Spatial Poisson Mixtures: London Burglary Case Study

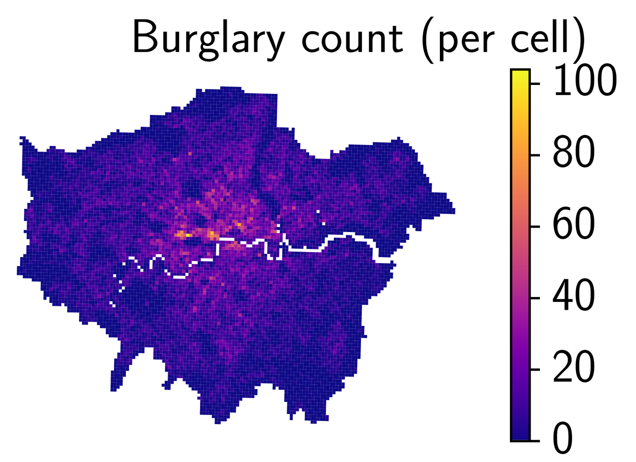
Jan Povala
Imperial College London

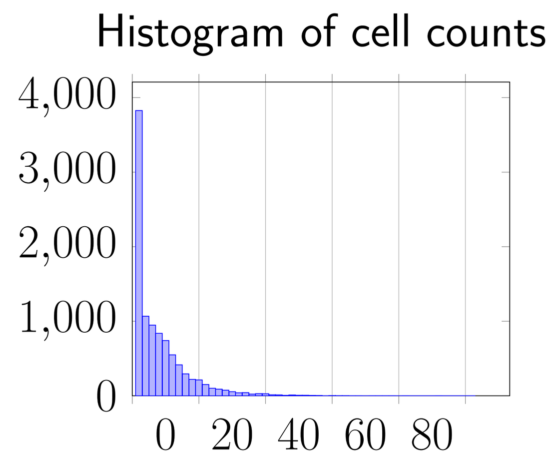
(this is a joint work with **Mark Girolami** and **Seppo Virtanen** recently accepted at The Journal of the Royal Statistical Society. Series C: Applied Statistics)

This work has been funded by EPSRC
jp2011/spatial-poisson-mixtures

Motivation

We are interested in modelling the intensity of the point pattern of burglaries and using the model to provide socioeconomic insights into criminal behaviour.





As is clear from the plots above, this data exhibits two common phenomena:

- **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.

Common approaches

- The go-to model for modelling spatial dependence of point patterns is the log-Gaussian Cox process model (Diggle et al. 2013).
- 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).

The approaches suffer from limited scalability, they often focus only on one of the two phenomena above, or provide limited interpretability.

Proposed model

We discretise the point pattern over a specific period of time into a regular grid and model the number of crime occurrences in each cell, y_n, as a Poisson random variable whose log-intensity is a linear component that is conditioned on the assignment to a mixture component k, with $k = 1, \ldots, K$. Mixture allocation for each cell is given by a categorical random variable with event probabilities $\boldsymbol{\pi}_b$ where b refers to the encompassing block of a cell. The blocks span the whole study region and each block is a group of contiguous cells.

$$
\begin{aligned}
\mathsf{y}_n | \mathsf{z}_n = k, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \boldsymbol{X}_n &\sim \text{Poisson}\left(\exp\left(\boldsymbol{X}_n^\top \boldsymbol{\beta}_k\right)\right) \\
\mathsf{z}_n | \boldsymbol{\pi} &\sim \text{Cat}(\pi_{1,b[n]}, \ldots, \pi_{K,b[n]}) \\
\boldsymbol{\pi}_{k,b} | f_k &= \frac{\exp(\mathsf{f}_{k,b[n]})}{\sum_{l=1}^{K} \exp(\mathsf{f}_{l,b[n]})} \\
f_k | \boldsymbol{\theta}_k &\sim \mathcal{GP}(0, \kappa_{\boldsymbol{\theta}_k}(\cdot, \cdot)) \\
\boldsymbol{\theta}_k &\sim \text{kernel-dependent prior} \\
\beta_{k,j} | \sigma^2_{k,j} &\sim \mathcal{N}(0, \sigma^2_{k,j}) \\
\sigma^2_{k,j} &\sim \text{InvGamma}(1, 0.01).
\end{aligned}
$$

Experiment

To formulate the hypotheses about the linear component of the log-intensity, we exploit the existing criminology studies on the target selection process for burglaries. We work under the framework where the offender is maximising the reward, minimising the risk and the effort. The predictors we consider are: household density, POI density, residential turnover, household income, real estate prices, transport accessibility, and other socioeconomic measures.

We aggregate the one-year point pattern from 2015 over a grid. Number of mixture components, K, ranges from 1 to 8. For the blocking structure we use existing census output areas (sensitivity study of this choice is in the paper). Additionally, we make a simplifying assumption that the mixture probabilities for each block, $\boldsymbol{\pi}_b$'s, are independent and distributed as $\text{Dirichlet}(1/K, \ldots, 1/K)$. We estimate the model using a Metropolis-within-Gibbs scheme.

Performance evaluation

As the baseline for our comparison we use the log-Gaussian Cox where the log-intensity is a Gaussian process with the mean specified by a linear model that includes covariates discussed above. We asses the performance using the RMSE on the one-period-ahead data. We also asssess the ability to predict hotspots using measures which were introduced by criminologists: predictive accuracy index (PAI) and predictive efficiency index (PEI).

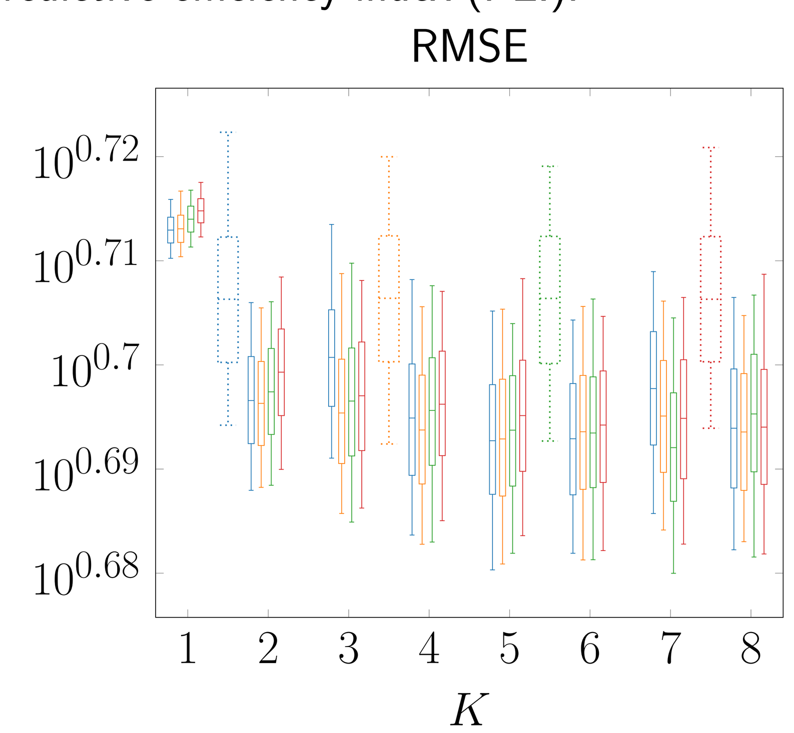
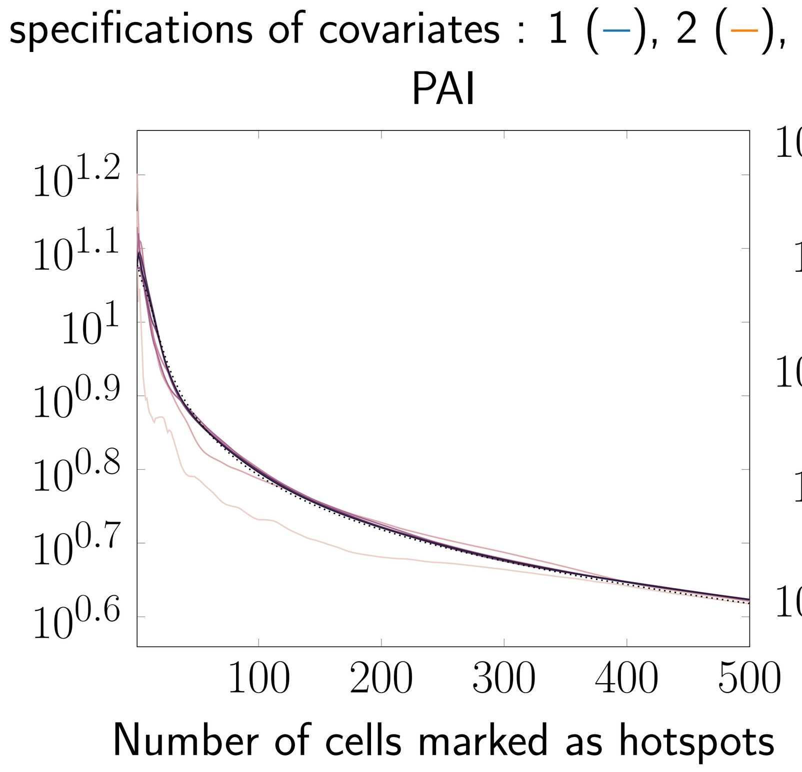


Figure 1: Proposed model (–) vs LGCP (···). We show the results for different specifications of covariates : 1 (–), 2 (–), 3 (–), 4 (–).



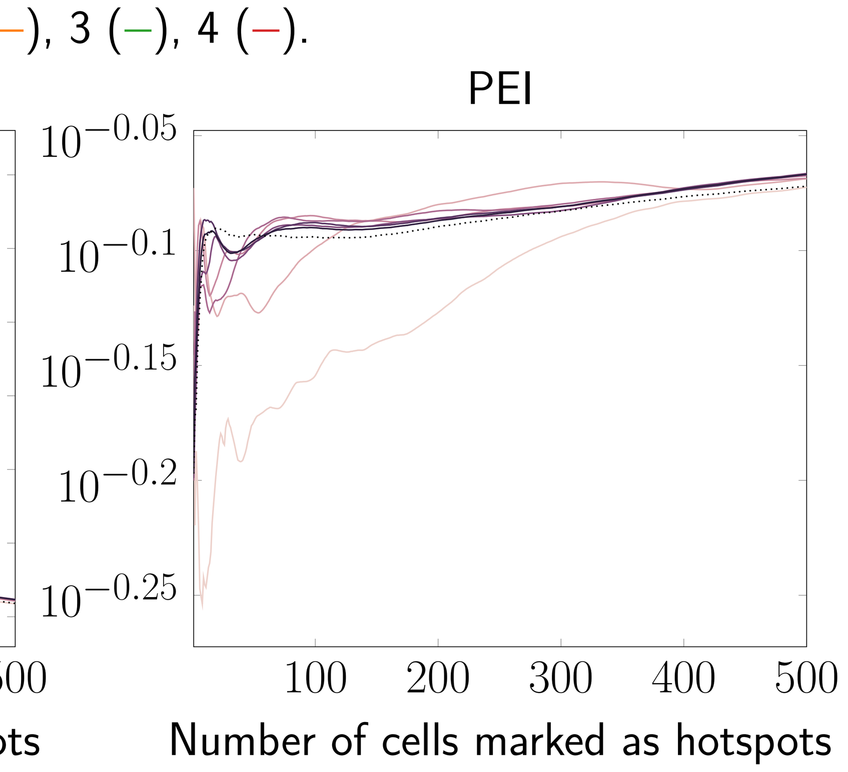
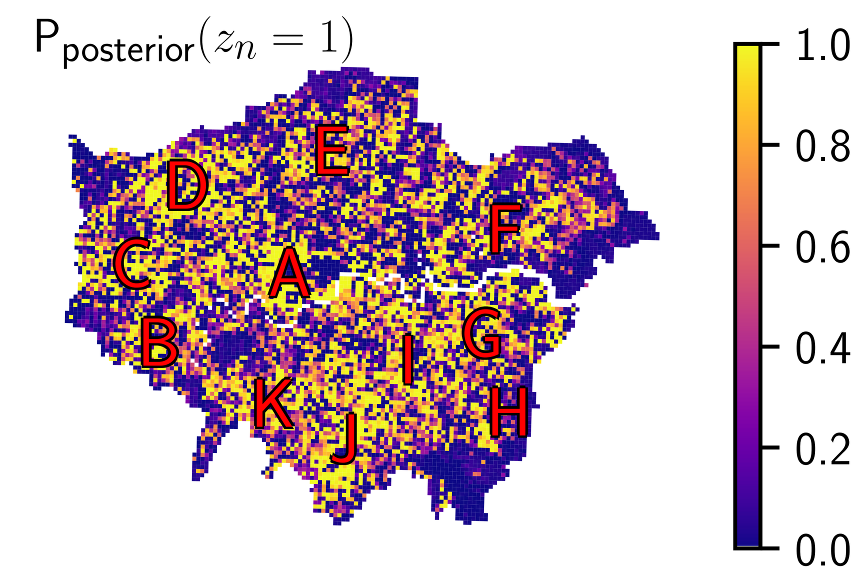


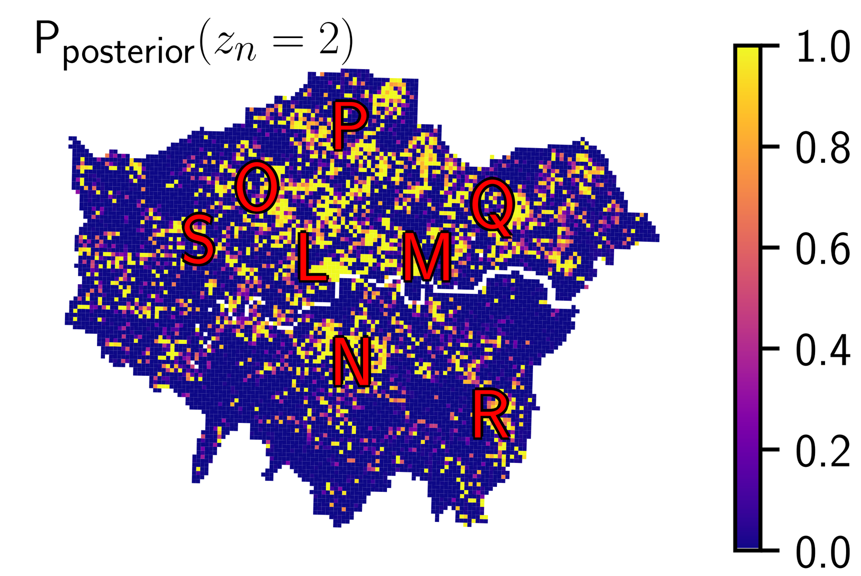
Figure 2: Proposed model (–) vs LGCP (····) with specification 4. $K = 1$(–), $K = 2$(–), $K = 3$(–), $K = 4$(–), $K = 5$(–), $K = 6$(–), $K = 7$ (–).

Interpretation of results

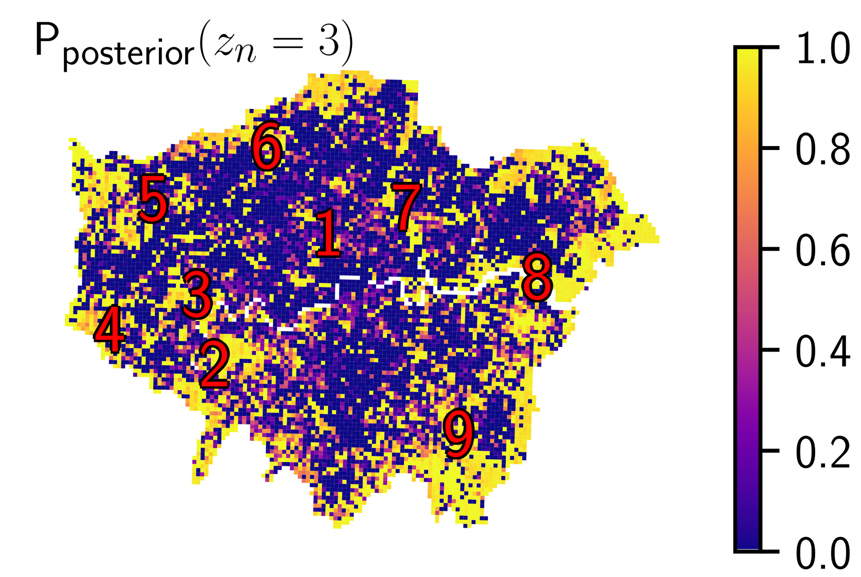
We show the posterior allocation of the cells to one of the $K = 3$ components. For each component, we assess the importance of each covariate using the measure that corresponds to the degradation of the fit if the covariate was removed from the model. It is clear that clusters of residential locations ($k = 1$), city centre and high street ($k = 2$), and non-urban areas ($k = 3$) have been inferred with the importance of each covariate changing for the components.



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)	+



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)	+



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)	+

Discussion

Using stationary GPs is not enough to effectively model point patterns in large urban domains. The blocking approach can significantly reduce computation time. 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 proposed model achieves the performance comparable to LGCP, is interpretable, and provides useful criminological insights.