ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

Applied Statistics
Series C

Appl. Statist. (2020)
69, *Part* 5, *pp.* 1067–1090

Burglary in London: insights from statistical heterogeneous spatial point processes

Jan Povala

Imperial College London and Alan Turing Institute, London, UK

and Seppo Virtanen and Mark Girolami

University of Cambridge and Alan Turing Institute, London, UK

[Received October 2019. Final revision June 2020]

Summary. To obtain operational insights regarding the crime of burglary in London, we consider the estimation of the effects of covariates on the intensity of spatial point patterns. Inspired by localized properties of criminal behaviour, we propose a spatial extension to mixtures of generalized linear models from the mixture modelling literature. The Bayesian model proposed is a finite mixture of Poisson generalized linear models such that each location is probabilistically assigned to one of the groups. Each group is characterized by the regression coefficients, which we subsequently use to interpret the localized effects of the covariates. By using a blocks structure of the study region, our approach enables specifying spatial dependence between nearby locations. We estimate the proposed model by using Markov chain Monte Carlo methods and we provide a Python implementation.

Keywords: Bayesian mixture model; Burglary; Crime model; Poisson process; Spatial heterogeneity; Spatial statistics

1. Introduction

Use of statistical models for understanding and predicting criminal behaviour has become increasingly relevant for police forces, and policy makers (Felson and Clarke, 1998; Bowers and Hirschfield, 1999; PredPol, 2019). Whereas short-term forecasting of criminal activity has been used to allocate policing resources better (Taddy, 2010; Mohler *et al.*, 2011; Aldor-Noiman *et al.*, 2016; Flaxman *et al.*, 2019; PredPol, 2019), understanding the criminal behaviour and target selection process through statistical models has potential to be used for designing policy changes and development programmes (Felson and 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 rate of detection 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 that are made by using these models help us to understand the underlying mechanisms of burglary. The main contribution of this work is the integration of statistical methods in spatial modelling with 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

Address for correspondence: Jan Povala, Department of Mathematics, Imperial College London, South Kensington Campus, London, SW7 2AZ, UK.
E-mail: jan.povala11@imperial.ac.uk

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 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 and Nelder, 1998). In the context of spatial data, it has been widely recognized 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 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 and Nelder, 1989). 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 will be different from the properties in a residential area. Formalizing these two concepts and incorporating them into modelling methodology results in more accurate spatial models (Anselin *et al.*, 2000).

Log-Gaussian Cox processes (LGCPs) (Møller *et al.*, 1998; Møller and Waagepetersen, 2007) have 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, LGCP models with stationary covariance functions handle spatial dependence but do not account for spatial heterogeneity.

Mixture-based approaches have been adopted as a way of enriching the collection of probability distributions to account for spatial heterogeneity that is often observed in practice (Green, 2010; Fernández and Green, 2002). Notably, Knorr-Held and Raßer (2000), Fernández and Green (2002) and Green and Richardson (2002) used mixtures for modelling the elevations of disease prevalence. Although these methods improve the model fit by accounting for spatial heterogeneity as well as spatial dependence, they provide little interpretation about why the level is elevated in certain areas. Also, these three methods have been tested only at a modest scale. 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 Hildeman *et al.* (2018) 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 that was discretized into a grid with 2461 cells.

A very different approach to controlling for spatial heterogeneity was taken by Gelfand *et al.* (2003) who allowed regression coefficients to vary across the spatial region. The method treats the coefficients of the covariates as a multivariate spatial process. The process is, however, very challenging to fit and is often limited to two or three covariates (Banerjee *et al.* (2015), page 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 must be specified *a priori* or learned through cross-validation.

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

More specifically, our approach builds on the mixtures of generalized linear models (Grün and Leisch, 2008), in which observations are modelled as a mixture of different models. We cater for spatial dependence by using an approach that was inspired by Fernández and Green (2002) and Knorr-Held and Raßer (2000). Our model probabilistically assigns each location to a particular mixture component, 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 component. 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 census tracts, but our method allows defining custom tracts. We further model spatial dependence of the blocks by using latent Gaussian processes, following Fernández and Green (2002). The posterior inferences for the individual components consisting of regression coefficients and the assignments of locations to clusters are used to draw conclusions and to provide insights about the heterogeneity of the spatial point pattern across the study region.

In contrast with Fernández and Green (2002) and Green and Richardson (2002), this work considers including the covariates in each mixture component, rather than having intercept-only components. Compared with the approach of Hildeman *et al.* (2018) who modelled the log-intensity of a point pattern as a mixture of Gaussian random fields, our model is more constrained but provides better scalability.

We show that the proposed methodology effectively models burglary crime in London. By comparing our approach with an LGCP, which is a standard model for spatial point patterns (Diggle *et al.*, 2013), we show that our method outperforms the LGCP and is more computationally tractable. Lastly, 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 a discussion of model choices that are specific to our application. The results obtained are discussed in Section 4. Section 5 concludes the paper.

2. Modelling methodology

It is widely recognized that burglary crime is spatially concentrated (Brantingham and Brantingham, 1981; Clare *et al.*, 2009; Johnson and 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 burglaries for structural reasons. To model burglary effectively, these phenomena need to be accounted for by using *spatial effects*. The two important spatial effects are *spatial dependence* and *spatial heterogeneity* (Anselin *et al.*, 2000).

For our modelling framework, we choose the Bayesian paradigm because it enables us to formalize 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 discretize the point pattern onto a grid of N cells by counting the points in each cell. Although any form of discretization is allowed, throughout this paper, we work with a regular grid.

We model the count of points in a cell n, y_n, conditioned on the mixture component k as a Poisson-distributed random variable, 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

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

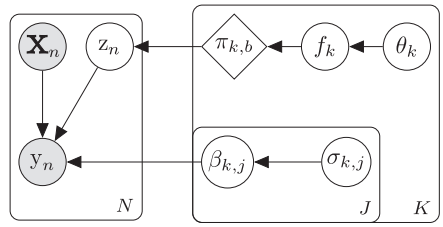


Fig. 1. Summary of SAM-GLM and its graphical representation

towards 0. For each coefficient, we set $\beta_{k,j} \sim \mathcal{N}(0, \sigma^2_{k,j})$, where $\sigma^2_{k,j} \sim \text{InvGamma}(1, 0.01)$. 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, $\boldsymbol{\pi}_{b[n]}$. The value of $\boldsymbol{\pi}_{b[n]}$ 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. The block-specific event probabilities will express the belief that the effect of the covariates is the same within the block unless evidence from the observed data outweighs this information.

To model the mixture weights prior for block b, $\boldsymbol{\pi}_b = (\pi_{1,b}, \ldots, \pi_{K,b})$, we allow for different choices provided that $\pi_{k,b} \geqslant 0$ and $\Sigma_{k=1}^{K} \pi_{k,b} = 1$, i.e. it is a valid probability measure. One possible choice which also takes into account the spatial dependence between the blocks is to model the mixture weights prior for block b and mixture component k as

$$\pi_{k,b} = \frac{\exp(f_{k,b})}{\sum_{l=1}^{K}\exp(f_{l,b})},$$

where $f_{k,b}$ is the evaluation of f_k at the centroid of block b and f_k is an independent zero-mean Gaussian process with hyperparameters $\boldsymbol{\theta}_k$. The prior for $\boldsymbol{\theta}_k$ is specified depending on the kernel function that is used. We shall use the squared exponential kernel throughout this work (Rasmussen and Williams, 2006).

We refer to the model proposed as SAM-GLM: a *spatially aware mixture of Poisson generalized linear models* (GLMs). The formulation is summarized in the equation and the graphical representation that is shown in Fig. 1. In the model proposed, we handle *spatial heterogeneity* by using the mixture components, each of which specifies a set of J regression coefficients $\boldsymbol{\beta}_k$. *Spatial dependence* is considered first within each block and also through interblock dependence imposed by K Gaussian processes. Modelling the spatial dependence by using Gaussian processes at the block level instead of cell level enables more efficient estimation procedures as we discuss later.

2.1. Excess of 0s; overdispersion

Two common challenges that are encountered when modelling count data by using standard Poisson GLMs are *excess of 0s* and *overdispersion* (McCullagh and Nelder, 1998; Breslow, 1984). The former refers to the presence of 0s that are structural, rather than due to chance. In

the context of burglary, structural 0s 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 what would be expected on the basis of a particular statistical model. The standard Poisson GLM for the burglary point pattern, which is a special case of our model ($K = 1$), suffers from overdispersion for different specifications of the covariates term—see section A in the on-line supplementary material. The flexibility of our proposed model can account for the excess of 0s by identifying a low count component to which areas of low intensity will be assigned. Similarly, introducing mixtures can reduce overdispersion. Two cells with similar values for the covariates, but with very different observed counts, are likely to have a similar expected count under the standard Poisson GLM. Under the mixture model, each cell would be allowed to follow a different model.

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 the Markov chain Monte Carlo (MCMC) method to sample from the posterior distributions (Gelman *et al.*, 2013).

Firstly, the scale parameter for the regression coefficients, σ^2_{kj}, is analytically integrated out to simplify the inference (see equation (14) in the on-line supplementary material). The quantities of interest are the allocation vector $\mathbf{z}$, regression coefficient vector for each mixture component, $\boldsymbol{\beta}_k$, unnormalized mixture weights priors at the centroids of the blocks, $f_{k,b}$, and its hyperparameters. For brevity, let $\boldsymbol{\beta}$ be a $K \times J$ matrix of regression coefficients for all mixture components and each covariate, $\mathbf{X}$ be an $N \times J$ matrix of all covariates for each location, $\mathbf{F}$ be a $B \times K$ matrix such that $\mathbf{F}_{b,k} = f_{k,b}$ and $\boldsymbol{\theta}$ the vector of kernel hyperparameters for all f_ks. The unnormalized joint posterior probability distribution is given as

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

We employ the Metropolis-within-Gibbs scheme (Geman and Geman, 1984; Metropolis *et al.*, 1953) and sample from the posterior in three steps.

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

$$p(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}, \mathbf{z}) \propto p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{X}, \mathbf{z})\, p(\boldsymbol{\beta}). \tag{2}$$

Equation (2) is sampled by using the Hamiltonian Monte Carlo method (Duane *et al.*, 1987), for which efficient sampling schemes are available, e.g. Girolami and Calderhead (2011).

(b) Mixture allocation is sampled cell by cell directly by using the equation

$$p(z_n = k|\mathbf{z}^{\bar{n}}, \alpha, \mathbf{X}_n, \boldsymbol{\beta}, \mathbf{y}, \mathbf{F}) \propto p(y_n|z_n = k, \mathbf{X}_n, \boldsymbol{\beta}_k) \frac{\exp(f_{k,b[n]})}{\sum_{l=1}^{K} \exp(f_{l,b[n]})}. \tag{3}$$

(c) We sample all K functions with the Gaussian process prior and their hyperparameters jointly by using the Hamiltonian Monte Carlo method. The joint posterior density is proportional to the expression

$$p(\mathbf{F}, \boldsymbol{\theta}|\mathbf{y}, \mathbf{z}) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} \left(\frac{\exp(f_{k,b[n]})}{\sum_{l=1}^{K} \exp(f_{l,b[n]})} \right)^{I(z_n = k)} \prod_{k=1}^{K} p(f_k|\boldsymbol{\theta}_k)\, p(\boldsymbol{\theta}_k), \tag{4}$$

where $I(\cdot)$ is the indicator function.

For the full expansion of the conditional distributions in equations (2)–(4), see section C in the on-line supplementary material.

In terms of computational tractability, equation (2) takes $\mathcal{O}(N+J)$ steps, equation (3) requires $\mathcal{O}(N \times K)$ steps and equation (4) requires $\mathcal{O}(B^3 \times K)$ steps because of matrix inversions of size $B \times B$ for each of the K components. To contrast it with a standard model for spatial point patterns, one sample from an LGCP involves matrix inversions that require $\mathcal{O}(N^3)$ steps (Diggle *et al.*, 2013). Thanks to blocking, the inference requires inversions of smaller matrices.

2.3. Special case: independent blocks

The model and the associated inference that were introduced above provide a very flexible framework for modelling the spatial dependence of cells via blocks that are also spatially dependent. However, this comes at a high cost—inferring posterior distributions over K Gaussian processes that are combined by using the logistic function is challenging at scale as each sample requires $\mathcal{O}(B^3 \times K)$ operations.

If we assume that the mixture weights priors $\boldsymbol{\pi}_b$ for all blocks are independent and, conditioned on α, distributed as

$$\boldsymbol{\pi}_b | \alpha \sim \text{Dirichlet}(\alpha, \ldots, \alpha), \tag{5}$$

the inference becomes more tractable. Specifically, equation (4) is not needed anymore, equation (2) stays the same and equation (3) is replaced by

$$p(z_n = k | \mathbf{z}^{\bar{n}}, \alpha, \mathbf{X}_n \boldsymbol{\beta}, \mathbf{y}) \propto p(y_n | z_n = k, \mathbf{X}_n \boldsymbol{\beta}_k) \frac{c^{\bar{n}}_{b[n]k} + \alpha}{K\alpha + \sum_{i=1}^{K} c^{\bar{n}}_{b[n]k}}. \tag{6}$$

As a result, the time complexity to take one sample from the unknown quantities is dominated by resampling z_ns in equation (6), which can be computed in $\mathcal{O}(N \times K)$ steps. For the full derivation of equation (6), see section C.2.4 in the on-line supplementary material.

In the literature, $\alpha = 1/K$ is a recommended choice; see, for example, Alvares *et al.* (2018). This prior formulation induces sparsity and can cancel out components in an overfitted mixture (Rousseau and Mengersen, 2011). In the experiments we compare the trade-off between computational complexity and modelling flexibility.

2.4. Identifiability

Specifying a mixture model means that the model likelihood is invariant under the 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, $p(\mathbf{y}|\mathbf{z}, \mathbf{X}, \boldsymbol{\beta})$ is invariant under the relabelling of $\boldsymbol{\beta}_k$ and f_ks, which are the component-specific model parameters.

Exploration in high dimensional spaces is in general difficult for an MCMC sampler. 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. For the sampler to switch to a different mode, it would have to pass 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 trace plot of the Markov chain for each identifiable parameter to assert that relabelling is not present when interpreting the mixtures.

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 on line by the UK police forces (Police.uk, 2019), are provided monthly as a spatial point pattern over the area of 1572 km^2 of both residential and non-residential burglaries. Non-residential burglary refers to instances where the target is not a dwelling, e.g. commercial or community properties. We discretize our study area into a regular grid by counting the number of burglaries within each cell. We choose a grid for computational reasons when comparing with competing methods (see section B in the on-line supplementary material). Given our focus on spatial modelling, we temporally aggregate the point pattern into two data sets: the 1-year data set, starting January 2015 and ending December 2015, with 70234 burglaries, and the 3-year data set, starting January 2013 and ending December 2015, with 224747 burglaries.

Our analysis uses land use data, socio-economic census data from 2011 and points-of-interest (POIs) 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 (OAs). The OAs have been designed to have similar population sizes and to be as socially homogeneous as possible, based on the tenure of households and dwelling types. Each of the 25053 OAs in London has between 100 people or 40 households and 625 people or 250 households. The OAs are aggregated into 4835 lower super-output areas (LSOAs), which in turn are aggregated into 983 middle super-output areas (MSOAs). An LSOA has at least 1000 people or 400 households and at most 3000 people or 1200 households. For an MSOA, the minimum is 5000 people or 2000 households, and the maximum is 15000 people or 6000 households. The POIs 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. The method assumes that the data are 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 to formulate hypotheses about burglary target selection. The target choice is a decision-making process of maximizing *reward* with minimum *effort*, and managing the *risk* of being caught (a process that is analogous to optimal foragers in wildlife (Johnson and Bowers, 2004)). Therefore, we categorize the explanatory variables into these three categories: reward, effort and risk.

3.2.1. Reward; opportunities; attractiveness

Theoretically supported by rational choice theory (Clarke and Cornish, 1985), offenders seek to maximize 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 and Nieuwbeerta, 2005; Clare *et al.*, 2009; Townsley *et al.*, 2015, 2016). *Real estate prices* and *household income* have been used in previous work as a proxy for the attractiveness of targets. The significance of their positive effect on residential burglary victimization rate has been mixed and varied depending on the study region and the statistical method that was used (Bernasco and Luykx, 2003; Bernasco and 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 and Wasilchick, 2010). Other measures of affluence that have been used include *house ownership rates* (Bernasco and Luykx, 2003).

With regard to non-residential burglary, the literature is more sparse. An analysis of non-residential burglary in Merseyside in the UK by Bowers and Hirschfield (1999) showed that non-residential facilities have a higher risk of both victimization and repeat victimization. In particular, sport and educational facilities have a disproportionately higher risk of being targeted compared with 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 shall use the POIs database from the Ordnance Survey which includes 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 and Brantingham, 1993) and routine activity theory (Cohen and 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 that were performed using the data on detected residential burglaries unanimously agree that areas that are *close to the offender's home* are more likely to be targeted (Bernasco and 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) argued that other awareness spaces than their residence play a significant role in target selection. These include previous addresses and 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 and Brantingham (1975) have shown that houses in the interior of a block are less likely to be burgled. Similarly, Townsley *et al.* (2015) and Bernasco and Nieuwbeerta (2005) showed that *single-family dwellings* are more vulnerable to burglaries than multifamily 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 and Bowers (2010) showed that main street segments are more likely to become a target for burglary. Clare *et al.* (2009) and Bernasco *et al.* (2015) showed that the presence of connectors such as train stations increases the likelihood of being targeted, whereas the so-called barriers such as rivers or highways decrease it.

3.2.3. Risk; likelihood of completion

In the social disorganization theory of crime (Shaw and McKay, 1942; Sampson and 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 a crime. Second, it decreases the chance of a successful completion once an offender has chosen to do so. This theory focuses on the effect 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 (to minimize effort). The attraction to 'prosperous targets' applies mostly to the local context

(to maximize gain). In contrast, when a neighbourhood has high social cohesion (also known as 'collective efficacy'), there is mutual trust between 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 rates of burglary (Sampson and Groves, 1989; Bernasco and Nieuwbeerta, 2005; Bernasco and Luykx, 2003; Clare *et al.*, 2009). *Residential turnover* is another measure of collective efficacy. Although Bernasco and Luykx (2003) documented a positive relationship between residential turnover and rates of burglary, results in Bernasco and Nieuwbeerta (2005) and 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) and Johnson and Summers (2015)), but it was either not considered or shown to be insignificant in the studies on burglary that we have reviewed. Other indicators of social disorganization and their effect on general crime rates (not only burglary) are the high rate of single-parent households and one-person households as well as households with younger members; Bernasco (2014), Sampson *et al.* (1997) and Andresen (2010).

3.3. Covariate selection

On the basis of the criminological overview above and the availability of covariates, we form four model specifications, from very rich to sparse representations. Table 1 shows the covariates that were 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 by using the index of variation that was introduced in

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

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

Agresti and 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 standardized to have zero mean and a standard deviation of 1.

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 with 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 semidetached. We remove *house dwellings* because of high correlation with (semi)detached houses and stronger theoretical backing for the latter (e.g. Bernasco and Nieuwbeerta (2005)). We remove the *urbanization level* because little empirical evidence was 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 POIs variables. We remove *single-parent households* because of a high correlation with social housing and unemployment rate, which are preferable indicators of social disorganization.

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 because of weak evidence in previous studies and other measures of collective efficacy already present; ethnic and socio-economic heterogeneity. The covariates *one-person households* and *accommodation POIs* are removed because of weak empirical evidence from previous studies. The variable *mean household income* is removed because of insufficient evidence from previous studies and an already present and more preferable measure of affluence—house price. The covariate *social housing* is removed because of weak empirical evidence and a high correlation with unemployment.

In the last specification, *specification 4*, we additionally remove *unemployment rate* because of 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.

4. Results

After discussing the modelling choices and experimental settings, we compare SAM-GLM with an LGCP, based on the out-of-sample generalization and crime hot spot prediction. For the LGCP, we use the standard formulation with a Matérn covariance function (see section B in the on-line supplementary material for full details). Lastly, we interpret the results that were obtained by using the method proposed and show the relevance for obtaining criminological insights.

4.1. Evaluation and interpretation

4.1.1. Out-of-sample performance

Firstly, we evaluate the performance of the proposed and competing models by using the Poisson likelihood of one-period-ahead data given the model parameters that were obtained from training data. The likelihood denotes how likely the observed data are for given parameters. For a given sample from the posterior distribution of the model parameters, $\phi^{(s)}$, the average pointwise *held-out log-likelihood* is defined as

$$\text{held-out log-likelihood} = \frac{1}{N}\sum_{n=1}^{N} \log\{p(\tilde{y}_n|\boldsymbol{\phi}^{(s)})\}, \tag{7}$$

where $p(\cdot)$ is the Poisson density function and $\tilde{y}_n$ is the realized next-period value. The log-likelihood is a relative measure used for model comparison and can only be used to compare models within the same family of models: in our case, Poisson-based models. A higher value indicates superior predictive power.

Next, we use the *root-mean-square error* (RMSE) metric. It is independent of the model and is measured on the same scale as the target variable. Given a sample from the posterior distribution of the model parameters, $\boldsymbol{\phi}^{(s)}$, we obtain a sample from the joint predictive probability distribution for the counts at all N locations, $\mathbf{y}^{(s)}$, using the sampling distribution of the data, $p(\mathbf{y}|\boldsymbol{\phi}^{(s)})$. Then, using the realized next-period value $\tilde{\mathbf{y}} = (\tilde{y}_1, \ldots, \tilde{y}_N)$, the RMSE is defined as

$$\text{RMSE} = \sqrt{\left\{\frac{1}{N}\sum_{n=1}^{N}(y_n^{(s)} - \tilde{y}_n)^2\right\}}. \tag{8}$$

A lower value of the RMSE indicates a better predictive performance.

4.1.2. Hot spot prediction

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

PAI, introduced by Chainey *et al.* (2008), assesses the ability to capture as many instances of crime as possible with as little area as possible. For a given size of the area to be marked as hot spots, a, it is defined as

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

where A is the total area of the study region, c_a is the number of crimes in the flagged hot spots with the total area a and C is 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 with what a perfect model would predict for a given size of the area to be marked as hot spots, a (Hunt, 2016). It is defined as

$$\text{PEI} = c_a/c_a^*,$$

where c_a is the number of crimes in the hot spots of size a flagged by the model, and c_a^* is the maximum number of crimes that could have been captured by using an area of size a.

In our context of a regular grid, we use both measures to compare competing models when up to n cells have been flagged as hot spots. For a given n, a higher value indicates better hot spot prediction ability.

4.1.3. Interpretation of results

Estimating the effects of different spatial covariates helps us to understand the underlying mechanisms of the point pattern. In the mixtures-of-regressions literature, the interpretation of the individual regression coefficients is of no interest, or the focus is on reporting the regression coefficients $\boldsymbol{\beta}_k$ for each component and quantifying their uncertainty so that their significance can be assessed (Frühwirth-Schnatter *et al.* (2019), chapter 8). To interpret the coefficients

further, one could look at each mixture component specifically and interpret the coefficients in a classical way, conditionally on the partitioning of observations. For example, for a GLM with the exponential link function, increasing a covariate by 1 unit multiplies the mean value of the observed variable by the exponential of the regression coefficient for that covariate, provided that other covariates are held constant. However, this approach allows only component-specific conclusions as it depends on the distribution of the covariate for the associated component. For example, one mixture component may be active in areas with very small values for a specific covariate, whereas some other component is active in areas with high values. Comparing regression coefficients for that covariate across different components would not be appropriate.

Instead, to be able to compare the covariates across mixture components, we derive a covariate importance measure IMP that is motivated by the coefficient of determination, R^2. The objective of this measure is to assess the magnitude and the sign (positive or negative) of the effect of a covariate for a specific mixture component on the data fit. 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 1. For a component k and a covariate j,

$$\mathrm{IMP}_{kj} = 1 - \frac{\sum_n I(z_n = k)(y_n - \hat{y}_{n\tilde{\boldsymbol{\beta}}})^2}{\sum_n I(z_n = k)(y_n - \hat{y}_{n\tilde{\boldsymbol{\beta}}^j})^2}, \tag{9}$$

where $I(z_n = k)$ is the indicator function of whether cell n is allocated to component k, $\hat{y}_{n\tilde{\boldsymbol{\beta}}}$ is the predicted count by using the full vector of inferred regression coefficients and $\hat{y}_{n\tilde{\boldsymbol{\beta}}^j}$ is the predicted count by using the regression coefficients with the jth coefficient set to 0. The magnitude of IMP is interpreted as a measure of the relative importance of the corresponding covariate for the model fit. A value of IMP closer to 1 represents that removing the corresponding covariate is more detrimental to model fit.

We determine the sign of IMP for a given covariate and a mixture component by inspecting the distribution of the covariate for the given component. We need to be careful with negative values as our covariates are centred on zero and standardized. To obtain the sign, 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}-estimate.

4.2. Simulation study details

For the methodology that was developed in Section 2, we need to choose the grid size, blocking structure, number of mixture components, K, and model specification.

4.2.1. Model choices

To choose the grid size, we take into account the precision of the burglary point pattern. The published data have been anonymized by mapping exact locations to predefined (snap) points (Police.uk, 2018). We follow the recommendations in Tompson *et al.* (2015) who assessed the accuracy of the anonymization method by aggregating both the original and the obfuscated data to areal counts at different resolutions and looking at the difference. They showed that the aggregation at LSOA level does not suffer from the bias that is introduced by the anonymization process. Therefore, for our cell size, we approximately match an average-size LSOA to avoid the

loss of precision that is caused by the anonymization process. As a result, our grid has $N = 9824$ cells, each of which corresponds to an area of 400×400 m^2.

For the blocking structure, we take advantage of the existing census output areas, that are designed to group homogeneous groups of households and people together (Office for National Statistics, 2019). Given that our grid is approximately at the LSOA level, we choose MSOAs as the blocking structure. We assess the sensitivity of this choice in Section 4.4.

The number of components, K, is a crucial parameter of our model. We run our model for varying K and use the performance measures that were introduced above to decide on the optimal number of components. From our experience, after a certain number of components, interpretation becomes more difficult whereas performance does not significantly improve.

We choose the model specification on the basis of the four options that were mentioned in Section 3.3.

4.2.2. Dependence of blocks

In Section 2 we proposed two possible formulations for the prior on the mixture weights: the multinomial logit transformation of K Gaussian random fields and independent Dirichlet random variables. To assess whether assuming block dependence has a major effect on the quality of the model, we compare the out-of-sample performance for both variants of the model. For this comparison, we set the blocking scheme to MSOAs, use model specification 4 and estimate the model on the burglary 2015 data set. To fit the model with dependent blocks, we use the squared exponential kernel (Rasmussen and Williams, 2006) where we choose the length scale parameter by optimizing out-of-sample RMSE by using grid search. Table 2 shows the mean and the standard deviation of the samples of held-out log-likelihood and RMSE for both variants of the model, and for various values of K. The italics signify which method performed better for the given K and for the given metric. The 'double-dagger' symbol indicates statistical significance with p-value less than 10^{-3} obtained from a two-sample t-test of samples of each metric for each variant of the model.

The results in Table 2 show that the model with dependent blocks does not consistently lead to improved performance. This indicates that block dependence structure in the burglary point pattern data that we consider is not a major effect. These findings highlight some aspects of the data structure in terms of capturing these effects and suggest that the point pattern data at a higher precision would be needed to uncover these effects, if they are present. For this reason,

Table 2. Model performance comparison of two variants of the model—dependent blocks using the logistic transform of K Gaussian processes, and independent blocks with Dirichlet prior†

K	*Held-out log-likelihood*		*RMSE*	
	Independent	*Dependent*	*Independent*	*Dependent*
2	-2.607 ± 0.010	*-2.605 ± 0.010‡*	*4.999 ± 0.028‡*	5.010 ± 0.028
3	-2.598 ± 0.012	*-2.593 ± 0.011‡*	4.973 ± 0.036	*4.950 ± 0.031‡*
4	*-2.588 ± 0.011‡*	-2.606 ± 0.012	*4.964 ± 0.034‡*	4.988 ± 0.031

†Reported values are a mean and standard deviation obtained from MCMC samples. Blocking, MSOAs; training data, burglary 2015; test data, 2016; model specification 4.
‡$p < 10^{-3}$.

in the rest of the paper we consider only independent blocks with Dirichlet prior weights as described in Section 2.3.

4.2.3. Identifiability

As mentioned in Section 2, the trace plot of the log-likelihood can be inspected for label switching. From our experience, the sampler would choose one of the $K!$ modes, that are a consequence of the likelihood invariance, and is unlikely to switch to another mode because of the high dimensionality of the parameter space.

4.3. Performance of SAM-GLM

Figs 2 and 3 report the performance for the 2015 and 2013–2015 data sets respectively. In Figs 2(a) and 3(a), we report the boxplots of the posterior distribution of the average held-out

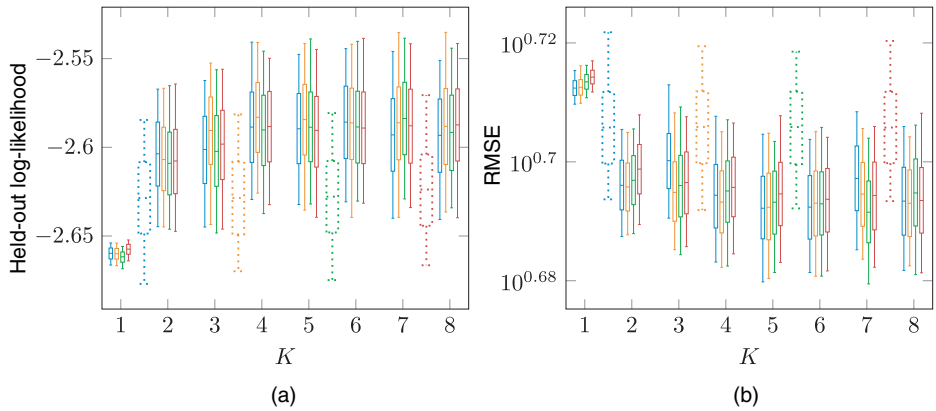


Fig. 2. Evaluation of the performance of SAM-GLM (———), compared with the LGCP model (· - - - - ·) for the 1-year data set: (a) log-likelihood and (b) RMSE for the held-out data for model specifications 1 (———), 2 (———), 3 (———) and 4 (———) (blocking, MSOAs; training data, burglary 2015; test data, burglary 2016; note that the axis with the value of K does not apply to the LGCP results)

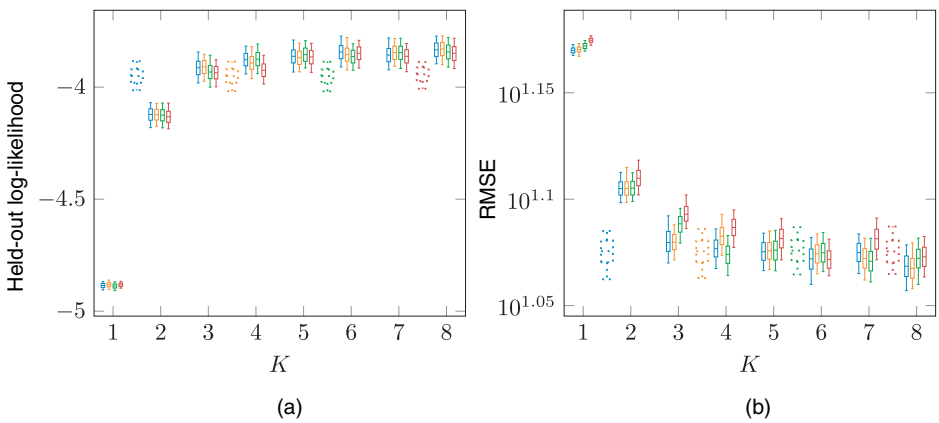


Fig. 3. Evaluation of the performance of SAM-GLM (———), compared with the LGCP model (· - - - - ·) for the 3-year data set: (a) log-likelihood and (b) RMSE for the held-out data for model specifications 1 (———), 2 (———), 3 (———) and 4 (———) (blocking, MSOAs; training data, burglary 2013–2015; test data, burglary 2016–2018; note that the axis with the value of K does not apply to the LGCP results)

log-likelihood. We show boxplots for different model specifications for both SAM-GLM with an increasing number of components, K, and the LGCP model. In Figs 2(b) and 3(b), we report analogous plots for the RMSE metric.

For the 1-year data set, SAM-GLM matches the predictive performance of the LGCP model for $K = 2$ components on both metrics. For the 3-year data set, $K = 3$ components are enough to match the LGCP model using the held-out log-likelihood, but at least $K = 4$ components are required for the RMSE. The extra components that are required to match the performance of the LGCP could be explained by the fact that the 3-year point pattern will naturally be smoother and thus easier to interpolate non-parametrically by using the Gaussian random-field part of the LGCP. The probability distribution for both metrics and for all models are more concentrated for the 3-year data set. For the 1-year data set, it is clear that $K = 2$ or $K = 3$ is the optimal number of components. For the 3-year counterpart, the range between three and five components would

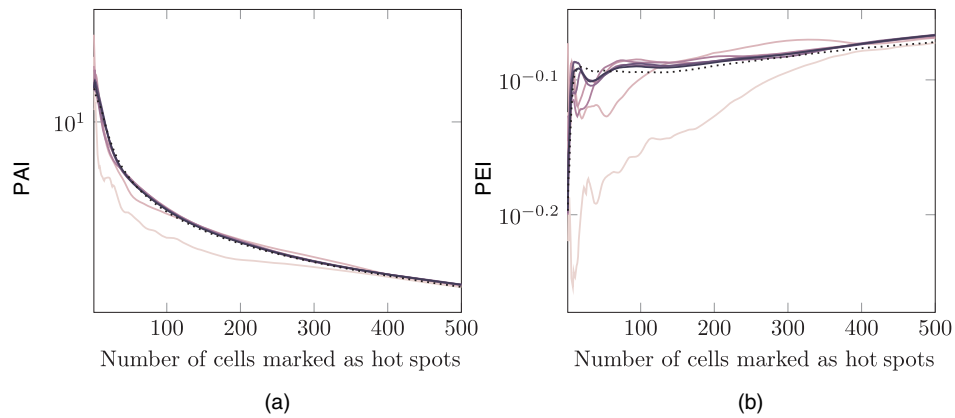


Fig. 4. (a) PAI- and (b) PEI-performance of SAM-GLM (———) and the LGCP model (· - - - - ·), using specification 4: for the SAM-GLM results, the colour of the line represents the number of components, $K = 1$ (———), $K = 2$ (———), $K = 3$ (———), $K = 4$ (———), $K = 5$ (———), $K = 6$ (———) and $K = 7$ (———) (blocking, MSOAs; training data, burglary 2015; test data, burglary 2016)

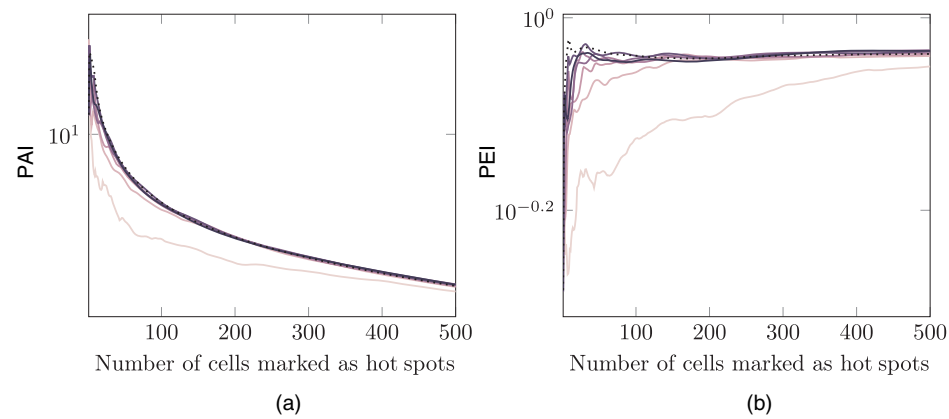


Fig. 5. (a) PAI- and (b) PEI-performance of SAM-GLM (———) and the LGCP models (· - - - - ·), using specification 4: for the SAM-GLM results, the colour of the line represents the number of components, $K = 1$ (———), $K = 2$ (———), $K = 3$ (———), $K = 4$ (———), $K = 5$ (———), $K = 6$ (———) and $K = 7$ (———) (blocking, MSOAs; training data, burglary 2013–2015; test data, burglary 2016–2018)

be an appropriate choice. For both data sets, the performance does not vary significantly for different model specifications. Consequently, in the following sections, we limit our attention to specification 4 because of its parsimony.

Whereas the out-of-sample performance, measured by the held-out log-likelihood or RMSE, takes into account all locations, practitioners might be interested only in predicting crime hot spots. For this, we evaluate PAI and PEI (see Section 4.1) as measures of hot spot prediction. Figs 4 and 5 show the plots of the PAI- and PEI-measures for both models with specification 4, using the 2015 and 2013–2015 data sets respectively. The plots show the score for when up to 500 cells (around 5% of the study region) are flagged as hot spots. Hot spots are chosen as the n cells with the highest expected value of burglaries. For the 1-year data set, SAM-GLM with $K = 2$ components is enough to outperform the LGCP on the PEI-measure when between 50 and 500 cells are flagged as hot spots. For the PAI-measure, no significant difference can be seen for $K > 2$. The results based on the 3-year data favour the LGCP model when up to 150 cells are flagged as hot spots and $K < 5$. After adding more components, SAM-GLM's performance

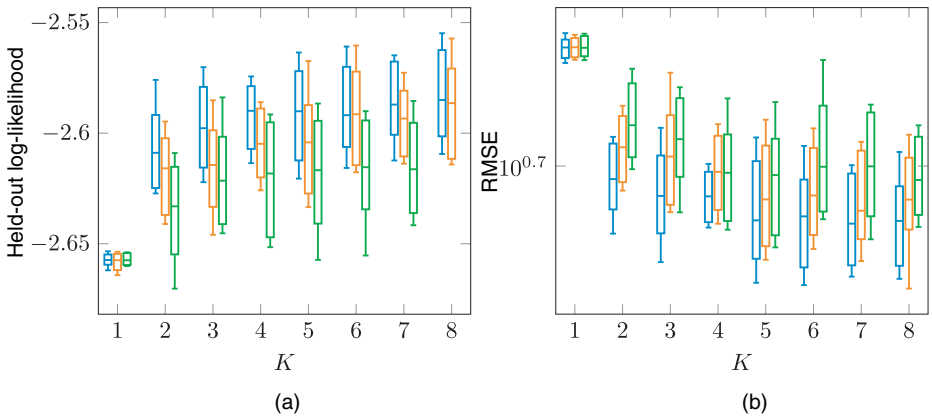


Fig. 6. (a) Log-likelihood and (b) RMSE for the held-out data for various block sizes: MSOAs (———); LADs (———); single blocks (———) (the error bars represent the standard deviation obtained from the MCMC samples; training data, 2015; test data, 2016; model specification 4)

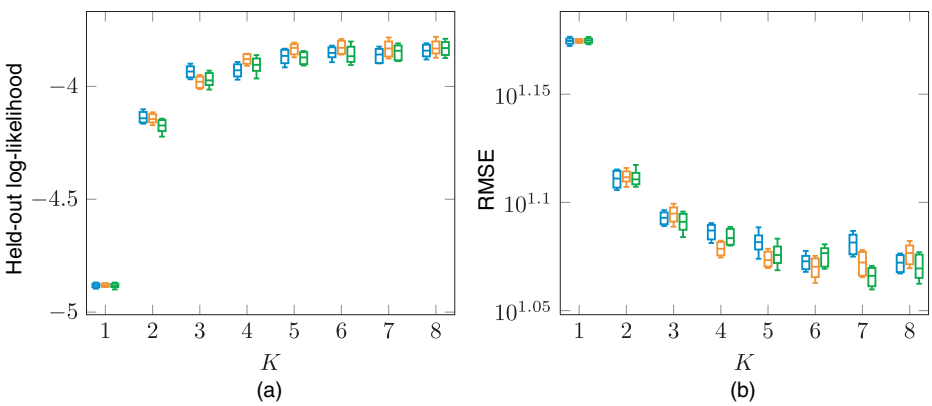


Fig. 7. (a) Log-likelihood and (b) RMSE for the held-out data for various block sizes: MSOAs (———); LADs (———); single blocks (———) (the error bars represent the standard deviation obtained from the MCMC samples; training data, 2013–2015; test data, 2016–2018; model specification 4)

matches that of the LGCP. When between 150 and 500 cells are flagged, $K \geqslant 3$ components are enough to outperform the LGCP. These results are consistent with the previous finding that to outperform the LGCP on the 3-year data set requires more components.

4.4. Block size sensitivity

The model proposed requires a specification of the blocking structure for the mixture weights prior. To assess the sensitivity of this choice, we compare with local authority districts (LADs), as well as a single block for the whole study region. In the latter case, the model reduces to a non-spatial mixture of Poisson GLMs. There are 946 MSOAs, and 33 LADs in the study region. The structure is hierarchical—multiple non-overlapping contiguous MSOAs constitute a single LAD region.

Figs 6 and 7 show the boxplots of the held-out log-likelihood and RMSE for the 1-year and the 3-year data sets respectively. The results for both metrics indicate that imposing spatial information using a more localized prior results in better out-of-sample performance for the 1-year data set. To confirm that the difference is statistically significant, we performed an unpaired two-sample t-test comparing RMSE samples obtained by using the MSOA blocking structure with those obtained by using LADs and single blocks. Table 3 summarizes the t-statistics and p-values. For the 3-year data set, there is no evident difference, and a spatial prior does not improve the predictive performance of the model. This is not surprising as the 3-year observation window will provide more information and thus the model is less likely to be overfitted even if we do not impose spatial dependence within the blocks.

4.5. Interpretation

For this analysis, we choose the 3-year data set because more data will lead to more robust inferences of the parameters. We choose specification 4 with $K = 3$ components because of its parsimony and the excellent performance that was shown above—for the 3-year data set and specification 4, there does not seem to be a significant improvement after $K > 3$ components. Fig. 8 shows the component allocation maps and the IMP-measure with the effect sign (positive or negative) for each covariate for all the three components. The allocation map for each component shows the proportion of the MCMC samples that a cell is allocated to that

Table 3. Sensitivity analysis of block sizes: p-values comparing whether the difference in RMSE performance is significant†

K	*Results for MSOA versus LAD*		*Results for MSOA versus single*	
	t-statistic	*p-value*	*t-statistic*	*p-value*
2	−68.732	$< 10^{-3}$	−115.042	$< 10^{-3}$
3	−76.260	$< 10^{-3}$	−87.534	$< 10^{-3}$
4	−39.016	$< 10^{-3}$	−35.207	$< 10^{-3}$
5	−26.858	$< 10^{-3}$	−52.991	$< 10^{-3}$
6	−41.913	$< 10^{-3}$	−76.152	$< 10^{-3}$
7	−12.173	$< 10^{-3}$	−56.847	$< 10^{-3}$
8	−31.547	$< 10^{-3}$	−66.688	$< 10^{-3}$

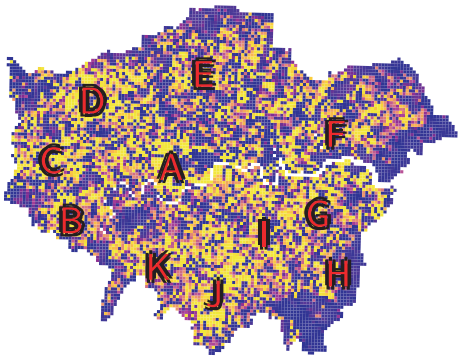
†Training data, burglary 2015; test data, burglary 2016; specification 4.

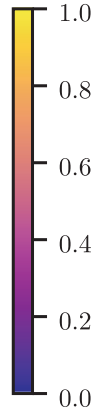
component. The alphanumeric labels on the allocation plots are used in the discussion below when referring to specific locations. IMP is computed for each sample and component separately and then averaged over the MCMC samples. We also report the standard deviation of the IMP-estimate in parentheses.

The first component is active throughout the study region, with large clusters around residential areas (Fig. 8(a)). These include areas around Kensington, Fulham and Shepherd's Bush (A), Hounslow, Kingston, Richmond and Twickenham (B), Hayes and Southall (C), Harrow and Edgware (D), East Barnet, Enfield, Walthamstow and Wood Green (E), Barking and Dagenham (F), Bexley (G), Orpington (H), Bromley (I), Croydon and Purley (J) and New Malden and Morden (K). In this component, the number of households and POIs have the strongest effect (excluding the intercept)—burglaries happen where targets are. Accessibility has also been inferred as an important covariate, which is consistent with past criminological studies. In this component, house price is inferred as having a positive effect on the intensity of burglary, suggesting that offenders choose attractive targets. The positive effect of ethnic heterogeneity confirms the hypothesis from social disorganization theory. The other indicators of social disorganization—occupation variation and residential turnover—are weaker but are consistent with the existing criminology literature. House price as a measure of reward and the proportion of houses that are detached and semidetached have low IMP-values.

Component 2 is active in the city centre and in the High Streets of neighbourhoods (Fig. 8(b)): Soho, Mayfair, Covent Garden, Marylebone and Fitzrovia (L), Shoreditch and Stratford (M), Streatham and Tooting Bec (N), Wembley and Brent (O), Enfield and Hampstead (P), Romford (Q), Orpington (R) and Wembley and Harrow (S). Burglary rates in these locations are largely driven by POIs and households. Compared with the first component (residential), the magnitudes of IMP-values for these covariates are different—POIs are more important for this component, and the number of households is more important for the first component. The accessibility measure is inferred to have high importance 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 activity theory which suggest that offenders choose locations that are part of their usual routine and in their awareness spaces. Ethnic heterogeneity and occupation variation have a strong positive effect and signify the lack of social cohesion. 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 and Luykx, 2003; Bernasco and 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 be less likely to 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 be targeted. This is contrary to the first component. A possible explanation that has been 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 and Wasilchick, 2010).

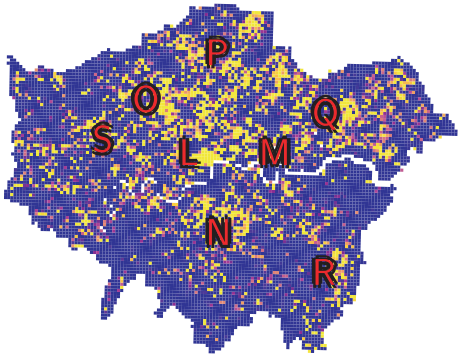
The last component is active in the areas of low intensity (Fig. 8(c)). These include Hyde Park, Regent's Park and Hampstead Heath (1), Richmond Park and Bushy Park (2), Osterley Park and Kew botanic gardens (3), Heathrow airport (4), RAF Northolt and parks near Harrow (5), Edgware fields (6), Lee Valley (7), the industrial zone in Barking and Rainham Marshes (8), parks around Bromley and Biggin Hill airport (9), and other non-urban

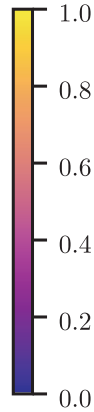




(a)

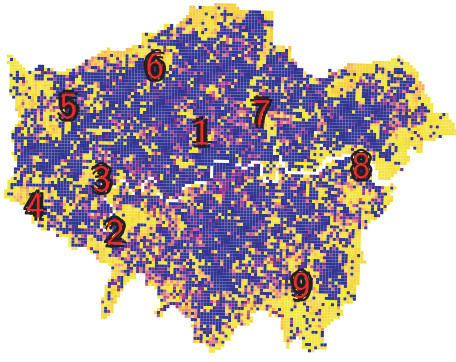
IMP, component 1		
intercept	0.947 (0.001)	+
log households	0.887 (0.002)	+
log POIs (all)	0.399 (0.022)	+
accessibility	0.225 (0.024)	+
log house price	0.100 (0.017)	+
ethnic heterogeneity	0.083 (0.016)	+
occupation variation	0.025 (0.011)	+
population turnover	0.011 (0.004)	+
(Semi-)detached houses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

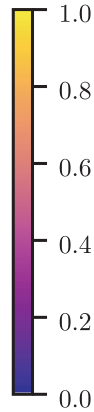




(b)

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)	+





(c)

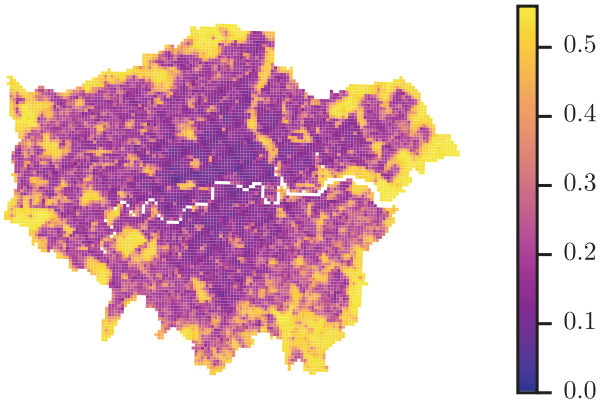
IMP, component 3		
log households	0.946 (0.003)	-
intercept	0.906 (0.004)	+
log POIs (all)	0.808 (0.015)	-
occupation variation	0.719 (0.060)	+
log house price	0.680 (0.027)	+
accessibility	0.435 (0.086)	+
ethnic heterogeneity	0.050 (0.024)	+
(Semi-)detached houses	0.002 (0.007)	+
population turnover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

Fig. 8. Mixture model, allocations and IMP-table for each mixture component (training data, 2013–2015; specification 4): (a) component 1 ($P_{\text{posterior}}(z_n = 1)$); (b) component 2 ($P_{\text{posterior}}(z_n = 2)$); (c) component 3 ($P_{\text{posterior}}(z_n = 3)$)

areas at the edges of the map. This component explains locations with little criminal activity, signified by negative IMP for the number of households and POIs. Occupation variation, as a measure of socio-economic heterogeneity, is strongly positive, which would support the hypothesis from social disorganization theory. However, this is more likely to be due to the very low population in those areas which results in a high occupation variation measure. The accessibility measure also has a positive effect on rates of burglary in these locations. This is expected and in line with the hypotheses from crime pattern theory. Other covariates have very small IMP-values.

The allocation of cells partitions the map into three clusters. By aggregating the number of observed crimes that occurred in each cluster we obtain that components 1, 2 and 3 cover 46%, 42% and 12% of all burglaries during the 2013–2015 period respectively. Official aggregated police data for this period make the split of 64% and 36% for residential and non-residential burglary (Police.uk, 2019). Our inference agrees that there is more residential burglary than non-residential burglary and that approximately 35–45% of burglaries are non-residential. It is unclear whether the crime in low count areas, which according to our model accounts for 12%, is residential or non-residential.

Support for spatial heterogeneity is further given by inspecting the inferences that were made by the LGCP model (for LGCP details see section B in the on-line supplementary material). Fig. 9(a) shows the standard deviation of the marginal posterior distribution 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 urbanized. In contrast, SAM-GLM has picked up this heterogeneity by allowing a separate component for it (see Fig. 8(c)). Fig. 9(b) shows IMP computed for all components of the LGCP model. The IMP-measure for the field component of the model is computed by treating it as a covariate with coefficient equal to 1. The IMP-value for the latent field component is the third highest, after the intercept and the number of households. A 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.



(a)

IMP, LGCP		
intercept	0.953 (0.001)	+
log households	0.937 (0.002)	+
field	0.810 (0.005)	+
log POIs (all)	0.728 (0.013)	+
accessibility	0.103 (0.058)	+
occupation variation	0.097 (0.052)	+
log house price	0.089 (0.043)	+
ethnic heterogeneity	0.061 (0.029)	+
population turnover	0.002 (0.027)	+
(Semi-)detached houses	0.000 (0.027)	-

(b)

Fig. 9. (a) Standard deviation of the posterior distribution of the latent field **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)) and (b) IMP-measure for the component of the LGCP model; for both panels, training data, 2013–2015, and model specification 4

4.6. Overdispersion; excess of 0s

The discussion of the inferences above shows that our model effectively handles an excess of 0s by allocating low count cells (non-urban areas) their own cluster, which has its own regression coefficients. Similarly, the mixture model proposed can reduce the overdispersion problem that is present in the standard Poisson GLM model (the special case of SAM-GLM, with $K = 1$). The mixture model may allocate each cell to a cluster that better describes the burglary count in that location. Inspecting the Pearson χ^2-statistic ($\chi^2 = \Sigma_{i=1}^{N}(\text{Observed}_i - \text{Expected}_i)^2/\text{Expected}_i$) provides supporting evidence for this. The introduction of two extra components has resulted in the 81% decrease, from 106942.43 to 20028.99, showing a better model fit. This is further confirmed by a scatter plot of expected *versus* observed counts for the Poisson GLM model and the proposed model with $K = 3$ as shown in Fig. 2 in the on-line supplementary material.

5. Conclusions

Spatial point patterns on large spatial regions, such as metropolitan areas, often exhibit localized behaviour. Motivated by this, we propose 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 different locations to follow a different model; for example, in the urban context, less urbanized locations can assume a different model from that of the city centre. Each component is an instance of the GLM which includes covariates. We account for spatial dependence through the mixture allocation part. The allocation of each location to one of the components is informed by both the data and the prior information. By utilizing existing blocks structure, or defining a custom structure, the prior supports 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 borrowing statistical strength for locations within the same block. Additionally, the model allows for spatial dependence between the blocks. Following the Bayesian framework, we present an MCMC sampler to infer the posterior distributions. Inspection of the posterior distributions of the model parameters enables us to learn new insights about the underlying mechanisms of the point pattern.

Our results show that the London burglary data are effectively modelled by the method proposed. Using out-of-sample and crime hot spot prediction evaluation measures, we showed that our model outperforms the LGCP model (with Matérn covariance function), which is the default model for point processes and is more computationally tractable.

The focus of this work on burglary crime does not limit the potential uses of the model proposed. We believe that the model can be applied in a wider setting of analysing spatial point patterns that may show localized behaviour and heterogeneity.

Future analysis could consider several directions that were not explored in this work. Firstly, our inference scheme for the model with block dependence produces an $\mathcal{O}(B^3 \times K)$ algorithm. To reduce this complexity, one could consider K level sets of a single Gaussian random field for mixture weights, instead of K Gaussian fields, thus reducing the dimensionality (Hildeman *et al.*, 2018; Fernández and Green, 2002). Another approach is to assume a Markovian structure for the Gaussian random fields, resulting in sparse computational methods (Rue and Held, 2005). A different approach is to consider inference schemes that are less computationally demanding than MCMC sampling such as variational methods (Jordan *et al.*, 1999). 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 a strong prior belief for doing so. Another possible area of investigation is the spatially varying-coefficient processes method, which was proposed by Gelfand *et al.* (2003).

6. Implementation

The source code that implements the methodology and reproduces the experiments is available from `https://github.com/jp2011/spatial-poisson-mixtures`.

Acknowledgements

We thank the Associate Editor and the referees for their helpful comments and suggestions. All the authors were supported by Engineering and Physical Sciences Research Council grants EP/P020720/1 and EP/P020720/2 for 'Inference, computation and numerics for insights into cities' (`https://iconicmath.org/`). Jan Povala was also supported by Engineering and Physical Sciences Research Council grant EP/L015129/1. Mark Girolami was supported by Engineering and Physical Sciences Research Council grants EP/T000414/1, EP/R018413/2, EP/R034710/1 and 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 the Strategic Insights Unit at the Metropolitan Police Service in London. The authors are also grateful to Louis Ellam for insightful comments and suggestions, and to Pavol Povala for his feedback on the drafts of the manuscript.

References

Agresti, A. and Agresti, B. F. (1978) Statistical analysis of qualitative variation. *Sociol. Methodol.*, **9**, 204–207.
Aldor-Noiman, S., Brown, L. D., Fox, E. B. and Stine, R. A. (2016) Spatio-temporal low count processes with application to violent crime events. *Statist. Sin.*, **26**, 1587–1610.
Alvares, D., Armero, C. and Forte, A. (2018) What does objective mean in a Dirichlet-multinomial process? *Int. Statist. Rev.*, **86**, 106–118.
Andresen, M. A. (2010) The place of environmental criminology within criminological thought. In *Classics in Environmental Criminology*, pp. 21–44. Boca Raton: CRC Press.
Anselin, L., Cohen, J., Cook, D., Gorr, W. and Tita, G. (2000) Spatial analyses of crime. *Crim. Just.*, **4**, 213–262.
Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2015) *Hierarchical Modeling and Analysis for Spatial Data*, 2nd edn. Boca Raton: CRC Press.
Beavon, D. J., Brantingham, P. L. and Brantingham, P. J. (1994) The influence of street networks on the patterning of property offenses. *Crime Prevn Stud.*, **2**, 115–148.
Bernasco, W. (2014) Residential burglary. In *Encyclopedia of Criminology and Criminal Justice* (eds G. Bruinsma and D. Weisburd), pp. 4381–4391. New York: Springer.
Bernasco, W., Johnson, S. D. and Ruiter, S. (2015) Learning where to offend: effects of past on future burglary locations. *Appl. Geog.*, **60**, 120–129.
Bernasco, W. and Luykx, F. (2003) Effects of attractiveness, opportunity, and accessibility to burglars on residential burglary rates of urban neighbourhoods. *Criminology*, **41**, 981–1002.
Bernasco, W. and Nieuwbeerta, P. (2005) How do residential burglars select target areas? *Br. J. Crimin.*, **45**, 296–315.
Bowers, K. and Hirschfield, A. (1999) Exploring links between crime and disadvantage in north-west England: an analysis using geographical information systems. *Int. J. Geog. Inform. Sci.*, **13**, 159–184.
Brantingham, P. and Brantingham, P. (1981) Notes on the geometry of crime. In *Environmental Criminology*. Beverly Hills: Sage.
Brantingham, P. J. and Brantingham, P. L. (1975) The spatial patterning of burglary. *Howrd J. Crim. Just.*, **14**, no. 2, 11–23.
Brantingham, P. L. and Brantingham, P. J. (1993) Nodes, paths and edges: considerations on the complexity of crime and the physical environment. *J. Environ. Psychol.*, **13**, 3–28.
Breslow, N. E. (1984) Extra-Poisson variation in log-linear models. *Appl. Statist.*, **33**, 38–44.
Brunsdon, C., Fotheringham, A. S. and Charlton, M. E. (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geog. Anal.*, **28**, 281–298.
Celeux, G., Hurn, M. and Robert, C. P. (2000) Computational and inferential difficulties with mixture posterior distributions. *J. Am. Statist. Ass.*, **95**, 957–970.
Chainey, S., Tompson, L. and Uhlig, S. (2008) The utility of hotspot mapping for predicting spatial patterns of crime. *Secty J.*, **21**, no. 1, 4–28.
Clare, J., Fernandez, J. and Morgan, F. (2009) Formal evaluation of the impact of barriers and connectors on residential burglars' macro-level offending location choices. *Aust. New Zeal. J. Crimin.*, **42**, 139–158.

Clarke, R. V. and Cornish, D. B. (1985) Modeling offenders' decisions: a framework for research and policy. *Crime Just.*, **6**, 147–185.
Cohen, L. E. and Felson, M. (1979) Social change and crime rate trends: a routine activity approach. *Am. Sociol. Rev.*, **44**, 588–608.
Diggle, P. J., Moraga, P., Rowlingson, B. and Taylor, B. M. (2013) Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. *Statist. Sci.*, **28**, 542–563.
Duane, S., Kennedy, A., Pendleton, B. J. and Roweth, D. (1987) Hybrid Monte Carlo. *Phys. Lett.* B, **195**, 216–222.
Evans, D. J. (1989) Geographical analyses of residential burglary. In *The Geography of Crime* (eds D. J. Evans and D. T. Herbert), pp. 86–107. London: Routledge.
Felson, M. and Clarke, R. V. (1998) Opportunity makes the thief. *Police Research Series, Paper 98*. Home Office, London.
Fernández, C. and Green, P. J. (2002) Modelling spatially correlated data via mixtures: a Bayesian approach. *J. R. Statist. Soc.* B, **64**, 805–826.
Flaxman, S., Chirico, M., Pereira, P. and 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". *Ann. Appl. Statist.*, **13**, 2564–2585.
Flaxman, S., Wilson, A. G., Neil, D. B., Nickisch, H. and Smola, A. J. (2015) Fast Kronecker inference in Gaussian processes with non-Gaussian likelihoods. In *Proc. 32nd Int. Conf. Machine Learning, Lille* (eds F. Bach and D. Blei), pp. 607–616.
Frühwirth-Schnatter, S., Celeux, G. and Robert, C. P. (eds) (2019) *Handbook of Mixture Analysis*. Boca Raton: CRC Press.
Gelfand, A. E., Kim, H.-J., Sirmans, C. F. and Banerjee, S. (2003) Spatial modeling with spatially varying coefficient processes. *J. Am. Statist. Ass.*, **98**, 387–396.
Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013) *Bayesian Data Analysis*. Boca Raton: Chapman and Hall–CRC.
Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattn Anal. Mach. Intell.*, **6**, 721–741.
Girolami, M. and Calderhead, B. (2011) Riemann manifold Langevin and Hamiltonian Monte Carlo methods (with discussion). *J. R. Statist. Soc.* B, **73**, 123–214.
GOV.UK (2017) Crime against businesses: findings from the 2017 Commercial Victimisation Survey. (Available from `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 *Handbook of Spatial Statistics* (eds S. Frühwirth-Schnatter, G. Celeux and C. P. Robert). Boca Raton: CRC Press.
Green, P. J. and Richardson, S. (2002) Hidden Markov models and disease mapping. *J. Am. Statist. Ass.*, **97**, 1055–1070.
Grün, B. and 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* (eds Shalabh and C. Heumann), pp. 205–230. Heidelberg: Physica.
Hildeman, A., Bolin, D., Wallin, J. and Illian, J. B. (2018) Level set Cox processes. *Spatl Statist.*, **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 DC.
Johnson, S. D. and Bowers, K. J. (2004) The stability of space-time clusters of burglary. *Br. J. Crimin.*, **44**, 55–65.
Johnson, S. D. and Bowers, K. J. (2010) Permeability and burglary risk: are cul-de-sacs safer? *J. Quant. Crimin.*, **26**, 89–111.
Johnson, S. D. and Summers, L. (2015) Testing ecological theories of offender spatial decision making using a discrete choice model. *Crime Delinq.*, **61**, 454–480.
Jordan, M. I., Ghahramani, Z., Jaakkola, T. S. and Saul, L. K. (1999) An introduction to variational methods for graphical models. *Mach. Learn.*, **37**, 183–233.
Knorr-Held, L. and Raßer, G. (2000) Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, **56**, 13–21.
McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.
Menting, B., Lammers, M., Ruiter, S. and Bernasco, W. (2019) The influence of activity space and visiting frequency on crime location choice: findings from an online self-report survey. *Br. J. Crimin.*, **60**, 303–322.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **21**, 1087–1092.
Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P. and Tita, G. E. (2011) Self-exciting point process modeling of crime. *J. Am. Statist. Ass.*, **106**, 100–108.
Møller, J., Syversveen, A. R. and Waagepetersen, R. P. (1998) Log Gaussian Cox Processes. *Scand. J. Statist.*, **25**, 451–482.
Møller, J. and Waagepetersen, R. P. (2007) Modern statistics for spatial point processes. *Scand. J. Statist.*, **34**, 643–684.

Office for National Statistics (2019) Census geography—Office for National Statistics. Office for National Statistics, Newport. (Available from `https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography`.)
Ordnance Survey (GB) (2018) Points of Interest [CSV geospatial data], Scale 1:1250, Items: 670887. Ordnance Survey. (Available from `https://digimap.edina.ac.uk`.)
Police.uk (2018) About | data.police.uk. (Available from `https://data.police.uk/about`.)
Police.uk (2019) Data downloads | data.police.uk. (Available from `https://data.police.uk/data/`.)
PredPol (2019) Aiming to reduce victimization keep communities safer. Predpol. (Available from `https://www.predpol.com/about/`.)
Rasmussen, C. E. and Williams, C. K. I. (eds) (2006) Gaussian processes for machine learning. In *Adaptive Computation and Machine Learning*. Cambridge: MIT Press.
Rengert, G. F. and Wasilchick, J. (2010) The use of space in burglary. In *Classics in Environmental Criminology* (eds M. A. Andresen, P. J. Brantingham and B. J. Kinney), pp. 257–272. Boca Raton: CRC Press.
Rousseau, J. and Mengersen, K. (2011) Asymptotic behaviour of the posterior distribution in overfitted mixture models. *J. R. Statist. Soc.* B, **73**, 689–710.
Rue, H. and Held, L. (2005) *Gaussian Markov Random Fields: Theory and Applications*. Boca Raton: Chapman and Hall–CRC.
Sampson, R. J. and Groves, W. B. (1989) Community structure and crime: testing social-disorganization theory. *Am. J. Sociol.*, **94**, 774–802.
Sampson, R. J., Raudenbush, S. W. and Earls, F. (1997) Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science*, **277**, 918–924.
Serra, L., Saez, M., Mateu, J., Varga, D., Juan, P., Díaz-Ávalos, C. and Rue, H. (2014) Spatio-temporal log-Gaussian Cox processes for modelling wildfire occurrence: the case of Catalonia, 1994–2008. *Environ. Ecol. Statist.*, **21**, 531–563.
Shaw, C. R. and 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: University of Chicago Press.
Smith, K., Taylor, P. and Elkin, M. (2013) Crimes detected in England and Wales 2012/13. *Statistical Bulletin 02/13*. Home Office, London.
Taddy, M. A. (2010) Autoregressive mixture models for dynamic spatial Poisson processes: application to tracking intensity of violent crime. *J. Am. Statist. Ass.*, **105**, 1403–1417.
Tobler, W. R. (1970) A computer movie simulating urban growth in the Detroit region. *Econ. Geog.*, **46**, 234–240.
Tompson, L., Johnson, S., Ashby, M., Perkins, C. and Edwards, P. (2015) UK open source crime data: accuracy and possibilities for research. *Cart. Geog. Inform. Sci.*, **42**, 97–111.
Townsley, M., Birks, D., Bernasco, W., Ruiter, S., Johnson, S. D., White, G. and Baum, S. (2015) Burglar target selection: a cross-national comparison. *J. Res. Crime Delinq.*, **52**, 3–31.
Townsley, M., Birks, D., Ruiter, S., Bernasco, W. and White, G. (2016) Target selection models with preference variation between offenders. *J. Quant. Crimin.*, **32**, 283–304.

Supporting information

Additional 'supporting information' may be found in the on-line version of this article:

'Supplementary material for "Burglary in London: insights from statistical heterogeneous spatial point processes"'.