chapman and Hall/CRC.

Geman, S. & Geman, D. (1984), 'Stochastic relaxation, gibbs distributions, and the bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **PAMI-6**(6), 721–741.

Gneiting, T. & Raftery, A. E. (2007), 'Strictly Proper Scoring Rules, Prediction, and Estimation', *Journal of the American Statistical Association* **102**(477), 359–378.

Green, P. J. & Richardson, S. (2002), 'Hidden Markov Models and Disease Mapping', *Journal of the American Statistical Association* **97**(460), 1055–1070.

Hildeman, A., Bolin, D., Wallin, J. & Illian, J. B. (2018), 'Level set Cox processes', *Spatial Statistics* **28**, 169–193.

Bibliography III

Hunt, J. M. (2016), Do crime hot spots move? Exploring the effects of the modifiable areal unit problem and modifiable temporal unit problem on crime hot spot stability, PhD Thesis, American University, Washington, D.C.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), ‘Equation of State Calculations by Fast Computing Machines’, *The Journal of Chemical Physics* **21**(6), 1087–1092.

Tobler, W. R. (1970), ‘A computer movie simulating urban growth in the detroit region’, *Economic Geography* **46**, 234–240.