

ORIGINAL ARTICLE

Spatio-temporal mixed membership models for criminal activity

Seppo Virtanen¹ | Mark Girolami^{2,3}¹University of Leeds, Leeds, UK²University of Cambridge, Cambridge, UK³The Alan Turing Institute, London, UK**Correspondence**Seppo Virtanen, University of Leeds,
Leeds, UK.

Email: s.virtanen@leeds.ac.uk

Funding informationEPSRC, Grant/Award Number: EP/
P020720/1**Abstract**

We suggest a probabilistic approach to study crime data in London and highlight the benefits of defining a statistical joint crime distribution model which provides insights into urban criminal activity. This is achieved by developing a hierarchical mixture model for observations, crime occurrences over a geographical study area, that are grouped according to multiple time stamps and crime categories. The mixture components correspond to spatial crime distributions over the study area and the goal is to infer, based on the observations, how and to what degree the latent distributions are shared across the groups.

KEYWORDS

Bayesian statistics, high-dimensional data, latent factor models, multi-view modelling, spatial and temporal methods

1 | INTRODUCTION

Increasing numbers of the population now concentrate in cities. Understanding the behaviour of crime and underlying urban structures that explain city-based crime is of high relevance for improving societal well being, city planning, developing smart cities and policy making.

It is well known that crime is concentrated or clustered in space and time. Each offence involves the presence of both offender and victim or target, and a report to the police, following a complicated target selection process and/or opportunity. Crime routine theory promotes in addition the absence of a capable guardian to prevent or deter crime from occurring (Cohen & Felson, 1979), and rational choice theory focuses on the target selection process (Clarke, 1997). However, not all crime is reported and some

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Journal of the Royal Statistical Society: Series A (Statistics in Society) published by John Wiley & Sons Ltd on behalf of Royal Statistical Society

of the offences may be missing from the data bases due to various reasons, such as, for sensitivity or privacy issues. Offenders may often operate near their residence, work place or places where they spend their time in general as well as routes connecting these to each other. The same principle applies to some targets or victims of crime. Crime pattern theory explains that crime can only occur when the spaces of offenders and targets overlap (Brantingham & Brantingham, 1993a). Notably, same offenders may commit different types of crime in their surroundings (Roach & Pease, 2014), leading to complex dependencies between different types of crime. So called hotspots experience elevated risk of crime; some of the hotspots may show little variation whereas others may emerge, move and disappear with varying spread, routes and decay respectively (Gorr & Lee, 2015). Environmental criminology aims to explore the nature of the hotspots by linking them with community-level demographic or socio-economic and place-level geographical features, such as, age structure, poverty, deprivation, ethnicity, health as well as places for retail, commerce and accommodation, respectively, for providing further criminological insights regarding the nature and causes of crime. Social disorganisation theory focuses on demographic variables and explains that deprived areas experience high crime rates (Shaw & McKay, 1942). Accordingly, when combined with the other crime theories, affluent areas surrounding or close to deprived areas may experience high risk of crime. Sampson et al. (1997) show that coherent communities with high level of social efficacy deter crime and vice versa. Often the theories may not conflict each other and some of the features may experience intricate interrelationships (Smith et al., 2000). In a similar vein, Cahill and Mulligan (2007) explore local as opposed to global effects of the features and Drawve and Barnum (2018); Barnum et al. (2017) and Piza et al. (2017) focus analysis on particular areas to account for locally relevant features. Near-repeat pattern theory explains that past crime events elevate the risk of future crime in the vicinity of the past events in both space and time (Bowers et al., 2004), building on the foraging analogy (Johnson et al., 2009). Cohen et al. (2007) follow this approach and propose to combine different types of criminal activity as features that explain crime.

Crime mapping forms the backbone of data-driven crime modelling for exploring and explaining crime, aiming to infer crime distributions or risk surfaces, based on the occurrences. Some of these techniques have been deployed by police (Perry, 2013), leading to significant decreases in crime (Braga et al., 2014). The statistical approaches may be divided into three classes, kernel density estimation (Chainey et al., 2008; Gorr & Lee, 2015; Johnson et al., 2009; Porter & Reich, 2012) and its probabilistic counterpart using Gaussian processes (Flaxman et al., 2015; Rodrigues & Diggle, 2012; Shirota et al., 2017), self-exciting point processes (Liu & Brown, 2003; Møller & Rasmussen, 2005; Mohler, 2013, 2014; Mohler et al., 2011; Taddy, 2010) and generalised linear models (Caplan et al., 2011; Cohen et al., 2007; Flaxman et al., 2018; Groff & Lockwood, 2014; Stucky & Ottensmann, 2009). Respectively, these classes build on the hotspot property (or clustering) of crime, near-repeat pattern theory and environmental criminology. A recent trend in regression modelling is to adopt extensive sets of features or covariates employing regularisation techniques (Bernasco & Block, 2011; Haberman & Ratcliffe, 2015; Jones & Pridemore, 2019).

Each crime occurrence is unique and depends on context; for instance, burglary can be either commercial or residential depending on the target. In the absence of original and confidential occurrence-specific crime records, high-level crime categorisation provided by the police (law) is useful but entails a significant loss of detailed information (Kuang et al., 2017). Also, a single category is unable to represent finer sub-categorisation or associations between different categories. In other words, crime is multi-faceted, categorisation may be ambiguous, and each offence may have several meanings. However, spatial and temporal context may provide clues to reveal such offence semantics. Existing crime mapping techniques and regression-based models are unable to infer underlying crime semantics based solely on crime occurrences. However, mixed membership modelling (Airoldi et al., 2014) has been shown to be able to uncover such semantics. For example, based on text documents,

the models group semantically similar terms into meaningful themes or topics based on term co-occurrences (Blei et al., 2003). However, this class of models has not been applied for crime occurrence data.

It is imperative to work on a fine-grained spatial resolution to uncover the underlying urban mechanisms in detail and prevent high harm but infrequent crime (Andresen & Malleon, 2011; Caplan et al., 2011; Chainey et al., 2008; Mohler et al., 2011; Weisburd, 2015). Unfortunately, as a consequence, model complexity increases and data become more sparse and higher dimensional. Inferring reliable and detailed crime distributions becomes challenging because of the high-dimensional sparse data. A common approach is to grid the study area but street or address – based segmentations employing different shapes may be more useful (Deryol et al., 2016; Rosser et al., 2017).

Burglary is a well-studied crime category; many prior studies and related criminal theories exist. On the other hand, it is of high interest to understand and prevent violent crime that may be sparse overall but has high harm. We hypothesise that other crime categories are related to burglary (and each other). For example, vehicle crime may take place in (semi) private parking areas or on streets close to residences, violent domestic crime and criminal damage to property may give clues about the locations where offenders reside and business-related robbery may be closely related to burglaries for commercial (i.e. non-residential) targets. Recent work has explored associations between different types of crime (Bowers, 2014; Ceccato et al., 2002; Haberman, 2017; Hipp, 2007; Kinney et al., 2008; Schreck et al., 2009; Weisburd et al., 1992).

In this work, we collect spatio-temporal crime occurrence data from the UK police for London for five crime categories including *burglary*, *robbery*, *violence and sexual offences*, *vehicle crime* and *criminal damage and arson*, as defined by the data provider. Based on the criminological theories, our motivation is that joint analysis of multiple crime categories can help detect where and when offenders operate and targets are found, providing a more complete holistic description of crime. We present a spatio-temporal extension of mixed membership modelling for crime data. Here, the crime occurrences are generated from a set of crime distributions (or maps) that we constrain as probability distributions over a partition of the geographical study area. We adopt techniques based on finite element method (Lindgren et al., 2011) to construct the partition such that fine geographical detail is assigned to high-crime areas for better accuracy, whereas low-crime areas are more coarse and inaccurate. Such adaptation is intuitive and computationally scalable. The maps are influenced by spatial covariates and may be intuitively visualised and interpreted in terms of the covariates, inferring which features explain crime. The covariate effects are map-specific capturing locally relevant features as opposed to global. The model is able to account for crime semantics based on crime co-occurrences. Even though crimes may often not co-occur in a strict sense at micro-level, depending on the target (such as, building vs. person), they may still be close to each other. We jointly model occurrences from multiple different crime categories. As such, models based on crime mapping or regression are difficult to formulate as joint models for multiple crime categories and they cannot fully detect shared relevant covariates. By combining data from multiple crime categories, our goal is to improve performance, sharing statistical strength to alleviate the sparse and high-dimensional data problem, and identifying crime hotspots that may attract more and other types of crime, potentially, in new (previously unseen) or close-by areas where there are little or no category-specific data potentially because of low crime reporting rates or privacy concerns. The crime maps are shared across different categories but each crime map is accompanied with temporal proportions specific for each category. We use Gaussian Markov random fields (GMRF) (Rue & Held, 2005), that provide a computationally scalable alternative to Gaussian processes, to account for both spatial and temporal smoothing, capturing close-by crime co-occurrences in space and time. We adopt Bayesian inference and use Bayesian model selection based on an approximation to model evidence to choose the number of crime maps.

In summary, we present a multi-view extension of mixed membership modelling combining data from multiple crime categories (or different *views* of crime) effectively increasing the amount of data to infer better crime maps, and inferring statistical dependencies between any subset of crime categories. We anticipate our model, that extends the family of successful applications of mixed membership modelling, to spark new interest for joint crime distribution modelling.

The paper is structured as follows. Section 2 explains the data. The proposed model is introduced in Sections 3 and 4 shows results. Section 5 concludes the paper.

2 | DATA

We collect data from the publicly available UK Police website, <https://data.police.uk/>, for the Metropolitan police force for London, that exhibits high crime rates and complex urban structure encoded in a wide range of spatial covariates, making it an interesting and relevant study area. We prefer to collect data from a single police force to avoid potential data duplication and management issues. We use data for a subset of London boroughs focusing primarily on Westminster, which has ample and heterogeneous criminal activity, as well as neighbouring or close-by boroughs, such as, Camden, Islington, Kensington and Chelsea, Hammersmith and Fulham, and Brent for a time interval from 2011 December to 2019 June.

The crime data are provided in a table, where each row represents a crime occurrence including a date (in year–month format), location coordinates and type of crime. For privacy concerns, actual crime occurrences are discretised based on offence categorisation as well as spatial and temporal partitions by the data provider. We focus on five crime categories, including, burglary, criminal damage and arson, violence and sexual offences, robbery and vehicle crime. The spatial partition contains a set of non-overlapping areas such that each area contains a centre point and an actual offence location is assigned to the area that has the smallest distance between the location and centre points, following the idea of Voronoi tessellation. The temporal partition is provided on a monthly basis. Accordingly, each area may contain more than one crime occurrence at some time point.

The number of provided of areas is 8804 and the average smallest distance between any two centre points is 63 m ranging between 1 and 718 m. To reduce computational load, following the idea of Voronoi tessellation and finite element methods (Lindgren et al., 2011), we construct a mesh (illustrated in Figure 1), providing a triangulation of the bounded area based on the set of centre points. We use INLA R-package, <https://www.r-inla.org/> for computing the mesh. We specify the boundary as a union of the selected boroughs with minimal distance of 100 m between any two mesh points resulting in 4519 mesh points; on average the smallest distance between any two mesh points is 126 m. This level of accuracy permits fine-grained and detailed analysis. The areas have relatively high number of neighbours, on average 5.7, ranging between 2 and 10, capturing the geometry well, as illustrated in Figure 1. To summarise, data for each mesh point and time stamp consist of crime occurrences closest to any of the mesh points.

Let $w_{t,n}^{(m)} \in \{1, 2, \dots, |S|\}$, for $m=1, \dots, M$, $t=1, \dots, T$ and $n = 1, \dots, N_t^{(m)}$, denote the location indices of areas for crime occurrences, taking values over a spatial partition of size $D=|S|$ of a finite region S . Here, M denotes the number of crime categories, T denotes the number of time stamps and $N_t^{(m)}$ denotes the total number of occurrences for the m th category at time stamp t . Each location index is associated with coordinates of the mesh points of the areas. Altogether, the data contain $T=91$ time stamps, $M=5$ crime categories, $D=4519$ mesh points and $\sum_{n,t,m} w_{t,n}^{(m)} = 6.5 \times 10^5$ crime occurrences.

Figure 2 shows average empirical monthly crime distributions based on the occurrences for each category. Based on the figure, we see that crime is sparse and clusters into hot and cold spots, indicated

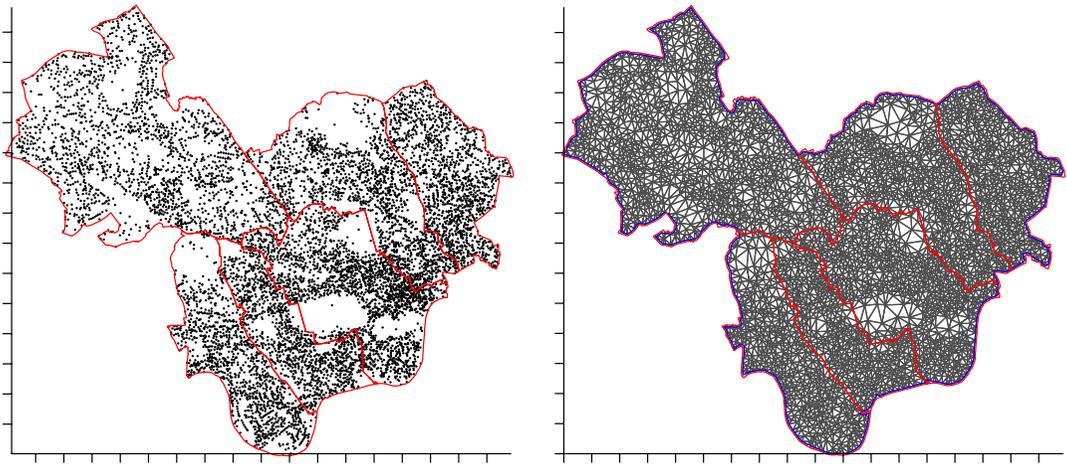


FIGURE 1 Set of centre points (left) and mesh graph with borough boundaries in red (right). The blue curve shows a non-trivial smoothed bound used for mesh construction. The nodes correspond to mesh centres and edges represent neighbourhood information. Large public parks, with little crime, are assigned a less dense mesh, whereas for relevant neighbourhoods with high crime rates the mesh is dense and roughly regular with small variation in area size. The study area spans roughly 18 times 13 km rectangle, as indexed by the tick marks

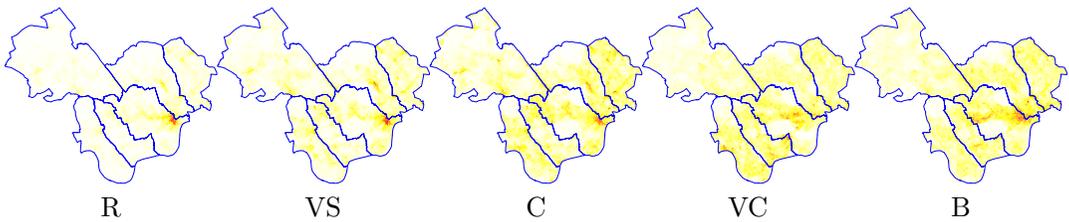


FIGURE 2 Empirical crime distributions for each category based on the total counts for robbery (R), violence and sexual offences (VS), criminal damage and arson (C), vehicle crime (VC) and burglary (B), respectively, from left to right, with the corresponding category abbreviations

TABLE 1 Data statistics, following the category abbreviations of Figure 2

	R	VS	C	VC	B
Total occurrences	53,514	266,767	76,321	136,308	117,925
Filling factor	0.85	0.89	0.88	0.89	0.88
Sparsity level	0.1	0.34	0.15	0.24	0.21
Mean and range	0.13 (0,18)	0.65 (0,30)	0.19 (0,25)	0.33 (0,25)	0.29 (0,21)
Gini index	0.84	0.6	0.75	0.65	0.68

by red and white areas respectively. All the distributions have long tails, as verified by large values for the monthly average Gini index based on the crime counts (Table 1), characteristic for power-law, Zipf and Pareto distributions, meaning that even though few hotspots account for most occurrences, there are many remaining areas that contribute significantly to the total. We note that for a visual illustration of the distributions only relative values are relevant instead of absolute values. Hence, the

colour bar is omitted from each such illustration. Table 1 shows the number of occurrences, filling factor (proportion of areas with non-zero occurrences over time), sparsity level (proportion of areas with non-zero occurrences over time stamps and areas) and average occurrences over areas and time stamps, including their range of values, for each category respectively. The table also verifies that crime is clustered but widely dispersed over the study region.

The targets or victims of crime vary depending on the category. Businesses and dwellings may be subject to robbery, burglary or property damage. On the other hand, persons may be subject to robbery, violence and sexual offences. Based on borough level crime occurrence aggregates¹, we are able to estimate that 39% of burglaries and only 6% of robberies target businesses respectively. Micro-scale areas may attract offences of mixed targets and categories. Based on the data, the average co-occurrence rate of pairs of different categories over areas and time stamps is 0.31, ranging from 0.14 between vehicle crime and robbery to 0.58 between robbery and violence, demonstrating substantial co-occurrence.

We collect spatial covariates using Census 2011 and Transport for London (TfL) accessibility index available from the London Data Store, <https://data.london.gov.uk>. We also collect points-of-interest (POI) data (EDINA Digimap Ordnance Survey Service, 2018). The accessibility index provides a fine-grained representation of how accessible each area is by public transport taking into account natural barriers and constraints affecting accessibility. We note that this covariate may substitute the use of actual road networks that are more complicated to work with. The Census provides a wide range of different socio-economic covariates, including more general covariates, such as, area, household and population size, at lower super output area (LSOA) or output area (OA) geographical resolution; these geographical partitions are also provided in the London Data Store. Our study area comprises of 773 LSOAs and 4314 OAs. Sections 2 and 3 of the online supplement of this paper show a subset of socio-economic and geographical covariates respectively. Altogether, the covariates consist of accessibility index, 107 socio-economic and 681 POI (this number includes groupings and categories of POIs, as provided by the system) variables.

We note that not all covariates are included in modelling because of increased computational load. Also some of the POI-based covariates are heavily collinear and hence it suffices to include only a subset of collinear covariates without (significant) loss of information. However, all of the covariates are used for model inspection, that is, for interpreting the components. We use area-weighted averaging to compute crime level at each LSOA for model inspection because most of the socio-economic covariates are available at the LSOA level and these covariates are naturally normalised in terms of social cohesion and population size. Similarly, we also average the LSOA/OA-based socio-economic covariates to the mesh areas for computing the model by constructing a covariate matrix \mathbf{X} of size $D \times C$, where C denotes the number of covariates. For the geographical covariates, we follow the idea of Voronoi tessellation and compute the number of relevant POIs at the mesh level directly and note that model inspection is performed also at the mesh level for the geographical covariates. The covariate matrix \mathbf{X} contains all the 107 socio-economic covariates, accessibility index and 4 geographical covariates, that combine POIs into retail, eating and drinking, transport and legal and financial groups, resulting in $C=112$ covariates.

3 | MODEL

A set of observed time-stamped and location-marked crime occurrences underlies analysis of temporal and spatial patterns of urban crime. The occurrences are often categorised by police/law to indicate

¹https://data.london.gov.uk/dataset/recorded_crime_summary#

the type of crime. Given a particular urban area, experts, such as, members of police or criminologists, may know which type of crime the area is associated with and to what extent at a particular time point. A city may contain a large set of such areas that may partially overlap with each other depending on the degree to which crime categories they are associated with and when. Experts may often expect that the areas and their associated categories change in time and that certain urban structure of the areas, encoded in spatial covariates, provides further information of the causes and nature of crime.

In this work, we present a model that is able to leverage such expert knowledge in a data-driven manner based on the crime occurrences. Following the notation set out in Section 2, the occurrences, that take discrete values over D areas, for category m at time stamp t , follow a multinomial distribution

$$\mathbf{w}_t^{(m)} = \{w_{t,1}^{(m)}, w_{t,2}^{(m)}, \dots, w_{t,N_t^{(m)}}^{(m)}\} \sim \text{Multinomial}\left(\sum_{k=1}^K \boldsymbol{\eta}_k \theta_{t,k}^{(m)}, N_t^{(m)}\right) \propto \prod_{n=1}^{N_t^{(m)}} p_{t,w_{t,n}^{(m)}}^{(m)}, \quad (1)$$

whose expectation parameter $\mathbf{p}_t^{(m)}$ is given by a decomposition of K components that include spatial crime distributions (*maps*) and *map* proportions, $\boldsymbol{\eta}_k$ and $\theta_{t,k}^{(m)}$, for $k=1, \dots, K$ respectively. The crime maps are distributions over D areas, whereas map proportions are distributions over the K maps, such that all elements are positive and sum to one. The crime maps are shared across crime categories but categories contribute differently via the proportions. Elements of $\boldsymbol{\eta}_k$ indicate which areas are relevant for the k th map. On the other hand, elements of $\theta_t^{(m)}$ indicate which maps are relevant for the m th category at t th time stamp. In other words, a map may contain one or more areas. Also, one or more crime maps may be responsible for the occurrences $\mathbf{w}_t^{(m)}$. The decomposition is essential to uncovering latent processes that explain criminal activity, as shown in the results (Section 4).

We reformulate (1) as a mixture model introducing map assignment variables $z_{t,n}^{(m)} \in \{1, 2, \dots, K\}$ for each occurrence. The assignment variable $z_{t,n}^{(m)}$ is generated from a categorical distribution

$$z_{t,n}^{(m)} \sim \text{Categorical}(\boldsymbol{\theta}_t^{(m)})$$

and the corresponding occurrence is generated from another categorical distribution

$$w_{t,n}^{(m)} \sim \text{Categorical}(\boldsymbol{\eta}_{z_{t,n}^{(m)}}), \quad (2)$$

conditioning on the assignment variable. This formulation highlights that the elements of the maps and proportions represent probabilities. The generative process is simple and intuitive and can be used to explain and simulate occurrences. Analytically marginalising out the crime map assignments, we get (1). The model formulation (1) emphasises the decomposition, whereas Equation (2) emphasises the mixture and leads to simplified posterior inference, as detailed in the following.

Applying spatial and temporal smoothing is key to capturing dependencies between close-by occurrences in both space and time. To this end, we use Gaussian Markov random fields (GMRF; Rue & Held, 2005) to construct $\boldsymbol{\eta}_k$, for $k=1, \dots, K$, explaining crime maps using spatial covariates, unstructured random effects and spatially structured random effects that capture deviation from the covariate or unstructured contribution. GMRFs are flexible, scalable and intuitive to specify. The construction of the crime maps is based on a so-called BYM model (Besag et al., 1991) and its extensions (Leroux et al., 2000; Riebler et al., 2016). To construct the maps, we introduce auxiliary variables that follow a variant of the BYM model and we further introduce non-trivial constraints for guaranteeing that the variables take values in a reasonable range, and that the posterior distributions for the corresponding hyperparameters are identifiable.

For the spatial partitions we assume two areas are neighbours if they are contiguous; Figure 1 illustrates the neighbourhood information. We introduce auxiliary Gaussian variables

$$\beta_k \sim \text{Normal}(\mathbf{X}\omega_k + \mu, \mathbf{Q}^{-1}),$$

for $k=1, \dots, K$, and assume $\eta_k \propto \exp(\beta_k)$, using element-wise exponentiation. We let spatial covariates \mathbf{X} to inform generation of the maps. Here, ω_k denote covariate regression coefficients that capture component-specific effects, μ is a mean parameter shared across the components, capturing overall crime rate for each area, and \mathbf{Q} is a precision (inverse covariance) matrix that captures both spatially structured and unstructured random effects via parameters λ and κ respectively. In more detail, elements of the precision matrix \mathbf{Q} for off-diagonal elements take value $-\lambda$ for any two neighbouring areas, otherwise they are zero, and the diagonal contains the total number of neighbours for each area times λ plus an additive constant κ . For numerical stability, we constrain $\beta_{k,1} \approx 0$ subtracting the contribution of $\mathbf{X}_{1,\cdot}\omega_k$ from the component-specific mean, setting $\mu_1 = 0$ and κ_1 to an arbitrarily large value, without loss of generality. The constraint is needed to ensure that the values remain in a computationally feasible range. We note that explicit conditioning of $\beta_{k,1} = 0$ provides an alternative approach but that is computationally more demanding because the sparsity structure of the corresponding precision matrix would be lost. For an increasing number of covariates C , under certain assumptions, the semi-parametric model becomes closer to a non-parametric approach based on Gaussian Processes (GPs). The key is to enforce sparsity for the regression coefficients to prevent over-fitting. For this reason, we follow so called automatic relevance determination (ARD) approach (MacKay, 1995; Neal, 2012) and assume $\omega_{k,c} \sim \text{Normal}(0, \delta_{k,c}^{-1})$ and $\delta_{k,c} \sim \text{Gamma}(a, b)$, for $k=1, \dots, K$ and $c=1, \dots, C$. For small values for a and b the distribution is sparsity-promoting. However, we allow the data to inform suitable amount of sparsity by inferring the parameters a and b based on the data. Finally, we assume $\mu_d \sim \text{Normal}(\mu_\mu, \tau_\mu^{-1})$, for $d=2, \dots, D$.

We take non-stationarity into account via time-varying proportions. For the temporal proportions $\theta_t^{(m)}$, for $t=1, \dots, T$, we also employ a flexible GMRF-based construction that is able to capture complex temporal dependencies. We introduce temporally dependent auxiliary variables

$$\alpha_{t,k}^{(m)} \sim \text{Normal}(\alpha_{t-1,k}^{(m)}, \tau_m^{-1}), \text{ for } t > 1, \text{ and, for } t = 1, \alpha_{1,k}^{(m)} \sim \text{Normal}(\mu_0, \tau_0^{-1}),$$

such that $\theta_{t,k}^{(m)} \propto \exp(\alpha_{t,k}^{(m)})$, for $k=1, \dots, K$. Here, τ_m , for $m=1, \dots, M$, denote category-specific parameters that affect smoothness of temporal variation. The corresponding $T \times T$ precision matrix for $\alpha_{\cdot,k}^{(m)}$ takes a simple tridiagonal form.

We assign weakly informative priors for the hyperparameters such that the model has a single parameter K to validate. For the hyperparameters λ, κ and τ_m , for $m=1, \dots, M$, we employ a two-parameter variant of generalised inverse Gaussian distribution,

$$x \sim \text{GIG}(\alpha_0, \beta_0) \propto x^{-(1+\alpha_0)} \exp\left(\left(\frac{1}{x} + x\right)\beta_0\right),$$

where $\alpha_0 \in \mathbb{R}$ and $\beta_0 > 0$. The distribution provides a convenient prior to constrain variables away from zero and infinity simultaneously; this is especially useful for the τ_m . We assume $\alpha_0 = \beta_0 = 10^{-3}$. For a, b and τ_μ we assume Gamma(ϵ, ϵ) distribution, for $\epsilon = 10^{-7}$, and set $\mu_0 = 0$ and $\tau_0 = 10$. For μ_μ we assume a normal distribution with mean zero and very large (infinite) variance. Section 1 of the online supplement of this paper summarises posterior inference.

3.1 | Related work

For $M=1$ and $K=1$ the model reduces to a stationary model, $\mathbf{p}_t^{(1)} = \boldsymbol{\eta}_1$, for $t=1, \dots, T$, and is related to (generative) crime mapping. Here, the benefit of the probabilistic approach is that the spatial map parameters affecting smoothness may be inferred based on the observed data instead of expensive cross-validation or manual tuning. Kernel density estimation, provides a non-parametric alternative for crime mapping (Bernasco et al., 2015; Chainey et al., 2008; Johnson & Bowers, 2004); here the goal is to estimate a kernel function that depends on locations and time stamps for computing a crime distribution for some crime category at particular time. The absence of proper probabilistic formulation complicates kernel parameter estimation and model selection and comparison. Flaxman et al. (2015) propose to employ Gaussian processes (GPs) that may be interpreted as a probabilistic variant of kernel methods. Flaxman et al. (2018) adopt the GP setting presenting a computationally more scalable linear approximation.

Based on the data distribution 1 and assuming different likelihood functions, the model may be interpreted as a form of low-rank or factor model (Buntine, 2002; Heller et al., 2008). We note that inference for the marginalised form is more demanding because $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ are coupled; for our model the updates are conditionally independent given the assignments. We also note that the model focuses directly on capturing the crime densities and does not explicitly take into account the number of occurrences as alternative formulations based on Poisson likelihood function for count data. Importantly, the Poisson-based models would be unable to fully leverage occurrence-specific information.

Alternatively, based on the mixture model formulation (2), our model is related to topic modelling (Blei et al., 2003). Blei and Lafferty (2006b) present a dynamic extension noting that Blei et al. (2003) are unable to take into account temporal dependencies. Blei and Lafferty (2006a) assume a general (unconstrained) covariance matrix that could in theory be used to capture spatial dependencies, as considered in this work. However, this formulation would not be suitable for our application because it does not impose spatial smoothness and the computational load would be very high. Our model is inspired by Teh et al. (2005); we present a spatio-temporal extension necessary for the application, as verified quantitatively in the experiments (Section 4). For these models the data often constitute of words appearing in text documents. Analogously, the crime occurrences would correspond to words and documents group words or crime occurrences according to time stamps. Our motivation for this class of models is based on the ability of the models to capture semantics based on word co-occurrences; so called topics capture together semantically similar words corresponding to certain themes. In our application, the topics would correspond to the crime maps. Finally, we note that the dynamic topic model assumes sequentially grouped documents and the model is not as such suitable for our application because the groups would contain only one document. Topic modelling has been applied for textual crime category descriptions (Kuang et al., 2017).

An alternative approach to factor or topic modelling would impose dependencies directly for the expectation parameters $\mathbf{p}_t^{(m)}$ working with Poisson likelihood (Diggle et al., 2013; Lindgren et al., 2011; Møller et al., 1998; Taddy, 2010). This approach may be computationally very demanding, difficult to capture associations between crime categories (Liu & Zhu, 2017; Quick et al., 2018), and less intuitive to interpret, because the model provides no dimension reduction or interpretable components. Here, $\mathbf{p}_t^{(m)}$ may include a linear mapping based on fixed spatial covariates and (un)structured random effects, potentially using GMRFs. We note that for our model the covariates are at a higher level in the hierarchy, resulting in a more weakly supervised formulation, where the spatial component plays a key role and it is not meaningful to look at the covariate effects separately from the (un)structured contributions. In the regression setting, covariates are at a lower level in a hierarchy, adding the amount of supervision; here, the spatial effect is mainly interpreted as noise that we are not

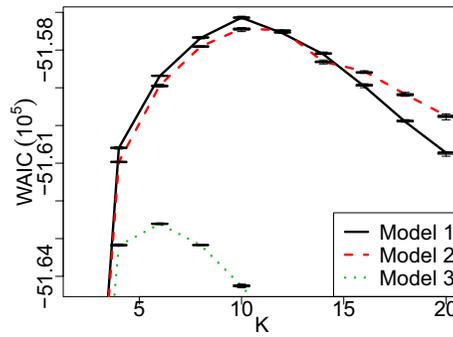


FIGURE 3 Model selection for the number of components; $K=10$ has maximal WAIC value. The performance is best for our model (Model 1)

particularly interested in. As such, extra care is needed for interpreting the covariate effects, possibly requiring feature selection to cope with multicollinearity. Also, these models may oversimplify the covariate contribution because they often capture global instead of local effects.

4 | ANALYSIS OF CRIMINAL ACTIVITY IN LONDON

4.1 | Model selection and comparison

The model contains one free parameter, the number of maps, K , that needs to be validated. Larger K implies a more complex model. In theory, the model is able to prune out irrelevant components but we prefer to train the model for a wide range of $K \in \{2, 4, \dots, 20\}$ for computational reasons. Our aim is to choose K such that the corresponding model explains the observations well without overfitting. For model selection and comparison, we adopt the Watanabe Information Criterion (WAIC; Gelman et al., 2013) that approximates model evidence, probability of the data given the model; we employ a formulation where higher values for WAIC are better, computing data log likelihood and subtracting the model complexity term. In more detail,

$$WAIC = \sum_{m,t} \left(\log \left(\frac{1}{S} \sum_{s=1}^S p \left(\mathbf{w}_t^{(m)} \mid \Omega^{(s)} \right) \right) - V_{s=1}^S \left[\log p \left(\mathbf{w}_t^{(m)} \mid \Omega^{(s)} \right) \right] \right),$$

where $\Omega^{(s)}$ denotes posterior samples², for $s=1, \dots, S$, and $V[\cdot]$ denotes sample variance. The model posterior distribution contains multiple modes and it is essential to explore local modes with high posterior probability density. To this end, we use 10 chains with well-dispersed different random initialisations. Because of multi-modality it is not appropriate to combine samples across multiple chains. To assess significance of the WAIC values, we also show box plots for WAIC values computed for a set of 10 replicates using partial re-sampling without substitution. We run the sampler for 2×10^5 iterations, that we find to suffice for convergence using standard diagnostics, collecting the last quarter of posterior samples to compute the WAIC score.

²When computing the likelihood we analytically marginalise out the assignments to increase robustness.

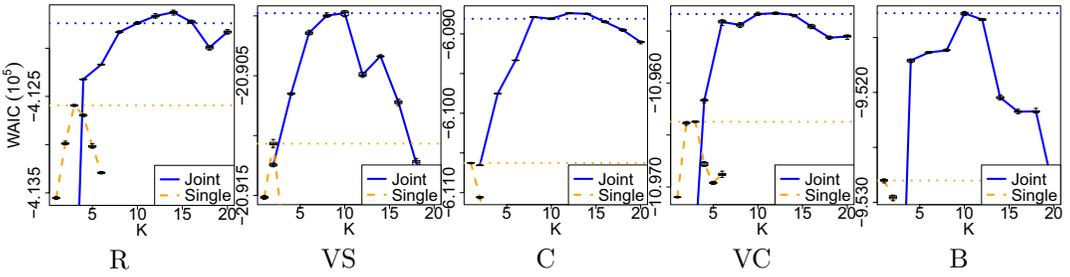


FIGURE 4 Comparison between joint and single-view models. Joint model provides best performance for $4 \leq K \leq 16$ for all categories

Figure 3 shows model selection results based on WAIC, the higher the value the better performance; the optimal number of components is $K=10$. We compare our approach, Model 1, to two related alternatives. First, we assess the utility of the covariates by comparing against a model that does not include covariates (referred to as, Model 2). Second, we compare against a standard model variant that discards spatial and temporal smoothing, assuming independent Dirichlet priors for the crime maps and proportions, referred to as, Model 3 (Teh et al., 2005). Based on the figure, we see that the covariates contain useful descriptive power and that smoothing is essential. Our model has best performance for $4 \leq K \leq 10$; the difference to Model 2 is statistically significant for this range of K (Wilcoxon one-sided test, $p < 10^{-3}$).

We also compare quantitatively our multi-view model to single-view models for each crime category. Each single-view model is a special of our model for $M=1$ using data for each category. For the single-view models, we followed a similar process to choose K , although, limiting to $K \in \{1, 2, \dots, 6\}$. Figure 4 shows that the WAIC score is best for our joint/multi-view model for all categories for $4 \leq K \leq 16$ (the results are statistically significant for $K \in \{4, 6\}$; one-sided Wilcoxon, $p < 10^{-3}$). Performance for the single-view models is optimal for $K \in \{1, 2, 3\}$ and performance decreases for $K \geq 4$ for each category. For the joint model we use only data related for each crime category when computing the score for each category. The dotted horizontal lines indicate best model for single-view models and for our model for $K=10$. The results verify that joint modelling is able to infer better category-specific models. For criminal damage and arson, and burglary the best performance is for $K=1$; the model reduces to a more simple statistical crime mapping approach that is unable to take dynamics or decomposition into account. For these categories, there are not enough data to support more complex models; for example, for burglary the single-view model is unable to separate commercial and residential burglaries into separate components. High dimensionality and sparsity of the data pose significant problems and it is very difficult to capture descriptive models without leveraging other related crime occurrences. For the remaining crime categories performance for single-view models is best for $K > 1$ showing that mixed membership modelling performs better than standard crime mapping. For our model, performance varies smoothly as a function of K , and is roughly maximal for $K=10$ for all categories and adding more components does not improve performance for any category.

4.2 | Model interpretation

We compute posterior averages for the maps and temporal patterns using the same set of posterior samples used to compute the WAIC scores for our model for the component number ($K=10$) with the maximal WAIC score. We are aware of the label switching problem inherent for mixture modelling

TABLE 2 Component activations \mathbf{A} for each category and component

	1	2	3	4	5	6	7	8	9	10
B	0.41	0.27	0.19		0.03	0.01			0.01	0.06
VC	0.01	0.15	0.09						0.34	0.41
C	0.03	0.06	0.04	0.11	0.14	0.45	0.06		0.04	0.07
VS			0.03	0.41	0.17	0.22	0.12	0.04		
R	0.04	0.05	0.03	0.01	0.02	0.06	0.38	0.39		0.02

but we found this not to be a problem for our application by tracking cross-component similarities between consecutive posterior samples for the maps.

Concentration of probability mass of the posterior maps η_k , for $k=1, \dots, K$, reveal areas that have high crime rates (i.e. crime hotspots). The maps may be naturally visualised to uncover the hotspots with high values. The corresponding posterior proportions $\theta_{t,k}^{(m)}$, for $t=1, \dots, T$ and $m=1, \dots, M$, indicate temporal patterns, uncovering which crime maps out of K are relevant (or active) for each crime category and revealing when and how much the maps contribute. Some of the shared maps may be active for a subset of two or more categories while category-specific maps are active only for a single category. Visual inspection of the $\theta_{t,k}^{(m)}$, for $t=1, \dots, T$, across multiple categories $m=1, \dots, M$ summarise the temporal patterns and relevance for each category, for the k th component.

To summarise the category-component activations, we average the proportions over time stamps $A_{m,k} = \frac{1}{T} \sum_{t=1}^T \theta_{t,k}^{(m)}$ for each category and component. Each row of the matrix sums to one. Table 2 illustrates the activations, matrix \mathbf{A} , showing subsets of components that are active for each crime category. Based on the activations, rows of \mathbf{A} , we see that the first 3 (of 10) components are clearly active for burglary, the next three components are active for both criminal damage and violent crime (4–6), the next two for robbery (7–8) and the last two for vehicle crime (9–10). We label the components accordingly and also use the bold symbol in Table 2 for visual clarity to illustrate the labelling. We note that, even though the number of crime categories is five, a relatively large number of components is needed to correctly decompose variation into components that are active for any subset of two or more crime categories and components that are mostly active for only a single category, explaining category-specific (marginal) variation. By inspecting the columns of \mathbf{A} , we see which of the categories are active for each component.

To associate relevant and significant covariates for the crime maps we compute pairwise correlations in log-domain between (posterior average) maps and a large set of (773) covariates. Here, the maps may be directly interpreted as expected crime counts per area. Figures 1 and 2 of the online supplement of this paper show correlations for socio-economic and a subset of place-based variables respectively. We carry out permutation testing to assess statistical significance of the pairwise correlations (all of the correlations are statistically significant, $p < 10^{-3}$).

In the following, for each component, we show temporal patterns (proportions) over all crime categories, $\theta_{t,k}^{(m)}$, for $m=1, \dots, M$ and $t=1, \dots, T$, and associated spatial maps η_k , for $k=1, \dots, K$. To save space we shorthand the crime categories: robbery (R), burglary (B), violence and sexual offences (VS), criminal damage and arson, (C) and vehicle crime (VC). For clarity, we order the proportions across categories in decreasing order of the activations.

Burglary is decomposed into three main components (Figure 5) capturing commercial burglary and residential burglary for houses and flats, respectively, as detailed in the following. Figure 5 (top) illustrates a component that captures commercial (non-residential) burglary. The component is active for robbery and criminal damage but inactive for vehicle and violent crime. The spatial map

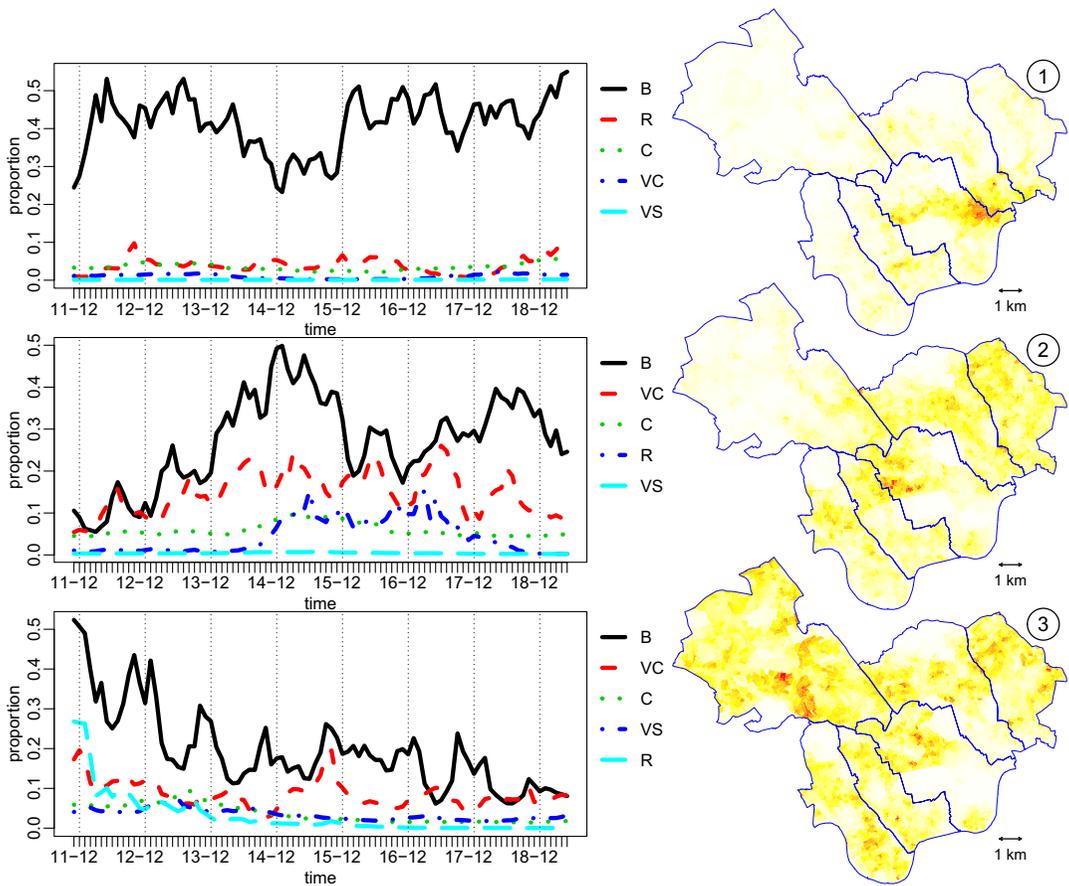


FIGURE 5 Burglary is decomposed into three main components (labels 1–3): (top) captures commercial, (middle) residential flat and (bottom) residential house burglary respectively. Residential burglary is associated with vehicle crime. The maps capture distinct hotspots with little overlap or similarity between them

shows crime concentrated in central commercial and affluent areas and high streets, including; Covent Garden, Soho, Oxford Street and Brunswick shopping centre. The geographical covariates verify that the map is highly associated with commercial activity in central areas with good transport accessibility; proximity to transport hubs, such as, train stations and main roads. Here, the opportunity for crime in these less residential and well accessible areas is very high, increasing criminal attractiveness. The temporal pattern shows mostly constant activity with small monthly variation without evident temporal trends. Commercial burglary accounts for 41% of all burglaries; this number is close to the empirical estimate (39%). Accordingly, the empirical distribution of burglary in Figure 2 strongly resembles map 1 meaning that the hotspot of commercial burglary dominates inferences based on the empirical distribution, failing to discover hotspots related to house or flat burglary. However, in the following, we show that our model is able to also focus on the long tails of the empirical distributions via the decomposition, providing new meaningful insights.

Figure 5 (middle) shows a component related to residential flat (non-house) burglary in affluent central areas. The temporal patterns (left) show that the component is active also for vehicle crime, criminal damage and robbery but inactive for violence. For vehicle crime there is an evident trend peaking at summer time; burglary exhibits a similar but less clear trend with large and frequent

fluctuations. Robbery is active only for a certain time frame. The covariates show associations with high education, ethnicity, low poverty/unemployment/deprivation, couples without children, single-person households, private renting and number of flats. These associations suggest that the component is related to so called 'young professionals' in the age range of 30–44 with high education and occupation in information and communication and professional, scientific and technical activities, and finance, insurance and real estate activities, and live in flats. The place-based covariates show minor associations, indicating that these areas are residential in nature.

Figure 5 (bottom) shows the third component that captures residential house burglary in semi-affluent less central (non-commercial) areas. The temporal patterns show a clear trend, the component is more active in winter time when the amount of daylight is low, suggesting that offenders may target low-density residential areas (i.e. houses), when residents are away from home at work during day-time with less day-light. Also, the component is active for vehicle crime and other crime categories to a lesser extent, including violence, as opposed to the two other burglary components. The covariates show the component is associated with the number of houses (equivalently, low number of flats), private transport (poor transport accessibility), (married) couples with dependent children, house ownership (with mortgage or loan), ethnicity, benefit claiming and provision of unpaid care. Population wise the component is associated with age levels 65+, 45–64 and 0–15, suggesting that households constitute of older couples with children (i.e. families). Occupation focuses on a wide range of different work types including mining, quarrying and construction, manufacturing, wholesale and retail trade, administrative and support service activities, transport and storage, human health and social work. Also education levels (2, 1 and apprenticeship) support these occupations. High level of private transport suggest that (i) work places may be situated further away from home, (ii) the need to take children to school or day-care and (iii) poor public transport, for instance. Furthermore, the component is associated with lower household income and house price, indicating opportunities with lower risk. The map captures criminally attractive areas which are accessible, especially at specific times of the year. The POI covariates show evident suburban residential features; positive covariate effects include community centres, playgrounds, schools and care homes with no effects for eating and drinking facilities, banks or 'high streets', for instance.

Figure 6 captures three components mainly associated with violence and sexual offences, and criminal damage and arson, and robbery to a smaller extent with distinct temporal and spatial patterns. We note that, non-violent crime such as burglary and vehicle crime are mostly inactive for these components, promoting a distinction between violent and non-violent areas. Figure 6 (top) captures mostly violent criminal activity in a small number of crime hotspots in central areas, for example, in Covent Garden, Camden and North Kensington. Figure 6 (middle) shows a component with a rapid increasing shift for both violent crime and criminal damage for a larger number of hotspots, after around 2014 December, showing that crime becomes more concentrated in these hotspots and explaining the declining trends for the other two components (top and bottom), respectively. Figure 6 (bottom) shows a component that is mostly active for criminal damage but is also active for robbery for a large number of isolated small hotspots (in deprived less central areas). Again, co-occurring robbery shows temporal fluctuations with peaks in winter, whereas violence peaks at summer. Maps for 4 and 5 share two close-by hotspots (in Covent Garden and Camden), whereas 5 has a larger number of hotspots, in general. For 6 the hotspots are more distinct. The covariates show that these components are jointly strongly associated with indicators related to deprivation and poverty, including, ethnicity, lone parent (and divorced) households (and low number of couple households), ill health, unemployment, low or no education, social renting, benefit claiming, crowding, low amount of residence ownership (low number of houses), high proportion of students and young population between 16 and 29. Occupations concentrate in accommodation and food services and transport and storage. POI-related covariates

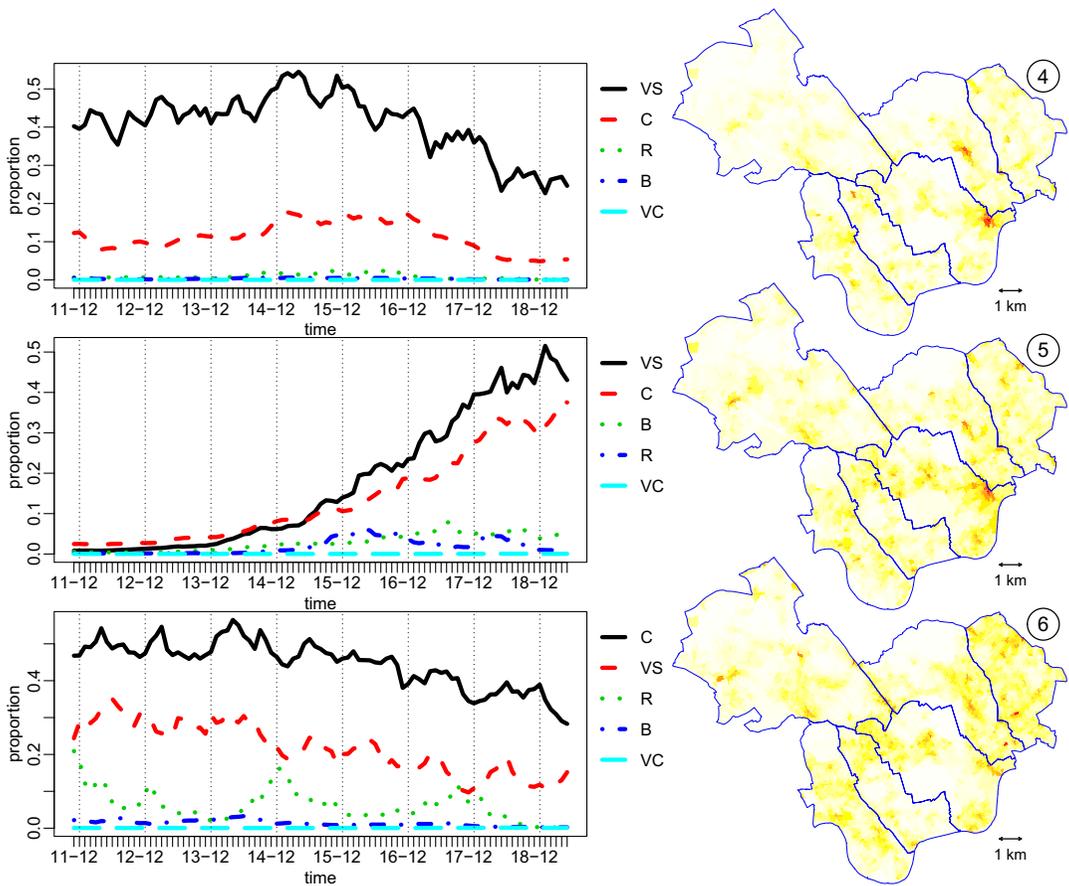


FIGURE 6 Violent and sexual crime, and criminal damage and arson are decomposed into three main components (labels 4–6). The temporal patterns show interesting trends and maps show a large number of small isolated hotspots in deprived less central neighbourhoods

show retail and commercial activity (similarly to commercial burglary showing association with local high streets and ambient population) but also show positive effects for more specific facilities, such as, bus stops, playgrounds, community and sport centres, schools, police stations, social care activities, libraries and hospitals, with smaller effects for attractions. Comparatively, the covariate effect size is stronger for component 6 (that focuses on criminal damage) than for the components 4–5 that show similar effect sizes between each other. The specific facilities provide targets for criminal damage, whereas bus stops promote street-level crime. When combined with large youth concentration, the components (especially 6) may capture vandalism (gang graffiti, for instance) and intra-youth violence and robberies involving knives and targets with an increasing number of smartphones. According to crime pattern theory, young offenders would operate, for instance, near schools and related (public) transport routes from and to home. Our list of the specific POIs strongly associate with routes and places of youths. The hotspots for 4–6 are temporally stable; tensions between local criminal gangs, that aim to criminalise suitable young individuals, may explain this property to some extent. We note that the hotspots around Covent Garden for maps 4 and 5 are close to but do not overlap with that of commercial burglary (1). On a larger scale, we also note that the socio-economic covariate effects differ significantly between 4–6 and 1–3, for burglary, further evidencing the difference between the

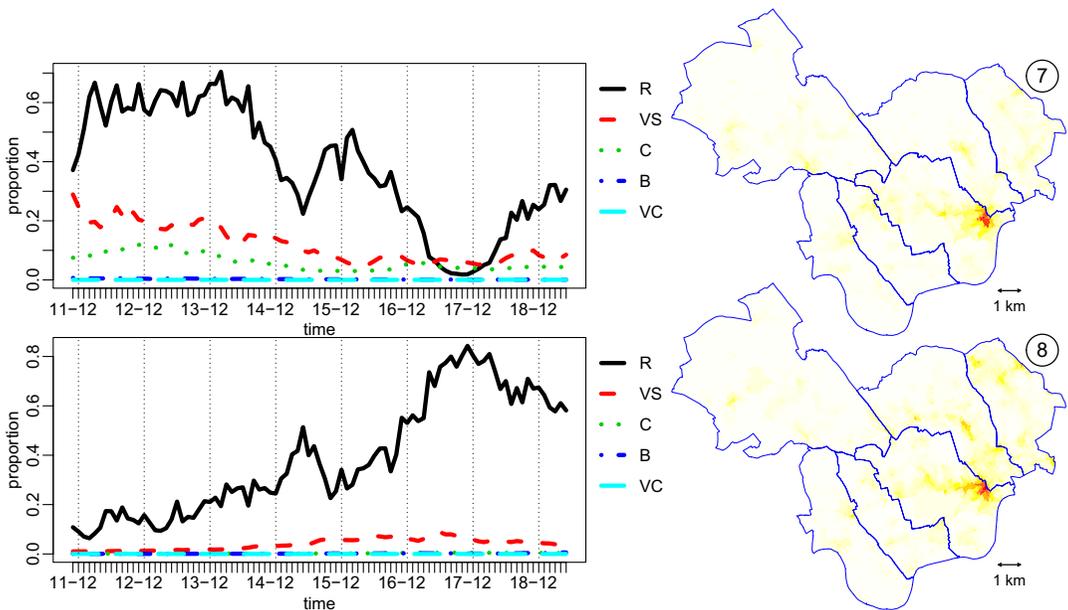


FIGURE 7 Robbery is decomposed into two main components (labels 7–8): robbery in Soho and central areas (top) and robbery near main high streets; Oxford and Piccadilly streets (bottom). The maps focus on single large connected hotspots that are close to each other. The temporal patterns show fluctuations between these two components

near-by hotspots. Although burglary is slightly active for 5–6, but given the high concentration (small total area of hotspots), the association becomes relatively significant, explaining residential burglary in deprived areas targeting young, unemployed and single household occupiers. Overall, criminal damage is active for almost all components thus bearing a strong association to crime in general and may be interpreted as a measure of physical disorder. Contrasting with the empirical distributions (Figure 2), we see that map 4 resembles the empirical distribution for violence, which fails to uncover the shift in the hotspots covered in map 5. Similarly the empirical distribution for criminal damage overly emphasises the hotspot in the city centre; our map 6 is able to better focus on the set of smaller hotspots.

Figure 7 shows two components related to robbery in central highly accessible areas, focusing mostly on single large hotspots and capturing 80% of total robberies. These hotspots are active for violence but not for burglary or vehicle crime. The trends show a fluctuation between the components indicating shifts between the near-by (but different) hotspots. Component 7 is also active for criminal damage. POI-based covariates for these maps focus on commercial and retail activity; areas, where targets with cash or other valuables are plenty. Especially, alcohol-related premises, nightclubs, theatres and concert halls, attractions, currency conversion, banks, adult venues, gambling, underground entrances and stations, cinemas, jewellery accessories, sports facilities, casinos, arcades, subways, bus stops, chemists, petrol stations, cheque cashing and department stores are associated with these components. The covariates link strongly to ambient population both during daytime (tourist attractions) and nighttime (adult venues and nightclubs) and also provide plenty of targets for business robbery. For instance, underground stations represent places with a constant flow of people providing plenty of targets for personal robbery. On the other hand, theatres and concert halls attract large numbers of people but less frequently. Socio-economic covariates show that component 7 is similar to components 4–6, promoting young population, unemployment, ethnicity, crowding, deprivation and poverty. On

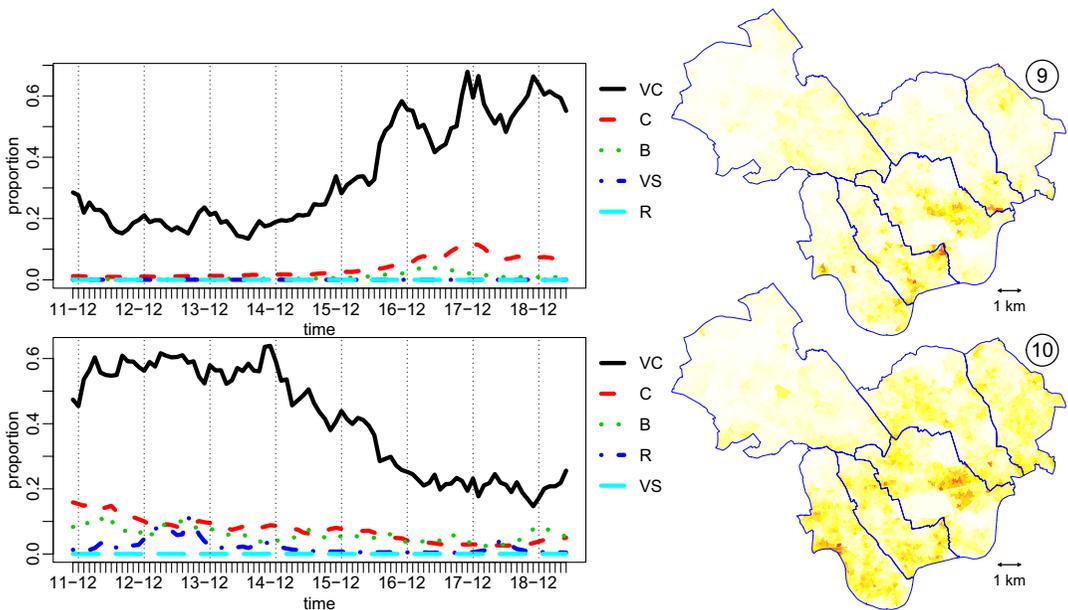


FIGURE 8 Vehicle crime is decomposed into two main components (labels 9–10), showing vehicle crime in commercial central areas and vehicle crime near residential areas and work places respectively

the other hand, component 8 distinguishes from 7 by private renting, lower level of deprivation, better transport accessibility, higher education (and occupation), less poverty, more single-person households and less unemployment; the component shares some aspects to components 2 and 1 for flat and commercial burglary respectively. The absence of criminal damage for 8, possibly due to the presence of a ‘guardian’ for property damage, suggests a difference in the level of physical disorder between the hotspots. The related violent crime may link to common assaults driven by alcohol consumption. Similarly to criminal damage, overall robbery is active for almost all components to varying degrees, suggesting another link to the concept of disorder to explain crime in general. Intricate temporal proportions show that hotspots for robbery are highly dynamic; they shift focus or emerge and then disappear. We note that it is difficult to distil apart business and personal robberies because business robbery is infrequent and may often occur in similar locations as personal robbery. The empirical distribution (Figure 2) for robbery provides trivial findings, whereas our model is able to provide finer dynamic details by exploring the long tails properly.

Figure 8 shows two components that capture mostly vehicle crime (75% of total) in non-violent hotspots. Both components are also active for criminal damage. In addition, component 10 is also active for burglary, and for robbery during a particular time interval. The components capture a change point around 2015; the first is mainly active after 2015, whereas the second before respectively. Figure 8 (top) shows a component that captures few hotspots in wealthy (commercial and central) areas, such as, Knightsbridge and Chelsea, for example. The component exhibits, similarly to house burglary, an evident winter trend after 2015, suggesting that little daylight may attract criminal activity. Figure 8 (bottom) shows a component that is spatially more dispersed (with larger hotspots) in affluent residential and work place areas, such as, Mayfair, Marylebone and in Fulham near the river. The place-based covariates for both components 9–10 associate with central commercial and retail centres (‘high streets’), and resemble commercial burglary and robbery components, 1 and 7–8

respectively. Component 9 associates additionally with car parks. Although, overall the associations are lower in magnitude possibly because of large spatial coverage. However, there are no effects for underground entrances and stations, playgrounds and department stores, and relatively smaller effects for bus stops, logically distinguishing from public transport. The socio-economic covariates for both components 9–10 evidence affluence of these areas; no deprivation, terraced houses, high education, couple households and high employment, house price and household income. They have similar covariate effects for commercial and flat burglary, components 1 and 2, respectively, showing vehicle activity due to journey origins and destinations, that is, residences, work places and high streets, for example. Despite these similarities for the effects, the spatial maps show distinct hotspots. Given that (residential) burglary is more active for component 10, we suggest that these hotspots explain more vehicle crime near residences (terraced houses require street parking, increasing accessibility). On the other hand, the proximity of hotspots of map 9 to commercial centres and car parks (providing plenty of targets), suggest vehicle crime related to commerce or commuting traffic. Finally, we suggest that the low amount of burglary in the hotspots of map 10 may be due to high risk in these very affluent areas; vehicles are more accessible targets in this case. Similarly to component 10, flat and house burglary components are also active for vehicle crime possibly because of vehicle crime on streets close to residences. The empirical distribution (Figure 2) for vehicle crime is more dispersed in the absence of few dominating hotspots. Our model is able to decompose the variation into several maps in a meaningful manner providing useful insights.

5 | DISCUSSION

We develop and apply useful and important improvements for the ubiquitous crime mapping approach that is currently widely in use both in academic research and in practice, highlighting that our work has high impact for research and society. Our model builds on the hotspot property of crime, near-repeat pattern theory and regression modelling (environmental criminology). The main novelty of the method and analysis is the decomposition of criminal activity into a weighted linear mixture of components. We anticipate that our findings based on London would be transferable to other major cities, noting that the developed methodology as such is directly applicable for other cities or study areas.

For the analysis of more than one crime categories, a common approach is to first analyse or model each category separately and then combine the analyses to capture inter-category associations (Bowers, 2014; Haberman, 2017; Hipp, 2007; Weisburd et al., 1992). As opposed to such two-step analysis, we jointly model crime occurrences over several categories at micro-scale uncovering associations between different types of crime. We show that the joint approach has better performance than following the dominant practice of analysing each crime category separately. Category-specific models are severely biased by the long tail (hotspot) property focusing only on the few most relevant hotspots at the cost of neglecting the long tails. However, the joint model combines crime occurrences over different types of crime, and is able to focus better on the tails. Weisburd et al. (1992) find that (i) property damage, auto theft, burglary, personal robbery and violent crime, (ii) commercial burglary and business robbery, (iii) personal robbery, auto theft and business robbery and (iv) violent crime, business robbery and residential burglary are associated with each other respectively. Overall, we verify most of these findings, noting that in our application we do not have direct access to more fine-grained crime categories that would directly separate, for example, different types of burglary from each other. Instead, we infer the finer sub-categories based on the data. In other words, the model is able to account for the information loss due to crime categorisation.

Our model uses an extensive set of both socio-economic and geographical spatial covariates to individually inform the crime distribution (map) of each component, taking locality naturally into account. We verify that the covariates contain useful information and a model that incorporates the covariates results in better performance. We infer shared and specific effects of covariates and components, providing further insights into dependencies between different types of crime. We note that there is no single covariate that explains crime distributions; rather sets of covariates are needed thus focusing on interactions.

For model inspection, we suggest to compute massive univariate pairwise correlations between maps and covariates to detect (potentially collinear and interacting) covariates that are positively or negatively associated for each component. Following regression modelling, we could alternatively inspect the actual covariate weights, but this would omit the contribution of the Gaussian field and mean, further, limiting the number of covariates. We show how a large set of associated covariates provide meaningful insights to the nature of the relationship that would be impossible to uncover based on a small set of covariates. Further sparsity-promoting approaches may prefer to choose one of many collinear covariates biasing analyses based on inspecting the regression weights.

Our model captures inherent spatio-temporal crime patterns via the decomposition. The model is directly relevant for crime hotspot prediction at micro places (Caplan et al., 2011; Chainey et al., 2008; Mohler et al., 2011) and guiding police resource allocation targeting the questions of what type of crime and where and when crime will occur (Perry, 2013). Our crime maps indicate a set of different hotspots each with its specific dynamics (or temporal trends), associations with any subset of crime categories and covariate effects. Such flexibility may be very useful for accurate forecasting and proactive resource allocation for specific type of crime. Our model is able to go beyond the most evident hotspots and explore the long tails of crime distributions. Future work is needed to evaluate the model for the task of crime hotspot prediction for both short- and long-term time frames and for different study areas. In this work, we focus on analysing and interpreting the findings for criminal activity in London and note that a full scale prediction evaluation is out of scope of this work.

In this work, we have adopted Bayesian model selection to choose the number of components in the decomposition. In particular, the adopted WAIC provides an approximation to model evidence and importantly is closely related to model predictive performance. Cross-validation would be a more computationally intense alternative that discards some of the data for inference. However, this may lead to a model selection bias; smaller data may support a simpler model. Furthermore, temporal dependence poses further bias; data for sequential time stamps may not satisfy the iid assumption. Future work is needed to evaluate the model for the task of crime forecasting; here model complexity may be chosen to optimise performance with respect to crime-specific evaluation measures (Adepeju et al., 2016; Chainey et al., 2008).

Crime is widely acknowledged to be highly clustered both spatially and temporally. Our results support this finding; the spatial maps are highly clustered and the temporal patterns show frequent and significant fluctuations. Johnson and Bowers (2004) and Bernasco et al. (2015) propose that crime follows so called repeat or near repeat victimisation pattern theories; frequently targeted areas attract more crime especially within short time intervals after recent criminal activity. This theory may partly explain the fluctuations or sudden bursts for the temporal patterns as well as the evident clustering property of the maps. The theory may further lead to power law crime distributions and thus relate to more abstract and widely applicable theories of least effort and preferential attachment or rich-get-richer property, for instance. Our model provides a data-driven Bayesian alternative for self-exciting process based models, that build on this theory, (Liu & Brown, 2003; Møller & Rasmussen, 2005; Mohler, 2013, 2014; Mohler et al., 2011; Taddy, 2010) without the need to specify particular decay functions. Instead, the dynamics of the components of our model are inferred based on the data using a flexible model formulation.

Our approach is related to the regression model by Cohen et al. (2007) that uses other types of crime as explanatory variables. The main difference is that we formulate a joint distribution for different types of crime, whereas in regression conditional distributions are used. Joint modelling provides more flexibility; often it is not evident which crime type to predict and which to use as explanatory variables. Also, in a similar context, our model also provides a principled alternative for local area regression modelling (Cahill & Mulligan, 2007) as opposed more traditional global modelling (Caplan et al., 2011). For our model the specific areas are inferred based on the data instead of manually choosing (several) small scale areas to model individually (Barnum et al., 2017; Drawve & Barnum, 2018; Piza et al., 2017). Often the hotspots are not known in advance.

Social disorganisation theory proposes that social (or group) coherence is related to explaining crime (Andresen, 2010; Shaw & McKay, 1942). Socially less coherent areas may experience more crime compared to more coherent areas. Here, higher coherence implies that individuals participate implicitly or explicitly in preventing crime. In general, we note that our list of crime-related covariates is strongly related to social cohesion (Sampson & Groves, 1989). For instance, a large proportion of young population is related to high population turn-over and low social coherence. Schreck et al. (2009) and Chamberlain and Hipp (2015) find that the theory applies to violent crime. Our findings agree with these findings and show how a large young population is a significant covariate for explaining crime in deprived and central areas.

Recent studies agree that areas close to residences of offenders may have higher risk of burglary (Bernasco & Nieuwebeerta, 2004; Clare et al., 2009; Menting et al., 2019; Townsley et al., 2015). Menting et al. (2019) find that, in general, the risk also increases for areas the offenders become familiar with. It is also widely assumed that most offenders live in deprived areas. Crime pattern theory suggests that areas close to deprived areas (an approximation for offender residence and operating environment) have high risk of burglary (and crime in general) (Andresen, 2010; Brantingham & Brantingham, 1993b). Our findings tentatively agree with these works; some offenders operate directly in the deprived areas or in the nearby less deprived and more attractive areas with higher reward but suitable risk. The hotspots for violence and criminal damage are connected to deprivation and thus offenders' residences, based on the theory. We show that the hotspots are nearby some of the hotspots for house and flat burglary and vehicle crime. However, future work is needed to study this connection further.

Proximity to transportation hubs, such as, train stations, has shown to increase the risk of burglary (Bernasco et al., 2015; Clare et al., 2009). Beavon et al. (1994) and Johnson and Bowers (2010) find that overall accessibility (or exposure) of an area has a positive effect on the crime rate. Also, Brantingham and Brantingham (1975) show that areas with decreased exposure (e.g. interior of a residential area) are less likely targets for burglary. Our findings agree with these studies. We also show that these effects are relevant for explaining commercial burglary.

High ethnic diversity of an area has been shown to explain burglary (Bernasco & Luykx, 2003; Bernasco & Nieuwebeerta, 2004; Clare et al., 2009; Sampson & Groves, 1989). Our findings strongly agree with these works, showing in more detail that certain ethnicities are *positively* or *negatively* associated with crime depending on location context.

House price and household income naturally reflect attractiveness and reward for residential burglary. Existing findings are, however, inconclusive: Bernasco and Luykx (2003) and Townsley et al. (2015) find either or both of these effects to be positively related to crime, whereas Bernasco and Nieuwebeerta (2004) and Clare et al. (2009) find these effects to be irrelevant. Our findings shed light to these mixed findings showing that affluence is locally relevant. We find that in residential areas house price and household income are *negatively* related to burglary, whereas, in more central areas, the effect is *positive*. However, affluence alone is not sufficient to explain crime; several very wealthy

areas such as Marylebone, Mayfair and Kensington and Chelsea are missing from our crime maps, suggesting that these areas deter burglars potentially because of high failure rate and risk despite high reward. Townsley et al. (2016) make similar findings based on offender reports.

Different occupations are related to socio-economic variability. Sampson et al. (1997) and Johnson and Summers (2015) find such socio-economic variability to be relevant for explaining crime (although not burglary). Our findings show that this hypothesis is also relevant for burglary. For example, we show positive effect for occupations in transport and storage, accommodation and food service, health and social work and education activities as well as high proportion of students; however, the effect is negative for finance, insurance and real estate activities.

Bernasco (2014); Sampson et al. (1997) and Andresen (2010) find that lone parents and single and young households are positively associated with crime. Our findings agree with these works, further showing that they are also relevant for burglary.

Hipp and Yates (2011) show poverty increases risk of crime. Especially, when combined with the presence of nearby alcohol-providing premises (Stucky & Ottensmann, 2009; Wheeler, 2019). Our results partly support these findings, noting that the concentration of pubs, bars and inns collates with the presence of restaurants, cafes and fast-food and takeaway outlets that are hence also associated with crime (Groff & Lockwood, 2014).

We note that the mesh-based approach for partitioning the study area has some advantages over grid or LSOA/OA-based alternatives. For the grid-based approach the number of grids grows quickly for smaller grid areas and the number of neighbours is low, wasting computational resources. Furthermore, gridding may suffer from boundary effects. On the other hand, even though LSOA/OA-based areas capture by design socio-economically similar areas of roughly same population and household sizes, the geometry of these areas does not take into account actual crime data. The mesh-based approach concentrates the mesh points for areas with high crime counts permitting fine detail where necessary to improve accuracy and explaining low crime areas in a computationally scalable manner. In addition, we can specify complex boundaries (based on natural barriers or holes, e.g.) for the study area, meaning that all the inferences are strictly based on the study region, as opposed to the grid-based approach, which becomes more accurate when the grid sizes approach zero posing significant computational issues.

Tompson et al. (2015) study the same data with varying levels of geographical resolution separately for each crime category raising issues for data analysis due to the data anonymisation process. In this work, our findings based on category-specific models at micro-level agree with these concerns. However, we show that joint modelling of crime data over several categories at micro-level produces meaningful findings and performs better than category-specific approach.

The dynamics of urban structure present challenges for data quality control. The changes should affect the design of centre points. For example, the recent development of King's Cross area is largely underrepresented, requiring update of the data base. The dynamics also affect socio-economic and geographical covariates.

We work with point-of-interest data at micro-level based on POI centre points. As shown in this work, the approach is successful when combined with smoothing for the spatial maps but the approach may not be appropriate for large-sized POIs. For instance stadium or department store centre points may be a poor representation. This connects to a more general question of which geographical partition (or proximity) is useful for certain place based covariates.

While the spatial resolution of the data studied can provide very fine details, the temporal resolution is more coarse. In the future, it would be interesting to apply the developed framework for more fine-grained temporal resolution consisting, for example, of weekly or daily crime occurrences. We expect such data would support more complex temporal patterns (and correspondingly different spatial maps) and higher model complexity, the number of effective components.

Our method supports modelling of individual offences providing a flexible approach for including occurrence-specific information, for instance, about the offender and/or target/victim and their behaviour, or crime records by police, or crime severity capturing money loss or level of injury, or the presence of alcohol or drugs. Furthermore, adding more data in the analysis may provide further detailed findings providing novel insights and bridging crime mapping and victimisation/offender research. In addition, labelling each offence as either based on planning or opportunity would give further important insight.

ACKNOWLEDGEMENTS

The authors were supported by the EPSRC grant EP/P020720/1, Inference Computation and Numerics for Insights into Cities (ICONIC), <https://iconicmath.org/>.

REFERENCES

- Adepeju, M., Rosser, G. & Cheng, T. (2016) Novel evaluation metrics for sparse spatio-temporal point process hotspot predictions—a crime case study. *International Journal of Geographical Information Science*, 30(11), 2133–2154.
- Airoldi, E.M., Blei, D., Erosheva, E.A. & Fienberg, S.E. (2014) *Handbook of mixed membership models and their applications*. Boca Raton: CRC Press.
- Andresen, M.A. (2010) The place of environmental criminology within criminological thought. In: Andresen, M.A., Brantingham, P.J. & Kinney, J.B. (Eds.) *Classics in Environmental Criminology*. Boca Raton: CRC Press, pp. 21–44.
- Andresen, M.A. & Malleson, N. (2011) Testing the stability of crime patterns: Implications for theory and policy. *Journal of Research in Crime and Delinquency*, 48(1), 58–82.
- Barnum, J.D., Caplan, J.M., Kennedy, L.W. & Piza, E.L. (2017) The crime kaleidoscope: A cross-jurisdictional analysis of place features and crime in three urban environments. *Applied Geography*, 79, 203–211.
- 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*. New York: Springer, pp. 4381–4391.
- Bernasco, W. & Block, R. (2011) Robberies in Chicago: A block-level analysis of the influence of crime generators, crime attractors, and offender anchor points. *Journal of Research in Crime and Delinquency*, 48(1), 33–57.
- Bernasco, W. & Luyckx, F. (2003) Effects of attractiveness, opportunity and accessibility to burglars on residential burglary rates of urban neighborhoods. *Criminology*, 41(3), 981–1002.
- Bernasco, W. & Nieuwebeerta, P. (2004) How do residential burglars select target areas? A new approach to the analysis of criminal location choice. *British Journal of Criminology*, 45(3), 296–315.
- Bernasco, W., Johnson, S.D. & Ruiter, S. (2015) Learning where to offend: Effects of past on future burglary locations. *Applied Geography*, 60, 120–129.
- Besag, J., York, J. & Mollié, A. (1991) Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1), 1–20.
- Blei, D. & Lafferty, J. (2006a) Correlated topic models. In: *Advances in Neural Information Processing Systems*.
- Blei, D.M. & Lafferty, J.D. (2006b) Dynamic topic models. In: *International Conference on Machine Learning*.
- Blei, D.M., Ng, A.Y. & Jordan, M.I. (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.
- Bowers, K. (2014) Risky facilities: Crime radiators or crime absorbers? A comparison of internal and external levels of theft. *Journal of Quantitative Criminology*, 30(3), 389–414.
- Bowers, K.J., Johnson, S.D. & Pease, K. (2004) Prospective hot-spotting: The future of crime mapping? *British Journal of Criminology*, 44(5), 641–658.
- Braga, A.A., Papachristos, A.V. & Hureau, D.M. (2014) The effects of hot spots policing on crime: An updated systematic review and meta-analysis. *Justice Quarterly*, 31(4), 633–663.
- 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. (1993a) Environment, routine and situation: Toward a pattern theory of crime. *Advances in Criminological Theory*, 5(2), 259–294.

- Brantingham, P.L. & Brantingham, P.J. (1993b) Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology*, 13(1), 3–28.
- Buntine, W. (2002) Variational extensions to EM and multinomial PCA. In: *European Conference on Machine Learning*.
- Cahill, M. & G. Mulligan (2007) Using geographically weighted regression to explore local crime patterns. *Social Science Computer Review*, 25(2), 174–193.
- Caplan, J.M., Kennedy, L.W. & Miller, J. (2011) Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting. *Justice Quarterly*, 28(2), 360–381.
- Ceccato, V., Haining, R. & Signoretta, P. (2002) Exploring offence statistics in Stockholm city using spatial analysis tools. *Annals of the Association of American Geographers*, 92(1), 29–51.
- 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.
- Chamberlain, A.W. & Hipp J.R. (2015) It's all relative: Concentrated disadvantage within and across neighborhoods and communities, and the consequences for neighborhood crime. *Journal of Criminal Justice*, 43(6), 431–443.
- 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.G. (1997) *Situational crime prevention*. New York: Criminal Justice Press Monsey.
- Cohen, L.E. & Felson, M. (1979) Social change and crime rate trends: A routine activity approach. *American Sociological Review*, 44(4), 588–608.
- Cohen, J., Gorr, W.L. & Olligschlaeger, A.M. (2007) Leading indicators and spatial interactions: A crime-forecasting model for proactive police deployment. *Geographical Analysis*, 39(1), 105–127.
- Deryol, R., Wilcox, P., Logan, M. & Wooldredge, J. (2016) Crime places in context: An illustration of the multilevel nature of hot spot development. *Journal of Quantitative Criminology*, 32(2), 305–325.
- Diggle, P.J., Moraga, P., Rowlingson, B. & Taylor, B.M. (2013) Spatial and spatiotemporal log-Gaussian Cox processes: Extending the geostatistical paradigm. *Statistical Science*, 28(4), 542–563.
- Drawve, G. & Barnum, J.D. (2018) Place-based risk factors for aggravated assault across police divisions in Little Rock, Arkansas. *Journal of Crime and Justice*, 41(2), 173–192.
- EDINA Digimap Ordnance Survey Service. (2018) Points of interest [CSV geospatial data], scale 1:1250, items: 421031, <https://digimap.edina.ac.uk>
- Flaxman, S., Wilson, A., Neill, D., Nickisch, H. & Smola, A. (2015) Fast Kronecker inference in Gaussian processes with non-Gaussian likelihoods. In *International Conference on Machine Learning*.
- Flaxman, S., Chirico, M., Pereira, P. & Loeffler, C. (2018) Scalable high-resolution forecasting of sparse spatiotemporal events with kernel methods: A winning solution to the NIJ "Real-Time Crime Forecasting Challenge". *arXiv preprint arXiv:1801.02858*.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B. (2013) *Bayesian data analysis*. Boca Raton: Chapman and Hall/CRC.
- Gorr, W.L. & Lee, Y. (2015) Early warning system for temporary crime hot spots. *Journal of Quantitative Criminology*, 31(1), 25–47.
- Groff, E.R. & Lockwood, B. (2014) Criminogenic facilities and crime across street segments in Philadelphia: Uncovering evidence about the spatial extent of facility influence. *Journal of Research in Crime and Delinquency*, 51(3), 277–314.
- Haberman, C.P. (2017) Overlapping hot spots? Examination of the spatial heterogeneity of hot spots of different crime types. *Criminology & Public Policy*, 16(2), 633–660.
- Haberman, C.P. & Ratcliffe, J.H. (2015) Testing for temporally differentiated relationships among potentially criminogenic places and census block street robbery counts. *Criminology*, 53(3), 457–483.
- Heller, K.A., Williamson, S. & Ghahramani, Z. (2008) Statistical models for partial membership. In *International Conference on Machine Learning*.
- Hipp, J.R. (2007) Income inequality, race, and place: Does the distribution of race and class within neighborhoods affect crime rates? *Criminology*, 45(3), 665–697.
- Hipp, J.R. & Yates, D.K. (2011) Ghettos, thresholds, and crime: Does concentrated poverty really have an accelerating increasing effect on crime? *Criminology*, 49(4), 955–990.
- Johnson, S.D. & Bowers, K.J. (2004) The stability of space-time clusters of burglary. *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.
- Johnson, S.D., Bowers, K.J., Birks, D.J. & Pease, K. (2009) Predictive mapping of crime by ProMap: Accuracy, units of analysis, and the environmental backcloth. In: Weisburd, D., Bernasco, W. & Bruinsma, G.J. (Eds.) *Putting crime in its place*. New York: Springer, pp. 171–198.
- Johnson, S.D., Summers, L. & Pease, K. (2009) Offender as forager? A direct test of the boost account of victimization. *Journal of Quantitative Criminology*, 25(2), 181–200.
- Jones, R.W. & Pridemore, W.A. (2019) Toward an integrated multilevel theory of crime at place: Routine activities, social disorganization, and the law of crime concentration. *Journal of Quantitative Criminology*, 35(3), 543–572.
- Kinney, J.B., Brantingham, P.L., Wuschke, K., Kirk, M.G. & Brantingham, P.J. (2008) Crime attractors, generators and detractors: Land use and urban crime opportunities. *Built Environment*, 34(1), 62–74.
- Kuang, D., Brantingham, P.J. & Bertozzi, A.L. (2017) Crime topic modeling. *Crime Science*, 6(1), 12.
- Leroux, B.G., Lei, X. & Breslow, N. (2000) Estimation of disease rates in small areas: A new mixed model for spatial dependence. In: Halloran, M.E. & Berry, D. (Eds.) *Statistical models in epidemiology, the environment, and clinical trials*. New York: Springer, pp. 179–191.
- Lindgren, F., Rue, H. & Lindström, J. (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4), 423–498.
- Liu, H. & Brown, D.E. (2003) Criminal incident prediction using a point-pattern-based density model. *International Journal of Forecasting*, 19(4), 603–622.
- Liu, H. & Zhu, X. (2017) Joint modeling of multiple crimes: A Bayesian spatial approach. *ISPRS International Journal of Geo-Information*, 6(1), 16.
- Møller, J. & Rasmussen, J.G. (2005) Perfect simulation of Hawkes processes. *Advances in Applied Probability*, 37(3), 629–646.
- Møller, J., Syversveen, A.R. & Waagepetersen, R.P. (1998) Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3), 451–482.
- MacKay, D.J. (1995) Probable networks and plausible predictions? A review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems*, 6(3), 469–505.
- 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*, 60(2), 303–322.
- Mohler, G. (2013) Modeling and estimation of multi-source clustering in crime and security data. *The Annals of Applied Statistics*, 7(3), 1525–1539.
- Mohler, G. (2014) Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *International Journal of Forecasting*, 30(3), 491–497.
- 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.
- Neal, R.M. (2012) *Bayesian learning for neural networks*, vol. 118. Berlin: Springer Science & Business Media.
- Perry, W.L. (2013) *Predictive policing: The role of crime forecasting in law enforcement operations*. Santa Monica: RAND Corporation.
- Piza, E., Feng, S., Kennedy, L. & Caplan, J. (2017) Place-based correlates of motor vehicle theft and recovery: Measuring spatial influence across neighbourhood context. *Urban Studies*, 54(13), 2998–3021.
- Porter, M.D. & Reich, B.J. (2012) Evaluating temporally weighted kernel density methods for predicting the next event location in a series. *Annals of GIS*, 18(3), 225–240.
- Quick, M., Li, G. & Brunton-Smith, I. (2018) Crime-general and crime-specific spatial patterns: A multivariate spatial analysis of four crime types at the small-area scale. *Journal of Criminal Justice*, 58, 22–32.
- Riebler, A., Sørbye, S.H., Simpson, D. & Rue, H. (2016) An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4), 1145–1165.
- Roach, J. & Pease, K. (2014) Police overestimation of criminal career homogeneity. *Journal of Investigative Psychology and Offender Profiling*, 11(2), 164–178.

- Rodrigues, A. & Diggle, P.J. (2012) Bayesian estimation and prediction for inhomogeneous spatiotemporal log-Gaussian Cox processes using low-rank models, with application to criminal surveillance. *Journal of the American Statistical Association*, 107(497), 93–101.
- Rosser, G., Davies, T., Bowers, K.J., Johnson, S.D. & Cheng, T. (2017) Predictive crime mapping: Arbitrary grids or street networks? *Journal of Quantitative Criminology*, 33(3), 569–594.
- Rue, H. & Held, L. (2005) *Gaussian Markov random fields: Theory and applications*. Boca Raton: Chapman and Hall/CRC.
- 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.
- Schreck, C.J., McGloin, J.M. & Kirk, D.S. (2009) On the origins of the violent neighborhood: A study of the nature and predictors of crime-type differentiation across Chicago neighborhoods. *Justice Quarterly*, 26(4), 771–794.
- Shaw, C.R. & McKay, H.D. (1942) *Juvenile delinquency and urban areas*. Chicago: University of Chicago Press.
- Shirota, S., Gelfand, A.E. et al. (2017) Space and circular time log Gaussian Cox processes with application to crime event data. *The Annals of Applied Statistics*, 11(2), 481–503.
- Smith, W.R., Frazee, S.G. & Davison, E.L. (2000) Furthering the integration of routine activity and social disorganization theories: Small units of analysis and the study of street robbery as a diffusion process. *Criminology*, 38(2), 489–524.
- Stucky, T.D. & Ottensmann, J.R. (2009) Land use and violent crime. *Criminology*, 47(4), 1223–1264.
- 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.
- Teh, Y.W., Jordan, M.I., Beal, M.J. & Blei, D.M. (2005) Sharing clusters among related groups: Hierarchical Dirichlet processes. In *Neural Information Processing Systems*.
- 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., Ruiters, S., Johnson, S.D., White, G. et al. (2015) Burglar target selection: A cross-national comparison. *Journal of Research in Crime and Delinquency*, 52(1), 3–31.
- Townsley, M., Birks, D., Ruiters, S., Bernasco, W. & White, G. (2016) Target selection models with preference variation between offenders. *Journal of Quantitative Criminology*, 32(2), 283–304.
- Weisburd, D. (2015) The law of crime concentration and the criminology of place. *Criminology*, 53(2), 133–157.
- Weisburd, D., Maher, L., Sherman, L., Buerger, M., Cohn, E. & Petrisino, A. (1992) Contrasting crime general and crime specific theory: The case of hot spots of crime. *Advances in Criminological Theory*, 4(1), 45–69.
- Wheeler, A.P. (2019) Quantifying the local and spatial effects of alcohol outlets on crime. *Crime & Delinquency*, 65(6), 845–871.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Virtanen S, Girolami M. Spatio-temporal mixed membership models for criminal activity. *J R Stat Soc Series A*. 2021;184:1220–1244. <https://doi.org/10.1111/rssa.12642>