Burglary in London: Insights from Statistical Heterogeneous Spatial Point Processes

Jan Povala[1,3], Seppo Virtanen[2,3], and Mark Girolami[2,3]

[1]Department of Mathematics, Imperial College London
[2]Department of Engineering, University of Cambridge
[3]The Alan Turing Institute

October 12, 2019

Abstract

To obtain operational insights regarding the crime of burglary in London we consider the estimation of effects of covariates on the intensity of spatial point patterns. By taking into account localised properties of criminal behaviour, we propose a spatial extension to model-based clustering methods from the mixture modelling literature. The proposed Bayesian model is a finite mixture of Poisson generalised linear models such that each location is probabilistically assigned to one of the clusters. Each cluster is characterised by the regression coefficients which we subsequently use to interpret the localised effects of the covariates. Using a blocking structure of the study region, our approach allows specifying spatial dependence between nearby locations. We estimate the proposed model using Markov Chain Monte Carlo methods and provide a Python implementation.

1 Introduction

Use of statistical models for understanding and predicting criminal behaviour has become increasingly relevant for police forces, and policymakers (Felson & Clarke 1998, Bowers & Hirschfield 1999, PredPol 2019). While short-term forecasting of criminal activity has been used to better allocate policing resources (Taddy 2010, Mohler et al. 2011, Aldor-Noiman et al. 2017, Flaxman et al. 2019, PredPol 2019), understanding the criminal behaviour and target selection process through statistical models has a potential to be used for designing policy changes, and development programs (Felson & Clarke 1998). In this work, we consider the problem of burglary crime in London. In the UK, burglary is a well-reported crime, but the detection rate remains at the 10-15% level (Smith et al. 2013). Rather than being concerned with short-term forecasting, we focus on understanding the effects of spatially varying explanatory variables on the target selection through descriptive regression models. Inferences made using these models help us understand the underlying mechanisms of burglary. The main contribution of this work is the integration of statistical methods in spatial modelling with the findings from the criminological literature.

Instances of burglary can be represented as a *spatial point pattern* – a finite or countably infinite set of points in the study region. Understanding the intensity of the occurrences through spatially varying covariates is the main objective of this work. The task of estimating the effects of the covariates on the intensity can be classified as a multivariate regression modelling, in which systematic effects of the explanatory variables are of interest while taking into account other random effects such as measurement errors and spatial correlation (McCullagh & Nelder 1998). In the context of spatial data, it has been widely recognised that multivariate regression modelling techniques which do not account for *spatial dependence* and *spatial heterogeneity* can lead to biased results and faulty inferences (Anselin et al. 2000). Spatial dependence refers to the Tobler's first law of geography: "everything is related to everything else, but near things are more related than distant things"(Tobler 1970). Spatial dependence manifests mostly in the spatial correlation of the residuals of a model. In non-spatial settings, the residuals are often assumed to be independent and identically distributed (McCullagh & Nelder 1998). Spatial heterogeneity is exhibited when the object of interest, in our case the intensity of a point pattern, shows location-specific behaviour. For example, properties of the burglary point pattern in a city centre are going to be different from the properties in a residential area. Formalising these two concepts and incorporating them into modelling methodology results in more accurate models (Anselin et al. 2000).

Log-Gaussian Cox process (Møller et al. 1998, Møller & Waagepetersen 2007) has been a common approach for modelling intensity of spatial point patterns (Diggle et al. 2013, Serra et al. 2014, Flaxman et al. 2015). The flexibility of the model is due to the Gaussian process part through which complex covariance structures, including spatial dependence and heterogeneity, can be accounted for. In practice, stationary covariance functions are used for computational reasons (Diggle et al. 2013). As a result, log-Gaussian Cox process models with stationary covariance functions handle spatial dependence but do not account for spatial heterogeneity.

To account for spatial heterogeneity, mixture based approaches have been adopted as a way of enriching the collection of probability distributions for modelling observed data (Green 2010, Fernández & Green 2002). Notably, Knorr-Held & Raßer (2000), Fernández & Green (2002), Green & Richardson (2002) use mixtures for modelling the elevations of disease incidence. While these methods improve the model fit by accounting for spatial heterogeneity as wells as spatial dependence, they provide little interpretation as to why the level is elevated in certain areas. Also, these three methods have been tested only on a modest scale, regions consisting of 544, 94, and 94 areal units, respectively. Following this line of work, Hildeman et al. (2018) proposed a method in which each mixture component can take a rich representation that may include covariates. Although this model is very rich in representation, the empirical study in the paper was limited to the case of two mixtures, with one of the components being held constant. Their study of a tree point pattern and its dependence on soil type was carried out on a region discretised into a grid with 2461 cells.

A very different approach to controlling for spatial heterogeneity has been taken by Gelfand et al. (2003) who allow regression effects to vary across the spatial region. The method treats the effects of the covariates as a multivariate spatial process. The process is, however, very challenging to fit and is often limited to 2 or 3 covariates (Banerjee et al. 2015, p.288). A simpler version of the same idea is geographically-weighted regression (Brunsdon et al. 1996), where the regression coefficients are weighted by a latent component whose properties have to be specified a priori or learned through cross-validation.

Motivated by the computational challenges and/or limited interpretability of the aforementioned approaches, we propose a mixture based method that takes into account spatial dependence, and is able to discover latent groups of locations and characterise each group by group-specific effects of spatially varying covariates. To effectively estimate the model parameters from limited data and to quantify the uncertainty of the estimates, we follow the Bayesian framework.

More specifically, our approach builds upon model-based clustering methods (Grün & Leisch 2008), in which each cluster is allowed to follow a different model. We cater for spatial dependence using an approach inspired by Fernández & Green (2002) and Knorr-Held & Raßer (2000). Our model probabilistically assigns each location to a particular cluster, while imposing spatial dependence through prior information. The prior information will suggest that locations that are close to each other are likely to belong to the same cluster. We define a pair of locations to be close if both of them are in the same block. We use the blocking structure predefined by the census tracts but our method allows defining custom ones. The posterior inferences for the individual clusters and the assignments of locations to clusters are used to draw conclusions and provide insights about the heterogeneity of the spatial point pattern across the study region.

We show that the proposed methodology effectively models burglary crime in London. By comparing our approach to log-Gaussian Cox process (LGCP), a standard model for spatial point patterns (Diggle et al. 2013), we show that our method outperforms LGCP, is more computationally tractable, and the interpretation of inferred quantities provides useful criminological insights.

The rest of the paper is structured as follows. Section 2 defines the model and details the inference method, section 3 elaborates on our application and gives the discussion of model choices specific to our application. The obtained results are discussed in section 4. Section 5 concludes the paper.

2 Modelling methodology

It is widely recognised that burglary crime is spatially concentrated (Brantingham & Brantingham 1981, Clare et al. 2009, Johnson & Bowers 2010). It is also apparent that some areas in the study region will exhibit extreme behaviour. For example, areas with no buildings such as parks will have no burglary for structural reasons. To effectively model burglary, these phenomena need to be accounted for using *spatial effects*. The spatial effects consist of *spatial dependence* and *spatial heterogeneity* (Anselin et al. 2000).

For our modelling framework, we choose the Bayesian paradigm because it allows us to formalise prior knowledge, and to quantify uncertainty in the unknown quantities of our model. In our application, burglary data are given as a point pattern over a fixed period of time. We discretise the point pattern

onto a grid of N cells by counting the points in each cell. Although any form of discretisation is allowed, throughout this paper we work with a regular grid.

We model the count of points in a cell n, y_n as a mixture of K Poisson-distributed random variables, with the logarithm of the intensity driven by a linear term, which is specific for each mixture component, indexed by $k = 1, \ldots, K$. The linear term is a linear combination of J covariates for cell n, $\mathbf{X}_n$, and the corresponding coefficients, $\boldsymbol{\beta}_k$. The covariates need to be specified for the application of interest and usually include the intercept. To specify the prior distribution for the regression coefficients we use a prior that shrinks the estimate towards zero. For each coefficient, we set as $\beta_{k,j} \sim \mathcal{N}(0, \sigma^2_{k,j})$, where $\sigma^2_{k,j} \sim \mathrm{InvGamma}(1, 0.01)$. We put the uniform prior on the intercepts, if present.

Each cell n is probabilistically allocated to one of the K components through an allocation variable, z_n, which is a categorical random variable with event probabilities given by the mixture weights prior, $\boldsymbol{\pi}_{b(n)} \sim \mathrm{Dirichlet}(\alpha, \ldots, \alpha)$. The prior, $\boldsymbol{\pi}_{b(n)}$, is shared for all locations within cell n's block, $b(n)$. The blocks for the study region are defined as non-overlapping spatial areas spanning the whole study region. In many practical applications, the block structure is already defined by administrative units or census tracts. Block $b(n)$ is the block that contains the centroid point of cell n. We set a log-uniform prior on α. The block-specific event probabilities will express the belief that the effect of the covariates is the same within the block unless strong evidence from the observed data outweighs this information.

We refer to the proposed model as *spatially-aware mixture of Poisson generalised linear models (SAM-GLM)*. The formulation is summarised in the equation and the graphical representation shown in figure 1.

$$
\begin{aligned}
\mathrm{y}_n | \mathrm{z}_n = k, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \mathbf{X}_n &\sim \mathrm{Poisson}\left(\exp\left(\mathbf{X}_n^\top \boldsymbol{\beta}_k\right)\right) \\
\mathrm{z}_n | \boldsymbol{\pi} &\sim \mathrm{Categorical}(\pi_{b(n),1}, \ldots, \pi_{b(n),K}) \\
\boldsymbol{\pi}_b | \alpha &\sim \mathrm{Dirichlet}(\alpha, \ldots, \alpha) \\
\beta_{k,j} | \sigma^2_{k,j} &\sim \mathcal{N}(0, \sigma^2_{k,j}) \\
\sigma^2_{k,j} &\sim \mathrm{InvGamma}(1, 0.01) \\
\log \alpha &\sim 1/\mathrm{Log\text{-}uniform}.
\end{aligned}
$$

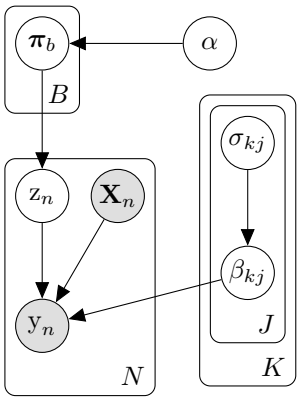


Figure 1: Summary of the SAM-GLM model and its graphical representation.

In the proposed model, we handle *spatial heterogeneity* using the mixture components, each of which specifies a set of J regression coefficients, $\boldsymbol{\beta}_k$. *Spatial dependence* is considered through mixture allocation components and the blocks structure. The blocks are used to specify the prior on localised behaviour.

2.1 Excess of zeros, overdispersion

Two common challenges encountered when modelling count data using standard Poisson regression models are *excess of zeros* and *overdispersion* (McCullagh & Nelder 1998, Breslow 1984). The former refers to the presence of zeros that are structural, rather than due to chance. In the context of burglary, structural zeros occur in locations with no buildings, e.g. parks. The latter issue refers to the situation when the variability of the observed data is higher than would be expected based on a particular statistical model. The standard Poisson regression model for the burglary point pattern, a special case of our model ($K = 1$), exhibits overdispersion for different specifications of the covariates term – see section A in the appendix. The flexibility of our proposed model can account for the excess of zeros by identifying a low-count component to which areas of low intensity will be assigned. Similarly, overdispersion can be alleviated by introducing another component to the mixture.

2.2 Inference

Statistical inference in the Bayesian setting involves inferring the posterior probability distribution for the quantities of interest. In this work, we choose Markov Chain Monte Carlo (MCMC) method to sample from the posterior distributions (Gelman et al. 2013).

To simplify the inference, σ^2_{kj} scale parameter and all $\boldsymbol{\pi}_b$'s are analytically integrated out. The quantities of interest are the allocation vector $\mathbf{z}$, and the regression coefficient vector for each mixture

component, $\boldsymbol{\beta}_k$. For brevity, let $\boldsymbol{\beta}$ be a $K \times J$ matrix of regression coefficients for all mixture components and each covariate, and let $\mathbf{X}$ be a $N \times J$ matrix of all covariates for each location. We employ Metropolis-within-Gibbs scheme (Geman & Geman 1984, Metropolis et al. 1953) and sample from the posterior in three 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), for which efficient sampling schemes are available, e.g. Girolami & Calderhead (2011).

2. Mixture allocation can be sampled cell by cell directly using the following equation

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

   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.

3. The hyper-parameter α is sampled using a random-walk Metropolis algorithm (Metropolis et al. 1953). The unnormalised conditional distribution for α is given as

$$p(\alpha|\boldsymbol{\beta}, \mathbf{y}, \mathbf{z}) \propto p(\mathbf{z}|\alpha)p(\alpha), \tag{3}$$

Full derivations of these equations can be found in the appendix (equation 23, equation 24, equation 25).

In terms of computational tractability, equation 1 takes $\mathcal{O}(N + J)$ steps, equation 2 requires $\mathcal{O}(N \times K)$ steps, and equation 3 requires $\mathcal{O}(B \times K)$ steps provided that component cell counts for each block are pre-computed. To contrast it with a standard model for spatial point patterns, one sample from log-Gaussian Cox process involves matrix inversions that require $\mathcal{O}\big(N^3\big)$ steps (Diggle et al. 2013), or $\mathcal{O}\big(N^{3/2}\big)$ when the grid structure and factorisable covariance functions are assumed (Flaxman et al. 2015). Using the approach of blocking, the inference is faster than it is for models which use Gaussian random fields to account for spatial dependence. However, it needs to be mentioned that this model handles spatial dependence only within the same block.

2.3 Identifiability

Specifying a mixture model means that the model likelihood is invariant under relabelling of the mixture components (Celeux et al. 2000). This issue is commonly referred to as lack of identifiability. In the context of SAM-GLM model, $p(\mathbf{y}|\mathbf{z}, \mathbf{X}, \boldsymbol{\beta})$ is invariant under the relabelling of $\boldsymbol{\beta}_k$ and $\boldsymbol{\pi}_b$'s, which are the component-specific model parameters.

Exploration in high dimensional spaces is much harder for an MCMC sampler in general. As the dimension of the parameter space for the mixture model increases, the sampler is likely to explore only one of the $K!$ possible modes. To switch to a different mode, the sampler would have to get past the area of low probability mass surrounding the chosen mode. However, note that as the number of mixture components increases, the chance of the sampler switching to a different mode increases as the shortest distance between a pair of component-specific parameters is likely to decrease.

Since the identifiability issue poses a problem only for the interpretation of the parameters, we inspect the traceplot of the Markov chain for each parameter to assert that relabelling is not present.

2.4 Evaluation and interpretation

We use a number of evaluation metrics to choose between different model specifications for the linear covariates term, as well as to compare it with other competing methods. Below, we discuss the evaluation metrics we use.

2.4.1 Watanabe-Akaike information criterion

Watanabe-Akaike information criterion (WAIC) is a measure of model fit penalised by model complexity which is often used in Bayesian statistics (Gelman et al. 2013). The criterion computes the negative of the sum of log pointwise predictive density and corrects for the effective number of parameters using pointwise variance of predictive log density as shown below (Gelman et al. 2013):

$$\text{WAIC} = -2\sum_{n=1}^{N}\log\left(\frac{1}{S}\sum_{s=1}^{S}p(\mathbf{y}_n|\boldsymbol{\theta}^{(s)}\right) + 2\sum_{n=1}^{N}V_{s=1}^{S}\left(\log p\left(\mathbf{y}_n|\boldsymbol{\theta}^{(s)}\right)\right), \tag{4}$$

where $V_{s=1}^{S}(\cdot)$ computes the sample variance, $\boldsymbol{\theta}$ is the vector of model parameters, of which S posterior samples $(\boldsymbol{\theta}^{(1)}, \dots \boldsymbol{\theta}^{(S)})$ have been obtained.

WAIC is a relative measure used for model comparison. A lower value indicates superior predictive power.

2.4.2 Forecast scoring

By assuming the same distribution for the one period ahead point pattern, we evaluate out-of-sample performance using forecast scoring rules. A forecast scoring rule is a function of the probability distribution of the forecast and the realised value (Gneiting & Raftery 2007). The computed score can subsequently be used to rank different models. For a thorough review of forecast scoring we refer the reader to Gneiting & Raftery (2007). Throughout this work we use the energy scoring rule.

Given samples from the posterior distribution of the model parameters, $\boldsymbol{\theta}^{(1)}, \dots \boldsymbol{\theta}^{(S)}$, we can obtain samples from the joint predictive probability distribution for the counts at all N locations, $\mathbf{y}^{(1)}, \dots, \mathbf{y}^{(S)}$, using sampling distribution of the data, $p(\mathbf{y}|\boldsymbol{\theta})$. Then, using the realised next-period value, $\tilde{\mathbf{y}} = (\tilde{y_1}, \dots, \tilde{y_N})$, the energy score is defined as follows:

$$\text{Energy score} = \frac{1}{S}\sum_{s=1}^{S}\|\mathbf{y}^{(s)} - \tilde{\mathbf{y}}\|_2^{\gamma} - \frac{1}{2S^2}\sum_{i=1}^{S}\sum_{j=1}^{S}\|\mathbf{y}^{(i)} - \mathbf{y}^{(j)}\|_2^{\gamma}, \tag{5}$$

where $\|\cdot\|_2$ is the Euclidean norm. We use $\gamma = 1$ throughout this paper.

2.4.3 Hotspot prediction

Given that burglary is our object of interest, we also evaluate models with respect to their ability to effectively model areas of high intensity, so called *hotspots*. The predictive accuracy index (PAI) and predictive efficiency index (PEI) are two standard approaches in criminology for assessing the ability to predict crime hotspots.

PAI, introduced by Chainey et al. (2008), assesses the ability to capture as many crime instances as possible with as little area as possible. It is defined as

$$\text{PAI} = \frac{c/C}{a/A},$$

where a is the area flagged as hotspots, A is the total area of the study region, c is the number of crimes in the flagged hotspots, and C it the total number of crimes in the study region.

However, for certain types of crime that are more serious and less frequent it is important that each instance of crime is captured. PEI measures how effective the model forecasts are compared to what would a perfect model predict (Hunt 2016). It is defined as

$$\text{PEI} = \frac{c}{c^*},$$

where c is the number of crimes in the flagged hotspots, and c^* is the maximum number of crimes that would have been captured by a perfect model using an area equal in size to the flagged hotspots.

In our context of a regular grid, we use both measures to compare competing models when up to n cells are flagged as hotspots. The higher value indicates better hotspot prediction ability.

2.5 Interpretation of results

Estimating the effects of different spatial covariates helps us understand the underlying mechanisms of the point pattern. Regression coefficients ($\boldsymbol{\beta}_k$) alone are not directly interpretable when using the mixture model. This is because the distribution for a covariate is component-specific. For example, one mixture component is active in areas with very small values for a specific covariate while some other component is active in areas with high values. Comparing regression coefficients for that covariate across different components would not be appropriate. For this reason, we define a measure of the effect of a covariate and call it `CovEffect` . The objective of this measure is to assess the magnitude and the sign (positive/negative) of the effect of a covariate for a specific mixture component. We measure the magnitude of the effect for a covariate j of the mixture component k as the ratio of the sum of squared residuals for the full model and the sum of squared residuals for the same model without covariate j, which is then subtracted from one.

$$\texttt{CovEffect}_{kj} = 1 - \frac{\sum_n I(\mathrm{z}_n = k)(\mathrm{y}_n - \hat{\mathrm{y}}_{n\tilde{\boldsymbol{\beta}}})^2}{\sum_n I(\mathrm{z}_n = k)(\mathrm{y}_n - \hat{\mathrm{y}}_{n\bar{\boldsymbol{\beta}}^j})^2}, \tag{6}$$

where, $I(\mathrm{z}_n = k)$ is the indicator function for whether cell n is allocated to component k, $\hat{\mathrm{y}}_{n\tilde{\boldsymbol{\beta}}}$ is the predicted count using the full vector of inferred regression coefficients, and $\hat{\mathrm{y}}_{n\bar{\boldsymbol{\beta}}^j}$ is the predicted count using the regression coefficients with jth component set to zero. The magnitude of `CovEffect` is interpreted as a measure of degradation of fit if the given covariate was excluded from the model.

We determine the sign of the effect for a given covariate and a mixture component by inspecting the distribution of the covariate for the given component. We take the mean of the covariate across the cells that are allocated to the given component and if that is positive, we take the sign of the corresponding β_{kj} estimate. Otherwise, we take the negative of the sign of the β_{kj}.

2.6 Related models

Our approach builds on top of Fernández & Green (2002), Green & Richardson (2002) in that it allows for including the covariates into each mixture component, rather than just having intercept-only components. Their methods of assigning each location to a mixture component are based on logistic transformations of Gaussian random fields and Potts model, respectively. Both formulations allow spatial dependence between nearby locations, but at a significant computational cost. In our case, we focus on estimating the effects of covariates on the intensity of a point pattern, so we choose spatial blocking as a more computationally-efficient, but less expressive, method of accounting for spatial dependence.

Hildeman et al. (2018) model the log-intensity of a point pattern as a mixture of Gaussian random fields and mixture allocation is handled using a grouped-continuous model based on a Gaussian random field. Although their approach is theoretically flexible, it has only been applied at a limited scale. Our model is more constrained but provides better scalability.

Another flexible, but computationally intensive approach that was mentioned in the introduction is spatially varying coefficient processes proposed by Gelfand et al. (2003). In this approach, each covariate has an accompanying Gaussian process as its coefficient. Our method, in contrast, infers a set of 'types of locations', such that locations within each type share the same regression coefficient.

3 Application: London burglary crime

3.1 Data description

The methodology above has been developed to enable the analysis of our application – burglary in London. The data, published online by the UK police forces (police.uk 2019), are provided monthly as a spatial point pattern of not only residential but also non-residential burglary occurrences. The non-residential burglary refers to instances where the target is not a dwelling, for example, commercial or community properties. We discretise our study area into a regular grid by counting the number of burglary occurrences within each cell. We choose a grid for computational reasons when comparing to competing methods (see section B in the appendix). Given our focus on spatial modelling, we temporally aggregate the point pattern into two datasets: one-year dataset starting 01/2015 and ending 12/2015, and a three-year dataset, starting 01/2013 and ending 12/2015.

Our analysis uses land use data, socioeconomic census data from 2011, and points of interest data from 2018 to estimate their effect on the intensity of the burglary point pattern. Land use data are

available as exact geometrical shapes, the census variables are measured with respect to census tracts, called output areas (OA). The OAs have been designed to have similar population sizes and be as socially homogeneous as possible, based on the tenure of households and dwelling types. Each of the 25,053 OAs in London has between 100 people or 40 households and 625 people or 250 households. The OAs are aggregated into 4,835 lower super output areas (LSOA), which in turn are aggregated into 983 middle super output areas (MSOA). An LSOA has at least 1,000 people or 400 households and at most 3,000 people or 1,200 households. For an MSOA, the minimum is 5,000 people or 2,000 households, and the maximum is 15,000 people or 6,000 households. The points of interest data are given as a point pattern. To project the data measured at non-grid geometries (the census and land use data) onto the grid we use weighted interpolation which assumes that the data is uniformly distributed across the OA. For cells that have an overlap with more than one OA, we compute the value for each such cell by combining the overlapping OAs and adjusting for the size of the overlap.

3.2 Criminology background

We use existing criminology studies to identify explanatory variables and formulate hypotheses about burglary target selection. The target choice is a decision-making process of maximising *reward* with minimum *effort*, and managing the *risk* of being caught (a process analogous to optimal foragers in wildlife (Johnson & Bowers 2004)). Therefore, we categorise the explanatory variables into these three categories: reward, effort, and risk.

3.2.1 Reward, opportunities, attractiveness

Theoretically supported by rational choice theory (Clarke & Cornish 1985), offenders seek to maximise their reward by choosing areas of many opportunities and attractive targets. Firstly, the *number of dwellings* is used in the literature as a measure of the abundance of residential targets (Bernasco & Nieuwbeerta 2005, Clare et al. 2009, Townsley et al. 2015, 2016). *Real estate prices* and *household income* have been used in previous works as a proxy for the attractiveness of targets. The significance of their positive effect on residential burglary victimisation rate has been mixed, and varied depending on the study region and the statistical method used (Bernasco & Luykx 2003, Bernasco & Nieuwbeerta 2005, Clare et al. 2009, Townsley et al. 2015, 2016). The finding that the effect of affluence was weak in some studies can be explained by the fact that most burglars do not live in affluent areas and hence are not in their awareness spaces, i.e. operating in an affluent neighbourhood is for them an unfamiliar terrain and the risk of being caught is higher (Evans 1989, Rengert & Wasilchick 2010). Other measures of affluence that have been used include *house ownership rates* (Bernasco & Luykx 2003).

With regard to non-residential burglary, the literature is more sparse. An analysis of non-residential burglary in Merseyside county in the UK by Bowers & Hirschfield (1999) shows that non-residential facilities have a higher risk of both victimisation as well as repeat victimisation. In particular, sport and educational facilities have a disproportionately higher risk of being targeted compared to other types of facilities. In the crime survey of business owners in the UK, the retail sector is the most vulnerable to burglaries (gov.uk 2017). For our application, we will use points of interest database from Ordnance Survey which include retail outlets, eating and drinking venues, accommodation units, sport and entertainment facilities, and health and education institutions (Ordnance Survey (GB) 2018)

3.2.2 Effort, convenience

Using the framework of crime pattern theory (Brantingham & Brantingham 1993) and routine activity theory (Cohen & Felson 1979), offenders will prefer locations that are part of their routine activities or are convenient to them, i.e. they are in their activity or awareness spaces. The studies performed using the data on detected residential burglaries unanimously agree that areas *close to the offender's home* are more likely to get targeted (Bernasco & Nieuwbeerta 2005, Townsley et al. 2015, Menting et al. 2019, Clare et al. 2009). In the study based on a survey of offenders, Menting et al. (2019) argue that other awareness spaces than their residence play a significant role in target selection. These include previous addresses, neighbourhoods of their family and friends, as well as places where they work and go about their recreation and leisure.

As confirmed by numerous studies, the spatial topology of the environment plays a significant role in the choice of a target. Brantingham & Brantingham (1975) have shown that houses in the interior of a block are less likely to get burgled. Similarly, Townsley et al. (2015), Bernasco & Nieuwbeerta (2005) showed that *single-family dwellings* are more vulnerable to burglaries than multi-family dwellings

such as blocks of flats. Beavon et al. (1994) studied the effects of the street network and traffic flow on residential burglary and found that crime was higher in *more accessible* and *more frequented* areas. Similarly, Johnson & Bowers (2010) show that main street segments are more likely to become a burglary target. Clare et al. (2009), Bernasco et al. (2015) showed that the presence of connectors such as train stations increases the likelihood of being targeted, while the so-called barriers such as rivers or highways decrease it.

3.2.3 Risk, likelihood of completion

In the social disorganisation theory of crime (Shaw & McKay 1942, Sampson & Groves 1989), it is argued that social cohesion induces collective efficacy. The effect of collective efficacy on crime is twofold. First, strong social control deters those who are thinking of committing one. Second, it decreases the chance of a successful completion once an offender has chosen to do so. This theory focuses on the impact that social deprivation, economic deprivation, family disruption, ethnic heterogeneity, and residential turnover have on the crime rates within an area. Most offenders live in disadvantaged areas and often commit a crime in their awareness spaces (minimise effort). The attraction to 'prosperous targets' applies mostly to the local context (maximise gain). On the other hand, when a neighbourhood has high social cohesion (also known as 'collective efficacy'), there is mutual trust among neighbours and residents are more likely to intervene on behalf of the common good (Sampson et al. 1997).

In the context of residential burglary, *ethnic diversity* has been shown to be positively related to burglary rates (Sampson & Groves 1989, Bernasco & Nieuwbeerta 2005, Bernasco & Luykx 2003, Clare et al. 2009). *Residential turnover* is another measure of collective efficacy. Although Bernasco & Luykx (2003) document a positive relationship between residential turnover and the burglary rates, results in Bernasco & Nieuwbeerta (2005), Townsley et al. (2015) do not confirm that hypothesis. *Socio-economic variation* among residents has been shown to be positively related to general crime rates (e.g. Sampson et al. (1997), Johnson & Summers (2015)), but it was either not considered or shown insignificant in the studies on burglary we have reviewed. Other indicators of social disorganisation and their effect on general crime rates (not only burglary) are the high rate of single-parent households, one-person households as well as younger households Bernasco (2014), Sampson et al. (1997), Andresen (2010).

3.3 Covariates selection

Based on the criminological overview above and the availability of covariates, we form four model specifications, from very rich representations to sparse ones. Table 1 shows the covariates used in each of the specifications.

Variables that represent density, i.e. given by the count per cell, are log-transformed to improve the fit. For the same reason, mean household income and mean house price are in log form. Indicators of heterogeneity are computed using the index of variation introduced in Agresti & Agresti (1978). These include ethnic heterogeneity and occupation variation within an area. Both are indicators of the lack of social cohesion. Subsequently, all variables were standardised to have zero mean and standard deviation of one.

The first specification, *specification 1*, is the richest representation and includes variables that are a proxy for the same phenomenon. For example, both household income and house price are a measure of affluence. This choice is deliberate as we use a shrinkage prior for the regression coefficients to choose the most relevant variables.

The second specification, *specification 2*, removes covariates that are strongly correlated to others or lack strong evidence in the criminological literature. We remove *owner-occupied dwellings* for its strong correlations with the house dwellings and the fraction of houses that are detached or semi-detached. We remove *house dwellings* due to high correlation with (semi-)detached houses and stronger theoretical backing for the latter (e.g.Bernasco & Nieuwbeerta (2005)). We remove *urbanisation level* because of little empirical evidence found in the literature. Naturally, it acts as a proxy for where buildings are, which is accounted for to a large extent by households and points of interest variables. We remove *single-parent households* due to a high correlation with social housing and unemployment rate, and the latter two being preferable indicators of social disorganisation.

In the third specification, *specification 3*, we exclude the following variables on top of those excluded in specification 2. *Median age*, as a proxy for collective efficacy, is removed due to weak evidence in previous studies and other measures of collective efficacy already present: ethnic and socio-economic heterogeneity. *One-person households* and *accommodation POIs* are removed because of weak empirical evidence from previous studies. *Mean household income* is removed due to insufficient evidence from previous studies

Table 1: Models specifications that are used throughout the evaluation of the proposed model.

	1	2	3	4
log households (count per cell)	x	x	x	x
log retail POIs (count per cell)	x	x	x	
log eating/drinking POIs (count per cell)	x	x	x	
log edu/health POIs (count per cell)	x	x	x	
log accommodation POIs (count per cell)	x	x		
log sport/entertainment POIs (count per cell)	x	x	x	
log POIs (all categories count per cell)				x
houses (fraction of dwellings)	x			
(semi-)detached houses (fraction of dwellings)	x	x	x	x
social housing (fraction of dwellings)	x	x		
owner-occupied dwelling (fraction of dwellings)	x			
single-parent households (fraction of households)	x			
one-person households (fraction of households)	x	x		
unemployment rate	x	x	x	
ethnic heterogeneity measure (index of variation)	x	x	x	x
occupation variation measure (index of variation)	x	x	x	x
accessibility (estimated by Transport for London)	x	x	x	x
residential turnover (ratio of residents who moved in/out)	x	x	x	x
median age	x	x		
log mean household income	x	x		
log mean house price	x	x	x	x
urbanisation index (proportion of urban area)	x			

and an already present and more preferable measure of affluence – house price. *Social housing* variable is removed because of weak empirical evidence and a high correlation with unemployment.

In the last specification, *specification 4*, we additionally remove *unemployment rate* due to weak empirical support from previous studies. This specification aggregates all POIs into a single variable (including accommodation POIs). This is to remove the strong correlations between them. As a single variable, it signifies the level of social activity: retail, education, entertainment, etc.

3.4 Simulation study details

3.4.1 Model choices

The methodology we developed in section 2 requires specifying the grid size, blocking structure, and the number of mixture components, K.

To choose cell size we take into account the precision of the burglary point pattern. The published data have been anonymised by mapping exact locations to predefined snap points (police.uk 2018). We follow the recommendations in Tompson et al. (2015) who assess the accuracy of the anonymisation method by aggregating both the original and obfuscated data to areal counts at different resolutions and looking at the difference. They show that the aggregation at lower super output area (LSOA) level does not suffer from the bias introduced by anonymisation process. Therefore, for our cell size, we approximately match an average-size LSOA to avoid the loss of precision caused by the anonymisation process. As a result, each cell corresponds to an area of 400×400 metres.

For the blocking structure, we take advantage of existing census output areas because they have been designed to group homogeneous groups of households/people together (Office for National Statistics 2019). Given that our grid is approximately at the LSOA level, we choose MSOAs as the blocking structure.

The number of components, K, is a crucial parameter of our model. We run our model for varying K and use the performance measures introduced above to decide on the optimal number of components. From our experience, after a certain number of components, adding more components becomes harder to interpret and does not bring significant performance improvements.

3.4.2 Identifiability

As already mentioned in section 2, the traceplot of the MCMC algorithm can be inspected for the label-switching behaviour. From our experience, the sampler would choose one of the $K!$ modes which are a consequence of likelihood invariance and was unlikely to switch to another mode due to the high dimensionality of the parameter space.

4 Results

We report results for the SAM-GLM model and compare them to the performance of the log-Gaussian Cox process (LGCP). For LGCP, we use the standard formulation with a Matèrn covariance function (see section B for full details). By varying the specification of the covariates (see section 3), we evaluate in-sample fit, out-of-sample generalisation, and hotspot prediction. Subsequently, we discuss the interpretation of the results obtained using the proposed method and show its relevance for obtaining criminological insights.

4.1 SAM-GLM performance

Figures 2 and 3 report the performance metrics for the 2015 and 2013-2015 datasets, respectively. On the left panels of the figures, we report in-sample performance, measured by WAIC, of the proposed method with an increasing number of components (K). On the right panels, we report the energy score for the one-period-ahead forecast, again with varying K. For both figures, we also show the performance obtained using LGCP which we use as the baseline. The results show that the in-sample predictive performance of the LGCP model is matched by SAM-GLM model with two components for the one-year dataset and three components for the three-year dataset. For the out-of-sample performance, as measured by the energy score, the LGCP model is outperformed with the three-component SAM-GLM model on the one-year point pattern. For the three-year dataset, 6 components are required to match the performance of LGCP. The extra three components required to match the energy score performance of LGCP could be explained by the fact that the three-year point pattern will naturally be more smooth and thus easier to non-parametrically interpolate using the Gaussian random field part of LGCP. For the one-year dataset, it is clear that $K = 2$ or $K = 3$ is the optimal number of components. This does not vary significantly per specification. For the three-year counterpart, the range between 3 and 6 components would be an appropriate choice. For this dataset, more expressive specifications perform better on both in-sample and out-of-sample performance. More data available in the three-year dataset will allow more descriptive specifications to effectively estimate weaker effects, which would otherwise be pruned out using shrinkage priors for the regression coefficients.

While out-of-sample performance, measured using the energy score, takes into account all locations, practitioners might only be interested in predicting hotspots. PAI and PEI measures have been used in the crime forecasting community (see section 2.4). Figures 4 and 5 show the plots of PAI and PEI measures for both models with specification 4, using the 2015 and 2013-2015 datasets, respectively. The plots show the score for when up to 500 cells (around 5% of the study region) are flagged as hotspots. For the one-year dataset, the SAM-GLM model with $K = 3$ components is enough to outperform LGCP on PEI measure when between 50 to 500 cells are flagged as hotspots. With $K = 4$, this is also true in the range 1-50. For PAI measure, no difference can be seen for $K > 2$. The results based on the three-year data favour LGCP model when up to 150 cells are flagged as hotspots and $K < 5$. After adding more components, the SAM-GLM performance matches that of LGCP. When between 150 and 500 cells are flagged, $K > 3$ components is enough to outperform LGCP. These results are consistent with the previous finding that outperforming LGCP on the three-year point pattern requires more components.

An advantage of the SAM-GLM model over the LGCP model is that within the mixture model, the components are of parametric form and interpretable without any penalty in performance. We discuss interpretation in a section below.

4.2 Interpretation

For this analysis, we choose the three-year dataset because more data usually leads to more robust inferences of the parameters. We choose specification 4 with $K = 3$ components because of its parsimony and the good performance shown above. Figure 6 shows the component allocation maps and the `CovEffect` measure with the effect sign (+/-) for each covariate for all of the three components. The allocation

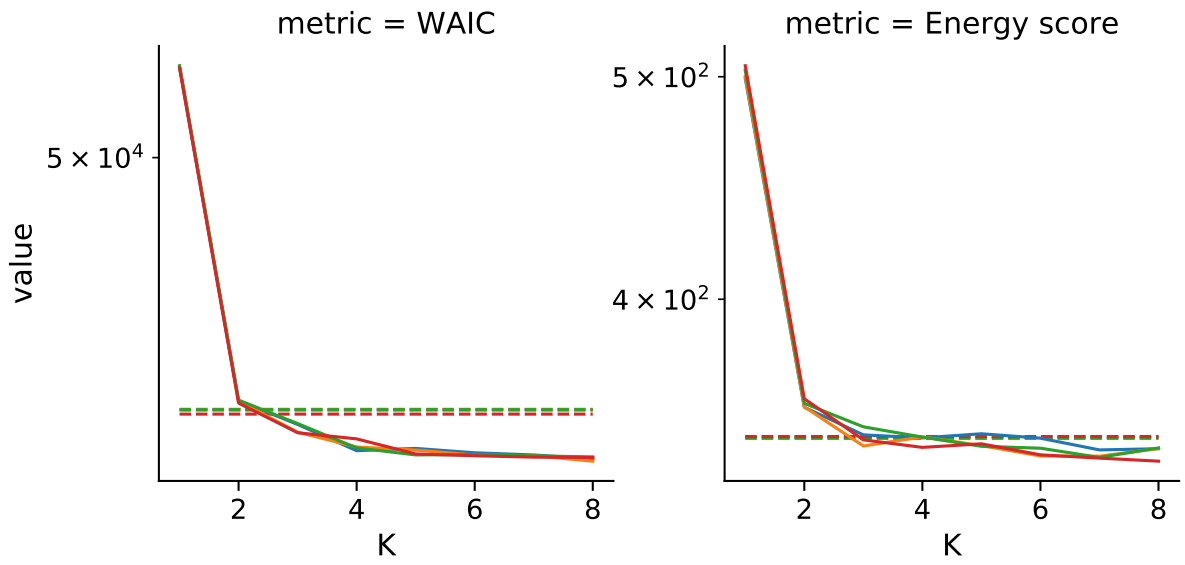


Figure 2: Evaluation of the performance of SAM-GLM (——), compared to LGCP (- - -). Results are shown for different model specifications: specification 1 (——), specification 2 (——), specification 3 (——), specification 4 (——). Training data: burglary 2015, test data: burglary 2016.

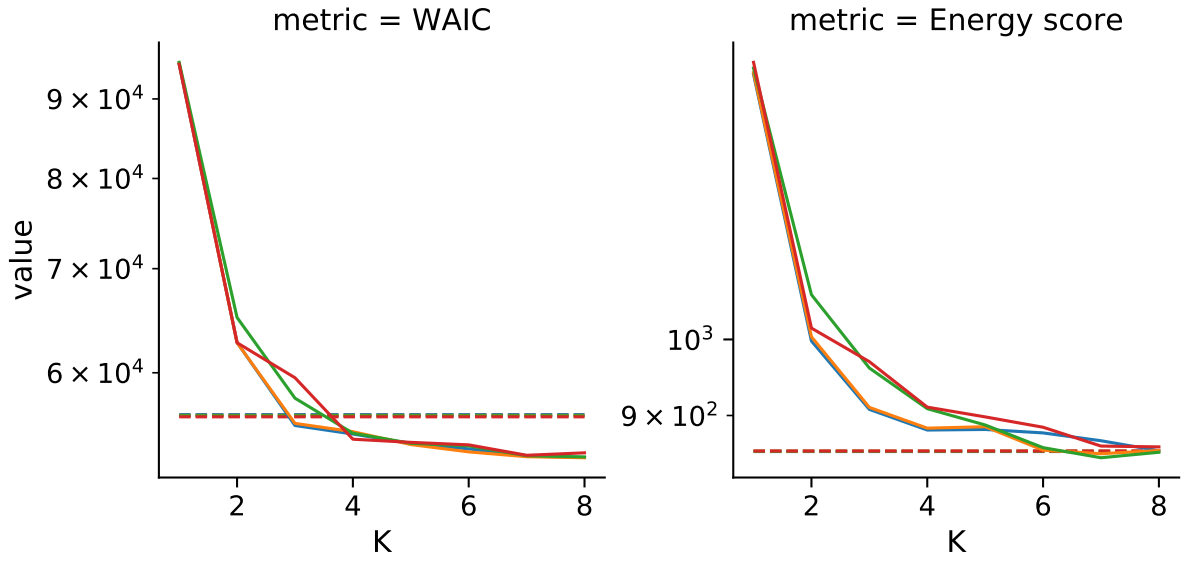


Figure 3: Evaluation of the performance of SAM-GLM (——), compared to LGCP (- - -). Results are shown for different model specifications: specification 1 (——), specification 2 (——), specification 3 (——), specification 4 (——). Training data: burglary 2013-2015, test data: burglary 2016-2018.

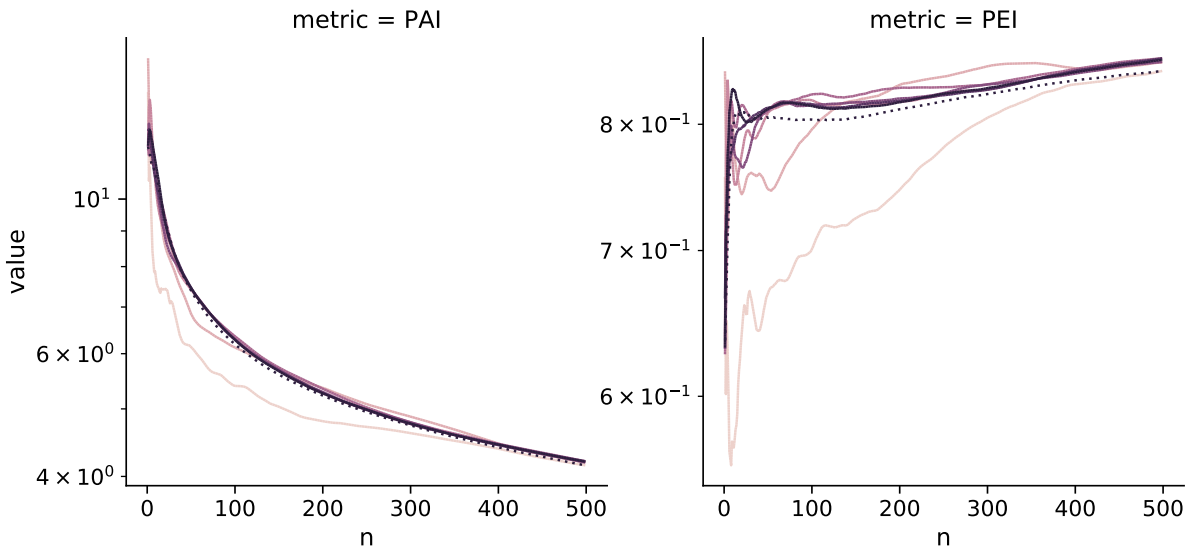


Figure 4: PAI/PEI performance SAM-GLM (——) and LGCP (·····) models, using specification 4. For the SAM-GLM results, the colour of the line represents the number of components: $K = 1$(——), $K = 2$(——), $K = 3$(——), $K = 4$(——), $K = 5$(——), $K = 6$(——), $K = 7$ (——). Training data: burglary 2015, test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

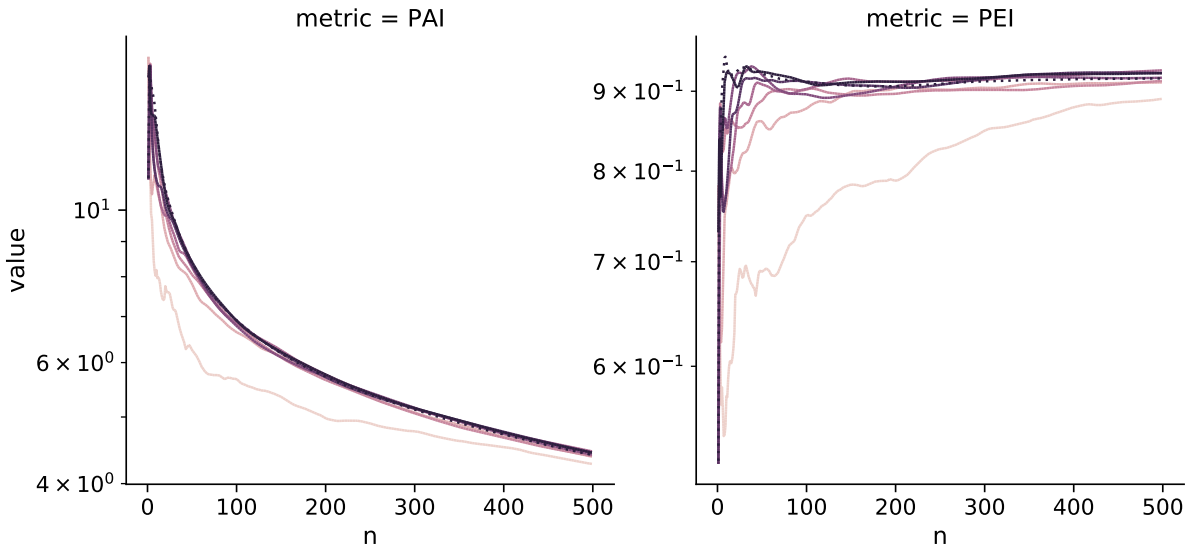


Figure 5: PAI/PEI performance SAM-GLM (——) and LGCP (·····) models, using specification 4. For the SAM-GLM results, the colour of the line represents the number of components: $K = 1$(——), $K = 2$(——), $K = 3$(——), $K = 4$(——), $K = 5$(——), $K = 6$(——), $K = 7$ (——). Training data: burglary 2013-2015,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

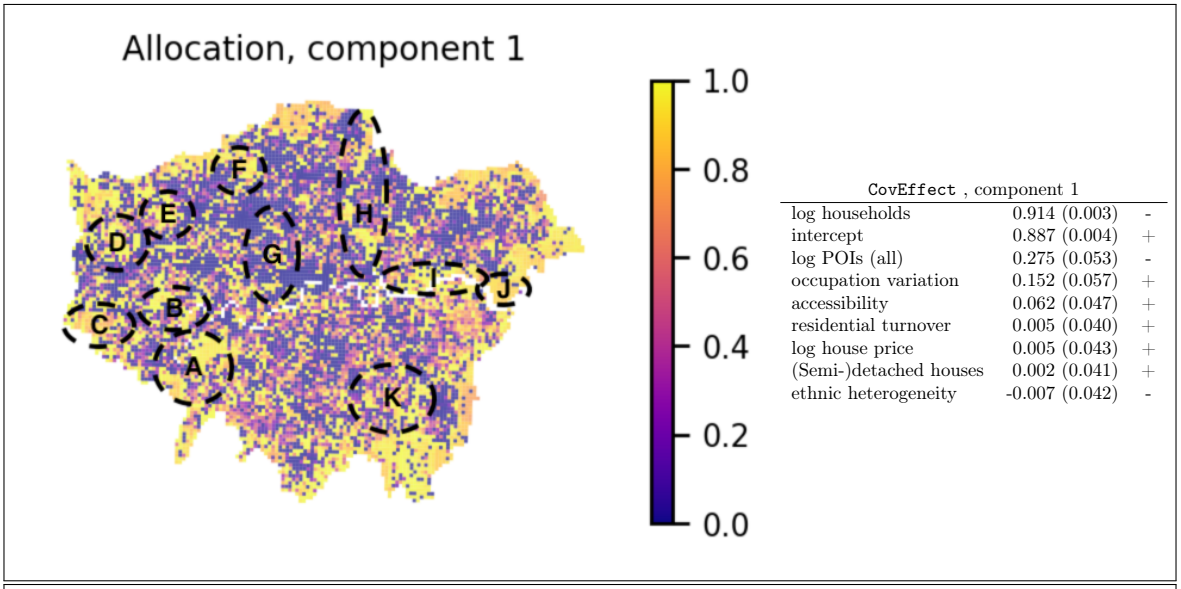
map for each component shows the proportion of times from the MCMC samples each cell is allocated to that component. The alphanumeric labels on the allocation plots are used in the discussion below when referring to specific locations. The `CovEffect` is computed for each sample and component separately and then averaged over the MCMC samples. We also report the standard deviation of the `CovEffect` estimate in brackets.

The first component is active in the areas of low intensity. These include Richmond and Bushy parks (A), Osterley Park and Kew botanic gardens (B), Heathrow airport (C), RAF Northolt base and nearby parks (D), parks near Harrow (E), green fields next to Edgware (F), Hyde Park, Regent's park, Hampstead Heath (G), Lee Valley (H), London City airport and the industrial zone in Barking (I), Rainham Marshes reserve (J), parks around Bromley (K), and other non-urban areas located on the edges of the map. This component explains locations with little criminal activity, signified by negative `CovEffect` for the number of households and points of interest. Occupation variation, as a measure of socioeconomic heterogeneity, is strongly positive. This would support the hypothesis from social disorganisation theory, however, this is more likely due to the very low population in those areas which results in high occupation variation measure. Accessibility measure also has a positive effect on burglary rates in these locations. This is expected and in line with the hypotheses from the crime pattern theory. Other covariates have very small `CovEffect` values.

The second component is active throughout the study region, with large clusters around residential areas. These include areas around Clapham, Balham, and Forrest Hill (L); Richmond (M); Southall (N); Ealing, Wembley, and Harrow (O); Chelsea and Kensington (P); Brent and Hampstead (Q); Edgware (R); East Barnet (S), Enfield (T); Haringey and Walthamstow (U); Stratford (V); Romford (W); Orpington (X); Purley (Y); and Twickenham (Z). In this component, the number of households and points of interest have the strongest effect (except the intercept) – burglaries happen where targets are. Similar to the previous component, accessibility has a relatively strong positive effect. The positive effect of ethnic heterogeneity confirms the hypothesis from the social disorganisation theory. The other indicators of social disorganisation – occupation variation, residential turnover – have not been inferred as significant. Similarly, house price as a measure of reward and fraction of detached and semi-detached houses have low `CovEffect` value.

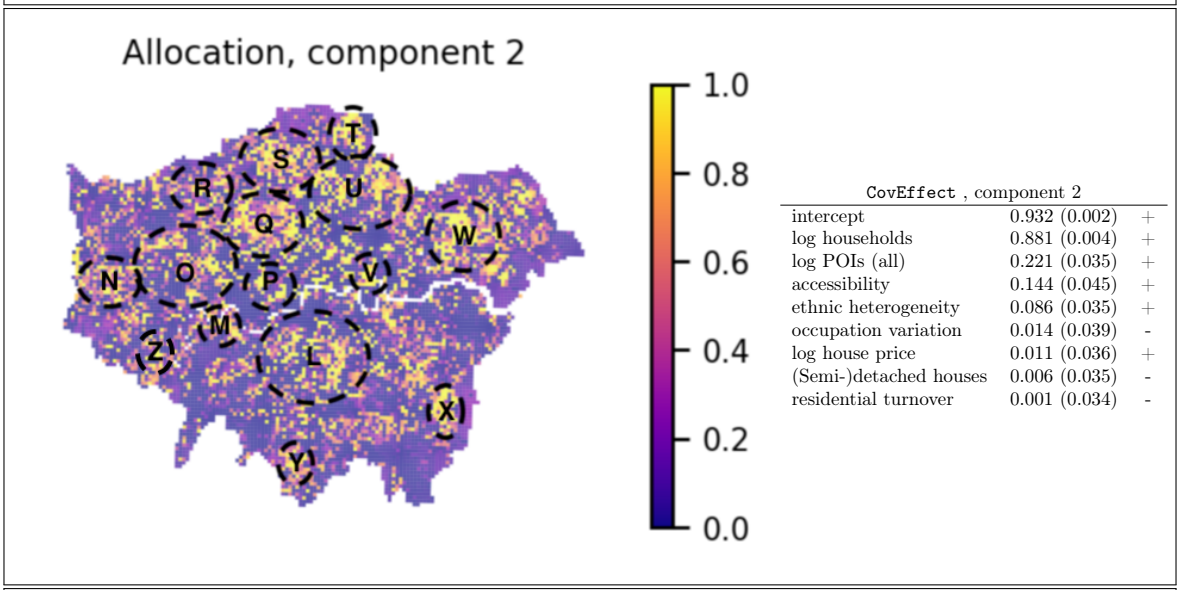
Component 3 is active in the city centre and in the high streets of neighbourhoods: Soho, Mayfair, Covent Garden, Marylebone, Fitzrovia, London Bridge, Shoreditch (1); Notting Hill and Holland Park (2); Earl's Court and Fulham (3); Hackney (4); Brent Cross (5); Wembley (6); Twickenham(7); Sutton (8); Croydon (9). Burglary rates in these locations are largely driven by points of interest, and households. Compared to the second component (residential), the order of `CovEffect` values for these covariates is reversed. Accessibility measure has the strongest effect in this component. This measure is high in the city centre and around the high streets which are usually well-connected to the public transport system. This confirms findings from crime pattern theory and routine activities theory which suggest that offenders choose locations that are part of their usual routine and in their awareness spaces. Accounting for the lack of social cohesion, both ethnic heterogeneity and occupation variation have a strong positive effect. Unexpectedly, our model infers a negative relationship between residential turnover and burglary intensity. Association of high residential turnover with the reduced risk of burglary apprehension has been shown as significant in only a few studies, and was limited to *residential* burglary (Bernasco & Luykx 2003, Bernasco & Nieuwbeerta 2005, Townsley et al. 2015). Areas that are less residential such as high streets have a higher proportion of flats. Dwellings with shared premises such as flats have been shown to less likely become a target than one-household buildings (Beavon et al. 1994). Another possible reason could be the staleness of the data for the covariates which are taken from the 2011 census. Also, house price has been inferred to have a negative effect, i.e. more affluent locations are less likely to get targeted. An explanation mentioned in previous studies is that offenders often live in disadvantaged areas and choose targets within their awareness spaces, which are less likely to be affluent areas (Evans 1989, Rengert & Wasilchick 2010).

The support for the presence of spatial heterogeneity is further given by inspecting the inferences made by the LGCP model (for LGCP details see section B). The left panel of figure 7 shows standard deviations of the marginal posterior distributions of the Gaussian random field component (f). It is clear that the variance of the field component is clustered, where the regions with higher values are easily identifiable as those less urbanised. In contrast, SAM-GLM model has pickled up this heterogeneity by allowing a separate component for it (see component 1 in figure 6). The right panel of figure 7 shows `CovEffect` computed for all component of the LGCP model. `CovEffect` measure for the field component of the model is computed by treating it as a covariate with the coefficient equal to one. The `CovEffect` value for the latent field component is the third-highest, after the intercept and the number of households. A



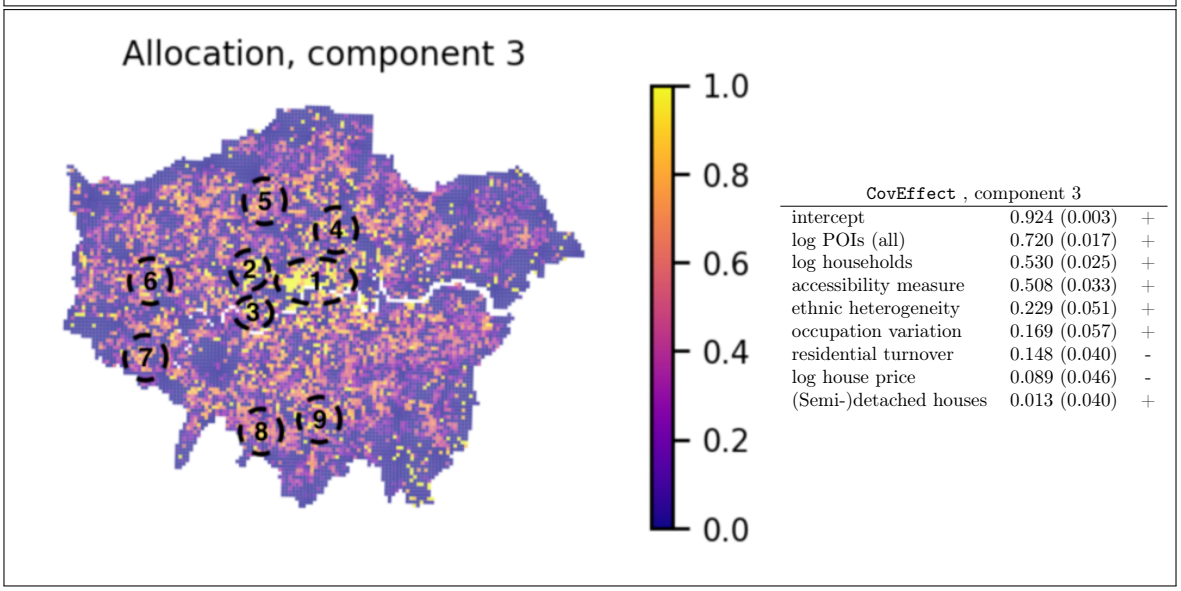
`CovEffect` , component 1

log households	0.914 (0.003)	-
intercept	0.887 (0.004)	+
log POIs (all)	0.275 (0.053)	-
occupation variation	0.152 (0.057)	+
accessibility	0.062 (0.047)	+
residential turnover	0.005 (0.040)	+
log house price	0.005 (0.043)	+
(Semi-)detached houses	0.002 (0.041)	+
ethnic heterogeneity	-0.007 (0.042)	-



`CovEffect` , component 2

intercept	0.932 (0.002)	+
log households	0.881 (0.004)	+
log POIs (all)	0.221 (0.035)	+
accessibility	0.144 (0.045)	+
ethnic heterogeneity	0.086 (0.035)	+
occupation variation	0.014 (0.039)	-
log house price	0.011 (0.036)	+
(Semi-)detached houses	0.006 (0.035)	-
residential turnover	0.001 (0.034)	-



`CovEffect` , component 3

intercept	0.924 (0.003)	+
log POIs (all)	0.720 (0.017)	+
log households	0.530 (0.025)	+
accessibility measure	0.508 (0.033)	+
ethnic heterogeneity	0.229 (0.051)	+
occupation variation	0.169 (0.057)	+
residential turnover	0.148 (0.040)	-
log house price	0.089 (0.046)	-
(Semi-)detached houses	0.013 (0.040)	+

Figure 6: Mixture model, allocations and `CovEffect` table for each mixture component. Training data: 2013-2015, specification 4.

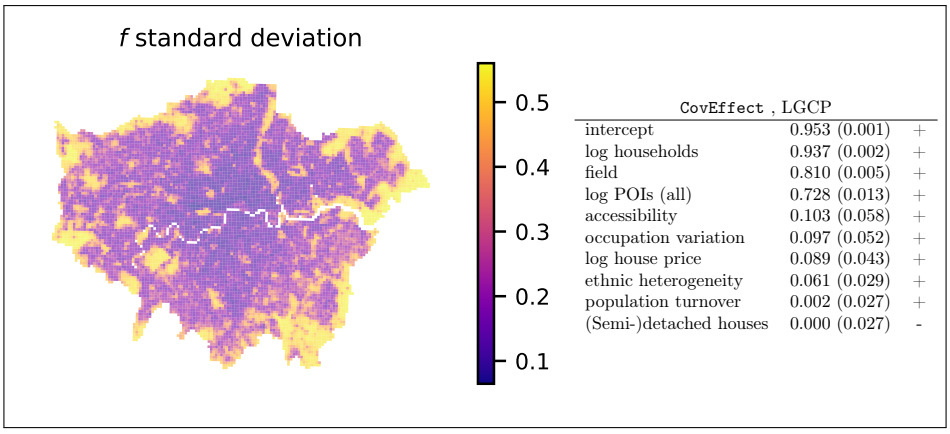


Figure 7: Left: Standard deviation of the posterior distribution of the latent field, $\mathbf{f}$, of the LGCP model. It is clear that, it is clustered and the elevated levels correspond to non-urban locations, airports, and parks (see the discussion above). Right: `CovEffect` measure for the component of the LGCP model. For both panels, training data: 2013-2015, covariates specification: 4.

large contribution from the latent component indicates that the linear term in the Poisson regression model cannot on its own sufficiently explain the variation in the intensity of burglary.

5 Conclusions

Spatial point patterns on large spatial regions, such as metropolitan areas, will most likely exhibit localised behaviour. Motivated by this, we proposed a mixture model that accounts for spatial heterogeneity, as well as incorporates spatial dependence. Each component of the mixture is a model in itself, and thus allows for different locations to follow a different model, e.g. in the urban context, less-urbanised locations can assume a different model from the city centre. Each component is an instance of the generalised linear model (GLM) which includes covariates. We account for spatial dependence through the allocation part. The allocation of each location to one of the components is informed by both the data and the prior information. By utilising existing blocks structure, or defining a custom one, the prior information in the model makes locations within the same block to come from the same component. This formulation attempts to find the right balance between the ability to model sharp spatial variations and the statistical strength borrowing for locations within the same block. Following the Bayesian framework, we developed an efficient Markov Chain Monte Carlo sampler to infer the posterior distributions. The posterior distributions of the parameter are interpretable, thus allowing us to learn new insights about the underlying mechanisms of the point pattern.

Our results show that London burglary data are effectively modelled by the proposed method. Using both in- and out-of-sample evaluation, we compare our model to the log-Gaussian Cox process with Matèrn covariance function as the default model for point processes. We show that our model outperforms it, is more computationally tractable, and is also interpretable. The interpretation provides insights into localised effects of different spatial variables that were chosen based on the criminology literature.

The focus of this work on burglary crime does not limit the potential uses of the proposed model. We believe that the model can be applied not only to other types of crime but also in the wider setting of spatial point patterns which are likely to show localised behaviour. The localised behaviour becomes an important factor to take into account especially when the study region is a large area with heterogeneity present throughout.

Future analysis could consider several options not explored in this work. Firstly, an alternative dependence structure could be obtained by introducing a latent component with spatial dependence within each mixture component, rather than a GLM. Similarly, the current method of imposing spatial dependence through component allocation within blocks could be replaced by a spatial stochastic process. Secondly, different options for specifying the term that involves covariates could be explored. One could consider forcing certain covariates to share the coefficients across all components if there is strong prior belief for doing so. Another possible area of investigation is spatially varying coefficient processes method, proposed by Gelfand et al. (2003).

6 Acknowledgements

All authors were supported by the EPSRC grants EP/P020720/1 and EP/P020720/2 for Inference, COmputation and Numerics for Insights into Cities (ICONIC, https://iconicmath.org/). Jan Povala was also supported by EPSRC (EP/L015129/1). Mark Girolami was supported by EPSRC grants EP/T000414/1, EP/R018413/2, EP/R034710/1, EP/R004889/1, and a Royal Academy of Engineering Research Chair in Data Centric Engineering. Parts of this work were carried out while Jan Povala was visiting Strategic Insights Unit at the Metropolitan Police Service in London. The authors are grateful to Louis Ellam for insightful comments and suggestions, to Tobias Fissler for his helpful suggestions on forecast scoring, and to Pavol Povala for his thorough feedback on the drafts of the manuscript.

References

Abramowitz, M. & Stegun, I. A. (1965), *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Vol. 55, Courier Corporation.

Agresti, A. & Agresti, B. F. (1978), 'Statistical Analysis of Qualitative Variation', *Sociological Methodology* **9**, 204.

Aldor-Noiman, S., Brown, L. D., Fox, E. B. & Stine, R. A. (2017), 'Spatio-temporal low count processes with application to violent crime events', *Statistica Sinica* .

Andresen, M. A. (2010), The place of environmental criminology within criminological thought, *in* 'Classics in environmental criminology', CRC Press, pp. 21–44.

Anselin, L., Cohen, J., Cook, D., Gorr, W. & Tita, G. (2000), 'Spatial analyses of crime', *Criminal justice* **4**(2), 213–262.

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.

Beavon, D. J., Brantingham, P. L. & Brantingham, P. J. (1994), 'The influence of street networks on the patterning of property offenses', *Crime prevention studies* **2**, 115–148.

Bernasco, W. (2014), Residential Burglary, *in* G. Bruinsma & D. Weisburd, eds, 'Encyclopedia of Criminology and Criminal Justice', Springer New York, New York, NY, pp. 4381–4391.

Bernasco, W., Johnson, S. D. & Ruiter, S. (2015), 'Learning where to offend: Effects of past on future burglary locations', *Applied Geography* **60**, 120–129.

Bernasco, W. & Luykx, F. (2003), 'Effects of Attractiveness, Opportunity, and Accessibility to Burglars on Residential Burglary Rates of Urban Neighbourhoods', *Criminology* **41**(3), 981–1002.

Bernasco, W. & Nieuwbeerta, P. (2005), 'How Do Residential Burglars Select Target Areas?', *The British Journal of Criminology* **45**(3), 296–315.

Bowers, K. & Hirschfield, A. (1999), 'Exploring links between crime and disadvantage in north-west England: an analysis using geographical information systems', *International Journal of Geographical Information Science* **13**(2), 159–184.

Brantingham, P. & Brantingham, P. (1981), Notes on the geometry of crime, *in* 'Environmental Criminology', Sage Publications, Beverly Hills, CA.

Brantingham, P. J. & Brantingham, P. L. (1975), 'The spatial patterning of burglary', *The Howard Journal of Criminal Justice* **14**(2), 11–23.

Brantingham, P. L. & Brantingham, P. J. (1993), 'Nodes, paths and edges: Considerations on the complexity of crime and the physical environment', *Journal of Environmental Psychology* **13**(1), 3–28.

Breslow, N. E. (1984), 'Extra-poisson variation in log-linear models', *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **33**(1), 38–44.

Brunsdon, C., Fotheringham, A. S. & Charlton, M. E. (1996), 'Geographically weighted regression: a method for exploring spatial nonstationarity', *Geographical analysis* **28**(4), 281–298.

Cameron, A. & Trivedi, P. K. (1990), 'Regression-based tests for overdispersion in the Poisson model', *Journal of Econometrics* **46**(3), 347–364.

Celeux, G., Hurn, M. & Robert, C. P. (2000), 'Computational and Inferential Difficulties with Mixture Posterior Distributions', *Journal of the American Statistical Association* **95**(451), 957–970.

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.

Clare, J., Fernandez, J. & Morgan, F. (2009), 'Formal Evaluation of the Impact of Barriers and Connectors on Residential Burglars' Macro-Level Offending Location Choices', *Australian & New Zealand Journal of Criminology* **42**(2), 139–158.

Clarke, R. V. & Cornish, D. B. (1985), 'Modeling Offenders' Decisions: A Framework for Research and Policy', *Crime and Justice* **6**, 147–185.

Cohen, L. E. & Felson, M. (1979), 'Social Change and Crime Rate Trends: A Routine Activity Approach', *American Sociological Review* **44**, 588–608.

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.

Evans, D. J. (1989), Geographical Analyses of Residential Burglary, *in* D. J. Evans & D. T. Herbert, eds, 'The Geography of Crime', Routledge, London, pp. 86–107.

Felson, M. & Clarke, R. V. (1998), 'Opportunity makes the thief', *Police research series, paper* **98**, 1–36.

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.

Flaxman, S., Chirico, M., Pereira, P. & Loeffler, C. (2019), 'Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: a winning solution to the NIJ "Real-Time Crime Forecasting Challenge"', *Revised and resubmit at Annals of Applied Statistics* .

Flaxman, S., Wilson, A. G., Neil, D. B., Nickisch, H. & Smola, A. J. (2015), Fast Kronecker Inference in Gaussian Processes with non-Gaussian Likelihoods, *in* 'Proceedings of the 32nd International Conference on International Conference on Machine Learning', Vol. 37 of *ICML'15*, JMLR.org, Lille, France, pp. 607–616.

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.

Girolami, M. & Calderhead, B. (2011), 'Riemann manifold Langevin and Hamiltonian Monte Carlo methods: Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(2), 123–214.

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

gov.uk (2017), 'Crime against businesses: findings from the 2017 Commercial Victimisation Survey', `https://www.gov.uk/government/statistics/crime-against-businesses-findings-from-the-2017-commercial-victimisation-survey`.

Green, P. J. (2010), Introduction to Finite Mixtures, *in* S. Frühwirth-Schnatter, G. Celeux & C. P. Robert, eds, 'Handbook of Spatial Statistics', Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, Florida.

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

Grün, B. & Leisch, F. (2008), Finite Mixtures of Generalized Linear Regression Models, *in* 'Recent Advances in Linear Models and Related Areas: Essays in Honour of Helge Toutenburg', Physica-Verlag HD, Heidelberg, pp. 205–230.

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

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.

Johnson, S. D. & Bowers, K. J. (2004), 'The Stability of Space-Time Clusters of Burglary', *The British Journal of Criminology* **44**(1), 55–65.

Johnson, S. D. & Bowers, K. J. (2010), 'Permeability and Burglary Risk: Are Cul-de-Sacs Safer?', *Journal of Quantitative Criminology* **26**(1), 89–111.

Johnson, S. D. & Summers, L. (2015), 'Testing Ecological Theories of Offender Spatial Decision Making Using a Discrete Choice Model', *Crime & Delinquency* **61**(3), 454–480.

Kleiber, C. & Zeileis, A. (2008), *Applied econometrics with R*, Springer-Verlag, New York. ISBN 978-0-387-77316-2.

URL: *https://CRAN.R-project.org/package=AER*

Knorr-Held, L. & Raßer, G. (2000), 'Bayesian Detection of Clusters and Discontinuities in Disease Maps', *Biometrics* **56**(1), 13–21.

McCullagh, P. & Nelder, J. A. (1998), *Generalized linear models*, number 37 *in* 'Monographs on statistics and applied probability', 2nd ed edn, Chapman & Hall/CRC, Boca Raton.

Menting, B., Lammers, M., Ruiter, S. & Bernasco, W. (2019), 'The Influence of Activity Space and Visiting Frequency on Crime Location Choice: Findings from an Online Self-Report Survey', *The British Journal of Criminology* p. In press.

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.

Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. & Tita, G. E. (2011), 'Self-Exciting Point Process Modeling of Crime', *Journal of the American Statistical Association* **106**(493), 100–108.

Møller, J., Syversveen, A. R. & Waagepetersen, R. P. (1998), 'Log Gaussian Cox Processes', *Scandinavian Journal of Statistics* **25**(3), 451–482.

Møller, J. & Waagepetersen, R. P. (2007), 'Modern Statistics for Spatial Point Processes*', *Scandinavian Journal of Statistics* **34**(4), 643–684.

Office for National Statistics (2019), 'Census geography - Office for National Statistics', `https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography`.

Ordnance Survey (GB) (2018), 'Points of Interest [CSV geospatial data], Scale 1:1250, Items: 670887'.
URL: *https://digimap.edina.ac.uk*

police.uk (2018), 'About | data.police.uk', `https://data.police.uk/about/#anonymisation`.

police.uk (2019), 'Data downloads | data.police.uk'.
URL: *https://data.police.uk/data/*

PredPol (2019), 'PredPol Mission | About Us | Aiming to reduce victimization keep communities safer'.
URL: *https://www.predpol.com/about/*

Rasmussen, C. E. & Williams, C. K. I. (2006), *Gaussian processes for machine learning*, Adaptive computation and machine learning, MIT Press, Cambridge, Mass. OCLC: ocm61285753.

Rengert, G. F. & Wasilchick, J. (2010), The Use of Space in Burglary, *in* 'Classics in environmental criminology', CRC Press, pp. 257–272.

Saatçi, Y. (2012), Scalable inference for structured Gaussian process models, PhD Thesis, Citeseer.

Sampson, R. J. & Groves, W. B. (1989), 'Community Structure and Crime: Testing Social-Disorganization Theory', *American Journal of Sociology* **94**(4), 774–802.

Sampson, R. J., Raudenbush, S. W. & Earls, F. (1997), 'Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy', *Science* **277**(5328), 918–924.

Serra, L., Saez, M., Mateu, J., Varga, D., Juan, P., Díaz-Ávalos, C. & Rue, H. (2014), 'Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994–2008', *Environmental and Ecological Statistics* **21**(3), 531–563.

Shaw, C. R. & McKay, H. D. (1942), *Juvenile delinquency and urban areas : a study of rates of delinquents in relation to differential characteristics of local communities in American cities*, Chicago, Ill. : The University of Chicago Press.

Smith, K., Taylor, P. & Elkin, M. (2013), Crimes detected in England and Wales 2012/13, Statistical Bulletin 02/13, Home Office, London.
URL: *https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/224037/hosb0213.pdf*

Stein, M. L. (1999), *Interpolation of spatial data: some theory for kriging.*, Springer, Place of publication not identified. OCLC: 968504419.

Taddy, M. A. (2010), 'Autoregressive Mixture Models for Dynamic Spatial Poisson Processes: Application to Tracking Intensity of Violent Crime', *Journal of the American Statistical Association* **105**(492), 1403–1417.

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

Tompson, L., Johnson, S., Ashby, M., Perkins, C. & Edwards, P. (2015), 'UK open source crime data: accuracy and possibilities for research', *Cartography and Geographic Information Science* **42**(2), 97–111.

Townsley, M., Birks, D., Bernasco, W., Ruiter, S., Johnson, S. D., White, G. & Baum, S. (2015), 'Burglar Target Selection: A Cross-national Comparison', *Journal of Research in Crime and Delinquency* **52**(1), 3–31.

Townsley, M., Birks, D., Ruiter, S., Bernasco, W. & White, G. (2016), 'Target Selection Models with Preference Variation Between Offenders', *Journal of Quantitative Criminology* **32**(2), 283–304.

A Poisson regression model: excess of zeros, overdispersion

In this section we demonstrate that the standard Poisson regression (McCullagh & Nelder 1998) is not a suitable model for the London burglary point pattern.

Firstly, the dataset consists of areas with no buildings in it, e.g. parks, airports, which results in counts equal to zero due to structure rather than due to chance. This is further supported by the plot of the observed count and the corresponding histogram, both shown in figure 8. This phenomenon is often referred to as *excess of zeros*.

Secondly, we fit Poisson GLM with all four specifications of covariates to the 2015 burglary dataset, as described in section 3. Then we use the overdispersion test proposed in Cameron & Trivedi (1990), and implemented in the AER package (Kleiber & Zeileis 2008). For the standard Poisson GLM model, $\mathrm{Var}(\mathrm{y}_n) = \mu_n$. The overdispersion test uses it as the null hypothesis, where the alternative is $\mathrm{Var}(\mathrm{y}_n) = \mu_n + c \times g(\mu_n)$, where $g(\cdot)$ must be specified. For our test, we choose $g(\cdot) = 1$. Table 2 shows the estimated c values and the p-values for each estimate, given that null hypothesis is $c = 0$. The data clearly show the presence of overdispersion in all four models.

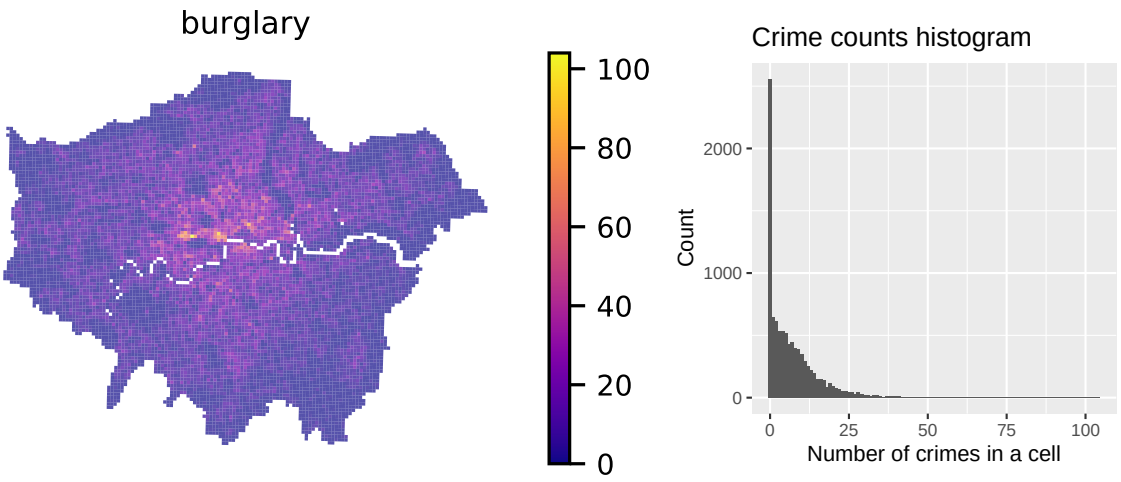


Figure 8: Observed count on the map (left) and the corresponding histogram (right) for the point pattern of burglary aggregated over the grid for the time period 1/2015-12/2015.

Table 2: Overdispersion test for Poisson GLM model.

Specification	c	p-value
1	1.905	2.2e-16
2	1.897	2.2e-16
3	1.910	2.2e-16
4	1.911	2.2e-16

B Log-Gaussian Cox process

Dicretising the spatial domain to a regular grid, the full Bayesian formulation of the model is given as follows:

$$\begin{aligned}
\mathrm{y}_n|\boldsymbol{\beta}, \mathbf{f}, \mathbf{X} &\sim \text{Poisson}\left(\exp(\mathbf{X}_n^\top \boldsymbol{\beta} + \mathrm{f}_n)\right) && (7)\\
f(\cdot)|\boldsymbol{\theta} &\sim \mathcal{GP}\left(0, k_{\boldsymbol{\theta}}(\cdot,\cdot)\right) && (8)\\
\beta_j &\sim \mathcal{N}(0, \sigma_j^2) && (9)\\
\sigma_j^2 &\sim \text{InvGamma}(1, 0.01) && (10)\\
\boldsymbol{\theta} &\sim \text{weakly-informative log-normal prior}, && (11)
\end{aligned}$$

where $n = 1, \ldots, N$ is the index over the cells on the map, $j = 1, \ldots, J$ is the index over the covariates, $f()$ is a zero-mean Gaussian process with covariance function $k_{\boldsymbol{\theta}}(\cdot,\cdot)$, and hyperparameters $\boldsymbol{\theta}$, f_n is the value of $f(\cdot)$ in the centre of cell n, $\mathbf{X}_n$ is the vector of the covariates at cell n, and $\boldsymbol{\beta}_j$ is the jth regression coefficient with a scale hyperparameter σ_{kj}^2. A plain Poisson generalised linear model (GLM) formulation assumes no spatial correlation, i.e. $\mathrm{f}_n = 0$ for all n. Compared to the Poisson GLM model, LGCP allows for modelling the variation in the intensity that cannot be explained by the covariates $\mathbf{X}$.

In order to allow for Kronecker product factorisation of the covariance matrix of the Gaussian process, we specify $k_{\boldsymbol{\theta}}(\cdot,\cdot)$ as a product of two Matérn covariance functions, one for the easting (E) coordinate, the other for the northing (N) coordinate. Matérn covariance function is a standard choice in spatial statistics as it allows specifying smoothness of the function (Stein 1999). It is given as follows

$$k_{\text{Matern}}(\boldsymbol{x}, \boldsymbol{x}') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}|\boldsymbol{x} - \boldsymbol{x}'|}{\ell} \right)^{\nu} K_\nu \left(\frac{\sqrt{2\nu}|\boldsymbol{x} - \boldsymbol{x}'|}{\ell} \right), \qquad (12)$$

where ℓ is the characteristic lengthscale, ν is the smoothness parameter, and K_ν is a modified Bessel function (Rasmussen & Williams 2006). It can be shown that that the Gaussian processes with Matérn covariance functions are k-times mean-square differentiable if and only if $\nu > k$. Abramowitz & Stegun (1965) show that if ν is a half-integer, i.e. for an integer p, $\nu = p + \frac{1}{2}$, the covariance function becomes

especially simple, giving

$$k_{\ell,\nu=p+1/2}(\boldsymbol{x},\boldsymbol{x}') = \exp\left(-\frac{\sqrt{2\nu}|\boldsymbol{x}-\boldsymbol{x}'|}{\ell}\right)\frac{\Gamma(p+1)}{\Gamma(2p+1)}\sum_{i=0}^{p}\frac{(p+i)!}{i!(p-i)!}\left(\frac{\sqrt{8\nu}|\boldsymbol{x}-\boldsymbol{x}'|}{\ell}\right)^{p-i}. \tag{13}$$

For this reason, we set $\nu = 3/2$. The final covariance function, including the σ^2 parameter to control the range of $f()$ therefore becomes

$$k_{\boldsymbol{\theta}}((x_{\mathrm{E}},x_{\mathrm{N}}),(y_{\mathrm{E}},y_{\mathrm{N}})) \quad = \quad \sigma^2 k_{\ell,\nu=3/2}(x_{\mathrm{E}},y_{\mathrm{E}}) \times k_{\ell,\nu=3/2}(x_{\mathrm{N}},y_{\mathrm{N}}), \tag{14}$$

where $\boldsymbol{\theta} = [\sigma^2,\ell]^\top$.

B.1 Inference

To infer posterior distribution of the regression coefficients, $\boldsymbol{\beta}$, latent field $\mathbf{f}$, and its hyperparameters $\boldsymbol{\theta}$, we use a Hamiltonian Monte Carlo sampler. The scale parameters $\sigma_1^2,\ldots,\sigma_J^2$ are analytically integrated out (see equation 20 in the appendix). Due to positivity constraint of the hyperparameters, we sample from $\boldsymbol{\phi} = \log\boldsymbol{\theta}$ (applied component-wise). The density function of the joint posterior distribution we are sampling from is proportional to the product of likelihood and the priors, i.e.

$$p(\mathbf{f},\boldsymbol{\beta},\boldsymbol{\phi}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{f},\boldsymbol{\beta})p(\mathbf{f}|\exp(\boldsymbol{\phi}))p(\boldsymbol{\beta})p_{\boldsymbol{\theta}}(\exp(\boldsymbol{\phi}))\prod_i\left|\frac{d}{d\phi_i}\exp(\phi_i)\right|. \tag{15}$$

In order to effectively use HMC sampler, log-likelihood of the posterior and its gradient need to be tractable. Thanks to the grid structure of our study region, we utilise Kronecker product structure that is present in the covariate matrix in $p(\mathbf{f}|\boldsymbol{\theta})$ if the covariance function $k_{\boldsymbol{\theta}}(\cdot,\cdot)$ is assumed to be a product of covariance functions, one per each dimension (For more details, see Saatçi (2012)). After expansion, the unnormalised log-density becomes

$$\begin{aligned}
\log p(\mathbf{f},\boldsymbol{\beta},\boldsymbol{\phi}|\mathbf{y}) &= \log p(\mathbf{y}|\mathbf{f},\boldsymbol{\beta}) + \log p(\boldsymbol{\beta}) + \log p(\mathbf{f}|\exp(\boldsymbol{\phi})) + \log p_{\boldsymbol{\theta}}(\exp(\boldsymbol{\phi})) + \sum_i \phi_i + \mathrm{const}_1 \\
&= \left(\mathbf{y}^\top\mathbf{X}\boldsymbol{\beta} + \mathbf{y}^\top\mathbf{f} - \exp(\mathbf{X}\boldsymbol{\beta}+\mathbf{f})\right) + \log p(\boldsymbol{\beta}) \\
&\quad + \left(-\frac{1}{2}\log|\boldsymbol{K}_{\boldsymbol{\theta}}| - \frac{1}{2}\mathbf{f}^\top\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}\mathbf{f}\right) + \log p_{\boldsymbol{\theta}}(\exp(\boldsymbol{\phi})) + \sum_i \phi_i + \mathrm{const}_1,
\end{aligned} \tag{16}$$

The gradients of the log posterior density w.r.t. quantities of interest are

$$\nabla_{\mathbf{f}}\log p(\mathbf{f},\boldsymbol{\beta},\boldsymbol{\phi}|\mathbf{y}) = (\mathbf{y}-\exp(\mathbf{X}\boldsymbol{\beta}+\mathbf{f})) + \left(-\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}\mathbf{f}\right) \tag{17}$$

$$\nabla_{\boldsymbol{\beta}}\log p(\mathbf{f},\boldsymbol{\beta},\boldsymbol{\phi}|\mathbf{y}) = \left(\mathbf{X}^\top\mathbf{y} - \mathbf{X}^\top\exp(\mathbf{X}\boldsymbol{\beta}+\mathbf{f})\right) + \nabla_{\boldsymbol{\beta}}\log p(\boldsymbol{\beta}) \tag{18}$$

$$\begin{aligned}
\nabla_{\phi_i}\log p(\mathbf{f},\boldsymbol{\beta},\boldsymbol{\phi}|\mathbf{y}) &= \frac{1}{2}\mathbf{f}^\top\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}\frac{\partial\boldsymbol{K}_{\boldsymbol{\theta}}}{\partial\theta_i}\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}\mathbf{f} - \frac{1}{2}\operatorname{tr}\left(\boldsymbol{K}_{\boldsymbol{\theta}}^{-1}\frac{\partial\boldsymbol{K}_{\boldsymbol{\theta}}}{\partial\theta_i}\right) \\
&\quad + \nabla_{\phi_i}\log p_{\boldsymbol{\theta}}(\exp(\boldsymbol{\phi})) + 1.
\end{aligned} \tag{19}$$

The expansion of un-normalised log-density and the gradients are derived in the appendix (see equation 21 and equation 22).

All operations involving $\boldsymbol{K}_{\boldsymbol{\theta}}$ can be sped up using Kronecker product factorisation. Given n^2 is the number of elements in the full matrix $\boldsymbol{K}_{\boldsymbol{\theta}}$, operations in equation 17 and equation 19 can be computed in $\mathcal{O}\left(n^{\frac{3}{2}}\right)$ time and space by utilising the Kronecker structure in matrix inversion and matrix-vector multiplication. For full details, see Saatçi (2012).

C Model derivations

C.1 Beta prior

Given a vector of J independent random variables $\boldsymbol{\beta}$, of which each component is distributed as follows

$$\begin{aligned}
\beta_j &\sim \mathcal{N}(0,\sigma_j^2), \\
\sigma_j^2 &\sim \mathrm{InvGamma}(a,b).
\end{aligned}$$

Let $\Psi = \left(\sigma_1^2, \ldots, \sigma_J^2\right)^\top$, then the prior for the coefficients is given by integrating out the nuisance parameter Ψ

$$\begin{aligned}
p(\boldsymbol{\beta}) &= \prod_j p(\beta_j) \\
&= \prod_j \int p(\beta_j|\Psi_j)p(\Psi_j)d\Psi_j \\
&= \prod_j \int \frac{1}{\sqrt{2\pi}}\Psi_j^{-1/2}\exp\left(-\frac{1}{2\Psi_j}\beta_j^2\right)\frac{b^a}{\Gamma(a)}\Psi_j^{-a-1}\exp\left(-\frac{b}{\Psi_j}\right)d\Psi_j \\
&= \prod_j \frac{b^a}{\sqrt{2\pi}\Gamma(a)}\int \Psi_j^{-a-\frac{1}{2}-1}\exp\left(-\frac{\frac{1}{2}\beta_j^2+b}{\Psi_j}\right)d\Psi_j \\
&= \prod_j \frac{b^a}{\sqrt{2\pi}\Gamma(a)}\frac{\Gamma\left(\frac{1}{2}+a\right)}{\left(\frac{1}{2}\beta_j^2+b\right)^{\frac{1}{2}+a}}
\end{aligned} \tag{20}$$

For the purposes of HMC, we derive both log-density and the gradient of log-density w.r.t. the each individual components. Log-density is given as

$$\log p(\boldsymbol{\beta}) = \sum_i -\left(\frac{1}{2}+a\right)\log\left(\frac{1}{2}\beta_i^2+b\right), \tag{21}$$

from which the gradient is equal to

$$\frac{\partial \log p(\boldsymbol{\beta})}{\partial \beta_i} = \frac{(-\frac{1}{2}-a)\beta_i}{\frac{1}{2}\beta_i^2+b}. \tag{22}$$

C.2 Conditional densities for SAM-GLM inference

The derivations below use the properties of the density function of the Dirichlet distribution and the following property of the Gamma function, $\Gamma(a+1) = a\Gamma(a)$.

C.2.1 Beta update

$$\begin{aligned}
p(\boldsymbol{\beta}|\alpha, \mathbf{X}, \mathbf{y}, \mathbf{z}) &\propto p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{X}, \mathbf{z})p(\boldsymbol{\beta}) \\
&\propto \left\{\prod_{k=1}^{K}\prod_{j=1}^{J} p(\beta_{k,j})\right\}\left\{\prod_{n=1}^{N} p(\mathrm{y}_n|\boldsymbol{\beta}, \mathbf{X}, \mathrm{z}_n)\right\} \\
&\propto \left\{\prod_{k=1}^{K}\prod_{j=1}^{J} p(\beta_{k,j})\right\}\left\{\prod_{n=1}^{N}\prod_{k=1}^{K} p(\mathrm{y}_n|\boldsymbol{\beta}_k, \mathbf{X})^{I(\mathrm{z}_n=k)}\right\} \\
&\propto \left\{\prod_{k=1}^{K}\prod_{j=1}^{J} p(\beta_{k,j})\right\}\left\{\prod_{n=1}^{N}\prod_{k=1}^{K}\left(\frac{\exp(\mathbf{X}_n^\top\boldsymbol{\beta}_k)^{\mathrm{y}_n}\mathrm{e}^{-\exp(\mathbf{X}_n^\top\boldsymbol{\beta}_k)}}{\mathrm{y}_n!}\right)^{I(\mathrm{z}_n=k)}\right\},
\end{aligned} \tag{23}$$

where $p(\boldsymbol{\beta})$ is expanded according to equation 20.

C.2.2 Mixture allocation update

$$
\begin{aligned}
p(\mathrm{z}_n = k|\mathbf{z}^{\bar{n}}, \alpha, \boldsymbol{\beta}, \mathbf{y}) &\propto p(\mathrm{y}_n|\mathrm{z}_n, \boldsymbol{\beta}) \int p(\mathrm{z}_n|\boldsymbol{\pi}_{b(n)}) p(\boldsymbol{\pi}_{b(n)}|\alpha, \mathbf{z}^{\bar{n}}) d\boldsymbol{\pi}_{b(n)} \\
&\propto p(\mathrm{y}_n|\mathrm{z}_n, \boldsymbol{\beta}) \int \prod_k \pi_{b(n),k}^{I(\mathrm{z}_n=k)} \frac{\Gamma(\sum_{j=1}^K B_{b(n),j})}{\prod_{j=1}^K \Gamma(B_{b(n),j})} \prod_{j=1}^K \pi_{b(n),j}^{B_{b(n),j}-1} d\boldsymbol{\pi}_{b(n)} \\
&\propto p(\mathrm{y}_n|\mathrm{z}_n, \boldsymbol{\beta}) \frac{\Gamma(\sum_{j=1}^K B_{b(n),j})}{\prod_{j=1}^K \Gamma(B_{b(n),j})} \frac{\prod_{j=1}^K \Gamma(B_{b(n),j} + I(j=k))}{\Gamma(\sum_{j=1}^K B_{b(n),j} + I(j=k))} \\
&\propto p(\mathrm{y}_n|\mathrm{z}_n, \boldsymbol{\beta}) \frac{B_{b(n),k}}{\sum_{j=1}^K B_{b(n),j}} \\
&\propto p(\mathrm{y}_n|\mathrm{z}_n = k, \boldsymbol{\beta}_k) \frac{c_{b(n)k}^{\bar{n}} + \alpha}{K\alpha + \sum_{j=1}^K c_{b(n)j}^{\bar{n}}}, \qquad (24)
\end{aligned}
$$

where $B_{b,k} = c_{b,k}^{\bar{n}} + \alpha$, and $c_{b,k}^{\bar{n}}$ is the number of cells in block b other than cell n that are assigned to component k.

C.2.3 Alpha update

$$
\begin{aligned}
p(\alpha|\boldsymbol{\beta}, \mathbf{y}, \mathbf{z}) &\propto p(\alpha) \int p(\mathbf{z}|\boldsymbol{\pi}) p(\boldsymbol{\pi}|\alpha) d\boldsymbol{\pi} \\
&\propto p(\alpha) \int \left(\prod_n p(\mathrm{z}_n|\boldsymbol{\pi}_{b(n)}) \right) \left(\prod_b p(\boldsymbol{\pi}_b|\alpha) \right) d\boldsymbol{\pi}_1 \ldots d\boldsymbol{\pi}_B \\
&\propto p(\alpha) \int \left(\prod_n \prod_k \pi_{b(n),k}^{I(\mathrm{z}_n=k)} \right) \left(\prod_b \left(\frac{\Gamma(\alpha K)}{\prod_k \Gamma(\alpha)} \right) \prod_k \pi_{b,k}^{\alpha-1} \right) d\boldsymbol{\pi}_1 \ldots d\boldsymbol{\pi}_B \\
&\propto p(\alpha) \left(\frac{\Gamma(\alpha K)}{\prod_k \Gamma(\alpha)} \right)^B \left(\prod_b \int \prod_k \pi_{b,k}^{S_{b,k}+\alpha-1} d\boldsymbol{\pi}_b \right) \\
&\propto p(\alpha) \left(\frac{\Gamma(\alpha K)}{\prod_k \Gamma(\alpha)} \right)^B \left(\prod_b \frac{\prod_k \Gamma(S_{b,k} + \alpha)}{\Gamma(\alpha K + \sum_k S_{b,k})} \right), \qquad (25)
\end{aligned}
$$

where $S_{b,k}$ is the number of cells in block b allocated to component k.

D Extra plots

We show the mean predicted intensity of burglary of all the classes of models we considered and compare it to the log of the observed count for the one period ahead. See figure 9 for the model trained on the 2015 data with specification 4. It is apparent that a simple GLM model oversmooths the data. The plots of log mean intensity for LGCP and SAM-GLM are indistinguishable.

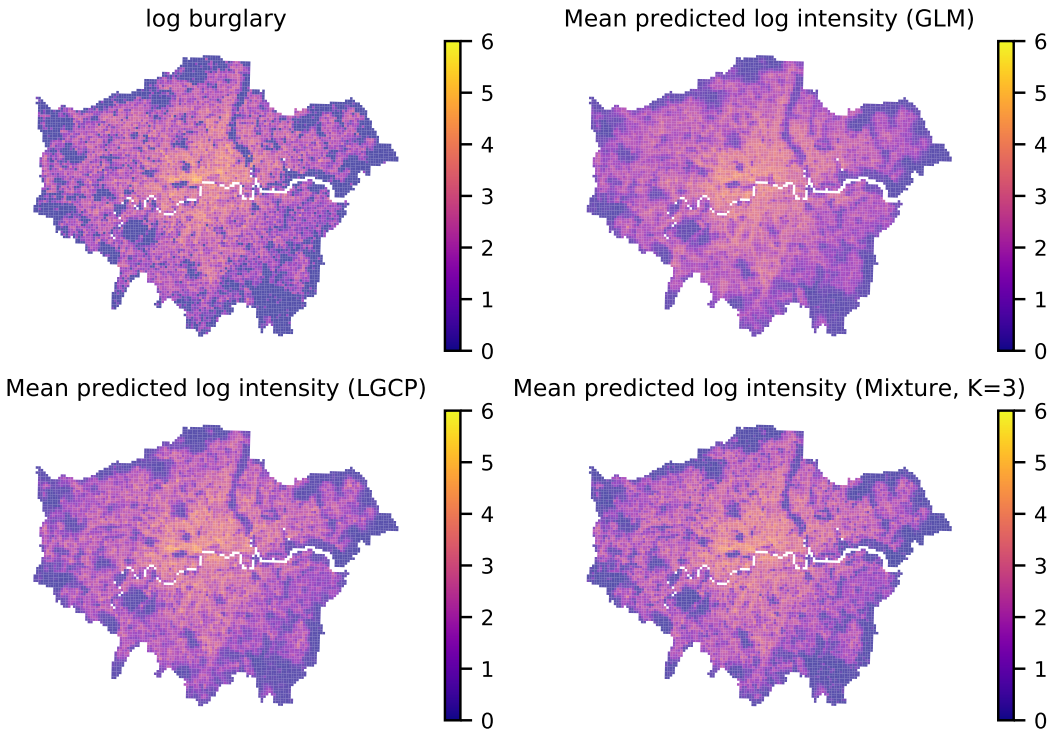


Figure 9: Observed count of burglaries for the period of 1/2016-12/2016, and the comparison of the mean of the predicted counts of three different models using the data from 1/2015-12/2015. Specification 4 is used for all three.